

Give your cells a competitive edge

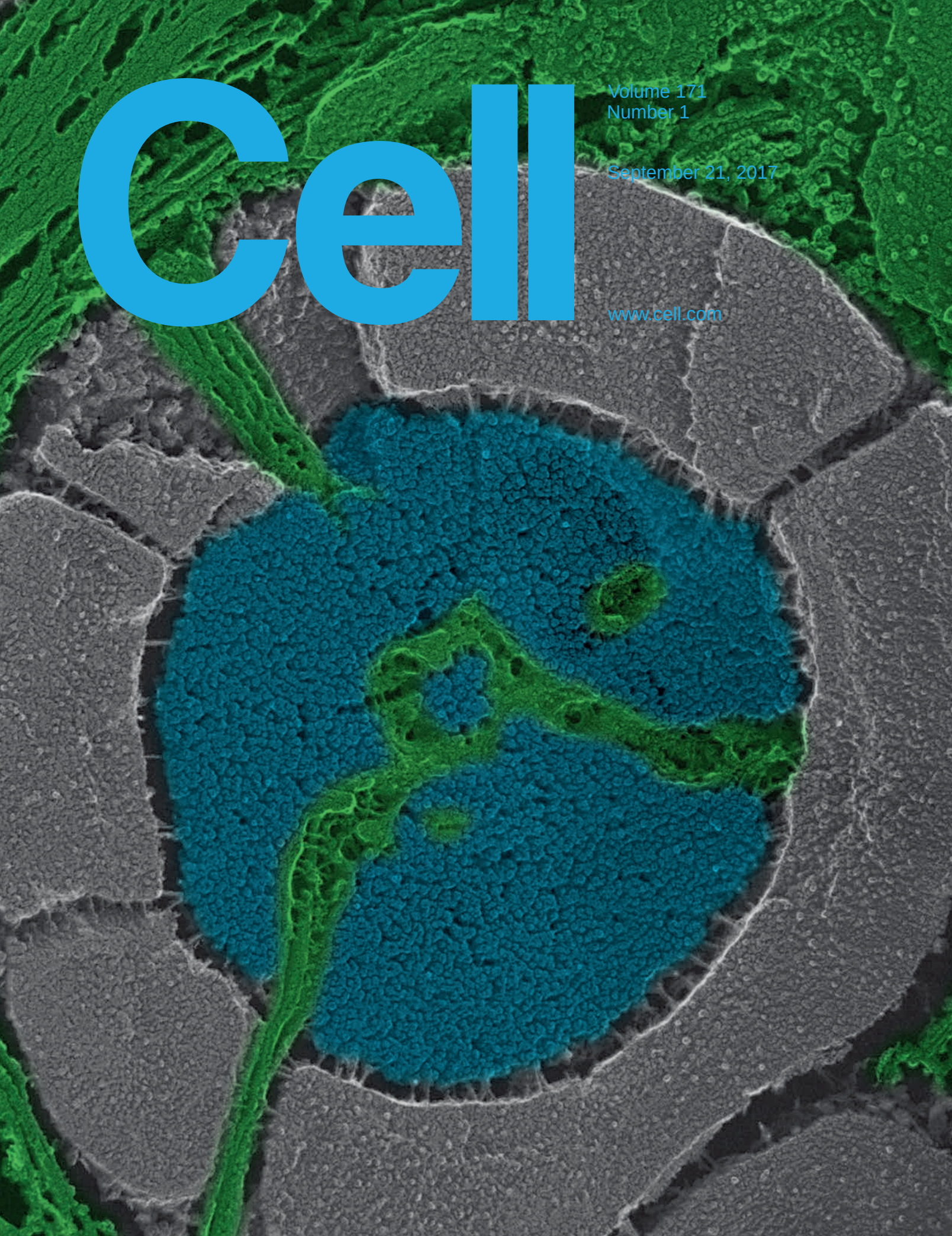
Sometimes it's the simplest of things that make the difference for cell health.

Anyone who has ever cultured stem cells in a lab knows that it requires a tremendous investment in time, energy, and expense. You just can't leave anything to chance. That's why so many researchers choose Biological Industries' **NutriStem® hPSC XF** and **MSC NutriStem® Media**. From optimal growth and expansion to weekend-free culture, NutriStem® is designed to promote superior growth in a flexible culture format so you can achieve more reliable, consistent cell health.

- **Defined, serum-free, and xeno-free**
- **FDA Drug Master File (DMF) available**
- **Produced under GMP**
- **Scale-up options available**

Visit **bioind.com** to learn more.

BI
Biological Industries
Culture of Excellence



Cell

Volume 171
Number 1

September 21, 2017

www.cell.com

Pipetting, Pipetting, Pipetting.



...Our Reputation is Built on Reliability

If you're tired of unreliable equipment, visit **drummondsci.com** because DRUMMOND means QUALITY. Go to **ergonomic.expert** to see how the superior ergonomic design of the Pipet-Aid® XL can change the way you think about lab hood pipetting.



DRUMMOND SCIENTIFIC COMPANY

Developers of the original Pipet-Aid®

500 Parkway, Box 700 • Broomall, PA 19008 • 800.523.7480



With Cell Press Webinars, our network is your network!

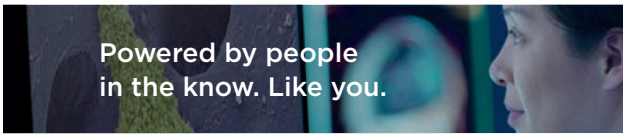
Cell Press Webinars give you access to hot topics in emerging research and the application of new technology.

Watch essential, need-to-know webcasts via live streaming or on demand from the comfort and convenience of your office, lab, or home.

Need-to-know topics, editorially curated

World-class presenters, experts in their field

Moderated by Cell Press editors



Tap into our network today!
Visit www.cell.com/webinars

CellPress
Webinars

ASHG 2017

ORLANDO • OCTOBER 17-21, 2017

SHARING DISCOVERIES. **SHAPING** OUR FUTURE.

ASHG is pleased to welcome two special guests to the

Presidential Symposium at ASHG 2017



Bill Gates
Bill & Melinda Gates Foundation



Francis Collins
National Institutes of Health

Wednesday, October 18, 2017
5:00-6:30 pm
South Hall B, Orange County Convention Center

REGISTER TODAY

ashg.org/2017meeting

ASGCT CALL FOR AWARD NOMINATIONS!



The **Outstanding Achievement Award (OAA)** recognizes an ASGCT Member who has achieved a pioneering research success, a specific high impact accomplishment, or a lifetime of significant scientific contributions to the field of gene and cell therapy.

The **Outstanding New Investigator Award (ONI)** recognizes ASGCT Members who are newly independent researchers, 10 years or fewer from their first independent research position, based on their contributions to the field.

The **Sonia Skarlatos Public Service Award (PSA)** will recognize a person or group that has consistently fostered and enhanced the field of gene and cell therapy through governmental agencies, public policy groups, public education, or non-governmental charitable organizations.

The deadline to submit nominations for the 2018 awards is December 15, 2017.

For more information on these awards and to submit a nomination, please visit **www.asgct.org**.

AMERICAN SOCIETY of GENE & CELL THERAPY

www.asgct.org

20817-306

Cell

Editor
Emilie Marcus

Senior Deputy Editor
Elena Porro

Deputy Editor
Robert Kruger

Acting Deputy Editor
Xiaohong Helena Yang

Scientific Editors
Sarah Geisler
Marta Koch
João Monteiro
Sri Devi Narasimhan
April Pawluk
Lara Szewczak
Jiaying Tan

Senior Managing Editor
Andy Smith

Managing Editor
Trina Arpin

Copyeditors
Erik Frenz
Tracy Campbell

Art Program Manager
Andrew A. Tang

Managing Designer and Illustrator
Yvonne Blanco

Illustrator
Kip B. Lyall

Journal Associate
Amanda Naples

Editorial Board
C. David Allis
James Allison
Geneviève Almouzni
Angelika Amon
Silvia Arber
Johan Auwerx
Richard Axel
Cori Bargmann
Yasmine Belkaid
Jeffrey Benovic
Wendy Bickmore
Elizabeth Blackburn
Joan Brugge
Lewis Cantley
Xuetao Cao
Joanne Chory
David Clapham
Andrew Clark
Hans Clevers
James Collins
Patrick Cramer
Benjamin Cravatt
George Daley
Jeff Dangl
Ted Dawson
Karl Deisseroth
Hongkui Deng
Pier Paolo di Fiore
Ivan Dikic
Julian Downward
Bruce Edgar
Steve Elledge
Anne Ephrussi
Ronald Evans
Marco Foiani
Elaine Fuchs
Daniel Geschwind
Yukiko Goda
Joe Goldstein
Douglas Green
Leonard Guarente
Taekjip Ha
Daniel Haber
Barton Haynes
Mark Hochstrasser
Erika Holzbaur
Tony Hunter
James Hurley
Trey Ideker
Thomas Jessell
Tarun Kapoor
Narry Kim
Mary-Claire King
David Kingsley
Frank Kirchhoff
Brian Kobilka
Jan Korbel
Mark Lemmon
Beth Levine
Wendell Lim
Jennifer Lippincott-Schwartz
Dan Littman
Richard Losick
Scott Lowe
Tom Maniatis
Matthias Mann
Kelsey Martin
Joan Massagué
Satyajit Mayor
Ruslan Medzhitov
Craig Mello
Tom Misteli
Tim Mitchison
Danesh Moazed
Alex Mogilner
Huck Hui Ng
Paul Nurse
Svante Pääbo
Roy Parker
Dana Pe'er
Erika Pearce
Kathrin Plath
Carol Prives
Lalita Ramakrishnan
Rama Ranganathan
Aviv Regev
Anne Ridley
Alexander Rudensky
Helen Saibil
Charles Sawyers
Joseph Schlessinger
Hans Schöler
Trina Schroer
Kevan Shokat
Nahum Sonenberg
Deepak Srivastava
Azim Surani
Keiji Tanaka
Sohail Tavazoie
Craig Thompson
Robert Tjian
Skip Virgin
Jonathan Weissman
Matthew Welch
Ian Wilson
Shinya Yamanaka
Richard Young
Marino Zerial
Xiaowei Zhuang
Juleen Zierath
Huda Zoghbi

***Cell* Office**
Cell, Cell Press, 50 Hampshire St., 5th Floor, Cambridge, MA 02139, USA
Phone: (+1) 617 397 2800, Fax: (+1) 617 661 7061, E-mail: celleditor@cell.com
Online Publication: http://www.cell.com

Cell (ISSN 0092-8674) is published biweekly by Cell Press, 50 Hampshire Street, 5th Floor, Cambridge, Massachusetts 02139. The institutional subscription rate for 2017 is $2,020 (US and Canada) or $2,277 (elsewhere). The individual subscription rate is $293 (US and Canada) or $431 (elsewhere). The individual copy price is $50. Periodicals postage paid at Boston, Massachusetts and additional mailing offices. Postmaster: send address changes to Elsevier Customer Service Americas, Cell Press Journals, 3251 Riverport Lane, Maryland Heights, MO 63043, USA.
♾The paper used in this publication meets the requirments of ANSI/NISO Z39.48-1992 (Permanence of Paper). Printed by Dartmouth Printing Company, Hanover, NH.

THE AMERICAN SOCIETY OF GENE & CELL THERAPY AND THE HEMOPHILIA FEDERATION OF AMERICA WEBINAR SERIES

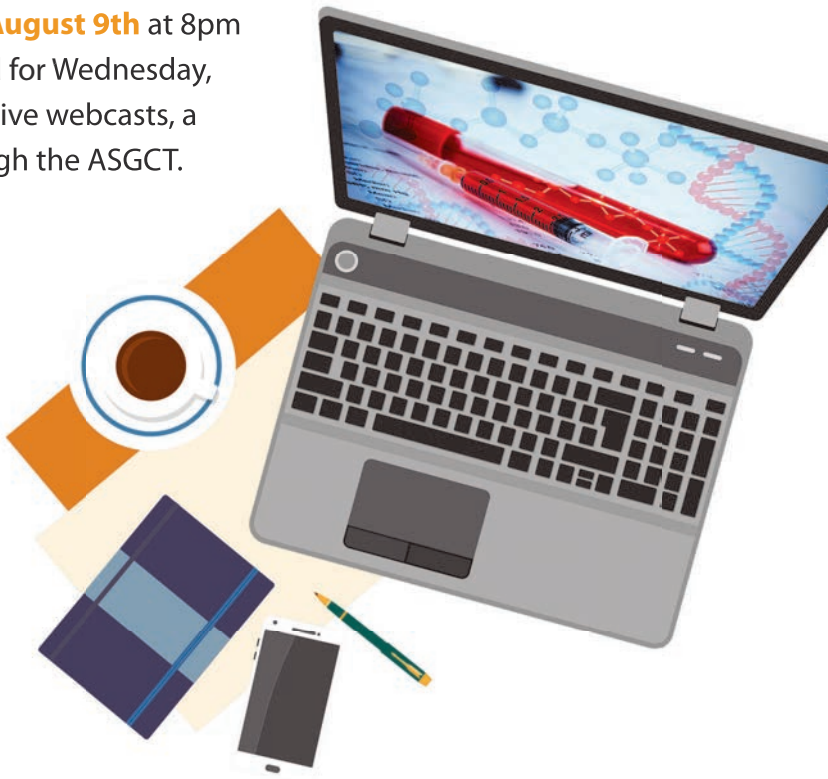
ASGCT and HFA are pleased to announce a series of two webinars on advances in gene therapy for hemophilia.

AMERICAN SOCIETY of GENE & CELL THERAPY

HFA Hemophilia Federation of America™

The first webinar will take place on Wednesday, **August 9th** at 8pm ET/ 5pm PT and the second webinar is scheduled for Wednesday, **August 30th** at 8pm ET/ 5pm PT. Following the live webcasts, a recording of the webinars will be available through the ASGCT.

Mark Kay, PhD of Stanford University School of Medicine will explain the basics of gene therapy to the hemophilia patient community and how gene therapy may be used to treat hemophilia.



Check the ASGCT website at asgct.org for more information.

Cell Press

CEO
Emilie Marcus

Vice President of Operations
Keith Wollman

Vice President of Editorial
Deborah Sweet

Vice President of Marketing and Commercial Sales
Jonathan Atkinson

Publishing Directors
Lakshmi Goyal
Peter Lee

Editorial Director, Content Development
Elena Porro

Director of Production
Meredith Adinolfi

Senior Product Manager
Gabriel Harp

Media Relations Manager
Joseph Caputo

Display Advertising
For advertising, sponsorship, and reprint opportunities, contact Jonathan Christison (jchristison@cell.com) or visit http://advertisers.cell.com/contact-us.

Recruitment Advertising
For recruitment advertising, visit http://recruitmentads.cell.com/contact-us.

©2017 Elsevier Inc. All rights reserved.
This journal and the individual contributions contained in it are protected under copyright by Elsevier Inc., and the following terms and conditions apply to their use:

Photocopying
Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for nonprofit educational classroom use.

Permissions
For information on how to seek permission, visit http://www.elsevier.com/permissions or call +44 1865 843830 (UK) / +1 215 239 3804 (US).

Derivative Works
Subscribers may reproduce tables of contents or prepare lists of articles including summaries for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution. Permission of the Publisher is required for all other derivative works, including compilations and translations (please consult http://www.elsevier.com/permissions).

Electronic Storage or Usage
Permission of the Publisher is required to store or use electronically any material contained in this journal, including any article or part of an article (please consult www.elsevier.com/permissions). Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Publisher.

Notice
No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence, or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

Reprints
Article reprints are available through Elsevier's reprint service; for information, contact Derrick Imasa (e-mail: d.imasa@elsevier.com; ph: +1 212 633 3874).

Subscription Orders and Inquiries
Mail, fax, or e-mail address changes to Elsevier Customer Service Americas, allowing 4–6 weeks for processing. Lost or damaged issues will be replaced, subject to availability, if Cell Press is notified within the claim period (US and airmail delivery: 3 months from issue date; surface delivery: 4 months from issue date). Periodical delivery in the US can take up to 3 weeks. Airmail delivery can take 2–4 weeks. Mailing address: Elsevier Customer Service Americas, Cell Press Journals, 3251 Riverport Lane, Maryland Heights, MO 63043, USA. Toll-free phone within USA/Canada: +1 866 314 2355; phone for outside US/Canada: +1 314 447 8880; fax: +1 314 447 8029; e-mail: subs@cell.com.

Funding Body Agreements and Policies
Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies, please visit https://www.elsevier.com/about/open-science/open-access/agreements.



The Ho-Am Prize ignites the passion and innovative spirit for scientific discovery

We welcome your recommendations for the 2018 Ho-Am Prize candidates.

- Award categories are Science, Engineering, and Medicine.
- Researchers of Korean heritage are eligible for the Prize.
- Deadline for submission is Oct. 31, 2017.

*Learn more about candidate recommendation at **www.hoamprize.org**



To continue Samsung Founder 'Ho-Am' Byung-chull Lee's (1910-1987) efforts to maximize both human potential and public interest, Chairman Kun-Hee Lee of Samsung established the Prize in 1990. The Prize is awarded annually to outstanding researchers of Korean heritage around the world who have made important contributions to the advancement of science.

THE HO-AM PRIZE

HOAM
THE HO-AM FOUNDATION

SAMSUNG

Leading Edge

Cell Volume 171 Number 1, September 21, 2017

SELECT

1 A New Model T on the Horizon? *X.H. Yang*

COMMENTARY

5 Artists Create Puzzles, Scientists Solve Them *J.L. Goldstein*

BENCHMARKS

10 TOR, the Gateway to Cellular Metabolism, Cell Growth, and Disease *J. Blenis*

14 A Prize for Cancer Prevention *H. Varmus*

STORIES

18 An Amazing Turn of Events *M.N. Hall*

CONVERSATIONS

23 Two Basic Scientists Walk into a Translational Space

PREVIEWS

28 The Pyrenoid: An Overlooked Organelle Comes out of Age *J.-D. Rochaix*

30 Prion-like Domains Program Ewing's Sarcoma *J. Shorter*

32 Huntingtin Fibrils Poke Membranes *P. Guedes-Dias and E.L.F. Holzbaur*

(continued)

Cell Symposia

Human Immunity

October 22–24, 2017 — Banff, Canada



Extended Abstract Submission Deadline:

August 14, 2017

Early registration deadline:

September 1, 2017

While we have experienced important advances in immune-based therapies for many types of diseases, the complexity of human biology and the diversity that makes each of us unique poses outstanding challenges to translate basic knowledge into effective treatments. The aim of this symposium is to bring together scientists at the forefront of this endeavour to examine and discuss a broad range of topics, from new approaches to map the human immune system to mechanisms to fight cancer and infections.

Keynote Speakers

Mark Davis, *USA*

Michel Nussenzweig, *USA*

Speakers

Galit Alter, *USA*
Dan Barouch, *USA*
Luis Barreiro, *Canada*
Yasmine Belkaid, *USA*
Marco Colonna, *USA*
Eran Elinav, *Israel*
Donna Farber, *USA*
Florent Ginhoux, *Singapore*
Adrian Hayday, *UK*
Antonio Lanzavecchia, *Switzerland*
Mike Lenardo, *USA*
Mihai Netea, *The Netherlands*
Virginia Pascual, *USA*
Dana Pe'er, *USA*
Sergio Quezada, *UK*
Lluis Quintana-Murci, *France*
Federica Sallusto, *Switzerland*
Shannon Turley, *USA*
Eric Vivier, *France*

Organizers

Tak Mak, *University of Toronto, Canada*
Tania Watts, *University of Toronto, Canada*
João Monteiro, Scientific Editor, *Cell*
Fabiola Rivas, Scientific Editor, *Immunity*

cell-symposia.com/human-immunity-2017

REVIEW

34 Genome Regulation by Polycomb and Trithorax: 70 Years and Counting — *B. Schuettengruber, H.-M. Bourbon, L. Di Croce, and G. Cavalli*

SNAPSHOT

258 N-Glycosylation Processing Pathways across Kingdoms — *C.-Y. Chung, N.I. Majewska, Q. Wang, J.T. Paul, and M.J. Betenbaugh*

(continued)

Cell Symposia

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Next Gen Immunology

February 11–14, 2018 — Rehovot, Israel



Abstract submission deadline:

Sept. 22, 2017

Early registration deadline:

December 8, 2017

We aim to stimulate exciting discussions, to foster new interdisciplinary collaborations, and to promote cross-pollination of ideas that can help us to move on to the next generation of immunology research.

Keynote Speakers

James P. Allison, *USA*

Ruslan Medzhitov, *USA*

Speakers

Ido Amit, *Israel*
David Artis, *USA*
Yasmine Belkaid, *USA*
Julie Magarian Blander, *USA*
Doreen Cantrell, *UK*
Marco Colonna, *USA*
Pascale Cossart, *France*
Jason Cyster, *USA*
Ana Domingos, *Portugal*
Eran Elinav, *Israel*
Brett Finlay, *Canada*
Michael Fischbach, *USA*
Richard Flavell, *USA*
Jorge Henao-Mejia, *USA*
Kenya Honda, *Japan*
Lora Hooper, *USA*
Guido Kroemer, *France*
Vijay Kuchroo, *USA*
Eicke Latz, *Germany*
Dan Littman, *USA*
Ofer Mandelboim, *Israel*
Alberto Mantovani, *Italy*
Diane Mathis, *USA*
Eric Pamer, *USA*
Virginia Pascual, *USA*
Erika Pearce, *Germany*
Fiona Powrie, *UK*
Richard Ransohoff, *USA*
Maria Rescigno, *Italy*
Philippe Sansonetti, *France*
Michal Schwartz, *Israel*
Robert Schreiber, *USA*
Eran Segal, *Israel*
Padmanee Sharma, *USA*

Organizers

Ido Amit, Weizmann Institute of Science
Eran Elinav, Weizmann Institute of Science
Bruce Koppelman, Scientific Editor, *Immunity*
João Monteiro, Scientific Editor, *Cell*

cell.com/symposia

Articles

Cell Volume 171 Number 1, September 21, 2017

59	Reconstructing Prehistoric African Population Structure	*P. Skoglund, J.C. Thompson, M.E. Prendergast, A. Mittnik, K. Sirak, M. Hajdinjak, T. Salie, N. Rohland, S. Mallick, A. Peltzer, A. Heinze, I. Olalde, M. Ferry, E. Harney, M. Michel, K. Stewardson, J.I. Cerezo-Román, C. Chiumia, A. Crowther, E. Gomani-Chindebvu, A.O. Gidna, K.M. Grillo, I.T. Helenius, G. Hellenthal, R. Helm, M. Horton, S. López, A.Z.P. Mabulla, J. Parkington, C. Shipton, M.G. Thomas, R. Tibesasa, M. Welling, V.M. Hayes, D.J. Kennett, R. Ramesar, M. Meyer, S. Pääbo, N. Patterson, A.G. Morris, N. Boivin, R. Pinhasi, J. Krause, and D. Reich*
72	The Kinetochore Receptor for the Cohesin Loading Complex	*S.M. Hinshaw, V. Makrantoni, S.C. Harrison, and A.L. Marston*
85	Dynamic Control of X Chromosome Conformation and Repression by a Histone H4K20 Demethylase	*K. Brejc, Q. Bian, S. Uzawa, B.S. Wheeler, E.C. Anderson, D.S. King, P.J. Kranzusch, C.G. Preston, and B.J. Meyer*
103	Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate	*T. Isoda, A.J. Moore, Z. He, V. Chandra, M. Aida, M. Denholtz, J. Piet van Hamburg, K.M. Fisch, A.N. Chang, S.P. Fahl, D.L. Wiest, and C. Murre*
120	Structure of an Intron Lariat Spliceosome from *Saccharomyces cerevisiae*	*R. Wan, C. Yan, R. Bai, J. Lei, and Y. Shi*
133	A Spatial Interactome Reveals the Protein Organization of the Algal CO_2-Concentrating Mechanism	*L.C.M. Mackinder, C. Chen, R.D. Leib, W. Patena, S.R. Blum, M. Rodman, S. Ramundo, C.M. Adams, and M.C. Jonikas*
148	The Eukaryotic CO_2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization	*E.S. Freeman Rosenzweig, B. Xu, L. Kuhn Cuellar, A. Martinez-Sanchez, M. Schaffer, M. Strauss, H.N. Cartwright, P. Ronceray, J.M. Plitzko, F. Förster, N.S. Wingreen, B.D. Engel, L.C.M. Mackinder, and M.C. Jonikas*
163	Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain	*G. Boulay, G.J. Sandoval, N. Riggi, S. Iyer, R. Buisson, B. Naigles, M.E. Awad, S. Rengarajan, A. Volorio, M.J. McBride, L.C. Broye, L. Zou, I. Stamenkovic, C. Kadoch, and M.N. Rivera*
179	In Situ Architecture and Cellular Interactions of PolyQ Inclusions	*F.J.B. Bäuerlein, I. Saha, A. Mishra, M. Kalemanov, A. Martínez-Sánchez, R. Klein, I. Dudanova, M.S. Hipp, F.U. Hartl, W. Baumeister, and R. Fernández-Busnadiego*
188	Load Adaptation of Lamellipodial Actin Networks	*J. Mueller, G. Szep, M. Nemethova, I. de Vries, A.D. Lieber, C. Winkler, K. Kruse, J.V. Small, C. Schmeiser, K. Keren, R. Hauschild, and M. Sixt*

(continued)

Cell Symposia

CRISPR: From Biology to Technology and Novel Therapeutics

October 22–24, 2017 — Sitges, Spain



Supporting Journals



Join us at this Cell Symposium to meet leading CRISPR researchers, learn about exciting unpublished work, and to gain practical insights into applying CRISPR technology to your own research.

Confirmed Speakers

Charlie Albright, *USA*
Chase Beisel, *USA*
Juan-Carlos Izpisua Belmonte, *USA*
Paula Cannon, *USA*
Jacob Corn, *USA*
Alan Davidson, *Canada*
Charles Gersbach, *USA*
Maria Jasin, *USA*
Martin Jinek, *Switzerland*
J. Keith Joung, *USA*
Nancy Maizels, *USA*
Andy May, *USA*
Francisco Mojica, *Spain*
Luigi Naldini, *Italy*
Tony Nolan, *UK*
John van der Oost, *The Netherlands*
Michel Sadelain, *USA*
Blake Wiedenheft, *USA*

Organizers

J. Keith Joung, Massachusetts General Hospital, Harvard Medical School, *USA*
Andrew May, Chan-Zuckerberg BioHub, *USA*
John Pham, Chief Editor, *Molecular Cell*
Jonathan Saxe, Editor, *Cell Stem Cell*

cell-symposia.com/crispr-2017

201 Regulatory Innate Lymphoid Cells Control Innate Intestinal Inflammation — *S. Wang, P. Xia, Y. Chen, Y. Qu, Z. Xiong, B. Ye, Y. Du, Y. Tian, Z. Yin, Z. Xu, and Z. Fan*

217 Sensory Neurons Co-opt Classical Immune Signaling Pathways to Mediate Chronic Itch — *L.K. Oetjen, M.R. Mack, J. Feng, T.M. Whelan, H. Niu, C.J. Guo, S. Chen, A.M. Trier, A.Z. Xu, S.V. Tripathi, J. Luo, X. Gao, L. Yang, S.L. Hamilton, P.L. Wang, J.R. Brestoff, M.L. Council, R. Brasington, A. Schaffer, F. Brombacher, C.-S. Hsieh, R.W. Gereau IV, M.J. Miller, Z.-F. Chen, H. Hu, S. Davidson, Q. Liu, and B.S. Kim*

229 A Human Bi-specific Antibody against Zika Virus with High Therapeutic Potential — *J. Wang, M. Bardelli, D.A. Espinosa, M. Pedotti, T.-S. Ng, S. Bianchi, L. Simonelli, E.X.Y. Lim, M. Foglierini, F. Zatta, S. Jaconi, M. Beltramello, E. Cameroni, G. Fibriansah, J. Shi, T. Barca, I. Pagani, A. Rubio, V. Broccoli, E. Vicenzi, V. Graham, S. Pullan, S. Dowall, R. Hewson, S. Jurt, O. Zerbe, K. Stettler, A. Lanzavecchia, F. Sallusto, A. Cavalli, E. Harris, S.-M. Lok, L. Varani, and D. Corti*

THEORY

242 A Unifying Theory of Branching Morphogenesis — *E. Hannezo, C.L.G.J. Scheele, M. Moad, N. Drogo, R. Heer, R.V. Sampogna, J. van Rheenen, and B.D. Simons*

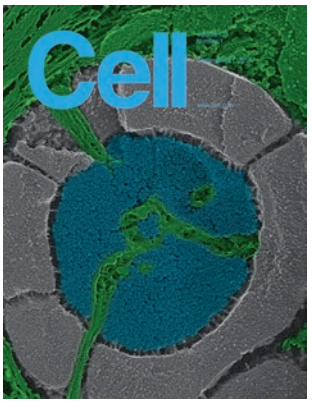
RETRACTION

256 Retraction Notice to: Calcium and SOL Protease Mediate Temperature Resetting of Circadian Clocks — *O. Tataroglu, X. Zhao, A. Busza, J. Ling, J.S. O'Neill, and P. Emery*

ANNOUNCEMENTS

POSITIONS AVAILABLE

On the cover: The pyrenoid is a proteinaceous, non-membrane-bound organelle that is found in the chloroplasts of most algae and is packed with the carbon-fixing enzyme Rubisco. Pyrenoids are estimated to mediate approximately one-third of global carbon fixation. In this issue, Freeman Rosenzweig et al. (148–162) show that the pyrenoid of the green alga *Chlamydomonas reinhardtii* is not crystalline but behaves as a liquid that is primarily inherited by fission and disperses and condenses during cell division. This image depicts a *Chlamydomonas* pyrenoid imaged by quick-freeze deep-etch electron microscopy. The pyrenoid matrix is false colored in blue; the thylakoid membranes, which transverse the matrix, are false colored in green; and the gray blocks are starch granules. Micrograph courtesy of Ursula Goodenough and Robyn Roth.



ANNOUNCING A NEW PUBLISHING PARTNERSHIP

beginning January 2017

American Society of Gene & Cell Therapy and Cell Press



Commencing January 2017, this premier collection of journals in the exciting areas of molecular and cellular therapeutics will be published alongside the high-quality, cutting-edge research that you've come to expect from Cell Press journals.

Learn more at
http://www.cell.com/molecular-therapy-family

AMERICAN SOCIETY of GENE & CELL THERAPY

CellPress

Cell

A New Model T on the Horizon?

If you are a cancer researcher, chances are you've been asked by your relatives and friends, or even that inquisitive passenger sitting next to you on an airplane, about cures for cancer. Is there one for this cancer? Does that drug really work? Needless to say, the recent success of immune checkpoint inhibitors ignites much discussion and hope. Lately, the curiosity turns to ''can you tell me what CAR-T is''?

As curious as it might be, CAR-T does not really have an automobile origin, let alone the Ford Model T. It is a type of gene- and cell-based immunotherapy that utilizes chimeric antigen receptor (CAR) T cells. But beyond that, let's just say that both the curiosity and the analogy are rather justifiable. Only 3 weeks ago, the U.S. Food and Drug Administration (FDA) approved Kymriah (tisagenlecleucel, formerly known as CTL019), a CAR-T therapy, for children and young adult patients with refractory or relapsing B cell precursor acute lymphoblastic leukemia (ALL). Calling it ''a historic action'' by the FDA, this approval not only ''brings the first gene therapy to the United States'' but also marks ''a new frontier in medical innovation with the ability to reprogram a patient's own cells to attack a deadly cancer,'' said FDA Commissioner Scott Gottlieb (U.S. Food & Drug Administration, 2017).

For the first time, we now have a medicine that is powered by a ''live engine''—living cells—and has a good number of moving parts built into it. CAR-T cells are living T cells genetically engineered to express synthetic, target-specific chimeric antigen receptors (CARs) on the cell surface. In the case of tisagenlecleucel, the CAR-T cells are autologous T cells coming from the patients themselves, modified to be CD19-targeting, a critical step that needs to be done in a manufacturing center. Engineered T cells are then sent back to the hospital where they can be reinfused to the same patient to perform the magic trick: recognize the target antigen present on tumor cells of the patient, proliferate, and kill the tumor cells upon antigen contact.

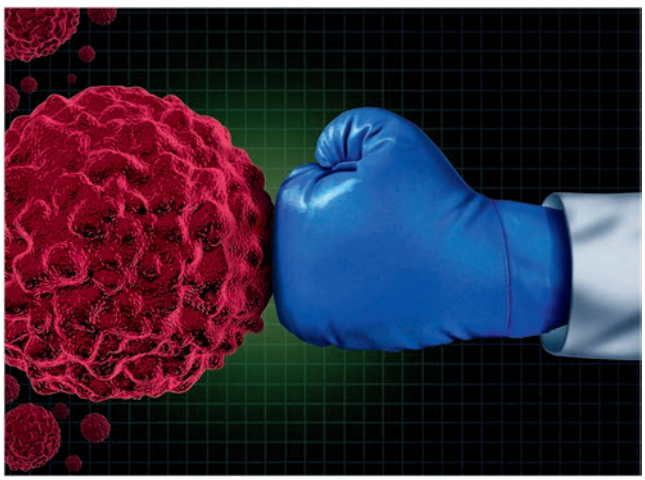


Image from iStock/wildpixel.

Why CD19? CD19 is a surface antigen highly expressed in many B cell leukemias and lymphomas, enabling the clinical efficacy of CD19-targeting CAR-T cells. Interestingly, complete remission of a brain metastasis was observed in a patient with primary refractory diffuse large B cell lymphoma (DLBCL) participating in a clinical trial of another CD19-targeting CAR-T therapy (JCAR017), the first case of a CAR-T response in a central nervous system lymphoma (Abramson et al., 2017). More remarkably, when a subcutaneous lesion developed two months after the initial infusion and a biopsy was performed, the anti-CD19 CAR-T cells spontaneously reexpanded in vivo and the tumor abated without further treatment. Identifying the underlying mechanism and designing strategies to reactivate CAR-T cells with prolonged antitumor activity will be of tremendous interest.

One unique and major bottleneck for the current CAR-T procedure is the isolation, storage, and manufacturing of the T cells, as mentioned earlier, whether they be autologous (patient's own) or allogenic (from healthy donors as an ''off-the-shelf'' product), and the safety and cost concerns that are associated with these steps. Matthias Stephan and colleagues recently made biodegradable polymer-based nanoparticles that can target T cells. When loaded with DNA designed to render host T cells to express a specific CAR, these nanocarriers can efficiently program the host's T cells and exert therapeutic efficacy in an immunocompetent mouse model of B cell ALL (Smith et al., 2017a). The new approach, therefore, has the potential to eliminate the need to isolate T cells from people, simplify the storage, and reduce treatment cost. Innovations in biomaterials and immunoengineering will be a key in the future development of CAR-T therapy and other ''living drugs.''

It also remains to be seen whether CAR-T can work equally well in solid tumors, which account for the majority of human cancers and whose complex microenvironment and heterogeneity present formidable challenges at multiple levels for the design and the effectiveness of the CAR-T cells. Brown et al. recently reported that a patient with recurrent multifocal glioblastoma showed regression of all intracranial and spinal tumors after multiple infusions of IL13Rα2-specific CAR-T cells (Brown et al., 2016). The clinical response also continued for 7.5 months after therapy initiation. In another clinical study of patients with recurrent glioblastoma, a single dose of autologous EGFRvIII-targeting CAR-T cells was intravenously administered to patients. The clinical experience of the first ten patients treated suggests that intravenous infusion achieves on-target activity in the brain with no evidence of off-tumor toxicity or cytokine release syndrome, which can be significant safety concerns. However, it also reveals that the microenvironment seems to become more immunosuppressive. Strategies to overcome the adaptive changes and the antigen heterogeneity observed are still needed in this setting (O'Rourke et al., 2017).

In another clever study, Matthias Stephan and colleagues tested a strategy to improve CAR-T's efficacy in solid tumors. Implantable biopolymer devices were designed to codeliverer

Cell Symposia

Neuroscience 2017 Satellite Meeting

Big Questions in Neuroscience

November 9–10, 2017 — Arlington, VA, USA



Extended abstract submission deadline:

Sept. 15, 2017

Early registration deadline:

October 6, 2017

This year's "Big Questions" span from how to understand the complexity of the human brain to the cellular features of neurons and non-neural cells and will feature visionary scientists with an eye to the future of the field.

Keynote Speaker

Stuart Firestein, *USA*

Speakers

Paola Arlotta, *USA*
Graeme Davis, *USA*
Dan Geschwind, *USA*
Chenghua Gu, *USA*
Tamas Horvath, *USA*
Elaine Hsiao, *USA*
Sabine Kastner, *USA*
Jennifer Lippincott-Schwartz, *USA*
Earl Miller, *USA*
Kia Nobre, *UK*
Josef Parvizi, *USA*
David Rowitch, *UK*
Erin Schuman, *Germany*
Paul Taylor, *USA*
Gina Turrigiano, *USA*

Organizers

Ann Goldstein, Scientific Editor, *Neuron*
Ulli Schridde, Scientific Editor, *Neuron*

cell-symposia.com/big-questions-in-neuroscience-2017/

both engineered CAR T cells directly to the surfaces of solid tumors and STING agonists to locally boost immune response. The goal is to combat solid tumors' immunosuppressive microenvironment and eliminate tumors cells that don't express the targeted antigen. In immunocompetent orthotopic mouse models of pancreatic ductal adenocarcinoma and melanoma, the strategy led to tumor regression and prolonged survival (Smith et al., 2017b).

A hundred years ago, Henry Ford's Model T forever transformed the American way of living and going places. He did it with innovations and the achievement of large-scale, assembly-line manufacturing capability. Time will tell whether CAR-T will be a new ''Model T'' on the horizon and with what we already witnessed, there is much to look forward to.

REFERENCES

Abramson, J.S., McGree, B., Noyes, S., Plummer, S., Wong, C., Chen, Y.B., Palmer, E., Albertson, T., Ferry, J.A., and Arrillaga-Romany, I.C. (2017). N. Engl. J. Med. *377*, 783–784.

Brown, C.E., Alizadeh, D., Starr, R., Weng, L., Wagner, J.R., Naranjo, A., Ostberg, J.R., Blanchard, M.S., Kilpatrick, J., Simpson, J., et al. (2016). N. Engl. J. Med. *375*, 2561–2569.

O'Rourke, D.M., Nasrallah, M.P., Desai, A., Melenhorst, J.J., Mansfield, K., Morrissette, J.J.D., Martinez-Lage, M., Brem, S., Maloney, E., Shen, A., et al. (2017). Sci. Transl. Med. *9*, eaaa0984.

Smith, T.T., Stephan, S.B., Moffett, H.F., McKnight, L.E., Ji, W., Reiman, D., Bonagofski, E., Wohlfahrt, M.E., Pillai, S.P.S., and Stephan, M.T. (2017a). Nat. Nanotechnol. *12*, 813–820.

Smith, T.T., Moffett, H.F., Stephan, S.B., Opel, C.F., Dumigan, A.G., Jiang, X., Pillarisetty, V.G., Pillai, S.P.S., Wittrup, K.D., and Stephan, M.T. (2017b). J. Clin. Invest. *127*, 2176–2191.

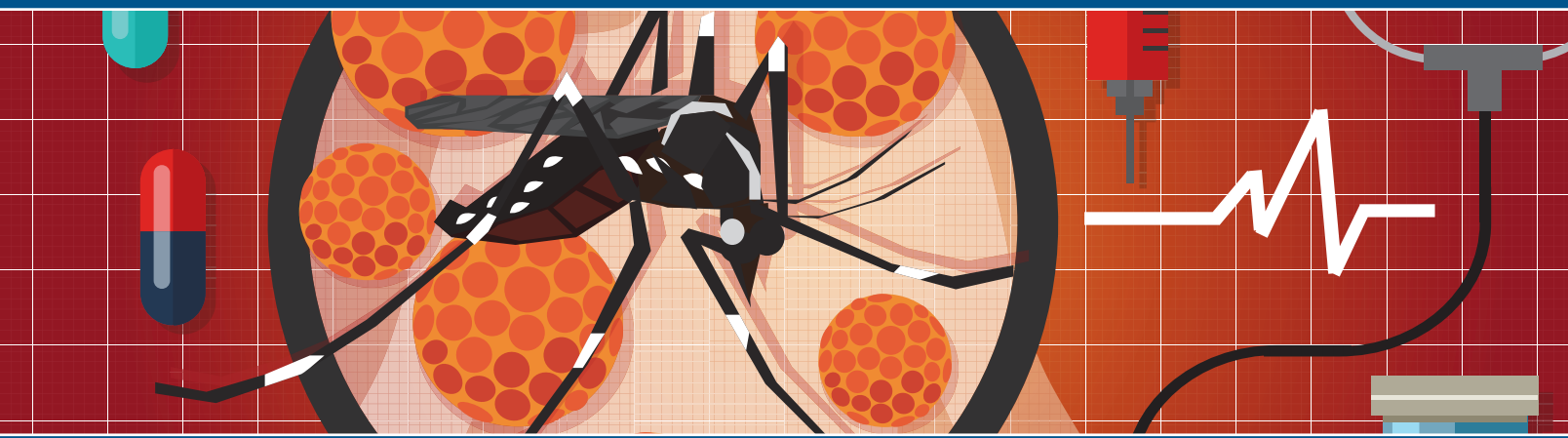
U.S. Food & Drug Administration (2017). FDA approval brings first gene therapy to the United States [Press Release]. Retrieved from **https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm574058.htm**.

Xiaohong Helena Yang

Cell Symposia

Emerging and Re-emerging Viruses

October 1–3, 2017 — Arlington, VA, USA



Extended abstract submission deadline:

July 7, 2017

Early registration deadline:

August 21, 2017

By focusing on a range of human pathogenic viruses, this symposium will center on understanding and establishing common principles to help combat current and future viral threats. It will bring infectious-disease researchers of many stripes together with scientists interested in the host response to these viruses from immunological, cell-biological, and pharmacological perspectives.

Keynote Speaker

Anthony S. Fauci, *USA*

Special Guest Speaker

Richard Preston, *Author of the best seller 'The Hot Zone'*

Speakers

Dan Barouch, *USA*
Sara Cherry, *USA*
James Crowe, *USA*
Peter Daszak, *USA*
Michael Diamond, *USA*
George Gao, *China*
Adolfo Garcia-Sastre, *USA*
Akiko Iwasaki, *USA*
Yoshihiro Kawaoka, *USA*
Sheemei Lok, *Singapore*
Luciano Moreira, *Brazil*
Johan Neyts, *Belgium*
Erica Ollmann Saphire, *USA*
Gustavo Palacios, *USA*
Oliver Pybus, *UK*
Zhengli Shi, *China*
Nancy Sullivan, *USA*
Linfa Wang, *Singapore*

Organizers

George Fu Gao, *Chinese Academy of Sciences, China*
Akiko Iwasaki, *Yale University, USA*
Lakshmi Goyal, Editor, *Cell Host & Microbe*
Sri Devi Narasimhan, Senior Scientific Editor, *Cell*

cell-symposia.com/emerging-viruses-2017

Cell

Artists Create Puzzles, Scientists Solve Them

Joseph L. Goldstein[1,2,*]
[1]Chair, Lasker Awards Jury
[2]Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
*Correspondence: joe.goldstein@utsouthwestern.edu
http://dx.doi.org/10.1016/j.cell.2017.08.022

The Spanish artist Diego Velázquez created a puzzle-painting 360 years ago that to this day remains unsolved, but still mystifies and intrigues. Unlike artists who get their thrills by creating puzzles that stimulate the imagination, scientists get their kicks by solving puzzles that advance biomedical research.

Kicks and Thrills

According to the philosopher of science Thomas Kuhn, the successful scientist is a passionate puzzle-solver who relishes the challenge to solve problems in which "the outcome remains very much in doubt." The Nobel Prize-winning physicist Richard Feynman expressed a similar view when he proclaimed that puzzle-solving is the primary motivation for becoming a scientist. To quote Feynman, "The prize is the pleasure of finding the thing out, the kick in the discovery." Feynman's "kick in the discovery" is like the thrill of having a eureka moment.

One difference between scientists and artists is that scientists like to solve puzzles by putting the pieces together, while artists get their kicks by creating the pieces. The Belgian surrealist René Magritte (1898–1967) was the consummate creator of puzzle-paintings. During his 40-year career, Magritte painted more than 1,600 visual puzzles that juxtaposed two or more ordinary objects in an unusual context that gave new meaning to familiar things. One of Magritte's puzzle-paintings, *Les Idées Claires* (*Clear Ideas),* is relevant to scientific discovery (Figure 1). Here, you see a rock over the ocean and underneath a cloud. The cloud represents our murky ideas. When these murky ideas crystallize and solidify, our whole being, like the rock, is uplifted into the sky, producing the kick and thrill of discovery.

Velázquez's *Las Meninas*: The Ultimate Puzzle-painting

Magritte may hold the record for the largest number of puzzle-paintings, but it is the 17th century Spanish artist Diego Velázquez who created the ultimate puzzle-painting—the famous *Las Meninas* (Portuguese for *The Maids of Honor*), completed in 1656 (Figure 2). Few paintings in the history of art have generated so much critical analysis with so many different interpretations. The fascination with *Las Meninas* derives from its large canvas (10.5 by 9 feet) with life-sized figures, the complex composition of its cast of characters, the seductive nature of its riddles and puzzles, and the artist's ability to create a visual world that seems palpable to the viewer (Carminati, 2012).

At first glance, the subject of *Las Meninas* seems straightforward. The setting takes place in Velázquez's studio in the royal palace in Madrid. At the center of the painting stands the young princess, the 5-year-old Infanta Margarita, surrounded by her entourage consisting of two maids of honor, two dwarfs, and a dog (Figure 2). Just behind the maid at the right is a lady-in-waiting chatting with a bodyguard. At the left, Velázquez portrays himself in front of an enormous canvas, glancing out beyond the pictorial space. On the center back wall, the Queen's chamberlain stands in a lighted doorway. To the left of the doorway is a mirror that reflects the blurred faces of the King and Queen of Spain, who are apparently standing just outside the pictorial space in the same place where museum viewers would stand.



Figure 1. *Les Idées Claires* (*Clear Ideas*)
René Magritte (1958). Oil on canvas. 20 × 24 in. Private Collection. Copyright 2017 C. Herscovici/Artists Rights Society (ARS), New York.

The Mystery of *Las Meninas*: Hidden Riddles and Puzzling Paradoxes

The mystery of *Las Meninas* that has been debated for 360 years is: "What is being painted on the hidden side of the canvas"? (Carminati, 2012). There are three leading theories. The oldest theory is that Velázquez is painting a life-sized portrait of the Infanta. In this formulation, the King and Queen are passing through the studio to watch Velázquez painting their daughter, and the Queen's



Figure 2. *Las Meninas* (*Maids of Honor*)
Diego Velázquez (1656). Oil on canvas. 10.5 × 9 ft. Prado Museum, Madrid.

chamberlain stands in the open doorway waiting to escort the royal couple back to their chambers. The faces of the King and Queen are reflected in the mirror on the back wall.

A second theory is that Velázquez is painting a portrait of the King and Queen. The Infanta and her retinue are stopping by to watch the royal parents sitting for the court painter. In this formulation advanced in 1981 by the art critic Leo Steinberg, the painted faces of the King and Queen are reflected from the canvas to the mirror (Steinberg, 1981). In 2009 two scientists, David Stork and Yasuo Furuichi, built a virtual 3-D computer graphic model of *Las Meninas* that permits the viewer to analyze the painting from any angle. Their results support the idea that the faces of the royal couple on the mirror originate from the painted faces on the hidden side of the canvas, not from the King and Queen watching Velázquez at work (Stork and Furuichi, 2009).

Other than the King and Queen, is there anything else that Velázquez is painting on the canvas? This bring us to a third theory, the most provocative of all: Velázquez is painting a picture of himself who is painting a picture of the painting that we and the King and Queen are looking at. The artist has created a deliberate ambiguity, tricking us into seeing the same picture twice. Is he telling us that art and life are both an illusion?

The art critic Blake Gopnik recently pointed out that the true subject of *Las Meninas* ''is the heroic moment of its own making, recorded 'live' in a real place, as no history painting had ever been before.'' Gopnik goes on to suggest that the title be changed to *The Immaculate Reflection* since the picture is an authentic mirror of nature itself like no other painting in the history of art (Gopnik, 2010).

Velázquez's creation of *Las Meninas* with its hidden riddles and puzzling paradoxes brings to mind the famous quip of James Joyce, who when asked to explain the success of his novel *Ulysses* responded: ''I've put in so many enigmas and puzzles that it will keep the professors busy for centuries arguing over what I meant, and that's the only way of insuring one's immortality.'' (Ellmann, 1982).

Picasso Attempts to Solve the Mystery

By the James Joyce standard, Velázquez achieved immortality. Not only has *Las Meninas* kept the professors and art historians busy for more than 350 years, it also kept Pablo Picasso busy. In 1957 at age 75, Picasso secluded himself for five months in his studio, devoting full time to a radical deconstruction of Velázquez's masterpiece. Picasso was determined to unearth its mystery by recasting *Las Meninas* in terms of his own contemporary cubist style (Rafart i Planas, 2001). As Picasso was fond of saying, ''Great artists don't copy, they steal.''

Picasso painted 46 different versions of *Las Meninas*. The entire series is on permanent display at the Picasso Museum in Barcelona. Picasso's first version of *Las Meninas* is one of his largest canvases, 6.5 by 8.5 feet, second in size only to his *Guernica*. And like *Guernica,* Picasso's first version of *Las Meninas* is painted in black and white and in a horizontal orientation (Figure 3). But the most radical change is the cubist rendition of the figures, which creates a chaotic scene, severing the perceptual connection between the figures in the painting and the viewers outside the pictorial space. We, as well as the King and Queen, are no longer right there in the room to watch the Infanta and the painter.

In his typical imaginative style, Picasso injects several humorous touches. As a tribute to Velázquez, he enlarges the height of the painter, making him a towering figure whose double-profiled face reaches the top of the ceiling (Figure 3). The chamberlain at the back

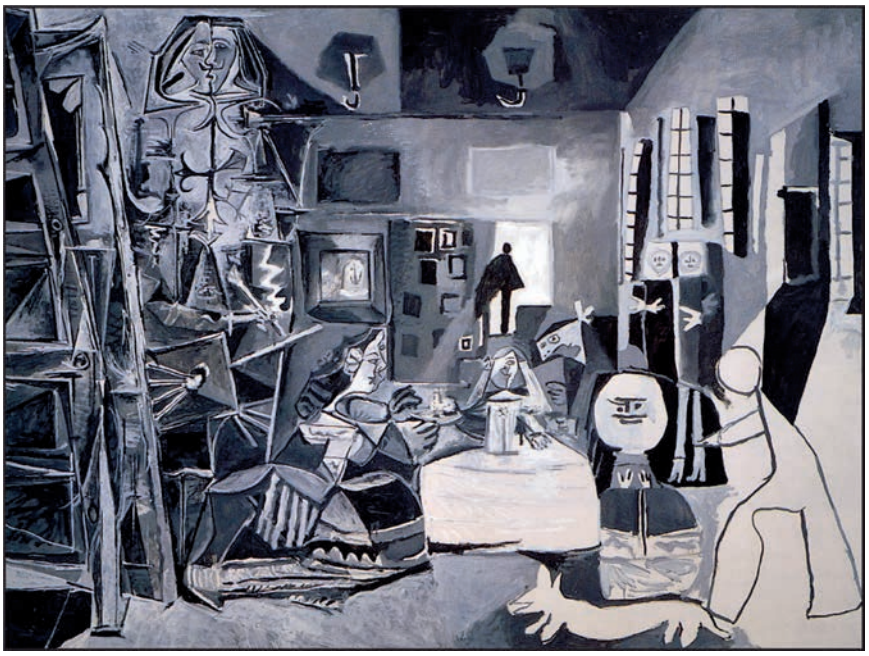


Figure 3. *Las Meninas after Velázquez*
Pablo Picasso (1957). Oil on canvas. 6.4 × 8.5 ft. Picasso Museum, Barcelona. Copyright 2017 Estate of Pablo Picasso/Artists Rights Society (ARS), New York.

of the room is given increased visibility as a shadow figure that is a symbolic representation of Picasso himself. The large dog in the original is replaced with a dwarfed dachshund, presumably to provide companionship for the dwarfed maid of honor.

Why did Picasso spend so much time on Velázquez's masterpiece? Even though he soon realized that he could not solve the mystery of what is being painted on the hidden side of the canvas, he became obsessed with showing the world how *Las Meninas* might have looked had it been painted in 1957 rather than in 1656. Picasso believed that great art will always be great no matter when it was conceived. "To me, there is no past or future in art. If a work of art cannot live always in the present, it must not be considered at all." In this sense, Picasso was subscribing to T. S. Eliot's philosophy of great poetry. According to Eliot (1985), "The perpetual task of poetry is to *make all things new* (sic). Not necessarily to make new things."

A Contemporary Puzzle Painting Inspired by *Las Meninas*

The latest artist "to make *Las Meninas* new" is the contemporary painter Kelly James Marshall. Marshall is widely acclaimed to be one of the leading African-American painters in the U.S. today. For the last 35 years, he has chronicled the African-American experience in a series of large narrative paintings that feature only black figures. The Metropolitan Museum recently sponsored a retrospective of Marshall's work, featuring 72 of his large painting done over three decades (Alteveer et al., 2016).

The highlight of this exhibit was a painting called *Untitled (Studio)*, completed in 2014. *Studio* resembles *Las Meninas* in its composition, subject matter, and large size (7 by 10 feet) with life-sized figures (Figure 4). At first glance, *Studio* appears straightforward. At the center of the painting is a model with salmon-colored trousers who sits in front of a red backdrop. A woman in green high heels is adjusting the model's head. At the left is an easel with an unfinished painting of the torso of the female model. In the background are two men, one behind the painting and the other behind the red backdrop. On the back wall are three open window panes reminiscent of the open door in *Las Meninas*. A yellow dog sits underneath the table.

On second viewing, things in *Studio* are not so straightforward. The mystery here is not what is being painted on the canvas, but where and who is the artist (Figure 4). Is it the man behind the red backdrop who is putting a yellow jacket over his arm? Or is it the shadowy, nude male figure standing behind the easel? Or is it the woman in the green high heels with the paint-smeared dress whose gaze is directed outside the pictorial space. Is she looking out at a mirror to help her adjust the model's profile? Or is she looking for confirmation from the artist who is outside the frame? Or maybe her gaze is meant for us, the viewers, who would be standing in the same place as the artist?

Marshall has created a puzzle-painting that is provocative, complex, and subject to different interpretations much like Velázquez's *Las Meninas*. *Studio* differs from *Las Meninas* in several ways. The visual illusion of depth in *Las Meninas* is missing in *Studio* where the figures and objects appear flat. Marshall's focus is not on perspective; it is on color. All the figures except for the dog are jet black, and this blackness heightens the intensity of the reds, the yellows, the blues, and the salmon color of the model's trousers. By focusing on color through his literalization of blackness, Marshall has achieved originality in the same way that Velázquez did with perspective and Picasso with cubism.

2017 Lasker Awards: Solving Puzzles that Advance Biomedical Science and Public Health

Unlike Velázquez, Picasso, and Kelly James Marshall, who created puzzles that stimulate the imagination, this year's Lasker Basic and Clinical Awardees solved puzzles that advance the biomedical sciences. This year's Public Service Awardee has advanced the health of women throughout the world for 100 years.

The **Lasker Basic Research Award** honors Michael N. Hall (University of Basel) for his discoveries concerning the nutrient-activated TOR proteins and their central role in the metabolic control of cell growth. Hall's original work, carried out in the late 1980s, began as a genetic analysis of how the immunosuppressant drug rapamycin inhibits the growth and survival of yeast cells. The result was the identification of TOR (target of rapamycin)



Figure 4. ***Untitled* (*Studio*)**
Kerry James Marshall (2014). Acrylic on PVC panel. 7 × 10 ft. Metropolitan Museum of Art, New York City.

and the two TOR complexes, TORC1 and TORC2. Hall's discovery of what at the time was a novel class of kinases laid the foundation for the current explosion in the mTOR field in which scientists soon identified the single mammalian ortholog of the two yeast TORs and elucidated the upstream and downstream mechanisms of TOR signaling. These studies revealed how mammalian cells use mTORC1 as a molecular integrator that senses multiple signals arising from growth factors, hormones, amino acids, and oxygen status. Once activated, mTORC1 regulates pathways required for cell proliferation and growth, involving the synthesis of proteins, lipids, and nucleotides. Research on TOR has also paved the way for the recent development of mTOR inhibitors to treat certain forms of cancer and has shed light on understanding mechanisms involved in the control of blood glucose and the aging of cells.

The **Lasker~DeBakey Clinical Research Award** honors two scientists, Douglas R. Lowy and John T. Schiller (National Cancer Institute, NIH), for technological advances that enabled development of HPV vaccines for prevention of cervical cancer and other tumors caused by human papillomaviruses. The identification of HPV as the infectious agent responsible for cervical cancer in the early 1980s by Harald zur Hausen (Nobel Laureate in Physiology or Medicine, 2008) provided a strong stimulus for research that would lead to an HPV prophylactic vaccine. In 1992, Lowy and Schiller published two observations that stood the test of time: (1) the L1 major structural protein of papillomaviruses (produced as a recombinant protein in insect cells) could self-assemble into non-infectious virus-like particles (VLPs) without the need for other viral proteins, and (2) such L1 VLPs possessed the ability to induce high-titer neutralizing antibodies. Moreover, in 2001 Lowy and Schiller, in collaboration with scientists at Johns Hopkins University, carried out the first phase I clinical trial of an HPV VLP vaccine containing the L1 VLP from type 16 HPV, which is the most common cause of cervical cancer. The vaccine, which was given to 36 normal volunteers, was well tolerated, proved safe, and was highly immunogenic in all subjects.

The aforementioned pioneering findings were followed by innovative contributions from scientists at Merck & Co. and GlaxoSmithKline (GSK), who developed the first commercially available prophylactic L1 HPV vaccines. These vaccines were approved by the FDA for

human use in 2006 (Merck) and 2009 (GSK). Since 2006, approximately 50 million women worldwide have been vaccinated. In the U.S.~60% of adolescent girls and ~50% of adolescent boys have received one or more doses of the three-dose HPV vaccine series.

The **Lasker~Bloomberg Public Service Award** honors Planned Parenthood for providing health services and reproductive care to millions of women. Planned Parenthood, which began as a birth control clinic in 1916, is now a vibrant international organization. In addition to providing contraceptive services, it provides examinations and testing for breast and cervical cancer, testing and treatment for sexually-transmitted diseases, pregnancy testing, and sex education and counseling.

For a more detailed account of the discoveries at this year's Lasker Awards, please refer to the Lasker website at www.laskerfoundation.org and to the articles in this issue of *Cell.*

REFERENCES

Alteveer, I., Molesworth, H., Roelstroete, D., and Winograd, A. (2016). Kerry James Marshall: Mastry (New York: Skira Rizzoli).

Carminati, M. (2012). Velazquez's Las Meninas: Art Mysteries (Milan: Ore Cultura S.R.L.).

Eliot, T.S. (1985). Tradition and the practice of poetry with an introduction and afterword by A. Walter Litz. (A typescript of the Dublin Lecture delivered by Eliot at University College, Dublin on January 24, 1936). The Southern Review *21*, 873–888.

Ellmann, R. (1982). James Joyce, Revised Edition (New York: Oxford University Press), p. 521.

Gopnik, B. (Dec. 24, 2010). The Velázquez painting 'Las Meninas': An encyclopedia of artistic greatness. Washington Post http://www.washingtonpost.com/wp-dyn/content/article/2010/12/23/AR2010122301932.html.

Rafart i Planas, C. (2001). Picasso's Las Meninas (Barcelona: Editorial Meteora).

Steinberg, L. (1981). Velázquez' "Las Meninas". October. *19*, 45–54.

Stork, D.G., and Furuichi, Y. (2009). Computer graphics synthesis for inferring artist studio practice: an application to Diego Velázquez's *Las Meninas*. Proc. SPIE *7238*, The Engineering Reality of Virtual Reality. http://dx.doi.org/10.1117/12.806264.

TOR, the Gateway to Cellular Metabolism, Cell Growth, and Disease

John Blenis[1,*]
[1]Meyer Cancer Center and Department of Pharmacology, Weill Cornell Medicine, 413 East 69th Street, New York, NY 10021, USA
*Correspondence: jblenis@med.cornell.edu
http://dx.doi.org/10.1016/j.cell.2017.08.019

Michael N. Hall is this year's recipient of the Lasker Basic Medical Research Award for the identification of the target of rapamycin, TOR. TOR is a master regulator of the cell's growth and metabolic state, and its dysregulation contributes to a variety of diseases, including diabetes, obesity, neurodegenerative disorders, aging, and cancer, making the TOR pathway an attractive therapeutic target.

In order to survive, differentiate, and grow, cells must evaluate their energy status and oxygen availability, take inventory of their surrounding nutrients, hormones, cytokines, and growth factors, and then integrate this information to decide what the cell's next step is. Amazingly, a signaling system has evolved that is capable of doing all of the above. The centerpiece of this pathway is the mTOR protein kinase.

Rapa Nui

The study of TOR originated in the 1960s with an expedition to Easter Island (known by the island inhabitants as Rapa Nui), with the goal of identifying natural products from plants and soil with possible therapeutic potential. In 1972, Suren Sehgal identified a small molecule, from a soil bacterium *Streptomyces hygroscopicus*, that he purified and initially reported to possess potent anti-fungal activity. He appropriately named it rapamycin, noting its original source and activity (Sehgal et al., 1975). However, early testing revealed that rapamycin also had potent immunosuppressive and cytostatic anti-cancer activity. Unfortunately, rapamycin did not initially receive significant interest from the pharmaceutical industry until the 1980s, when Wyeth-Ayerst supported Sehgal's efforts to further investigate rapamycin's effect on the immune system. This eventually led to its FDA approval as an immunosuppressant following kidney transplantation. However, prior to its FDA approval, how rapamycin worked remained completely unknown.

The Discovery of TOR

An important outcome of the renewed interest in rapamycin was that large quantities of rapamycin were produced and made available to the academic community. Taking advantage of its anti-fungal activity, Michael Hall and Joseph Heitman designed a genetic screen in *Saccharomyces cerevisiae* to identify resistant mutants (Heitman et al., 1991). Three major hits from the screen provided the foundation for the identification of the target of rapamycin. First, the yeast ortholog of the cellular receptor for rapamycin, FKBP12 (*FPR1*), was identified. FKPB12 was known to be a cis-trans prolyl isomerase that bound rapamycin, but Stuart Schreiber and colleagues had demonstrated that inhibition of this activity was not linked to rapamycin's biological activity (Schreiber, 1991), so it remained unclear if FKBP12 was the cellular receptor mediating rapamycin's effects. Critically, Hall's genetics showed that mutants lacking FKBP12 were completely resistant to rapamycin, demonstrating that it is the drug receptor responsible for its activity. However, this alone did not provide insight into how rapamycin worked. That was to come from the identification of two novel and related mutants identified in the screen that were named *TOR1* and *TOR2*, for targets of rapamycin. The *TOR1* and *TOR2* genes were later found to encode large kinases that resemble PI3 kinase (Kunz et al., 1993) and were the founding members of the atypical PI-kinase-like protein kinase (PIKK) family. Importantly, the TOR mutations identified in the yeast genetics screen are in a single codon that is required for FKBP12-rapamycin binding to and inhibition of TOR. The discovery of the *TOR* gene set the stage for 25-plus years of investigations into how TOR is regulated and signals, the biological consequences of its normal as well as its inappropriate regulation, and investigations into the therapeutic potential of targeting this pathway (Figure 1).

The TOR Pathway Takes Shape

Following the Hall discovery, the use of rapamycin helped scientists identify the proteins acting both up and downstream of this novel kinase. In 1992, we and others discovered that rapamycin potently inhibited activation of S6 protein kinase 1 (S6K1) by growth-factor-stimulated receptor tyrosine kinases, the Src oncogene, heterotrimeric G proteins, tumor promoter phorbol ester (activator of protein kinase C, PKC), or stress stimuli, etc. We also demonstrated that rapamycin delayed S-phase entry linking TOR to G1 cell cycle progression. Inspired by the Hall paper, we demonstrated that the rapamycin receptor, FKBP12, was required for inhibition but that S6K1 was not directly inhibited by the drug (Chou and Blenis, 1995). However, rapamycin did not inhibit activation of ERK-MAP kinase and RSK, which we had shown to be activated downstream of Ras (Wood et al., 1992).

Two years later, we found that PI3-kinase was required for growth factor- and insulin-dependent activation of S6K1 and that this activation was also potently blocked by rapamycin (Chung et al., 1994). This suggested that a mammalian

 © 2017 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

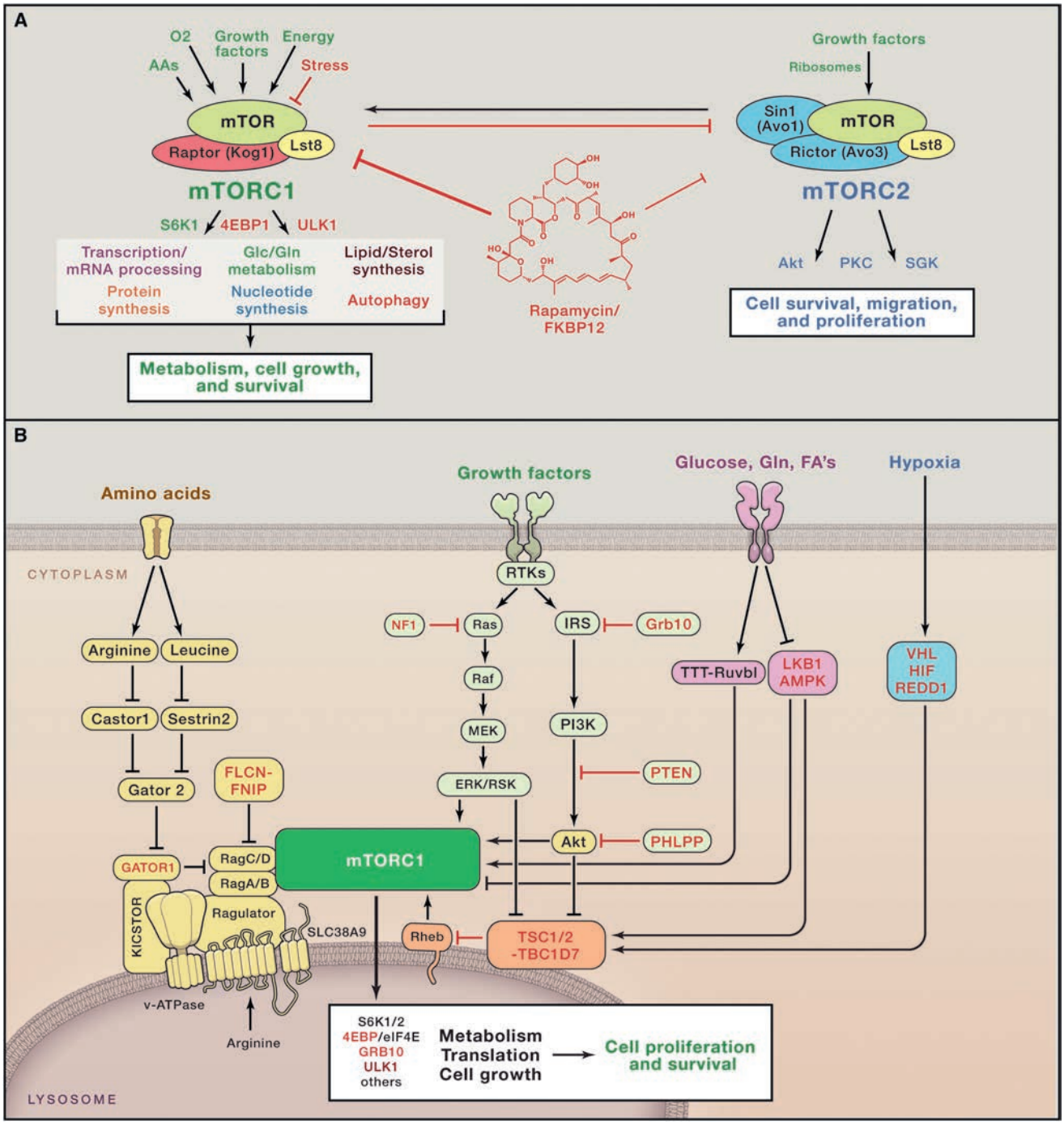


Figure 1. From Rapamycin to the mTOR Pathway and Disease

(A) The core components of mTOR complex 1 (mTORC1) and mTORC2. mTORC1 is acutely sensitive to rapamycin, whereas mTORC2 is occasionally observed to have reduced signaling upon long-term treatment. mTORC1 is regulated by multiple inputs. This information is deftly integrated to promote mTORC1-dependent regulation of a variety of growth-associated biological processes. mTORC2 is regulated downstream of PI3-kinase, and ribosomes participate in its activation.

(B) Over the past 25-plus years since the discovery of yeast TOR, the molecular details of mTORC1 regulation have been investigated. This has resulted in the discovery of multiple novel mechanisms for signal transduction, as well as revealing potential next-generation therapeutic targets and biomarkers. In red are putative or known tumor suppressors, where loss of expression/function occurs in cancer. In green are positive regulators of mTORC1 that are often activated by mutation or overexpression in cancer. mTORC1 has been estimated to be activated in 60%–80% of all cancers. For detailed description of the regulation of the pathway components see Saxton and Sabatini (2017) and Shimobayashi and Hall (2014).

TOR was mediating signaling from PI3K to S6K1. The following year, Akt/protein kinase B (PKB) was shown to be regulated downstream of PI3K and upstream of S6K1 by Coffer and Tsichlis (Chan et al., 1999). Since Akt was also not directly inhibited by rapamycin, there remained a gap between PI3K/Akt and S6K1. Based on the work of Hall and the above studies, we believed that a mammalian TOR was acting upstream of S6K1 and downstream of Akt, although how was unclear. Moreover, our studies suggested that TOR and S6K1 participated in a novel growth (G1 cell cycle) promoting pathway (Figure 1B).

Mammalian TOR

To realize the importance of Hall's discovery of yeast TOR, it was essential to demonstrate its conservation in higher eukaryotes as he had predicted. In 1994, the mammalian homolog of TOR was biochemically purified and found to be the ortholog of yeast TOR by Snyder, Schreiber, Berlin, and Abraham (Saxton and Sabatini, 2017; Shimobayashi and Hall, 2014). Although initially assigned several names, mTOR or mammalian TOR was accepted in recognition of its discovery in yeast by Hall (now also referred to as mechanistic TOR). As mentioned, rapamycin was known to suppress cell proliferation, and once discovered, mTOR was linked to cell cycle progression by regulating cell growth, a process where cells essentially double their cellular material as a result of dramatic and coordinated changes is cellular metabolism before proceeding to cell division (proliferation). The Hall lab suggested a link between yeast TOR and cell growth and also provided evidence that this was related to the ability of TOR to regulate protein synthesis (Barbet et al., 1996). mTOR was then shown to phosphorylate and inhibit the repressor of CAP-dependent translation, 4EBP1, and to phosphorylate and contribute to the activation of S6K1 (Burnett et al., 1998), major players in translational control.

mTORC1 and mTORC2

What was to follow in the early 2000s were several elegant studies utilizing biochemistry, cell biology, and genetics, predominately by the Hall and Sabatini labs, to demonstrate that mTOR existed in two highly conserved, large molecular complexes, termed mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2) (Figure 1A). mTORC1 consists of mTOR and the core components Raptor (Kog1) and mLst8 (Lst8), and mTORC2 consists of mTOR and Rictor (Avo3), Sin1 (Avo1), and mLst8 (Lst8). Rapamycin:FKBP12 binds to and acutely inhibits mTORC1, but not mTORC2. Interestingly, long-term treatment with rapamycin can also suppress mTORC2 in some cell types, and this is possibly due to sequestration of mTOR in the inhibited mTORC1 complex and its decreased availability for incorporation into new mTORC2 complexes. Additional unique components for these complexes have also been identified that likely contribute to specific functions in yeast versus mammals (Saxton and Sabatini, 2017; Shimobayashi and Hall, 2014). Initially, the main targets for mTORC1 were S6K1 (Sch9) and 4EBP1, and for mTORC2, the main targets were AGC kinases Akt, SGK (Ypk1), and PKC. We now know that there are many potential mTORC1 and mTORC2 targets based largely on the recent mass-spectrometry-generated phosphoproteomes from the Hall lab, the Sabatini lab, and my lab (Shimobayashi and Hall, 2014). These datasets are likely to reveal new functions of mTORC1 and mTORC2. Indeed, several of these targets have led to new insights into how mTORC1, in particular, is linked to negative feedback loops, metabolism, transcription, mRNA processing, and translation.

During this time, the convergence of human genetics associated with the rare autosomal dominant disorder tuberous sclerosis complex, combined with genetic studies in mouse, yeast, *Drosophila*, *C. elegans*, mammalian cell biology, and biochemistry revealed the complex relationship between the TSC1 (hamartin) and TSC2 (tuberin) tumor suppressor complex, its GAP activity toward the Rheb GTPase, and the ability of Rheb to activate mTORC1 and promote cell growth. Another key finding during this time was the demonstration that Akt directly phosphorylated TSC2, antagonizing its ability to suppress mTORC1 signaling and providing the link between mitogen signaling and mTORC1 (Saxton and Sabatini, 2017; Shimobayashi and Hall, 2014). Additional insights into mTORC1 selective sensitivity to FKBP12-rapamycin, its regulation, and substrate specificity came from structural analysis of mTORC1 from the labs of Drs. Sabatini, Pavletich, Hall, and colleagues (Saxton and Sabatini, 2017). The activation of mTORC2 has remained more enigmatic; however, recent studies suggest a role for ribosomes and PI3K signaling (Liu et al., 2015; Zinzalla et al., 2011).

A Signaling Platform for mTORC1

The importance of spatial and temporal regulation of several signaling pathways, such as the Ras/ERK-MAP kinase cascade, had been previously established, raising the question of whether the mTORC1 pathway was similarly regulated. Since the Rheb GTPase is prenylated, one guess was that it was regulated at intracellular membranes, similar to other Ras and Rho family GTPases. We now know, from the work of Sabatini, Kun-Liang Guan, and others, that the lysosome is the major intracellular membrane location for activation of mTORC1. Furthermore, as demonstrated by Sabatini and others, the lysosome serves as the landing pad for the complex amino-acid-dependent Rag GTPase activation, resulting in the recruitment of mTORC1, where it is activated by the lysosome-associated Rheb GTPase (Saxton and Sabatini, 2017; Shimobayashi and Hall, 2014). Critically, the TSC1-TSC2-TBC1D7 tumor suppressor GAP complex, which inactivates Rheb, dissociates from the lysosome following Akt-dependent phosphorylation of TSC2, allowing Rheb to assume a GTP-bound state and activate mTORC1 (Manning and Toker, 2017). Furthermore, we are still learning how the combination of amino acids, nutrients, energy and oxygen status, stress, mitogens, and other inputs are sensed by specific mechanisms and coordinated to exquisitely regulate mTORC1 signaling (Figure 1B).

Beyond Basic Research: mTOR in Physiology and Disease

Due to its complex regulation, it was anticipated that mTOR would have a significant role in controlling metabolic homeostasis at the organismal level. This has been confirmed through the use

of rapamycin in preclinical and clinical investigations and by the generation of a variety of tissue-specific mouse genetic models, where mTOR or its regulatory subunits have been manipulated. These exciting investigations are on-going, particularly in the areas of development, diabetes, obesity, neurodegenerative diseases, aging, and cancer (Saxton and Sabatini, 2017; Shimobayashi and Hall, 2014).

With regards to cancer, as illustrated in Figure 1, many proto-oncogenes and tumor suppressors participate in mTOR signaling. Thus, activating mutations, overexpression, or loss of function or expression of many positive or negative regulators of the mTOR pathway result in inappropriate activation of mTORC1 signaling. These include several growth factor receptors, cytoplasmic tyrosine kinases, PI3-kinase, Akt, mTOR, folliculin-FNIP, PTEN, LKB1, TSC1, TSC2, etc. Indeed, it is estimated that between 60%–80% of all cancers have activated mTORC1 signaling. In the last 10 years, rapalogs have received FDA approval for renal cell carcinoma, neuroendocrine carcinomas, angiomyolipomas associated with tuberous sclerosis complex and lymphangioleiomyomatosis, advanced breast cancer, and some rare children's brain cancers, with other cancers being investigated in monotherapy and combination therapy approaches and investigations into various dosing options. Furthermore, preclinical and clinical validation of additional pathway inhibitors is being pursued in academia and the pharmaceutical industry. As happened with rapamycin, the availability of small molecule inhibitors to these pathway participants will hasten the characterization of these candidate drugs and identify potential side-effects, feedback, or compensatory mechanisms. Such thorough studies are needed for the continued evolution of drug development and therapeutic strategies. Rapamycin is an on-going poster child for such an effort.

By identifying the rapamycin target, mTOR, and playing a major role in defining its function in cell signaling and disease, the legacy of Michael Hall's original discovery is being further defined as mTOR inhibitors and inhibitors of other pathway components are winding their way through preclinical and clinical trials for cancer, diabetes, obesity, heart disease, neurodegenerative disorders, and aging. We are already witnessing some limited success in cancer with rapalogs, but additional progress will continue to be realized as the mechanisms of feedback control, metabolic dependencies, compensatory mechanisms, and combination therapies are identified and incorporated into therapeutic strategies.

ACKNOWLEDGMENTS

Michael Hall's TOR discovery has provided the nucleus upon which many scientists have built upon. I therefore apologize that I could not reference important studies that have contributed to Figure 1. I would like to thank members of my lab—Ana Gomes, Yasir Ibrahim, Didem Ilter, Gina Lee, Anders Mutvei, Michal Nagiec, and Bethany Schaffer—for their comments and help with the review and figure.

REFERENCES

Barbet, N.C., Schneider, U., Helliwell, S.B., Stansfield, I., Tuite, M.F., and Hall, M.N. (1996). Mol. Biol. Cell *7*, 25–42.

Burnett, P.E., Barrow, R.K., Cohen, N.A., Snyder, S.H., and Sabatini, D.M. (1998). Proc. Natl. Acad. Sci. USA *95*, 1432–1437.

Chan, T.O., Rittenhouse, S.E., and Tsichlis, P.N. (1999). Annu. Rev. Biochem. *68*, 965–1014.

Chou, M.M., and Blenis, J. (1995). Curr. Opin. Cell Biol. *7*, 806–814.

Chung, J., Grammer, T.C., Lemon, K.P., Kazlauskas, A., and Blenis, J. (1994). Nature *370*, 71–75.

Heitman, J., Movva, N.R., and Hall, M.N. (1991). Science *253*, 905–909.

Kunz, J., Henriquez, R., Schneider, U., Deuter-Reinhard, M., Movva, N.R., and Hall, M.N. (1993). Cell *73*, 585–596.

Liu, P., Gan, W., Chin, Y.R., Ogura, K., Guo, J., Zhang, J., Wang, B., Blenis, J., Cantley, L.C., Toker, A., et al. (2015). Cancer Discov. *5*, 1194–1209.

Manning, B.D., and Toker, A. (2017). Cell *169*, 381–405.

Saxton, R.A., and Sabatini, D.M. (2017). Cell *169*, 361–371.

Schreiber, S.L. (1991). Science *251*, 283–287.

Sehgal, S.N., Baker, H., and Vézina, C. (1975). J. Antibiot. *28*, 727–732.

Shimobayashi, M., and Hall, M.N. (2014). Nat. Rev. Mol. Cell Biol. *15*, 155–162.

Wood, K.W., Sarnecki, C., Roberts, T.M., and Blenis, J. (1992). Cell *68*, 1041–1050.

Zinzalla, V., Stracka, D., Oppliger, W., and Hall, M.N. (2011). Cell *144*, 757–768.

A Prize for Cancer Prevention

Harold Varmus[1,*]
[1]Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA
*Correspondence: varmus@med.cornell.edu
http://dx.doi.org/10.1016/j.cell.2017.08.020

This year's Lasker-DeBakey Prize for Clinical Research to Douglas Lowy and John Schiller celebrates the science behind one of the greatest advances in the history of cancer research: the development of vaccines that prevent infection and thus prevent tumor induction by pathogenic strains of human papilloma virus (HPV).

Before describing the science that produced HPV vaccines, it is useful to contemplate the sometimes-paradoxical relationship between cancer and prevention.

Prevention: The Backseat Driver of Improved Cancer Statistics

Cancer-related journalism is dominated today by advances in cancer treatment, especially immunotherapies and drug therapies based on mutant cancer genes. Still, prevention strategies are responsible for a major portion of the recent steady decline, about 1.5% per year, in the overall death rate for cancers in the US (Jemal et al., 2017). Reduced use of tobacco accounts for much of that decline; the incidence rates for lung cancers, the major cause of cancer mortality here and globally, have been falling for males and now females for several years, with a delayed relationship to smoking practices. Avoidance of occupational and environmental exposures to asbestos, UV and X-irradiation, and other recognized carcinogens have also contributed to the declines. And detection of certain kinds of pre-cancerous lesions and early cancers—especially cervical (by Pap smears), colorectal (by fecal blood tests and endoscopies), melanoma (by skin exams), and breast (by mammography)—have prevented deaths from those cancers, too.

Cancer prevention is nevertheless underappreciated; we will always hear more pleas for cures than for prevention. Preventing cancer does not produce survivors who know the bullet they've dodged, so we lack grateful patients whose possible tumors have been prevented. The protected individuals and their families are not making donations to cancer research or marching for better prevention. Furthermore, for many kinds of cancers, no clear strategy for prevention can yet be envisioned, in part because we have learned that cancers commonly arise from our inherently mutation-prone machinery for DNA replication, DNA repair, and cell division.

In contrast, prevention has always dominated thinking about the control of infectious diseases, especially acute viral infections for which vaccines are the first port of call. But even in those not-so-rare situations in which infectious agents—viruses, bacteria, parasites—contribute to carcinogenesis, establishing a causal connection between cancers and microbes, especially those as prevalent as HPV, Epstein-Barr virus, *Helicobacter,* or schistosomes, can be difficult; only a subset of infected individuals may develop a cancer, and the latency is likely to be many years in duration.

But when the connection to an infectious cause is made, the result can be powerful, especially if there is a route to an effective vaccine. This was first demonstrated by the now-universally used vaccine against hepatitis B virus (HBV)—cheap, effective, and non-toxic—that has already reduced the incidence of HBV-associated hepatoma, formerly one of the most common lethal cancers worldwide. And now, we have HPV vaccines with the potential to reduce the still-high incidence of cervical and several other potentially lethal types of cancer (see Figure 1A) associated with infection by certain strains of HPV.

The Strange History and Properties of Papilloma Viruses

The remarkable potency and effectiveness of the HPV vaccines being celebrated by this year's clinical award seem all the more extraordinary in view of some of the unusual features of papillomaviruses: the many genetically and antigenically distinct strains of HPVs, now numbering more than 100, with varying carcinogenic potency; transmission through mucosal surfaces, a route of infection potentially refractory to immune protection; and the failure to propagate HPVs in conventional cell cultures, usually a requirement for making viral vaccines.

This virus class was first studied long ago by three giants in tumor virology: Peyton Rous (famed for his discovery of the iconic Rous sarcoma virus), Richard Shope, and J.W. Beard. A filterable factor in extracts from benign papillomas (warts) found in cottontail rabbits induced papillomas upon injection into naive rabbits; the warts sometimes turned into squamous carcinomas, especially when exposed to chemical carcinogens (Rogers and Rous 1951). Particles found in human, bovine, rabbit, and other papillomas appeared to be essentially indistinguishable symmetrical particles, about 50–60 nm in diameter and composed of 72 pentamers of the major capsid protein (Baker et al., 1991). These papilloma viruses, including the most commonly studied bovine and human versions (BPV and HPV), contain circular, double-stranded DNA genomes of about 8,000 base pairs. We now know that the papillomavirus genomes are organized in a stereotypic manner, encoding a few "early" proteins required for DNA replication and at least two "late" proteins, L1 and L2, that assemble to form the coat of the virus particles.

Epidemiological observations had long hinted that carcinoma of the uterine cervix might be a sexually transmitted disease. For those of us old enough to remember

 © 2017 The Author. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

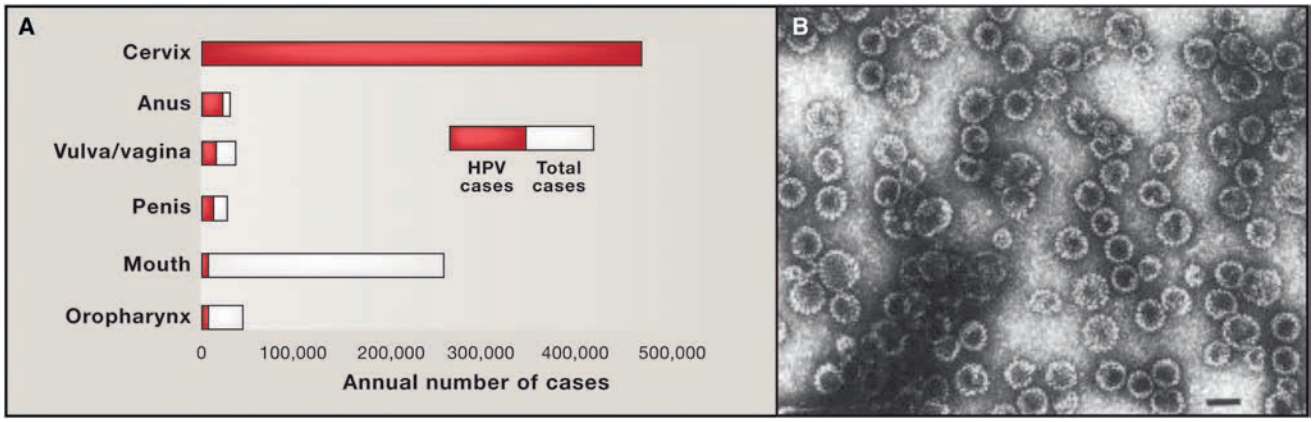


Figure 1. Preventing Cancer Associated with Human Papilloma Virus Infection
(A) The global incidences of human cancers associated with human papilloma virus (HPV) (horizontal bars); the red coloring denotes the number of cases directly attributable to infection with oncogenic strains of HPV. Adapted from Parkin, 2006.
(B) Electron microscopic view of virus-like particles composed of L1 protein from HPV-16. Micrograph from Kirnbauer et al., 1993.

the uncertainties about its possible infectious cause (for many years, a herpes virus was viewed as the leading contender), Harold zur Hausen's demonstration that many cervical cancers contain DNA belonging to one of the many types of HPV was a bombshell (Dürst et al., 1983). The likely conclusion—that more than half of the tumors were caused by a single strain of HPV, type 16—was reached despite features of HPV that have continued to complicate the study of these viruses.

HPVs have never been efficiently propagated in culture, so there has been no ready means to classify them with traditional serological methods. But because their relatively small DNA genomes differ substantially, they could be grouped, well before DNA sequencing became routine, by DNA hybridization. Thus, zur Hausen's finding, that one type of HPV was commonly found in cervical carcinomas, was a powerful indicator of a requirement for infection by select strains to produce cervical cancer. As discussed later, variations in the regional prevalence and carcinogenic potency of the many types of HPV continue to influence epidemiological and prevention strategies.

Despite zur Hausen's compelling evidence for a causative role of HPV in cervical cancer, it was not apparent how to put that information to use for patient benefit. In an era in which large armies of investigators were trying to understand oncogenes found in RNA and DNA tumor viruses—especially the retroviruses, polyomaviruses, and adenoviruses—it was natural for papilloma virologists to seek and study the oncogenic loci in HPV and other papillomavirus genomes. Using cell-based assays for oncogenesis after DNA transfer, viral genes—mainly the "early" genes E6 and E7—were implicated in transformation. Strikingly, these two genes were found to do what other DNA tumor virus oncogenes do: interfere with the actions of the now-well-known mammalian tumor suppressor genes, *P53* (Scheffner et al., 1990) and *Retinoblastoma-1* (Dyson et al., 1989). These findings were fascinating, but they did not provide obvious avenues to prevent or treat cervical cancer.

Envisioning a Vaccine Based on Virus-like Particles

John Schiller and Doug Lowy, working on papilloma viruses (mainly BPV, but also HPV) in the National Cancer Institute's intramural program, had also devoted most of their efforts to papilloma virus oncogenes (Schiller and Lowy, 2011). But because BPV could be studied in cell culture (by transforming rodent cell lines with infectious virus, not just by DNA transfection), Lowy and Schiller were positioned to perform quantitative assays, measure neutralizing antisera, and think about the structural attributes of papillomavirus proteins or particles that induce a protective immune response (Schiller and Lowy 2011).

In turning their attention to prevention of virus infection, they recognized that a traditional viral vaccine—live, attenuated, or killed—would be impractical or unethical (Schiller and Lowy 2011); HPV could not be grown in culture to make a conventional vaccine, and a virus particle containing oncogenes was not, in any case, likely be medically acceptable. So, they considered the option of making virus-like particles (VLPs) composed solely of papillomavirus protein, lacking viral nucleic acid. There were important precedents: Robert Garcea's group had made such particles by self-assembly of coat proteins from the distantly related polyomaviruses (Salunke et al., 1986), and an effective, widely used vaccine against HBV was composed of VLPs containing viral surface antigen produced in yeast (Valenzuela et al., 1982). There were also reasons to believe that properly assembled capsid proteins would be more antigenic and more likely to induce neutralizing antibodies than would individual viral proteins in solution.

These initial ideas about a papillomavirus vaccine were confirmed, using BPV as an effective model, in a remarkable paper published by the NCI group in 1992 (Kirnbauer et al., 1992). Expression of only the major BPV virion protein, L1, made in insect cells infected by a baculovirus vector, produced abundant, uniform, correctly sized particles; after injection into rabbits, those VLPs induced high titers of neutralizing antibodies that protected cultured cells from infectious BPV.

Of course, as beautiful as this was, the results did not ensure that oncogenic

strains of HPV, especially type 16, would behave in the same way, nor did they predict whether the immune response would protect against infection of mucosal surfaces or whether this approach would be amenable to scaled-up production or commercial viability.

Indeed, the first issue proved initially problematic: efforts to reproduce the BPV-based findings with the commonly used HPV-16 L1 DNA clone from a cervical cancer produced few and inappropriately sized virus-like particles, similar to earlier reports from an Australian team that had also included the L2 protein (Zhou et al., 1991). Lowy and Schiller were skeptical of this discrepancy, in part because a member of their laboratory group was able to make abundant VLPs by expressing an L1 gene from a rhesus monkey papilloma virus closely related to HPV-16 (Schiller and Lowy, 2011). When the original HPV-16 L1 clone was sequenced and compared with the L1 gene from other HPV-16 genomes obtained from non-malignant lesions, a single difference (Asp202His) was noted in the predicted L1 protein sequence. Use of the putative wild-type clones of HPV-16 L1 reassuringly restored production of VLPs to levels comparable to those observed with BPV (Kirnbauer et al., 1993).

Reduction to Practice: Making, Testing, and Improving HPV Vaccines

Despite these promising findings and the magnitude of the problem that cervical cancer poses to human health, it was not initially easy to find commercial partners to take on the scientifically difficult and commercially risky task of making a viable vaccine against a sexually transmitted pathogen. Schiller and Lowy (2011) have recounted a fateful meeting with the vaccine pioneer at Merck, Maurice Hilleman, who immediately embraced the concept and convinced the company to proceed with HPV vaccine development. Shortly thereafter, other companies, notably MedImmune and GlaxoSmithKline (GSK), were also able to take up the challenge, thanks to non-exclusive licensing practices at the NIH.

Many choices are required for the development of any vaccine. An HPV vaccine that was dependent on VLPs composed only of L1 protein, however simple in concept, was no exception. In this case, the choices included the following: the expression system for production of L1 and VLPs (Merck shifted to bakers yeast, *S. cerevisiae* [Mach et al., 2006]); the HPV strains to be included in the vaccine (another highly oncogenic strain, type 18, was used by both Merck and GSK, and Merck included two strains commonly found in genital warts, types 6 and 11); the recipe for presentation of the VLPs (Merck scientists reassembled VLPs in vitro from L1 pentamers [Mach et al., 2006]); the inoculation schedule (initially three doses, months apart); and the design of efficient clinical trials.

Safety trials quickly revealed that HPV VLPs were well tolerated and induced high titers of neutralizing antibodies in human subjects (Harro et al., 2001), but the choice of appropriate endpoints for large-scale, controlled efficacy trials was particularly vexing, involving many sectors: scientific, regulatory, and commercial. Since the major purpose of the vaccines was protection against cancer, looking simply for reduction in the frequency of infection by relevant types of HPV might not have been a reliable indicator of success (only a subset of infected women develop cervical cancer, and many infections regress naturally). But waiting to observe a reduction in cancer incidence would not only be slow, it would be ethically unacceptable; standard care dictates ablation of any pre-malignant lesions detected by regular Pap smears. As a compromise, it was generally agreed to follow two metrics: presence of viral DNA by HPV strain-specific polymerase chain reaction (PCR) assays and, more importantly, the appearance of intermediate or high-grade cervical intraepithelial neoplasias (CIN2 and CIN3).

The trials conducted with the two major HPV vaccines (Merck's quadrivalent and GSK's bivalent vaccine) in various patient populations have been uniformly, indeed dramatically, successful (reviewed by Schiller and Lowy, 2011, and Schiller et al., 2012) and have led to widespread licensure and use.

But the story does not end here: HPV vaccination has yet to fulfill its potential. Some of the remaining problems are inherent in the scientific plan; others have social origins and became quickly apparent.

Incomplete Protection

Even successfully vaccinated women remain at risk of disease caused by strains less commonly implicated in the causation of cervical cancer; HPV16 and HPV18 are together responsible for only about 70% of cervical cancers in the U.S. and for lower percentages in some other places. Therefore, it remains necessary to advise vaccinees, both here and abroad, to continue surveillance for early lesions. A nonavalent Merck vaccine that includes VLPs from seven oncogenic strains offers 90% protection and was recently approved (Petrosky et al., 2015), but this does not eliminate the dilemma. Designing a truly universal HPV vaccine (or at least one that protects against infection by all known disease-producing strains) continues to be a valuable quest.

Neglecting Males

Since HPV vaccines were developed principally to prevent cervical cancer, by far the most prevalent of the HPV-associated cancers (see Figure 1), public health campaigns and financial supporters have emphasized the vaccination of girls. But immunization of boys would restrict the frequency of virus transmission and enhance herd immunity. Furthermore, it would protect them from lethal cancers, including the HPV16-initiated oropharyngeal cancers that are rising in frequency in some places, including in the U.S. (Pytynia et al., 2014).

Inadequate Use

Rates of vaccination of the primary target population, adolescent females, have been disappointing, even in many advanced economies, including the U.S., and the vaccines are generally unused or rarely used in many developing countries. The failure to make universal use of a method that could prevent lethal cancers in hundreds of thousands of people worldwide each year may seem surprising, but many factors are at work: the cost (initially about $360 for a full three-injection course in the U.S.); the recommended multi-dose vaccination protocol, initially requiring three health care visits, independent of other vaccinations; inadequate education of parents and health care workers about the benefits of the vaccine; social resistance to a vaccine predicated on the assumption that 9- to 11-year-old

recipients will soon have sexual relationships; and more general opposition to immunization, based on fears of undocumented toxicities.

Efforts have been made to surmount these barriers with financial support for purchase of vaccines (e.g., through the Global Alliance for Vaccines and Immunizations [http://www.gavi.org/support/nvs/human-papillomavirus/]), shortening of the immunization schedule from three doses to two (https://www.cdc.gov/media/releases/2016/p1020-hpv-shots.html), and reports from authoritative groups, like the President's Cancer Panel (https://deainfo.nci.nih.gov/advisory/pcp/annualreports/hpv/index.htm), but opportunities to save lives are lost every day.

A Final Word about Immunization and Cancer Prevention

This Lasker Prize is more than just a reward to two individuals for good science and for the design of a vaccine that prevents cancers and saves lives. It is also a prestigious endorsement of a global community of scientists and health care professionals attempting to harness the immune system to prevent infection and disease. This validation is sorely needed at a time when a grass-roots anti-vaccine movement and a powerful handful of national leaders are undermining confidence in immunization, placing the public at risk of contracting dangerous but fundamentally preventable infections.

It is impossible to identify the individuals who are the beneficiaries of any prevention strategy, including vaccination. But it is easy to recognize those who suffer the consequences of failing to use the vaccines that human ingenuity has produced. If this year's prize prompts greater attention to the value of vaccines, the Lasker Foundation will have done much more than reward two ingenious scientists.

ACKNOWLEDGMENTS

Readers should know that Doug Lowy served as Deputy Director of the National Cancer Institute during my tenure as Director (2010–2015). I am grateful to Joe Goldstein (University of Texas Southwestern) for guidance about the composition of this essay and Titia de Lange (Rockefeller University) for helpful suggestions.

REFERENCES

Baker, T.S., Newcomb, W.W., Olson, N.H., Cowsert, L.M., Olson, C., and Brown, J.C. (1991). Biophys. J. *60*, 1445–1456.

Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. (1983). Proc. Natl. Acad. Sci. USA *80*, 3812–3815.

Dyson, N., Howley, P.M., Münger, K., and Harlow, E. (1989). Science *243*, 934–937.

Harro, C.D., Pang, Y.Y., Roden, R.B., Hildesheim, A., Wang, Z., Reynolds, M.J., Mast, T.C., Robinson, R., Murphy, B.R., Karron, R.A., et al. (2001). J. Natl. Cancer Inst. *93*, 284–292.

Jemal, A., Ward, E.M., Johnson, C.J., Cronin, K.A., Ma, J., Ryerson, A.B., Mariotto, A., Lake, A.J., Wilson, R., et al. (2017). J. Nat. Cancer Inst. *109* http://dx.doi.org/10.1093/jnci/djx030.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D.R., and Schiller, J.T. (1992). Proc. Natl. Acad. Sci. USA *89*, 12180–12184.

Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Dürst, M., Gissmann, L., Lowy, D.R., and Schiller, J.T. (1993). J. Virol. *67*, 6929–6936.

Mach, H., Volkin, D.B., Troutman, R.D., Wang, B., Luo, Z., Jansen, K.U., and Shi, L. (2006). J. Pharm. Sci. *95*, 2195–2206.

Parkin, D.M. (2006). Int. J. Cancer *118*, 3030–3044.

Petrosky, E., Bocchini, J.A., Jr., Hariri, S., Chesson, H., Curtis, C.R., Saraiya, M., Unger, E.R., and Markowitz, L.E.; Centers for Disease Control and Prevention (CDC) (2015). MMWR Morb. Mortal. Wkly. Rep. *64*, 300–304.

Pytynia, K.B., Dahlstrom, K.R., and Sturgis, E.M. (2014). Oral Oncol. *50*, 380–386.

Rogers, S., and Rous, P. (1951). J. Exp. Med. *93*, 459–488.

Salunke, D.M., Caspar, D.L., and Garcea, R.L. (1986). Cell *46*, 895–904.

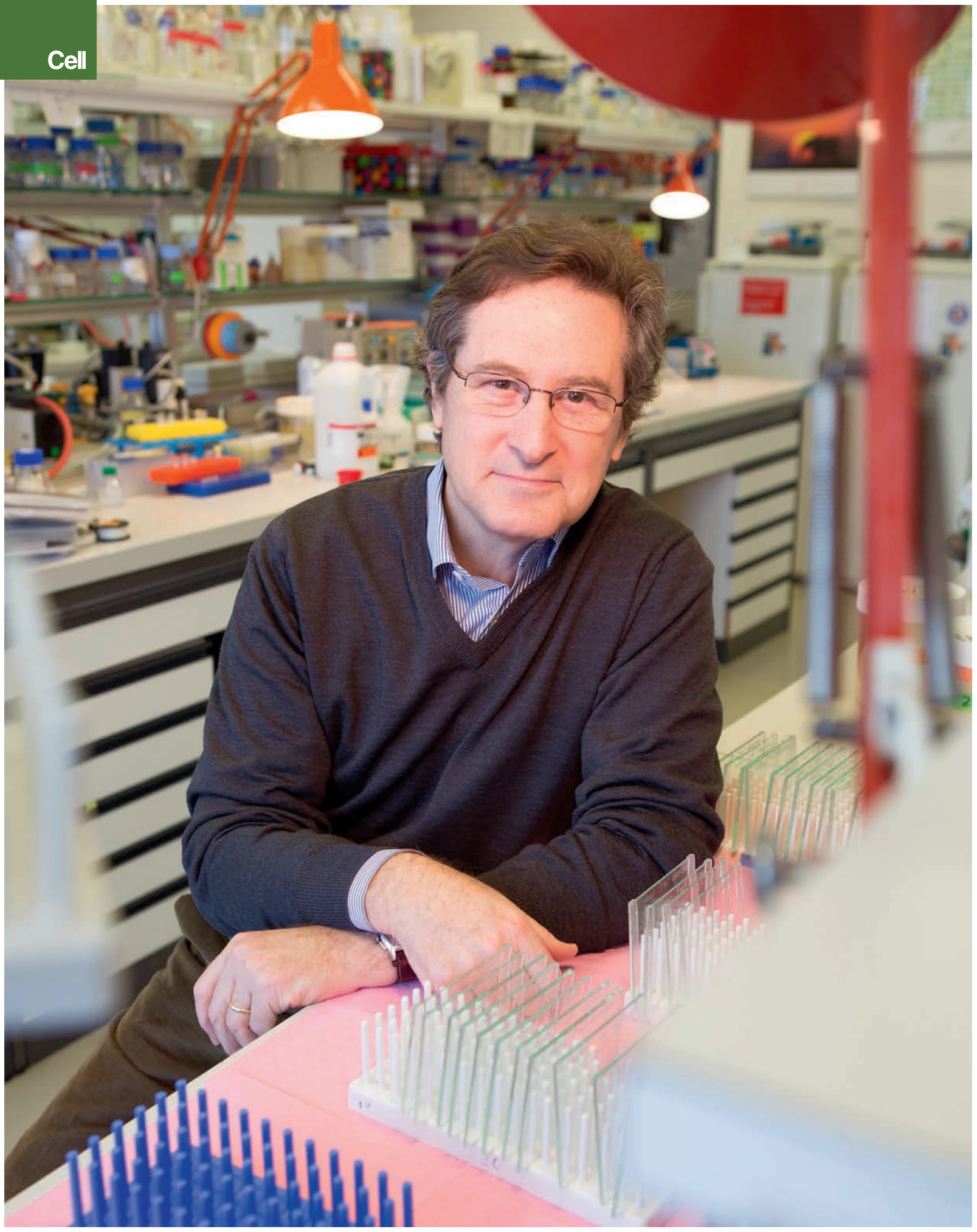
Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. (1990). Cell *63*, 1129–1136.

Schiller, J.T., and Lowy, D.R. (2011). History of Vaccine Development, S.A. Plotkin, ed. (Springer), pp. 265–284.

Schiller, J.T., Castellsagué, X., and Garland, S.M. (2012). Vaccine *30* (*Suppl 5*), F123–F138.

Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G., and Hall, B.D. (1982). Nature *298*, 347–350.

Zhou, J., Sun, X.Y., Stenzel, D.J., and Frazer, I.H. (1991). Virology *185*, 251–257.



Cell

An Amazing Turn of Events

Michael N. Hall

In 1965, a group of scientists from Montreal arrived on Easter Island, also known as Rapa Nui, to collect soil samples. This was the inauspicious beginning of an extraordinary and unpredictable series of events that make up a wonderful biomedical story. Distinct achievements and discoveries that shaped this story, in which I had the good fortune to be a protagonist, were recognized with a Lasker Award in 2012 and now in 2017.

"I liked the name TOR because it was easy to say and remember. Joe had a more romantic reason for favoring it."

The purpose of the scientists' journey to a remote island in the South Pacific was to prospect for exotic microbes that might produce novel metabolites that could be developed into drugs—in this case, antifungal drugs. This is how drugs were developed in those days. In 1975, they did indeed isolate such a metabolite, which they named rapamycin, after Rapa Nui. However, while developing rapamycin as an antifungal, they found that it had the undesirable side effect of suppressing the immune system. Rapamycin was therefore abandoned and largely forgotten until it caught the attention of two bold transplant surgeons. In the 1980s and 1990s, Roy Calne and Thomas E. Starzl exploited the formerly undesirable immunosuppressive properties of rapamycin and two other natural compounds (cyclosporin A and FK506) to advance transplantation from an experimental procedure to an accepted treatment for organ failure. For establishing the clinical utility of cyclosporin A, FK506, and rapamycin as immunosuppressive drugs, Calne and Starzl shared the 2012 Lasker Clinical Medical Research Award. The citation reads, "For the development of liver transplantation, which has restored normal life to thousands of patients with end-stage liver disease."

I had the pleasure of seeing Tom Starzl again a couple years ago, after a long hiatus, and we were soon reminiscing about the old days. At 89 years of age, his enthusiasm for science was undiminished. Shortly thereafter, I received a letter from Tom with reprints of some of his rapamycin papers from the early '90s. He ended the letter by writing, "Have these 2 papers passed or flunked the test of time?" Needless-to-say, they had passed.

Calne and Starzl's achievements were heroic and profoundly important, but the underlying mode(s) of action of the three drugs was unknown other than that they inhibited signaling in T cells. In 1989, Joe Heitman, an outstanding postdoc, joined my laboratory at the Biozentrum of the University of Basel after his M.D. and Ph.D. studies at Cornell and The Rockefeller University. Given the medical breakthrough made possible by the new immunosuppressive drugs, and thus the heightened sense of excitement surrounding them, he decided to study the drugs' molecular mechanism of action. This was a departure from our ongoing research, not least because we were primarily a yeast genetics laboratory, and Joe wanted to study drugs used, or being developed for use, in humans. Joe and Rao Movva, a collaborator and group leader at the local pharmaceutical company Sandoz (now Novartis), had the brilliant albeit risky idea to use yeast genetics to study the drugs. They were, of course, making the assumption that the drugs' molecular targets were conserved from yeast to human. I could justify this departure from our normal research by viewing the compounds as probes to identify novel signaling pathways. Nevertheless, it seemed to many that our experiments were physiologically irrelevant—giving human drugs to yeast cells?! Colleagues were rolling their eyes, and rightfully so.

Joe's initial experiments were with cyclosporin A and FK506. Although these two drugs had little to no effect on growth of our laboratory yeast strain, Joe characterized the yeast FK506-binding protein (FKBP) encoded by the gene he named *FPR1*. Rao later gave us some rapamycin, which Joe found arrested yeast cells in the G1 phase of the cell cycle—the same effect the drug had on T cells! This gave us confidence that our risky idea might have merit. Moreover, Joe found that FK506 antagonized rapamycin toxicity (structurally similar, FK506 and rapamycin competitively bind FKBP) and that strains in which he had deleted the *FPR1* gene were fully resistant to rapamycin. This suggested that an FKBP-rapamycin complex was the active, toxic agent in the cell, rather than rapamycin alone. To identify the target of the FKBP-rapamycin complex, Joe

Michael N. Hall in his lab



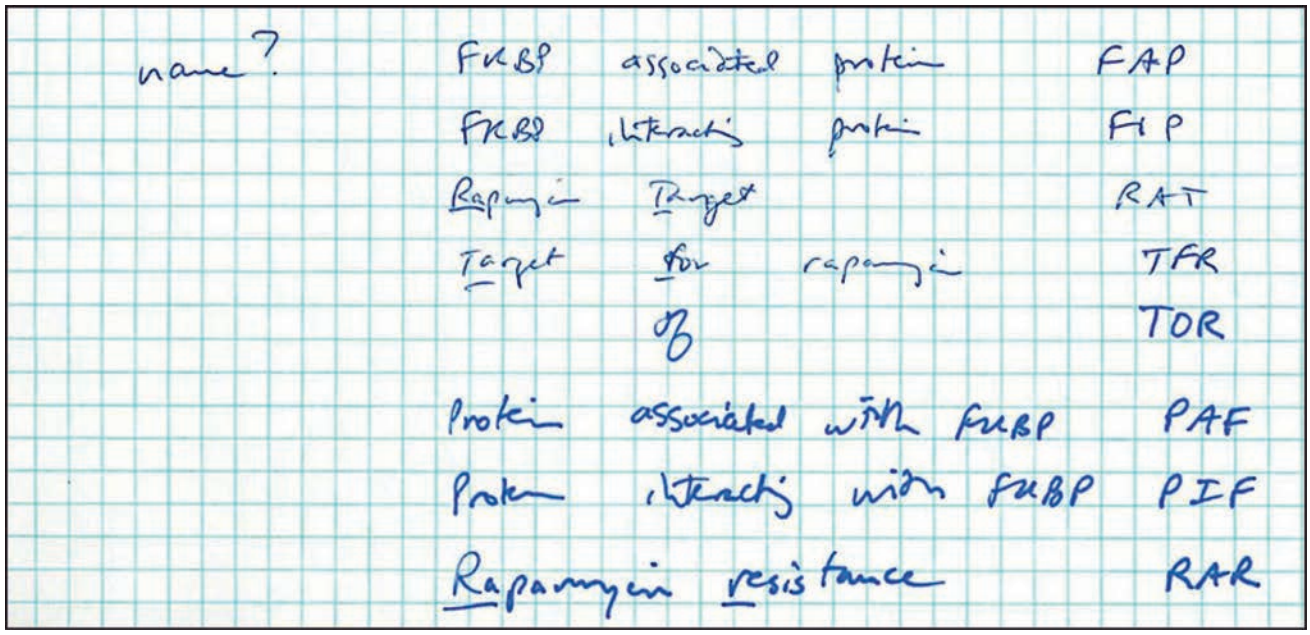
(L to R) Jeannette Kunz, early 1990s. Joe Heitman, 1990. Robbie Loewith, early 2000s

isolated several spontaneous rapamycin-resistant mutants. At this stage, Joe was furiously working around the clock to finish the characterization of his mutants before returning to New York City to complete clinical rotations. By early December 1990, he knew that the mutants were defective in the *FPR1* gene, as expected based on his earlier experiments, or that they harbored a dominant mutation in at least one other unknown gene. The December 2 entry from Joe's lab notebook shows his progress in characterizing two resistant mutants (R1 and R17) defective in the unknown gene(s). As indicated in his notes, he wondered whether the mutations were in the same gene or in different genes. Knowing there was at least one new gene ("gene X"), he began listing candidate gene names (*FAP*, *FIP*, *RAT*, *TFR*, *TOR*, *PAF*, *PIF*, *RAR*) that fit the yeast three-letter-acronym nomenclature convention. As is now well known, we chose *TOR*, for target of rapamycin. I liked the name TOR because it was easy to say and remember. Joe had a more romantic reason for favoring it. In German, TOR (*das Tor*) means "gate," and at the time, we viewed TOR as the entry to the cell cycle, much like the gates in the wall that formerly encircled Basel were the entry into the city. Joe was not deterred by the fact that the masculine form of the German noun *Tor* (*der Tor*) means "fool." To resolve the issue of one versus two new genes, Joe put the diploid from an R1 × R17 cross on sporulation medium ("spo plates") to eventually dissect apart the spores and determine whether the mutations were linked or segregated independently. He managed to complete this experiment, among many others, before leaving Basel in late December. He showed that the rapamycin resistance-conferring mutations in R1 and R17 were unlinked and thus in two distinct genes: *TOR1* and *TOR2*. This was the discovery of TOR, at least as genetic loci in the yeast genome. We published Joe's findings in *Science* on August 23, 1991. Coincidentally, on that same day, Stuart Schreiber published a paper in *Cell*, showing that the phosphatase calcineurin is a common target of cyclosporin A and FK506.

"To the students of today, it probably seems odd that we once thought cell growth is a spontaneous, passively controlled process that just happens when building blocks are available."

What did the two *TOR* genes encode? To answer this question, two very talented students in the lab, Jeannette Kunz and later Stephen Helliwell, cloned and sequenced the *TOR* genes. This was much more difficult than anticipated because the *TOR* genes are among the largest in yeast and were thus under-represented in the size-fractioned genomic libraries we had generated for cloning purposes (Hall, 2016, Mol. Biol. Cell *27*, 2804–2806). The sequencing revealed that the two TORs were novel, highly similar kinases and also that Joe's rapamycin-resistance-conferring mutations were missense mutations that, as shown later, prevented FKBP-rapamycin binding to TOR without otherwise affecting TOR activity. Rapamycin acts by forming a complex with FKBP/FPR1, which in turn directly binds and inhibits TOR. The mammalian TOR (mTOR) gene was cloned shortly thereafter by several other groups, including Stuart Schreiber's, David Sabatini in Solomon Snyder's, and Robert Abraham's groups, and was shown to encode a kinase similar to yeast TOR, hence the name mTOR. Schreiber's laboratory subsequently demonstrated that a reconstituted mTOR variant containing a missense mutation analogous to one of our yeast TOR mutations conferred rapamycin resistance in mammalian cells. Thus, the mode of action of rapamycin was conserved, as presumed by Joe and Rao when they initiated the project.

What is the cellular role of the TOR kinases other than being the target of rapamycin? Answering this question took several years and led to our most gratifying and possibly most important discovery (Hall, 2016, Mol. Biol. Cell *27*, 2804–2806). We originally thought that the role of TOR



Section of December 2, 1990, entry from Joe Heitman's lab notebook, the first appearance of the name TOR.

was to control the cell cycle. This model was based on the cell cycle arrest we observed upon TOR inhibition. However, during a fateful seminar I gave in Vienna in late 1993 to members of the Institute of Molecular Pathology, including the cell cycle expert Kim Nasmyth and his group, I was disabused of this notion. As I learned in Vienna, our kinase did not fit neatly into what was then known about cell cycle control. This view was reinforced as we began to discover that TOR controlled processes related to macromolecular synthesis and unrelated to the cell division cycle. We then began to consider that the true role of TOR was to control cell growth (increase in cell size/mass), rather than cell division (increase in cell number), in response to nutrients. This was a particularly important conceptual advance because it cast our hitherto confusing results in a new light. We suddenly understood our results. It was also a paradigm shift because cell growth was not thought to be actively controlled. To the students of today, it probably seems odd that we once thought cell growth is a spontaneous, passively controlled process that just happens when building blocks are available. Over the subsequent years, we and others showed that TOR activates several anabolic processes and inhibits catabolic processes, confirming that TOR controls mass accumulation and thereby cell growth. In 2000, we proposed that TOR is a central controller of cell growth. The misleading cell cycle arrest we had observed was an indirect consequence of a cell growth defect.

The next major problem was to elucidate the mechanism by which TOR controls cell growth. Jeannette Kunz had earlier knocked out the two *TOR* genes, finding that the two TORs are not functionally equivalent. More precisely, she had found that loss of *TOR1* had no effect on cell viability, whereas knockout of *TOR2* was lethal. When she knocked out both *TOR1* and *TOR2*, cells were not viable, but the cells displayed a G1 arrest not observed upon loss of *TOR2* alone. This suggested that TOR1 has one function, while TOR2 has two functions, one of which is redundant with the single TOR1 function. Further analyses, by several excellent students and postdocs, including Stephen Helliwell, Nic Barbet, Anja Schmidt, Marc Bickle, Thomas Beck, Estela Jacinto, José Luis Crespo, Tobias Schmelzle, Martin Dietmar, Vittoria Zinzalla, Alexandre Soulard, and Mitsugu Shimobayashi, revealed that the two TOR "functions" are two distinct signaling pathways. TOR1 signals via an effector pathway that activates protein synthesis and other processes. TOR2 can substitute for TOR1 in this pathway, but also has an independent function, signaling via a different effector pathway to control the actin cytoskeleton. These observations provided a logical conceptual framework—TOR signals via two pathways to integrate temporal and spatial control of cell growth—but we still did not understand the molecular basis of this signaling complexity. To investigate the complexity of TOR signaling, Robbie Loewith, an undaunted cold-tolerant Canadian postdoc, and my long-term technician Wolfgang Oppliger

entered the cold room to purify the TOR proteins from yeast. The idea was to develop a purification scheme that would be sufficiently gentle to allow co-purification and thereby identification of TOR interacting proteins. This was difficult because we did not know whether TOR even had interacting proteins and, thus, what the appropriate purification scheme might be. Robbie's biochemical characterization, also aided by Stephan Wullschleger, revealed that the TORs are part of two structurally and functionally distinct complexes, which we named TORC1 and TORC2. Either TOR1 or TOR2 assembles into TORC1, but only TOR2 assembles into TORC2. Robbie then showed that the two TORCs correspond to the two previously identified TOR signaling pathways, thereby providing a molecular basis for the complexity of TOR signaling. Estela Jacinto, Pazit Polak, Nadine Cybulski, Asami Hagiwara, Aaron Robitaille, Raúl Durán, Marion Cornu, and Verena Albert in our laboratory and others showed that the two complexes are structurally and functionally conserved in mammals, where they are known as mTORC1 and mTORC2. David Sabatini, in particular, played a key role in elucidating the mammalian TOR complexes.

Over the course of the last 15 years, it has become apparent that mTOR is implicated in a wide variety of diseases in addition to allograft rejection, including cancer and diabetes, and in aging. This, in turn, has spawned a great deal of activity in the pharmaceutical industry to develop new drugs that target the mTOR signaling network. Such drugs may restore normal life to even more patients. Thus, due to an amazing turn of events, rapamycin has had impact in medicine, fundamental biology, and the pharmaceutical industry. It all started with an unknown scientist bending over to pick up a handful of dirt in a faraway land and time.

Cell

Two Basic Scientists Walk into a Translational Space

When John Schiller first joined Douglas Lowy's lab at the National Cancer Institute of the NIH, he could have not predicted that their common interest in the molecular biology of oncogenes would set them in path for discoveries that ultimately enabled the development of a vaccine for the human papillomavirus, which causes the majority of cervical cancers worldwide. John and Doug, the recipients of the 2017 Lasker-DeBakey Clinical Award, have joined *Cell* editor João Monteiro in a Conversation about science, public health, and the joys and challenges of being basic scientists in a translational space. Annotated excerpts from this conversation are presented below.

João Monteiro: How long have you two worked together? Was that first encounter at the National Institutes of Health (NIH)?

Douglas Lowy: We have worked together for more than 30 years.

John Schiller: I came to Doug's lab as a postdoc and basically then just never left; we became colleagues. Eventually I ended up becoming independent and having my own group, but Doug and I continued to collaborate.

JM: Were you already working with the human papillomavirus (HPV) by that time, Doug?

DL: Actually, I was working on the bovine papillomavirus, an animal papillomavirus, rather than on HPV, because when we started, it was before the discovery of the importance of HPV in cervical cancer and other cancers.

JS: When I came to Doug's lab, I wanted to study how bovine papillomavirus, BPV, induces cellular transformation. This virus could make normal fibroblasts become more cancer-like. I was interested in understanding the molecular biology of that phenomenon. I remember very clearly that the second lecture I attended after joining the lab was given by Harald zur Hausen. This was in the beginning of 1983, and Harald came and gave this big lecture here at the NIH, saying, "Eureka, we found this papillomavirus in 50% of cervical cancers." All of a sudden we went from studying a basic model of tumor biology to studying a human virus that causes cancer. Things changed a lot when those discoveries came about.

JM: Doug, were you first interested in viruses or in cancer?

DL: My training after medical school was in a virology lab at NIH in the National Institute of Allergy and Infectious Diseases. That lab studied endogenous mouse retroviruses, so it was viruses plus cancer. When I was a medical student, my first research experience was at New York University with Jan Vilcek, who was studying interferon and how interferon inhibited virus infection and replication. So you could argue I first did viruses, but as soon as I went to the NIH, I ended up doing research on cancer viruses.

"All of a sudden we went from studying a basic model of tumor biology to studying a human virus that causes cancer."

JS: When I applied to Doug's lab, what he was probably best known for was some of his work on the Ras oncogene and endogenous retroviruses. I actually was applying to his lab to work on those sorts of things, not papillomaviruses, but then Doug said, "Well, you know we're starting to work on papillomaviruses," and I saw a lot of growth potential in that area, even before HPV was found to be associated with cervical cancer. It seemed like such a good idea to get involved with it, so I really jumped at the chance to do papillomavirus research in this bovine model. It was really an opportune time to get into the field.

JM: What do you consider your major discovery? Or perhaps the moment you realized that the work you were doing had the potential to change the way cancer is prevented and have big implications for public health?

DL: I think that for us in the lab, our "ah-ha" moment, if you will, was when we saw that the major structural protein of the papillomavirus would self-assemble and make virus-like particles and that those particles induced high levels of neutralizing antibodies. We thought that we had something potentially very useful as a vaccine. The second moment was when, in collaboration with colleagues at the Pasteur Institute in Paris who were working on the Shope papillomavirus model in rabbits, we saw that the vaccine was highly protective in the rabbit animal model. And the third big moment for me was not our research but actually the research from Merck when they did a study of a monovalent HPV 16 vaccine. In this randomized control trial there were 41 cases of HPV 16 infection between the placebo group and the women who got the vaccine. All 41 cases were in the placebo group.

JS: I agree entirely. The only other moment I would add was when we ran the first clinical trial of the VLP (virus-like particles) vaccine in people, in collaboration with our colleagues at Johns Hopkins. This was the first study showing that the vaccine was safe, that there weren't bad reactions, and everybody who got vaccinated generated really high levels of antibodies even if we didn't have any adjuvant in the vaccine. People were always telling us that it's easy to get antibodies in mice and rabbits, but

people aren't the same. Well, it turns out in this case people are like mice, and there was really no difference in the high levels of antibody responses that we saw in people as opposed to the animal models. Then, at least we knew that the vaccine was very immunogenic.

Doug is right, though, in the sense that until we actually did the clinical trial, until the efficacy trials came out, we weren't really sure that the vaccine would protect against genital infections. Because the animal models we were using did not involve intravaginal challenges. Later on we developed an assay in which we could test that. But at the time, we weren't really sure that the vaccine would protect against repeated exposure by the infected partner and protect long term. So, it was really important to actually have the clinical data.

DL: Let me go back to the beginning because there really were two linked observations. The first was that if you just express this single protein from the virus, . . . it would make particles, but what made it really interesting was that when you immunize the animal, you raise high titers of neutralizing antibodies.

JS: Just as an anecdote, the titers were so high that our postdoc, Reinhard Kirnbaur, who was doing this work at that time, had to repeat the assay three different times because the first time we only diluted the serum 100-fold and we got a complete blocking of infection at 100-fold. Then he did it at 10,000-fold, and we got complete blocking of infection even at 10,000-fold. And it wasn't until we diluted it a million fold in the third assay that we found an endpoint where we got infection.

JM: Wow!

DL: I should also mention that our previous work with BPV also helped us to develop the assays that we would need in order to test whether antibodies neutralize viral infection. So there really were two critical parts. One was having a ready source of infectious virus, and the second was having a tractable bio-assay so that you could actually measure neutralizing antibodies because the induction of neutralizing antibodies is the cornerstone of most preventive vaccines.

JM: After you started getting some of the initial clinical data, how did the vaccine transition then from the lab to the industry to become an actual product? Was it a collaborative effort?

"I think it is an ongoing struggle to get the vaccine to more women and men who need it. . . . You'd think that a couple of injections can protect your daughter from cervical cancer. Why aren't you taking it?"

DL: I think that we need to make a distinction between the early clinical trial that we did—which involved a contract laboratory making the virus like particles—and what the two commercial companies Merck and GlaxoSmithKline did. They developed their own vaccine. We had a little bit of input into those, but we were not co-authors on the development on either vaccine. They licensed the technology from the NIH, and they made the commercial vaccines.

JM: I see. Using the principle you had already developed.

JS: Once we had these sky-high neutralizing titers with our model system, we went around and talked to basically almost all the major vaccine manufacturers and all of them were very skeptical. Not of our data but whether a sexually transmitted disease vaccine could work, because they hadn't worked in the past. To this day, HPV is the only one that has worked, in fact. So there was reasonable skepticism. When we went to Merck, we actually first talked to Maurice Hilleman, who was by then Emeritus, but he was like the godfather of Merck vaccines. He was very enthusiastic; he said, "You know this is going to work, and Merck's going to do it. As soon as you go, I'm going to call the people in the licensing department and we're going to get started working on this." He was the first person we talked to in a major drug company that was enthusiastic about the project. At least I think to my recollection that's the way it was.

Then, another company approached us, MedImmune, which was just up the road here in Rockville. They decided that they wanted to do something. After a successful phase one trial, they then sublicensed their vaccine to GSK. We were very happy, because we wanted to have competition in the marketplace and so we wanted the vaccine to be licensed to more than one company. So we had communications with the companies, but I think what Doug said was correct. We weren't active collaborators with the companies during the development of their vaccines.

DL: They licensed the technology, and they went ahead and did it. There had been clinical trials in herpes simplex virus vaccines, which had worked in animal models but then did not work in people to give protection. So, from the point of view of most pharmaceutical companies, our data were analogous. They clearly showed that the HPV vaccine worked in an animal model, but how do you know that the animal model is predictive of what will happen in people? Merck and MedImmune, and later GSK, were the companies that were willing to take a chance.

JS: It was a leap of faith that this could work. The animal models were imperfect, and there were clear limitations, as the one I mentioned about not having a cervical vaginal challenge model at the time.

JM: Now that the vaccines are out there and being used in people, has the impact of the vaccine in public health and cancer prevention matched your expectations?

JS: I think it is an ongoing struggle to get the vaccine to more women and men who need it. I must say that personally, I've been disappointed with the uptake in the United States—maybe Doug has a different idea. You'd think that a couple of injections can protect your daughter from cervical cancer. Why aren't you taking it? It's now just over 50%, 60%. Some developed countries have done very well, like Australia and England. Scotland also has done very well with uptake of the



John Schiller (left) and Doug Lowry (right), National Cancer Institute, NIH. Photo Credit: National Cancer Institute.

vaccines. Other countries, like Japan, have basically abandoned it. So it's a mixed bag. The other big problem is the cost of the vaccine and of delivering three, or now two, intramuscular doses to young women in low-resource settings. There are places where I think there would be demand for the vaccine, and this certainly seems to be playing out in demonstration trials, but they just don't have access to the vaccine and the administration programs. Again, this is something that needs to be addressed. Perhaps, the way to put it is that this vaccine exceeded all expectations in terms of science, immunology, and these sorts of things. But, in terms of uptake, I think it's been underwhelming. What would you say, Doug?

"The good news is that the vaccine has had a dramatic impact on the incidence of infection and disease that can be monitored short term."

DL: The good news is that the vaccine has had a dramatic impact on the incidence of infection and disease that can be monitored short term. In other words, the goal is to try to prevent cancer, but it takes more time from the time of the infection until you get cancer than has elapsed up until now. So it's not realistic to think that you would see a decrease in incidence and mortality from cervical cancer and the other cancers that are caused by HPV. But, when you look at viral immunity, at the reduction in high-grade cervical dysplasia, the reduction in genital warts—all of that points to the notion that the vaccine is highly protective. On the other hand, for a variety of reasons, the vaccine has been underutilized. One point has been social or sociologic concerns in some places, such as the United States, another has been safety concerns. As John mentions, the cost of the vaccine has been an important factor as well, especially in low-resource settings, even though companies have done tiered pricing and the cost in the lowest income countries for the vaccine is relatively low. It takes a very long time to prevent cancer. For most vaccines that are preventive, you give the vaccine and you see a benefit from the vaccine in the very few years after vaccination. But that's not the situation with the HPV vaccine, which means that it is important for the cost of the vaccine to be as low as possible while still being safe and effective.

JS: One of the projects we are actively working on is whether we can reduce the number of doses of vaccine that are required to produce long-term protection. We're just starting a clinical trial based upon data that we generated in a previous NCI-sponsored trial. It's beginning to look like one dose is going to be enough, which again is unheard for a subunit vaccine. However, we need to demonstrate that this is true in a randomized clinical trial, what we are now beginning to do in a NCI trial co-sponsored by the Gates Foundation. We are actually going to test whether one dose works as well as two doses of both the Merck and the GCK vaccines. We think that this could really be transformative. If you could just give everybody one dose, you wouldn't have to keep track of who got it and things like that. It just would be an enormous cost saver just to give one dose. So we are working on this question. There's still a lot to be done with this vaccine.

JM: In those lines, what are other kinds of questions you are currently working on?

JS: So right now we think that the most practical thing to move forward with is to reduce the number of doses to one dose.

DL: Because we think that the decreased cost and the great simplification of logistics from one dose is key to worldwide uptake. And the reason we're testing the two different vaccines is largely because we want to make sure that both vaccines work well in the one-dose setting. Or, if one of them works well but not the other, we want to be sure that we know that, because there are different adjuvants that are used in the two different vaccines. And, therefore, it's at least theoretically possible that one vaccine might work better in the one-dose setting than the other. But, we certainly look forward to the possibility of both vaccines working well.

JS: I think the other thing I might add is we're actually trying on a more basic science level to really understand why this vaccine works so well. We think it has exceptional immunogenicity in terms of inducing antibody responses. This has not been seen for a subunit vaccine where not only everybody responds well but they have plateaued titer where, basically after a year or so, out to 6 or 7 . . . 8 years, the titers don't change. The other point is that in our animal models it looks like very low levels of antibodies are very effective at preventing infection. We're trying to understand from both the immunological and the biological point of view why these vaccines are so exceptional. With the idea that maybe we can translate what we've learned from the HPV vaccines to other vaccines, sexually transmitted and otherwise. But that's on a more basic science level. Doug and I both sort of came up as basic molecular biologists. You know, where we tried to understand basic virology, basic tumor biology. And it was only after we've been working together for about 10 years that we really got involved in vaccinology and what I would say really is more translational research. Even to this day, a large component of what we do is more basic research with the idea of trying to understand things. Then we ask if they would have translational implications. We aren't just translational researchers, we really have a very strong—we think—basic biology component to our studies, and we're continuing that with the vaccines.

JM: Do you think the fact that you were at the NIH influenced the way your research evolved and your ability to move through all these different steps?

DL: In many ways, the short answer is ''yes.'' One of the aspects of being in the intramural program is that you have a level of resources and you can use them for what you think are important problems. Then, you need to defend how you have used them at your lab site visits, which occur every 4 years. We were able to take some of our resources and use them to test, for instance, if expressing viral proteins in insect cells with recombinant baculovirus would produce viral particles. Later, when we were developing the early-phase clinical trials, it was quite helpful to be here because the National Cancer Institute has also a strong clinical research component and we were able to take advantage of that. I will also say that our colleagues of the intramural program in the National Institute of Allergy and Infectious Diseases actually greatly made it easier for us to do this research because they already were working with investigators at Johns Hopkins at the Center for Immunization Research. So we worked with them. More recently, we were involved in another clinical trial with our colleagues—medical epidemiologists who specialized in cervical HPV infection. Again that was made easier because these are our intramural colleagues and they basically did an efficacy trial in Costa Rica where historically cervical cancer has been the number one cancer of women. That trial was what told us that you don't need all three doses of the vaccine in order to get protection. And that has led to this second trial that John mentioned, where we are testing the hypothesis: will just one dose of the vaccine be sufficient in young adolescents?

JS: I think it is good to point out that when Doug and I started this vaccine work, we had no background whatsoever in vaccinology or immunology. We had been publishing good work on viral oncogenesis and some work on transcriptional regulation of the virus. We really had not done any of this type of translational work. Being at the NIH afforded us the opportunity to take a chance at what we saw as an opportunity to make progress in an area where very little progress had been made in the 10 years since Harald zur Hausen had discovered HPV 16. It allowed us to seize the moment but not have to ask anybody. Because quite frankly, if we would've made this proposal to a granting agency, or to a study section, they would have laughed at us. They would have said, ''Well, you guys study how the virus causes cancer, you don't know anything about vaccines or immunology or vaccinology. What are you trying to do here?'' And so it really was an enormous advantage for us to be in the intramural program, I feel.

JM: What an interesting journey!

JS: Indeed. It has been an amazing opportunity, for us coming from a virology and cancer backgrounds, to learn about the questions involved with translational research, being involved with the clinical trials and learning what's involved in those types of studies. It has been a very nice way of being involved in the whole spectrum of biomedical research by getting involved in the vaccine project.

DL: The way I think about it is that John and I are really basic scientists. But, our research took us in a translational direction,

and because that translational research was successful, it has enabled us to be involved in clinical research and now also in implementation research. It's quite exceptional for scientists who start out as basic researchers to have these kinds of other research experiences.

JS: It's not only gratifying to see your basic work leading to what this is—into a public health intervention that looks like it's going to save hundreds of thousands of lives. But, it's just interesting to be involved with really good people that do these types of research and learn an awful lot from being able to interact with them.

DL: I think the one final point that I would add is that the world should be grateful to the patient volunteers who volunteered to take these vaccines when they were still experimental. Because there was no guarantee that they would be beneficial, and whenever you do something that hasn't been tried in people, there's always a risk. Nonetheless, these people who volunteered in the clinical trials—the world needs to be grateful to them because there would not be a vaccine out there without them.

JM: Thanks again for talking to us and once more congrats on the Lasker Award!

The Pyrenoid: An Overlooked Organelle Comes out of Age

Jean-David Rochaix[1,*]
[1]Departments of Molecular Biology and Plant Biology, University of Geneva, Geneva, Switzerland
*Correspondence: Jean-David.Rochaix@unige.ch
http://dx.doi.org/10.1016/j.cell.2017.09.012

The pyrenoid is a membrane-less organelle that exists in various photosynthetic organisms, such as algae, and wherein most global CO_2 fixation occurs. Two papers from the Jonikas lab in this issue of *Cell* provide new insights into the structure, protein composition, and dynamics of this important organelle.

Most life on Earth depends on photosynthesis, the process whereby inorganic carbon is fixed and converted into organic carbon metabolites using solar energy. However, an important limitation in photosynthesis is the poor performance of the CO_2-fixing enzyme Rubisco (ribulose 1,5 bisphosphate carboxylase/oxygenase) due to its dual carboxylase and oxygenase activity. At ambient CO_2 levels, the carboxylase and oxygenase activities compete with each other, leading to a significant decrease in carbon fixation through photorespiration. Nearly all freshwater and marine photoautotrophs, including cyanobacteria and algae, have solved this problem by increasing the level of CO_2 in the vicinity of Rubisco through CO_2-concentrating mechanisms (CCM). In algae, CCM is mediated by several plasma and chloroplast-membrane inorganic carbon transporters, a set of carbonic anhydrases in strategic locations, and the pyrenoid, a membrane-less organelle within chloroplasts, where approximately one-third of global CO_2 fixation is estimated to occur (Figure 1) (Wang et al., 2015). The pyrenoid contains a starch sheath and a matrix in which CO_2 is concentrated together with Rubisco. The matrix is traversed by membrane tubules that are continuous with the thylakoid membranes (Engel et al., 2015). The CCM is highly regulated by CO_2 levels and by light and circadian signals (Mitchell et al., 2014), and many mutants affected in CCM have been identified based on their ability to grow photoautotrophically under high-CO_2 but not under low-CO_2 conditions (Wang et al., 2015).

Although the pyrenoid and the CCM system have been studied in organisms such as the model green alga *Chlamydomonas reinhardtii*, up until recently little was known about its protein composition, structure, and assembly. However, the situation has changed radically in recent years and spectacular advances are now reported in two studies from the group of Martin Jonikas in this issue of *Cell* (Mackinder et al., 2017; Freeman Rosenzweig et al., 2017). The Jonikas lab has invested a great deal of effort in establishing a reverse genetic system in *Chlamydomonas* by generating an indexed mapped mutant library (Li et al., 2016). Screening of these mutants for a CCM phenotype revealed several novel genes involved in CCM that provide a rich source for further investigations (Mackinder et al., 2016).

In the first study, Jonikas and colleagues (Mackinder et al., 2017) develop high-throughput techniques involving fluorescent protein tagging and affinity purification mass spectrometry to identify additional components and to generate a spatially defined network of the *Chlamydomonas* CCM. Using this platform, they identify 135 candidate CCM-associated proteins and determine physical interactors for 38 of them. This work provides a wealth of new insights into the CCM machinery, increasing the number of known pyrenoid proteins by more than 10-fold (from 7 to over 80). This set includes novel Rubisco-interacting proteins, photosystem I assembly factor candidates, and inorganic carbon flux components (Figure 1). This study also reveals that the pyrenoid structure is more complex than previously assumed, consisting of three distinct protein layers: a plate-like layer, a mesh layer, and a punctate layer.

In the second study, Jonikas and colleagues (Freeman Rosenzweig et al., 2017) further examine the structure and dynamics of the pyrenoid of *Chlamydomonas* during the cell cycle. Contrary to what was assumed earlier, the pyrenoid matrix is not crystalline or amorphous but appears to be liquid like, as revealed by the analysis of the Rubisco distribution by cryoelectron microscopy and by FRAP experiments, mixing internally and dispersing into the surrounding stroma during cell division and in response to high CO_2. Jonikas and colleagues also observe that proteins larger than 78 kDa are excluded from the pyrenoid matrix, which may be due to the surface tension generated by the proteins that produce the liquid phase (Bergeron-Sandoval et al., 2016). Previous work identified EPYC1 (essential pyrenoid component 1), a protein containing four nearly identical repeats that co-localizes with Rubisco throughout the pyrenoid matrix and interacts with its small subunit to form a lattice (Mackinder et al., 2016). EPYC1 is essential for normal pyrenoid size, number, morphology, Rubisco content, and efficient carbon fixation at low CO_2. Although the precise binding between Rubisco and EPYC1 is not yet known, this new study proposes that the liquid-like nature of the pyrenoid could be mediated through the binding of Rubisco to the repeats of EPYC1, forming a network with each EPYC1 binding four Rubiscos and each Rubisco binding eight EPYC1s. Thus, EPYC1 would act as a molecular glue holding Rubisco together in the pyrenoid.

Another fascinating aspect of the study by Freeman Rosenzweig et al., is

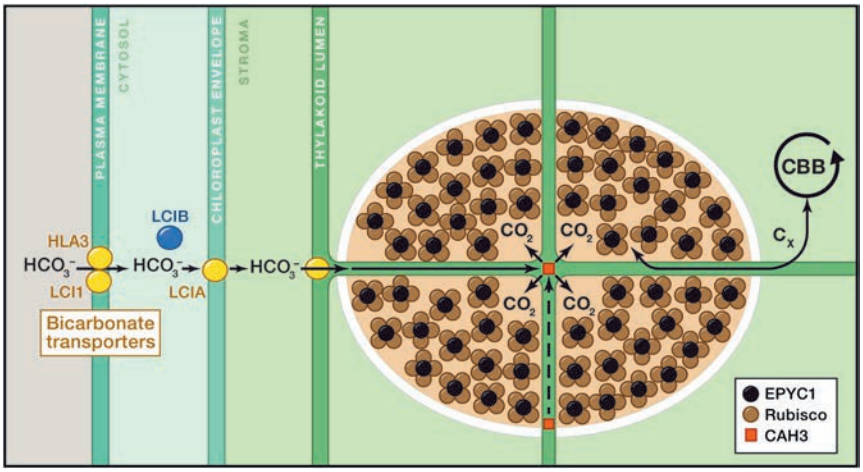


Figure 1. Simplified Model for CO_2-Concentrating Mechanisms in the Pyrenoid of *Chlamydomonas*

The liquid-like matrix of the pyrenoid is shown as a network of Rubisco aggregates held together by EPYC1 in which each EPYC1 protein (black sphere) binds four Rubiscos (brown sphere). Inorganic carbon is imported as HCO_3^- into the cell by the bicarbonate transporters LCI1 and HLA3, which form a complex on the plasma membrane. Alternatively, CO_2 may also enter the cell and be converted to HCO_3^- possibly by LCIB. Entry of HCO_3^- into the chloroplast is mediated by LCIA on the chloroplast envelope and into the thylakoid lumen by an unknown transporter. Finally HCO_3^- diffuses into the pyrenoid through the membrane tubules connected to the thylakoid network (dark green), where it is converted to CO_2 by CAH3 (orange square) and available for CO_2 fixation by Rubisco. The white layer around the pyrenoid represents the starch sheath. The cytosol and stroma are indicated. CBB, Calvin-Benson cycle; C_x, putative shuttle for Rubisco substrates and products.

the highly dynamic reorganization of the pyrenoid that occurs following changes in CO_2 concentration and during the cell cycle. Possibly, the relocalization of Rubisco from the pyrenoid to the stroma under high CO_2 or in darkness could reflect a phase transition of the pyrenoid matrix from an aggregated to a soluble phase, allowing for rapid reorganization of Rubisco to optimize CO_2 fixation under changing environmental conditions. The fluidity of the pyrenoid matrix may also facilitate the access of Rubisco activase chaperones to the Rubisco active sites within the matrix. A similar phase transition may occur during cell division when the pyrenoid undergoes regression, i.e., reduction in size and/or fission. This study shows that both phenomena can occur simultaneously in the same cell.

These two papers not only significantly further our knowledge on the pyrenoid but also open widely the field of CCM to new investigations by raising a number of intriguing questions. How is the pyrenoid positioned within the cell, and how does it divide? What is the minimal requirement for assembling a functional pyrenoid? Can the algal system be transferred to higher plants for improving crop productivity? Introducing CCM into higher plants is predicted to increase yields by up to 60% and to improve the efficiency of nitrogen and water use (Long et al., 2015). In this respect, it is encouraging that nearly all algal CCM proteins localize correctly in transgenic higher plants, indicating that the transfer of algal components is feasible (Atkinson et al., 2016). However, these engineering efforts were restricted because only few components of the algal CCM were known. By providing a detailed blueprint of the algal CCM, the work of Jonikas and colleagues reveals new targets for transfer into crop plants that could improve carbon fixation. The finding that EPYC1 interacts with a kinase and two 14-3-3 proteins raises questions about the role of phosphorylation in the regulation of the interactions between EPYC1 and Rubisco, which could modulate the matrix phase. Clearly, further functional studies of the pyrenoid are eagerly awaited and are likely to bring further surprises related to this rather unusual organelle.

REFERENCES

Atkinson, N., Feike, D., Mackinder, L.C., Meyer, M.T., Griffiths, H., Jonikas, M.C., Smith, A.M., and McCormick, A.J. (2016). Plant Biotechnol. J. *14*, 1302–1315.

Bergeron-Sandoval, L.P., Safaee, N., and Michnick, S.W. (2016). Cell *165*, 1067–1079.

Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., and Baumeister, W. (2015). eLife *4*, e04889.

Freeman Rosenzweig, E.S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H.N., Ronceray, P., Plitzko, J.M., Förster, F., et al. (2017). Cell *171*, this issue, 148–162.

Li, X., Zhang, R., Patena, W., Gang, S.S., Blum, S.R., Ivanova, N., Yue, R., Robertson, J.M., Lefebvre, P.A., Fitz-Gibbon, S.T., et al. (2016). Plant Cell *28*, 367–387.

Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Cell *161*, 56–66.

Mackinder, L.C., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O., Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., et al. (2016). Proc. Natl. Acad. Sci. USA *113*, 5958–5963.

Mackinder, L.C.M., Chen, C., Leib, R.D., Patena, W., Blum, S.R., Rodman, M., Ramundo, S., Adams, C.M., and Jonikas, M.C. (2017). Cell *171*, this issue, 133–147.

Mitchell, M.C., Meyer, M.T., and Griffiths, H. (2014). Plant Physiol. *166*, 1073–1082.

Wang, Y., Stessman, D.J., and Spalding, M.H. (2015). Plant J. *82*, 429–448.

Prion-like Domains Program Ewing's Sarcoma

James Shorter[1,*]
[1]Department of Biochemistry and Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA
*Correspondence: jshorter@mail.med.upenn.edu
http://dx.doi.org/10.1016/j.cell.2017.09.010

Prion-like domains have emerged as important drivers of neurodegenerative disease. Now, Boulay et al. establish that the translocated prion-like domain of the oncogenic EWS-FLI1 fusion protein enables phase-separation events, which inappropriately recruit chromatin-remodeling factors to elicit the aberrant transcriptional programs underlying Ewing's sarcoma.

Prion-like domains (PrLDs) are found in ~240 human proteins and resemble those that drive prionogenesis of specific yeast proteins. These low-complexity domains possess an unusual amino-acid composition enriched in glycine and uncharged polar amino acids, including glutamine, asparagine, tyrosine, and serine (March et al., 2016). Whether PrLDs enable human proteins to form prions remains unknown. However, PrLDs are found in an emerging cadre of human RNA-binding proteins, including TDP-43, FUS, TAF15, and EWSR1, which are linked to fatal neurodegenerative disorders such as frontotemporal dementia and amyotrophic lateral sclerosis (Couthouis et al., 2012; March et al., 2016). PrLDs render the proteins in which they are housed intrinsically aggregation prone. PrLDs also enable liquid-liquid phase separation, a condensation phenomenon that underpins the biogenesis of diverse membraneless organelles such as paraspeckles and stress granules. Liquid-liquid phase separation creates high local concentrations of PrLDs, which can promote nucleation of pathological cross-β fibrils (Lin et al., 2015; March et al., 2016). Similar to the portable nature of yeast prion domains, the ability of PrLDs to promote phase separation and aberrant fibrillization is also portable, and appending PrLDs to various proteins can transfer these properties (Kato et al., 2012; Kwon et al., 2013; Lin et al., 2015). Remarkably, in Ewing's sarcoma, the second most common pediatric bone cancer, a chromosomal translocation fuses a large portion of the PrLD of EWSR1 to the transcription factor FLI1 (Figure 1) (Tan and Manley, 2009; Toretsky and Wright, 2014). In this issue of *Cell*, groundbreaking studies by Rivera, Kadoch, and colleagues demonstrate that the PrLD of the EWS-FLI1 fusion protein phase separates and recruits BAF (BRG1/BRM-associated factor, also known as SWI/SNF) chromatin-remodeling complexes to tumor-specific enhancers, thereby activating the transcriptional events of Ewing's sarcoma. These findings establish that translocated PrLDs can elicit phase-separation events, which unleash aberrant transcriptional cascades that cause cancer (Boulay et al., 2017).

Using an unbiased mass-spectrometry approach, Boulay et al. establish that BAF complexes interact transiently with EWSR1 in a variety of cell types and also with EWS-FLI1 in Ewing's sarcoma (Boulay et al., 2017). EWS-FLI1 acts as a pioneer factor in mesenchymal stem cells, the probable cell of origin for Ewing's sarcoma (Riggi et al., 2014). Here, they discover that the pioneer activity of EWS-FLI1 to induce tumor-specific de novo enhancers at GGAA microsatellite repeats is aided by specifically recruiting BAF complexes, which remodel chromatin to spark transcription of EWS-FLI1 target genes. Like the EWS-FLI1 fusion, FLI1 could interact with BAF chromatin-remodeling complexes, but FLI1 could not recruit BAF complexes to tumor-specific de novo enhancers at GGAA microsatellite repeats. Thus, an emergent property provided by the translocated PrLD enables EWS-FLI1 pioneer activity. Moreover, EWS-FLI1 forms numerous nuclear foci, whereas FLI1 remains diffuse (Boulay et al., 2017). Thus, the ability of the PrLD in EWS-FLI1 to phase separate likely enables pioneer activity and recruitment of EWSR1 and BAF complexes to GGAA microsatellite repeats.

The authors further dissect the features of the PrLD that enable EWS-FLI1 to cause Ewing's sarcoma. More specifically, they establish that, *in vitro*, purified recombinant EWS-FLI1 and EWSR1 rapidly aggregate, whereas FLI1 remains soluble (Boulay et al., 2017; Couthouis et al., 2012). Thus, the translocated PrLD from EWSR1 renders EWS-FLI1 aggregation prone. Replacement of tyrosine residues with serine in PrLDs alters the molecular grammar of PrLD-mediated phase separation and can reduce fibrillization (Kato et al., 2012; Kwon et al., 2013). Indeed, replacement of 37 tyrosine residues with serine in the EWS-FLI PrLD ablates the ability of EWS-FLI to aggregate *in vitro*. Importantly, this tyrosine-replaced EWS-FLI1 variant is also unable to bind GGAA microsatellite repeats and to create active enhancers in mesenchymal stem cells. Thus, the ability of the EWS-FLI1 PrLD to phase separate is intimately connected to the ability to activate the transcriptional program of Ewing's sarcoma. Finally, fusion of short fragments of the EWSR1 PrLD to FLI1 yields proteins that are intrinsically aggregation prone *in vitro* and phenocopy EWS-FLI1 in inducing the gene-expression program of Ewing's sarcoma (Figure 1) (Boulay et al., 2017).

The elegant studies of Boulay et al. firmly connect the ability of the EWSR1 PrLD to phase separate to the EWS-FLI1-driven transcriptional program that underpins Ewing's sarcoma (Boulay et al., 2017). This important work raises several questions. For example, the material properties of the phase-separated EWS-FLI1 that enables the Ewing's sarcoma transcriptional program remain uncertain. Thus, further studies are needed to clarify whether EWS-FLI1 occupies a liquid phase, a hydrogel phase, or a solid fibril phase to recruit chromatin-remodeling factors and

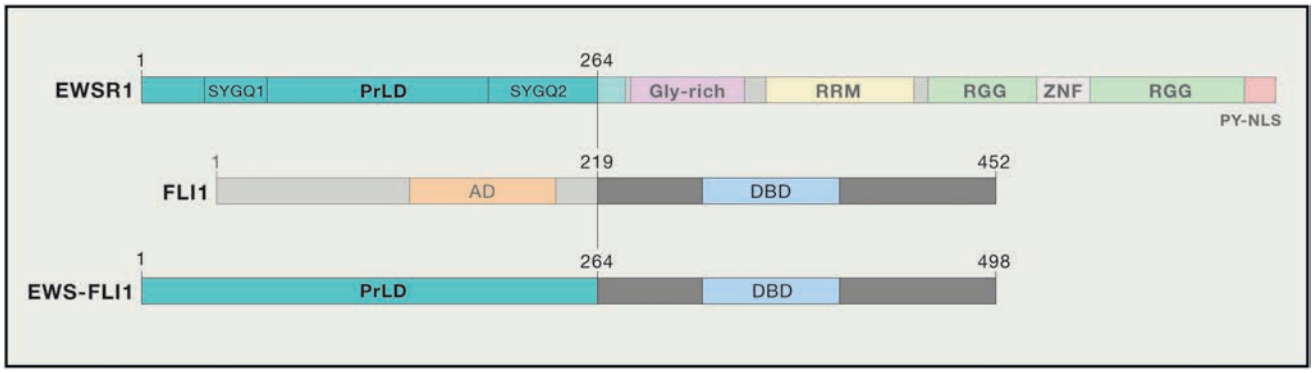


Figure 1. Domain Architecture of EWSR1, FLI1, and EWS-FLI1

EWSR1 contains an N-terminal prion-like domain (PrLD), a glycine-rich domain (Gly-rich), a RNA-recognition motif (RRM), two RGG domains separated by a zinc-finger domain (ZNF), and a C-terminal PY-nuclear localization sequence (PY-NLS). FLI1 harbors an N-terminal activation domain (AD) and a C-terminal DNA-binding domain (DBD). In the EWS-FLI1 (type 1) fusion protein, amino acids 1–264 of the EWSR1 PrLD are fused with residues 219–452 of FLI1. Note that the SYGQ1 and SYGQ2 portions of the EWSR1 PrLD could substitute for the larger EWSR1 PrLD fragment (residues 1–264) and induce the gene-expression programs of Ewing's sarcoma when fused to FLI1 (Boulay et al., 2017).

drive transcription. Precisely how the altered phase of EWS-FLI1 recruits BAF also remains uncertain. Previous studies indicate that EWSR1 PrLD fibrils can recruit RNA polymerase II to promote transcription (Kwon et al., 2013), but whether this proposed mechanism operates in Ewing's sarcoma is unclear. Agents that prevent or reverse phase separation by EWS-FLI1 such as protein disaggregases could have therapeutic utility for Ewing's sarcoma (March et al., 2016). Phosphorylation of multiple serine and threonine residues in the FUS PrLD promotes the liquid phase and prevents fibrillization (Monahan et al., 2017). Thus, it will be interesting to explore whether increased phosphorylation of the EWS-FLI1 PrLD antagonizes fibrillization and the EWS-FLI1 transcriptional program. Finally, portions of the PrLDs of other human RNA-binding proteins such as FUS and TAF15 also get fused to transcription factors via chromosomal translocations that cause other cancers (Tan and Manley, 2009). For example, a portion of the FUS PrLD gets fused to CHOP in liposarcoma and a portion of the TAF15 PrLD gets fused to Ciz in acute leukemia (Tan and Manley, 2009). It will be enlightening to elucidate whether the translocated FUS and TAF15 PrLDs also must phase separate to drive the aberrant transcriptional programs underlying these cancers.

REFERENCES

Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cell *171*, this issue, 163–178.

Couthouis, J., Hart, M.P., Erion, R., King, O.D., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H.J., Mojsilovic-Petrovic, J., Panossian, S., et al. (2012). Hum. Mol. Genet. *21*, 2899–2911.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell *149*, 753–767.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Cell *155*, 1049–1060.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Mol. Cell *60*, 208–219.

March, Z.M., King, O.D., and Shorter, J. (2016). Brain Res. *1647*, 9–18.

Monahan, Z., Ryan, V.H., Janke, A.M., Burke, K.A., Rhoads, S.N., Zerze, G.H., O'Meally, R., Dignon, G.L., Conicella, A.E., Zheng, W., et al. (2017). EMBO J. Published online August 8, 2017. http://dx.doi.org/10.15252/embj.201696394.

Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., et al. (2014). Cancer Cell *26*, 668–681.

Tan, A.Y., and Manley, J.L. (2009). J. Mol. Cell Biol. *1*, 82–92.

Toretsky, J.A., and Wright, P.E. (2014). J. Cell Biol. *206*, 579–588.

Huntingtin Fibrils Poke Membranes

Pedro Guedes-Dias[1] and Erika L.F. Holzbaur[1,*]
[1]Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6085, USA
*Correspondence: holzbaur@mail.med.upenn.edu
http://dx.doi.org/10.1016/j.cell.2017.09.009

A hallmark of Huntington's disease is the presence of intracellular aggregates of mutant huntingtin, the pathological significance of which has long been debated. Using cryo-electron tomography, Bauerlein et al. reveal the fibrillary structure of huntingtin aggregates in situ and show that huntingtin fibrils interact with the endoplasmic reticulum, distorting its morphology and dynamics.

Huntington's disease (HD) is the most common dominantly inherited neurological disorder and is characterized by uncoordinated body movement, cognitive dysfunction, psychiatric impairment, and premature death. HD is caused by an expanded CAG trinucleotide repeat sequence encoding a polyglutamine (polyQ) expansion within the huntingtin gene (*HTT*), resulting in a pathogenic polyQ stretch $\geq$36Q residues. Mutant huntingtin (Htt) is prone to aggregate, forming oligomers as well as large insoluble inclusion bodies (IBs) within neurons. Htt is known to interact with >700 proteins (Shirasaki et al., 2012) and is implicated in multiple cellular processes (Saudou and Humbert, 2016). At the cellular level, abnormal polyQ expansion results in a broad phenotype that includes transcription dysregulation (Kumar et al., 2014), mitochondrial dysfunction (Guedes-Dias et al., 2016), and defects in intracellular transport (Fu and Holzbaur, 2014).

Intracellular IBs of mutant Htt are a pathological hallmark of HD and were thought to be the main cytotoxic agent in HD. This idea was challenged in a seminal study showing that IB formation in neurons was associated with increased neuronal survival by decreasing levels of soluble mutant Htt (Arrasate et al., 2004). The pathophysiological significance of IBs remains, nevertheless, poorly understood. Do IBs confine or trap proteins and organelles? And do IBs interfere with normal cellular membranes and organelles? At the structural level, it is not clear how mutant Htt assembles to form IBs or which domains of Htt are important for this arrangement. In this issue of *Cell*, Bäuerlein et al., (2017) used state-of-the-art cryo-electron tomography (cryo-ET) to examine IBs formed in neurons and HeLa cells expressing a fragment of Htt that includes a pathogenic polyQ stretch, and they address some of these questions.

Solving the ultrastructure of macromolecular complexes in situ is technically challenging. Standard fixation can disrupt cellular morphology, imaging resolution decreases with sample thickness, and mechanical sectioning can induce distortion artifacts. Further, it can be difficult to locate structures of interest within the cell, and significant computational power is required to generate high-resolution 3D models. Recent advancements in cryo-ET have overcome many of the hurdles involved, enabling groundbreaking visualizations of macromolecular complexes in the unperturbed cellular environment (reviewed in Oikonomou and Jensen, 2017).

Bäuerlein et al. expressed a fluorescently labeled N-terminal fragment of Htt encoding exon 1 that includes a pathological polyQ stretch of 97Q in primary neurons and HeLa cells cultured on holey carbon electron microscopy (EM) grids. Correlative light microscopy was used to map fluorescent IBs within cells prior to EM. Focused gallium ion beams were used to generate 150 to 250 nm-thick sections with high precision, free of distortion artifacts associated with cryosectioning, and images were then acquired in a tilt series to generate a high-resolution tomogram.

In the resulting tomograms, spherical Htt97Q IBs in the cytosol and nucleus were found to be composed of a network of radially arranged amyloid-like fibers. The mechanical proprieties of these fibrils, including persistence length (stiffness) and Young's modulus (tensile elasticity), were similar to those of Tau fibrils, a pathological hallmark of Alzheimer's disease. Although the fibril network appears quite dense, Htt97Q fibrils occupied ~3% of the volume—possibly a consequence of the elevated stiffness that these fibrils show, as radius of gyration increases with polymer stiffness.

Strikingly, interactions were observed between fibrils at the periphery of IBs and organelles, including the endoplasmic reticulum (ER; Figure 1). Vesicles with irregular shapes and significant membrane curvature were found lodged within IBs. The authors speculate that these vesicles result from the disruptive effect of direct interactions of IBs with the ER membrane. Previous studies in HD cell models reported that ER stress was reduced once Htt oligomers aggregated to form an IB, suggesting a protective role for IB formation (Leitman et al., 2013). However, the observations of Bäuerlein et al. suggest that IBs formed from pathogenic Htt alter ER morphology through fibril-membrane interactions that induce remarkably high levels of curvature. These contacts locally disrupt ER dynamics and may negatively affect ER function and induce ER stress. Further work is required to determine the specificity of the observed interactions between Htt fibrils and the ER membrane, but these data support the hypothesis that IB formation may also contribute to, rather than ameliorate, the cytopathology of HD.

One limitation of the current study is that it relies on an overexpression paradigm and only investigates the consequences of expressing a truncated form of Htt that includes the first 17 amino acid residues followed by the polyQ chain. Endogenous Htt is a 3,144 amino acid residue protein, predicted to fold into multiple solenoid-like structures that scaffold intermolecular interactions for several protein complexes (Saudou and

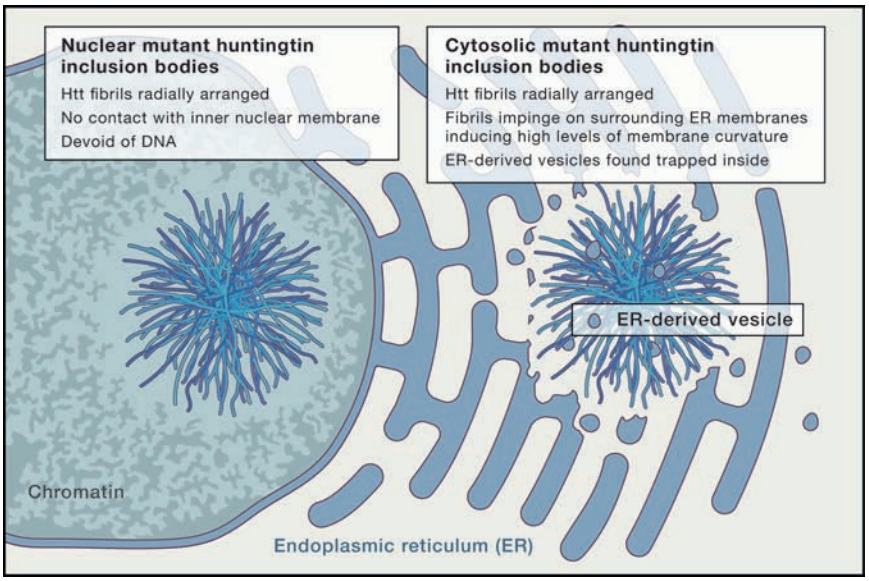


Figure 1. Mutant Huntingtin (Htt) Forms Fibrillary Inclusion Bodies (IBs) *In Situ*

Bäuerlein et al. use cryo-electron tomography (cryo-ET) to reveal the ultrastructure of mutant Htt IBs in primary neurons and HeLa cells expressing an N-terminal fragment of Htt encoding exon 1 that includes a pathological polyQ stretch of 97Q. The authors identify a similar structure of radially arranged fibrils in nuclear and cytosolic IBs and show that cytosolic IBs significantly interact and disrupt endoplasmic reticulum (ER) membranes. Their work provides evidence for a pathogenic role of mutant Htt IBs in Huntington's disease.

Humbert, 2016). In knockin mouse models and human HD patient samples, N-terminal fragments of Htt have been observed, potentially generated by aberrant splicing or proteolytic cleavage (Bates et al., 2015). However, it remains unknown whether IBs from HD patients include both full-length and N-terminal fragments of Htt and whether IB composition modifies the formation kinetics and biophysical proprieties of IBs and/or the resulting cytopathology. Further, the polyQ stretches used in this study (64Q and 97Q) are significantly longer than the expanded stretches associated with typical adult-onset HD. Therefore, future studies will be required to demonstrate the relevance of these findings to IB formation in HD patients.

An interesting observation from the work of Bäuerlein et al. is that, while the ultrastructure of nuclear IBs resembles that of cytosolic IBs, nuclear IBs do not contain vesicles nor contact the inner nuclear membrane. Live-cell super-resolution imaging studies have shown that chromatin is excluded from nuclear Htt94Q aggregates and that transcriptional dysregulation in HD stems from an increase in non-specific sampling by transcription factors before they reach cognate target sites due to the presence of large Htt aggregates (Li et al., 2016). In agreement with these findings, chromatin was not found to be trapped within nuclear IBs.

By taking advantage of the latest advancements in cryo-ET technology, Bäuerlein et al. reveal the fibrillary structure of mutant Htt IBs in situ and provide compelling evidence for a pathogenic role of IBs in HD by showing that fibril-membrane interactions can significantly disrupt both the organization and the dynamics of the ER. The potential of cryo-ET to solve the structure of macromolecular complexes in their native environment is extraordinary. By unveiling the molecular conformation of pathological Htt IBs and their interactions with cellular membranes, this study not only contributes to elucidating HD pathology, but also opens the door to the application of cryo-ET to investigate intracellular aggregates in other neurodegenerative disorders, including Alzheimer's and Parkinson's diseases.

REFERENCES

Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Nature *431*, 805–810.

Bates, G.P., Dorsey, R., Gusella, J.F., Hayden, M.R., Kay, C., Leavitt, B.R., Nance, M., Ross, C.A., Scahill, R.I., Wetzel, R., et al. (2015). Huntington disease. Nat. Rev. Dis. Primers *1*, 15005.

Bäuerlein, F.J.B., Saha, I., Mishra, A., Kalemanov, M., Martínez-Sánchez, A., Klein, R., Dudanova, I., Hipp, M.S., Hartl, F.U., Baumeister, W., and et al. (2017). Cell *171*, this issue, 179–187.

Fu, M.M., and Holzbaur, E.L. (2014). Trends Cell Biol. *24*, 564–574.

Guedes-Dias, P., Pinho, B.R., Soares, T.R., de Proença, J., Duchen, M.R., and Oliveira, J.M. (2016). Neurobiol. Dis. *90*, 51–57.

Kumar, A., Vaish, M., and Ratan, R.R. (2014). Drug Discov. Today *19*, 956–962.

Leitman, J., Ulrich Hartl, F., and Lederkremer, G.Z. (2013). Nat. Commun. *4*, 2753.

Li, L., Liu, H., Dong, P., Li, D., Legant, W.R., Grimm, J.B., Lavis, L.D., Betzig, E., Tjian, R., and Liu, Z. (2016). eLife *5*, 5.

Oikonomou, C.M., and Jensen, G.J. (2017). Annu. Rev. Biochem. *86*, 873–896.

Saudou, F., and Humbert, S. (2016). Neuron *89*, 910–926.

Shirasaki, D.I., Greiner, E.R., Al-Ramahi, I., Gray, M., Boontheung, P., Geschwind, D.H., Botas, J., Coppola, G., Horvath, S., Loo, J.A., and Yang, X.W. (2012). Neuron *75*, 41–57.

Cell

Leading Edge
Review

Genome Regulation by Polycomb and Trithorax: 70 Years and Counting

Bernd Schuettengruber,[1,*] Henri-Marc Bourbon,[2] Luciano Di Croce,[3,4,5,*] and Giacomo Cavalli[1,*]
[1]Institute of Human Genetics, UMR9002 CNRS, University of Montpellier, 141 Rue de la Cardonille, 34396 Montpellier Cedex 5, France
[2]Centre de Biologie du Développement, UMR5547 CNRS/UPS, Centre de Biologie Intégrative, Université de Toulouse, 31000 Toulouse, France
[3]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
[4]Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain
[5]ICREA, Pg. Lluis Companys 23, 08010 Barcelona, Spain
*Correspondence: bernd.schuttengruber@igh.cnrs.fr (B.S.), luciano.dicroce@crg.eu (L.D.C.), giacomo.cavalli@igh.cnrs.fr (G.C.)
http://dx.doi.org/10.1016/j.cell.2017.08.002

Polycomb (PcG) and Trithorax (TrxG) group proteins are evolutionarily conserved chromatin-modifying factors originally identified as part of an epigenetic cellular memory system that maintains repressed or active gene expression states. Recently, they have been shown to globally control a plethora of cellular processes. This functional diversity is achieved by their ability to regulate chromatin at multiple levels, ranging from modifying local chromatin structure to orchestrating the three-dimensional organization of the genome. Understanding this system is a fascinating challenge of critical relevance for biology and medicine, since misexpression or mutation of multiple PcG components, as well as of TrxG members of the COMPASS family and of the SWI/SNF complex, is implicated in cancer and other diseases.

Epigenetic regulation of gene expression is one of the key mechanisms regulating cell-fate choices and cell identity during development. One of the most prominent and enigmatic epigenetic regulatory systems involves the evolutionarily conserved Polycomb group (PcG) and Trithorax group (TrxG) components, acting antagonistically to orchestrate the expression of key genes in cell differentiation and developmental processes. Seventy years ago, the first PcG gene, *Polycomb (PC)*, was discovered in *Drosophila melanogaster* by Pamela Lewis (Lewis, 1947). Later, Ed Lewis determined that *Polycomb* mutations transform anterior embryonic segments into more posterior ones due to the ectopic expression of *Homeotic* (*Hox*) genes (Lewis, 1978). Subsequent genetic screens identified other genes whose mutations resulted in phenotypes similar to the loss of function of *Pc* mutations, leading to the definition of the Polycomb group (PcG) proteins. A few years after the discovery of PcG, the first TrxG gene, *Trithorax,* was isolated as a regulator of *Hox* gene expression. Its mutation causes embryonic segments to be transformed into more anterior ones by antagonizing PcG proteins (Ingham, 1983; Ingham, 1985b; Struhl and Akam, 1985). After additional genes were discovered that had mutant phenotypes counteracting PcG mutations and consistent with the loss of *Hox* function, the overall group was defined as the Trithorax group (TrxG) of proteins (Kennison and Tamkun, 1988). The observation that PcG and TrxG factors are required to maintain *Hox* gene expression after their initial transcriptional regulators disappear from the embryo gave rise to the hypothesis that PcG and TrxG proteins act as a cellular memory system (Ingham, 1985a). Soon after these early discoveries, some PcG and TrxG members were linked to proliferation, senescence (Jacobs et al., 1999a), and cancer (Djabali et al., 1992; van Lohuizen et al., 1991). Extensive research over the past decade has revealed that PcG and TrxG proteins are much more than epigenetic gatekeepers of *Hox* gene expression. They regulate a plethora of cellular processes, including X chromosome inactivation, genomic imprinting, cell cycle control, stem cell biology, and cancer. This functional diversity is achieved by a variety of PcG and TrxG complexes that are assembled in a developmental-stage- and cell-specific manner to modify chromatin at their target genes via histone-modifying or chromatin-remodeling activities. Although PcG and TrxG target genes are conserved during evolution (Schuettengruber et al., 2007), the way these complexes are targeted to chromatin can significantly differ between species. In *Drosophila*, Polycomb response elements (PREs) and Trithorax response elements (TREs) target PcG and TrxG complexes to chromatin, thus driving the epigenetic inheritance of silent or active chromatin states throughout development. In mammals, hypomethylated CpG islands (CGIs) represent PRE-like sequences that can recruit PcG and TrxG complexes. Research in the last decade has shown that PcG and TrxG complexes can regulate their target genes at multiple levels, from modifying local chromatin structure to regulating higher-order chromatin organization and global genome architecture. PcG and TrxG proteins are currently under intense study and, during the last decade, major progress was made toward understanding (1) their diversity and biochemical interactions, (2) their recruitment to chromatin, (3) the mechanisms by which they regulate genome function, (4) their role and mechanisms in conveying inheritance of chromatin states, and (5) their implications in normal and pathogenic cellular processes, particularly embryonic stem cell (ESC) biology and cancer (Figure S1).

PcG Complexes

Historically, based on biochemical purification experiments from *Drosophila melanogaster*, the PcG machinery has been

Table 1. Overview of PcG and TrxG Complexes and Their Core Subunits, Enzymatic Functions, and Functional/Structural Domains

PcG Complex Components		Characteristic Domain	(Epigenetic) Function
Mammals	Flies		
Core PRC1 Complex (common to all PRC1 Complexes)			
RING1A/B	dRing/Sce	RING finger	H2AK119 ubiquitylation
PCGF1–PCGF6	Psc/Suz(2)	RING finger, UBL (RAWUL)	H2AK119 ubiquitylation, oligomerization
Canonical PRC1			
CBX2, CBX4, CBX6–CBX8	Pc	chromodomain	H3K27me3 binding
PHC1–PHC3	Ph	sterile alpha motif (SAM)	oligomerization/protein-protein interaction
SCMH1/2	Scm	SAM	oligomerization/protein-protein interaction
Non-canonical PRC1			
RYBP/YAF2	dRybp	zinc finger	DNA binding
KDM2B	dKdm2	JmjC, CxxC	H3K36 demethylase, DNA binding
DCAF7	Wap/n.i.	WD40	scaffold factors
WDR5	Wds/n.i.	WD40	scaffold factors
Core PRC2 Complex			
EZH1/2	E(z)	SET, SANT	H3K27 methyltransferase, histone binding
SUZ12	Suz(12)	zinc finger	RNA/DNA binding
EED164	Esc/Escl	WD40	H3K27me binding
RBBP4/7	Nurf55/Caf1	WD40	H3K36me3 binding
PRC2 Accessory Proteins			
PCL1–PCL3	Pcl	Tudor, PHD finger	H3K36me3 binding
JARID2	Jarid2	zinc finger, ARID	H2Aub binding, RNA binding
AEBP2	Jing	zinc finger	DNA binding, H2Aub binding
EPOP/C17orf96	n.i.	N/A	modulating enzymatic activity
LCOR/C10orf12	n.i.	N/A	N/A
Core PR-DUB			
BAP1	Calypso	ubiquitin C-terminal hydrolase (UCH) N-terminal catalytic	ubiquitin C-terminal hydrolase
ASXL1/2	Asx	N/A	chromatin binding
PR-DUB Accessory Proteins			
FOXK1/2	FoxK/n.i.	Forkhead box	DNA binding
OGT1	Sxc/n.i.	N/A	O-GlcNAcylation
KDM1B	dLsd1/n.i.	amine oxidase	histone demethylation
MBD5/6	Sba/n.i.	methyl binding	DNA binding

TrxG Complex Components		Protein Domain	(Epigenetic) Function
Mammals	Flies		
Core COMPASS Components common to all COMPASS (-like) Complexes			
WDR5	Wds	WD40	histone binding
ASH2L	Ash2	zinc finger	DNA binding
RBBP5	Rbbp5	WD40	histone binding
DPY30	Dpy30	N/A	N/A
SET1/COMPASS			
SET1A/B	dSet1	SET	H3K4 methyltransferase
HCF1	Hcf1	Kelch	N/A

(*Continued on next page*)

Table 1. ***Continued***

TrxG Complex Components		Protein Domain	(Epigenetic) Function
WDR82	Wdr82	WD40	histone binding
CFP1	Cfp1	CxxC	DNA binding
MLL1/2 COMPASS-like			
MLL1/2	Trx	SET d	H3K4 methyltransferase
HCF1	Hcf1	Kelch	N/A
MENIN	Menin	N/A	N/A
MLL3/4 COMPASS-like			
MLL3/4	Trr	SET	H3K4 methyltransferase
NCOA6	Ncoa6	N/A	N/A
PA1	Pa1	N/A	N/A
UTX	Utx	JmjC	H3K27 demethylase
PTIP	Ptip	BRCT	N/A
ASH1			
ASH1L	Ash1	SET, bromodomain	H3K36 methyltransferase
CBP	dCbp	HAT, bromodomain	H3K27 acetyltransferase
SWI/SNF (BAF and PBAF) Complex			
BRM/BRG1	Brm	helicase, bromodomain	ATPase-dependent chromatin remodeling
BAF250A/B	Osa	ARID	possible DNA binding
BAF155/170	Mor	SWIRM, SANT, chromodomain	possible DNA and histone binding
BAF47	Snr1	winged helix	possible DNA binding
BAF45A-D	Sayp	PHD-finger	possible DNA binding
BAF53A/B	Bap55	actin-like	–
BAF180/BAF200	Bap180	polybromodomain	histone binding
BAF60A-C	Bap60	Swi-B	–
BAF57	Bap111	HMG	possible DNA binding
beta-ACTIN	Actin5C	N/A	N/A
BCL7A-C	Bcl7-like/n.i.	N/A	N/A
BRD7/9	CG7154/n.i.	N/A	N/A

Note that only TrxG complexes with a major function in counteracting PcG function are shown. n.i. indicates that homologous proteins exist, but they have not been purified as part of PcG/TrxG complexes.

subdivided into two main complexes: Polycomb repressive complex 1 (PRC1) and PRC2 (Table 1). This view has been greatly clarified and expanded in the last decade revealing that the PcG system is much more diverse in mammals than in *Drosophila*, especially for PRC1, which is subdivided into canonical (cPRC1) and non-canonical complexes (ncPRC1) (for a review, see Blackledge et al., 2015). All PRC1 complexes share a protein core that is conserved in the five major animal lineages and in plants, but not in fungi (Figure 1A). In mammals, this core is composed of RING1 proteins (RING1A or RING1B), which have E3 ubiquitin ligase activity mediating ubiquitylation of histone H2A on lysine 119 (H2AK119ub), and one of the six Polycomb group ring-finger domain proteins (PCGF1–PCGF6). cPRC1 complexes are assembled around PCGF2/4 and are specified by the presence of one chromobox protein (CBX2, CBX4, and CBX6–CBX8) that binds H3K27me3 and a Polyhomeotic (Ph) homologous protein (PHC1–PHC3), which contains a sterile alpha motif (SAM) domain essential for PcG-mediated repression (Robinson et al., 2012). In contrast, ncPRC1 possesses the zinc-finger domain and YY1-binding protein (RYBP) or its paralog, YAF2 (Wilkinson et al., 2010), which associates with PCGF1, PCGF3/5, or PCGF6 to form ncPRC1.1, ncPRC1.3/PRC1.5, or ncPRC1.6, respectively (Figure 1B). ncPRC1.1, which contains the H3K36-specific histone demethylase KDM2B, is the only non-canonical complex also purified in *Drosophila* (dRAF) (Lagarou et al., 2008). The enzymatic activity of each complex is defined by the choice of PCGF protein, which is stimulated by the presence of RYBP (Rose et al., 2016). The WD40 domain proteins DCAF7 and WDR5 act as central scaffold proteins of the ncPRC1.3/1.5/1.6 complexes (Hauri et al., 2016), whereas an additional set of subunits specific to each complex can modulate their DNA-binding affinities or regulatory functions. Interestingly, RYBP/YAF2 and KDM2B, but not the CBX, PHC, or SCMH proteins, are present in choanoflagellates, the closest unicellular relatives of animals, suggesting that ncPRC1 appeared earlier during evolution than cPRC1 (Figure 1A).

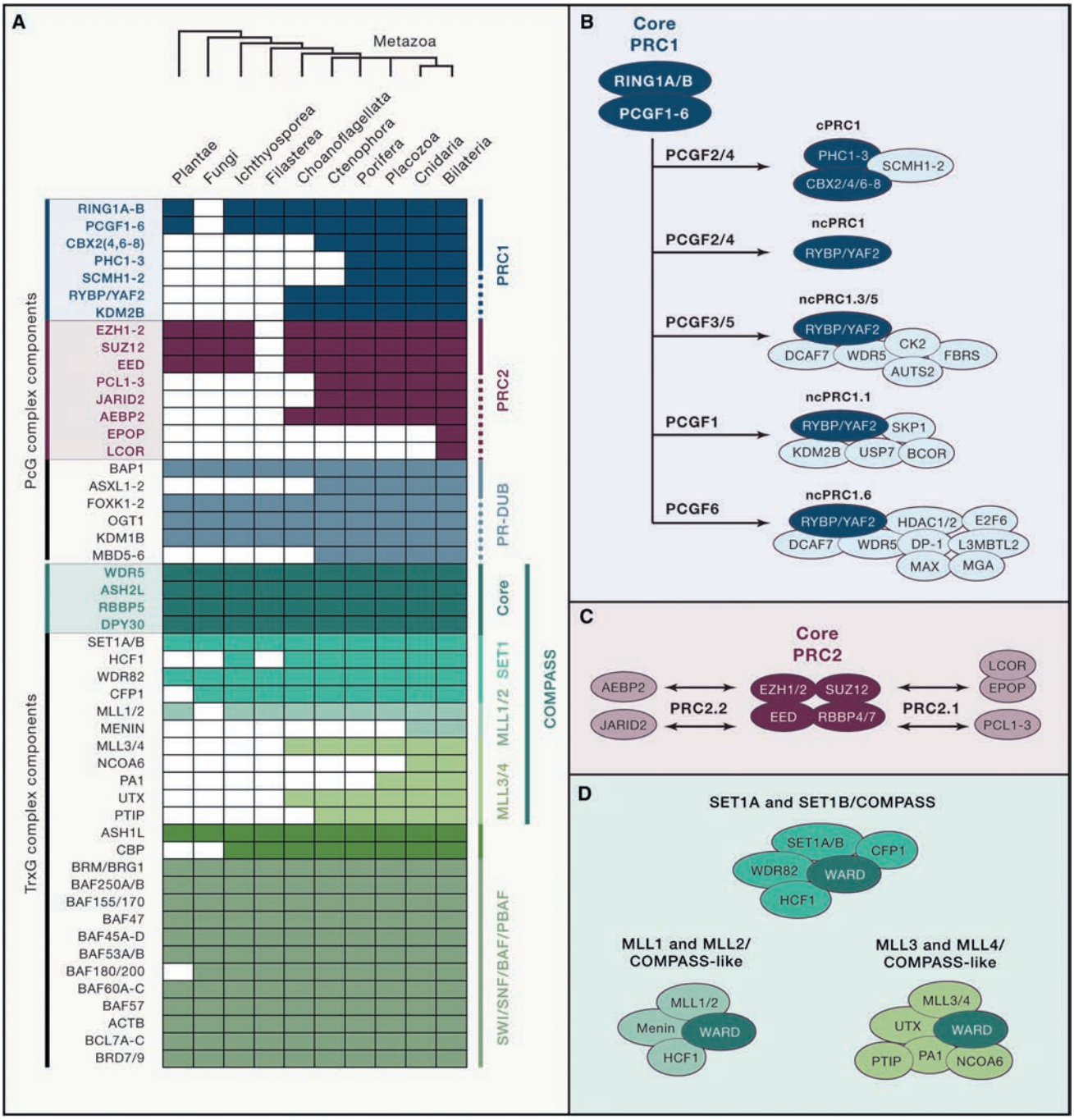


Figure 1. Composition and Evolution of PcG and TrxG Complexes

(A) Phylogenetic distribution of PcG and TrxG complexes. A broad spectrum of eukaryotes has been investigated, with an emphasis on holozoans, among which are the five major metazoans lineages (see Supplemental Information for experimental details on homolog searches). The presence of predicted proteins orthologous to bilaterian PcG/TrxG complex subunits is indicated by filled boxes. Accessory proteins (or non-canonical PRC1 components) are represented by lateral dashed lines. Note that the broadly conserved DCAF7 (ncPRC1), RBBP4/7 (core PRC2), and RBBP5 (COMPASS) WD proteins are not included in this analysis.

(B) Composition of PRC1 complexes. The core of PRC1 can associate with different PCGF proteins (PCGF1–PCGF6) defining canonical PRC1 complexes (cPRC1) and non-canonical (ncPRC1) PRC1 complexes. cPRC1 contains PHC and CBX proteins and the more loosely associated SCMH proteins. ncPRC1 complexes are defined by the presence of RYBP/YAF2 that associate with additional proteins.

(C) Composition of PRC2 complexes. The core of the PRC2 complex associates with different accessory proteins to define the PRC2.1 and PRC2.2 complexes, respectively.

(D) Composition of COMPASS(-like) complexes. A conserved core (WARD) composed of the proteins WDR5, ASH2, RBBP5, and DPY30 associates with the H3K4 methyltransferases SET1A/B, MLL1/2, and MLL3/4 as well as with additional regulatory proteins to produce COMPASS(-like) complexes.

The functional core of the mammalian PRC2 essential to mediate histone methyltransferase activity (for a review, see Blackledge et al., 2015) is composed of one of the SET-domain-containing histone methyltransferases enhancer of zeste (EZH2 or EZH1), embryonic ectoderm development (EED), suppressor of zeste (SUZ12), and the CAF1 histone-binding proteins RBBP4 and RBBP7 (Table 1). This PRC2 core is broadly conserved in the five major animal lineages as well as in plants and some fungi, including *N. crassa*, and the yeast *Cryptococcus neoformans* (Dumesic et al., 2015; Jamieson et al., 2013) (Figure 1A). PRC2 compositional diversity is achieved through association with non-stoichiometric subunits (hereafter called accessory proteins) that can modulate its enzymatic activities or chromatin target sites (Ciferri et al., 2012; Li et al., 2010; Pasini et al., 2010; Shen et al., 2009). Importantly, the emergence of accessory subunits of PRC2 follows the increasing cell and tissue complexity of metazoans (Figure 1A), suggesting that the regulation of PRC2 activity by these different regulatory subunits might be key for specifying cell identity during development and differentiation. A systematic proteomics approach in human cells revealed two fundamental alternative assemblies linked to the PRC2 core complex (Hauri et al., 2016) (Figure 1C). PRC2.1 is defined by its mutually exclusive binding of one of the three Polycomb-like homologs (PCLs) PHF1, PHF19, or MTF2. PHD finger protein 1 (PHF1) and its *Drosophila* homolog, Pcl, stimulate efficient trimethylation activity of EZH2 toward the H3K27me2 substrate (Nekrasov et al., 2007; Sarma et al., 2008). PRC2.2 is defined by the presence of the zinc-finger proteins AEBP2 and JARID2, which enhance enzymatic activity and regulate chromatin binding of the PRC2 complex. Finally, the two mammalian-specific proteins, C10ORF12/LCOR and C17ORF96/EPOP, also co-purify with PRC2 complexes (Beringer et al., 2016; Kloet et al., 2016).

Genetic screens in *Drosophila* further identified PcG proteins that are not components of PRC1/2 but constitute additional PcG complexes (Table 1). Pho-repressive complex (PhoRC) is composed of the zinc-finger protein Pleiohomeotic (Pho) and dSfmbt (Scm-related gene containing four MBT domains), which can bind to H3K9me1 and H3K20me2 (Klymenko et al., 2006). No enzymatic activity is associated with PhoRC. The Polycomb repressive deubiquitinase complex (PR-DUB) was initially identified in *Drosophila* and contains the ubiquitin C-terminal hydrolase Calypso (Bap1) and additional sex combs (Asx) (Scheuermann et al., 2010). It possesses H2A-specific deubiquitinase activity that is paradoxically required for PcG-mediated repression, suggesting that a tight balance between H2A ubiquitination and deubiquitination is essential for PcG-dependent silencing. Mammalian homologs of PR-DUB can be divided into PR-DUB1 and PR-DUB2, depending on whether the BAP1 partner is ASXL1 or ASXL2, respectively (Hauri et al., 2016). Both complexes share a similar set of accessory proteins, including FOXK1/2, the histone demethylase KDM1B, methyl-binding domain proteins MBD5/6, and the O-GlcNAc transferase OGT1. Notably, the *Drosophila* homolog Ogt/Sxc was previously identified as a PcG protein (Gambetta et al., 2009) but had been not co-purified with *Drosophila* PR-DUB. In summary, considerable progress has been made in determining the composition of PcG complexes in various species. However, much remains to be done in this field. In particular, the functional importance of many of the identified subunits—purified from diverse and sometimes highly aneuploid cell lines—in a developing organism has yet to be determined.

TrxG (COMPASS, SWI/SNF) Complexes

The complexity of TrxG complexes matches that of their PcG counterparts. TrxG is a heterogeneous group that plays a widespread role in transcriptional activation. Here, we will focus on TrxG complexes (the SWI/SNF complex and COMPASS family) that were shown to be key players in specifically counteracting PcG-mediated repression. Genetic studies in *Drosophila* first identified antagonistic links between PcG genes and the SWI/SNF (switch/sucrose non-fermentable) complexes (Kennison and Tamkun, 1988), originally purified and characterized in yeast. Subsequently, homologous complexes having ATP-dependent chromatin-remodeling activities were identified in flies and mammals (Dingwall et al., 1995; Khavari et al., 1993). Elegant biochemical, genetic, and molecular studies led to the discovery that mammalian homologs of SWI/SNF, BRM, and BRG1 form two distinct complexes, BAF and PBAF, each of which contains up to 15 additional subunits that are well conserved during evolution (Table 1; Figure 1A) and which regulate chromatin structure of a large number of genes implicated in the cell cycle, signaling, and proliferation and are often dysregulated in cancer (reviewed in Hodges et al., 2016; Kadoch and Crabtree, 2015).

A second group of TrxG complexes is associated with histone-modifying activities, including histone acetylation and methylation, thereby counteracting the action of the repressive marks deposited by PcG complexes. The histone methyltransferase complex SET1/COMPASS (Piunti and Shilatifard, 2016) was first isolated in yeast and is well conserved across all pre-metazoan and major animal lineages (Figure 1A). However, during evolution, this complex diverged to give rise to COMPASS-like complexes containing different SET-domain methyltransferases with unique functionalities (reviewed in Piunti and Shilatifard, 2016). The well-conserved protein core of all COMPASS complexes is required for their function and includes the proteins WDR5, ASH2, RBBP5, and DPY30 (sometimes abbreviated as WARD) (Figure 1D). SET1/COMPASS additionally contains HCF1, WDR82, and the DNA-binding protein CXXC1 (CFP1) and mediates bulk trimethylation of K4 on histone H3 (H3K4me3) (Figure 2B). MLL1/MLL2 COMPASS-like additionally contains the proteins MENIN, HCF1, and either MLL1 or MLL2 in a mutually exclusive manner (Hu et al., 2013). MLL2 is responsible for H3K4 trimethylation at bivalent promoters in ESCs (Figure 2B) (Denissov et al., 2014), whereas MLL1 is required for H3K4 trimethylation at only a small subset of genes, including *HOX* genes, emphasizing the non-redundant function of these two histone methyltransferases. The *Drosophila* homolog of MLL1, Trx, is dispensable for genome-wide H3K4me3 levels but has recently been found to dimethylate H3K4 at PREs to maintain the developmental expression pattern of its target genes (Rickels et al., 2016). In addition to the WARD subunits, MLL3/4 COMPASS-like includes the proteins NCOA6 and PA1 and the histone demethylase KDM6/UTX, which can remove the PcG-mediated repressive H3K27me3 mark (Figure 2A). Like the *Drosophila* homolog Trr, MLL3/4 are the major methyltransferases mediating monomethylation of H3K4 at enhancers (Figure 2B) (reviewed in Piunti

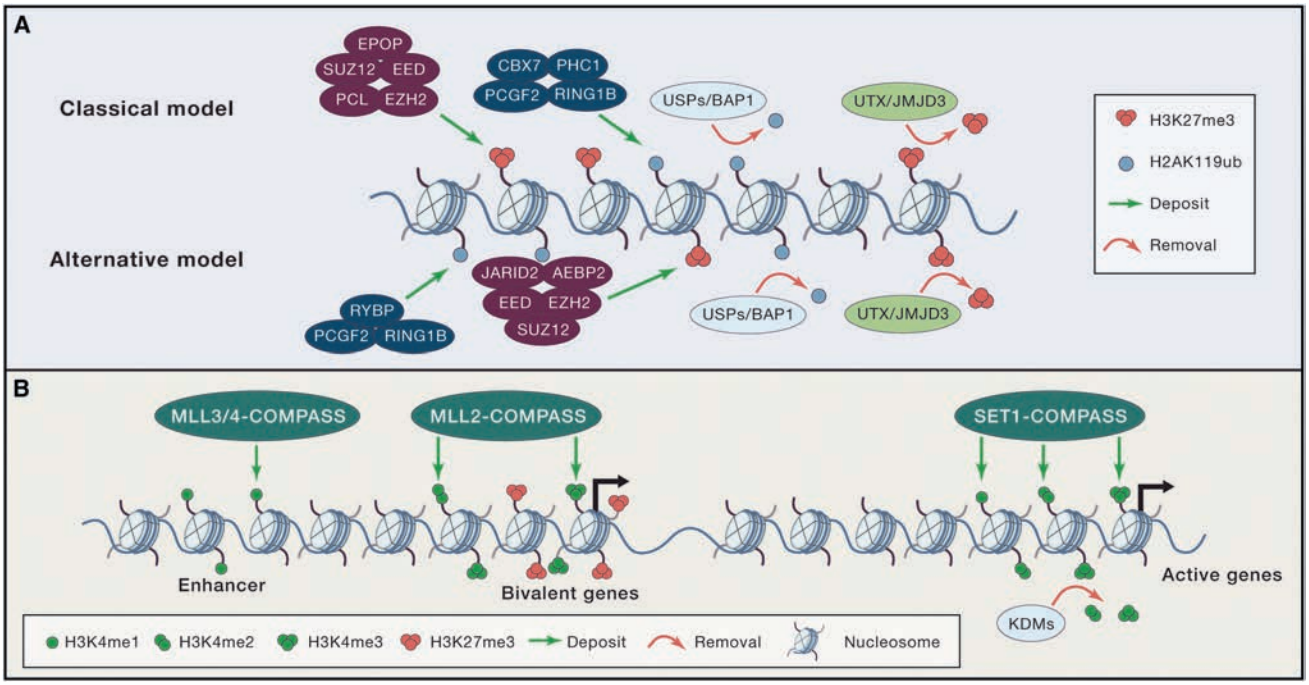


Figure 2. Deposition and Removal of PcG and TrxG-Mediated Histone Marks Contributes to Their Balanced Action on Chromatin
(A) According to the "classical model," targeting of the PRC2 complex leads to the methylation of H3K27. This modification is recognized by the Cbx subunit of the canonical PRC1 complex, which in turn catalyzes monoubiquitination of histone H2A at lysine 119 (H2AK119ub). An alternative recruitment mechanism posits that ncPRC1 is targeted to unmethylated CpG islands through the FBXL10/KDM2B subunit. ncPRC1 seems to be responsible for the majority of the H2AK119ub modifications present at PcG-target sites, which might facilitate the recruitment of PRC2. H3K27 methyl marks can be erased by histone demethylases, such as UTX/KDM6A and JMJD3/KDM6B, while several deuquibitinases (including BAP1, USP16 and USP21) can remove the monoubiquitin moiety from histone H2A. (B) Set1A/B COMPASS complexes catalyze mono-, di-, and trimethylation on H3K4 at active promoters. The activity of the partially redundant COMPASS complexes containing MLL3/KMT2C and MLL4/KMT2D leads to the deposition of H3K4me1 at enhancers, facilitating the recruitment of other activators such as CBP/p300. The deposition of methyl marks on H3K4 at bivalent regions is performed by MLL2/COMPASS. Multiple histone demethylases, including members of the KDM1/LSD, KDM2A/FBXL11, and KDM5/JARID families, are implicated in the removal of methyl groups on H3K4.

and Shilatifard, 2016). Intriguingly, the choanoflagellate *S. rosetta* (but not *M. brevicolis*) possesses MLL3/4 and UTX proteins (Figure 1A). This is of special interest, as *S. rosetta* can exist either as single cells or as multicellular rosette-shaped colonies, and it possesses homologs of several genes required for animal cell signaling and adhesion (Dayel et al., 2011). Further, distal enhancer elements are only present in metazoans that possess MLL3/4; in the filasterean *Capsaspora*, in which MLL3/4 are absent, regulatory regions are close to transcription start sites (Sebé-Pedrós et al., 2016). The correlation of the presence of MLL3/4 in metazoans with a corresponding increase in *cis*-regulatory genome complexity suggests that the emergence of enhancer-specific MLL3/4 activity was a key event in cell differentiation processes during pre-metazoan evolution. An additional histone demethylase (KDM) specific for H3K4me3 (little imaginal discs [Lid]) has been classified as a TrxG protein in *Drosophila* (Eissenberg et al., 2007; Lee et al., 2007b; Secombe et al., 2007), and its mammalian homolog, JARID1d, associates with PCL proteins regulating transcriptional initiation through H3K4 demethylation (Figure 2B) (Lee et al., 2007a).

Recruitment of PcG Components

Although PcG proteins and their target genes are highly conserved during evolution (Schuettengruber et al., 2007), the sequence requirements underlying PcG recruitment can diverge significantly. Early work in *Drosophila* identified PREs as the DNA regulatory elements that recruit PcG factors to chromatin and mediate epigenetic inheritance of silent chromatin states throughout development (Fauvarque and Dura, 1993; Müller and Bienz, 1991; Simon et al., 1993). *Drosophila* PREs frequently contain DNA motifs for sequence-specific DNA-binding proteins like Pho, Phol, Trl/GAF, Spps, Dsp1, and Cg. While these transcription factors (TFs) play important roles in PcG recruitment in *Drosophila*, none of them are sufficient to recruit PcG complexes on their own (reviewed in Entrevan et al., 2016). Initially, a hierarchical recruitment model was proposed in which TFs recruit PRC2, which subsequently recruits PRC1 via the interaction of Polycomb (Pc/CBX) with the PRC2-deposited H3K27me3 mark (Figure 2A) (Wang et al., 2004). However, genome-wide identification of PREs (Schuettengruber et al., 2009; Schwartz et al., 2006; Tolhuis et al., 2006) revealed that they are usually nucleosome-depleted regions and that PRC1-bound regions devoid of H3K27me3 exist (Loubière et al., 2016; Schaaf et al., 2013; Schwartz et al., 2006). This context-dependent relationship among TFs, PRC1, and PRC2 is cooperative rather than hierarchical (Figure 3A). In particular, PRC1 can stabilize binding of PhoRC by a positive feedback loop (Kahn et al., 2014; Schuettengruber et al., 2014) and supports PRC2 binding at a large

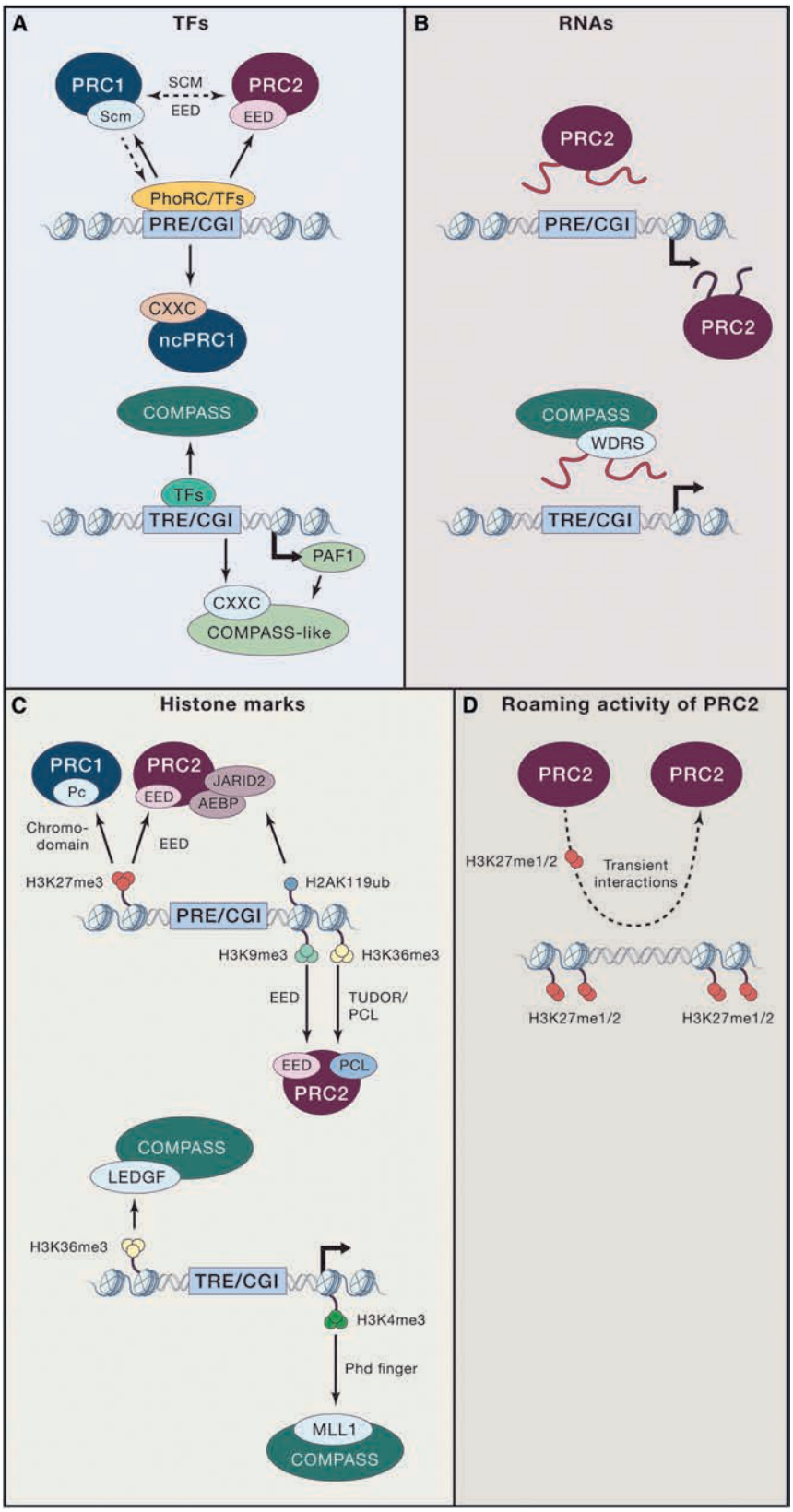


Figure 3. Analogous Factors Regulate the Recruitment and Stable Binding of PcG and TrxG Complexes

(A) Transcription factor (TF)-mediated recruitment. Cooperative interactions (indicated by the dashed arrows) between TFs and PRC1 and PRC2 complexes contribute to stable PcG binding. The PRC1 accessory protein Scm is a key protein in mediating PRC1 interaction and PhoRC. Similarly, SCMH1/2 and EED might mediate interactions between PRC1 and PRC2 complexes, thereby facilitating stable binding to chromatin. KDM2B, which is part of ncPRC1.1, binds to CGIs through its zinc-finger CxxC domain. Similarly, COMPASS(-like) complexes can bind sequence specifically to their target sites via the CxxC domain of MLL1/2 or can be recruited by TFs such as CFP1 or FOXA1 or via their interaction with PAF1.

(B) RNA-mediated recruitment. ncRNAs (like XIST, HOTAIR, or KCNQ1, shown in red), or short RNAs transcribed from repressed genes that form stem-loop structures (in black), can induce recruitment of PcG complexes. In a similar manner, HOTTIP ncRNA interacts with the WDR5 subunit of COMPASS to recruit it to the *HOXA* locus.

(C) Chromatin-mediated recruitment. H3K27me3 can increase the affinity of PRC1 to chromatin via interaction with the chromodomain of Pc/CBX. The EED subunit of PRC2 can bind to methylated H3K9 and H3K27, thereby stabilizing its binding to chromatin. H2AK119ub can induce PRC2 recruitment mediated by AEBP2 and JARID2. H3K36me3 can stabilize binding of PRC2 via the TUDOR domain of PCL proteins. TrxG complexes containing MLL1 can be recruited to their target sites via their interaction of the PHD finger with H3K4me3. H3K36me2, deposited by ASH1L, can promote binding of COMPASS complexes via the epigenetic reader protein LEDGF.

(D) In the absence of PRE-like sequences and PcG-recruiting components, PRC2 associates only transiently to mediate mono- and dimethylation of H3K27 (H3K27me1/me2).

subset of PREs. Moreover, both complexes can be recruited independently to chromatin in certain genome contexts (Kahn et al., 2016). PRC2-independent recruitment of PRC1 might be mediated by the PRC1 accessory component Scm, which acts as a molecular bridge connecting PhoRC and PRC1 via the oligomerization ability of the SAM domain (Frey et al., 2016).

In mammals, genome-wide mapping studies of PcG proteins have revealed a strong correlation with hypomethylated CGIs (Boyer et al., 2006; Tanay et al., 2007). This, together with the observation that an artificial GC-rich element devoid of activating TF-binding sites can ectopically recruit PcG proteins, gave rise to the idea that CGIs represent mammalian PREs (Farcas et al., 2012; Mendenhall et al., 2010; Riising et al., 2014). KDM2B binds to CGIs through its zinc-finger CxxC domain, contributing to PRC1.1 recruitment. However, most PcG-binding sites are not affected by loss of KDM2B, suggesting that PcG complexes can bind to CGIs in alternative ways. In addition, although KDM2B is bound to virtually all CGIs, it induces recruitment of PRCs at only ~30% of mammalian CGIs in ESCs (Bernstein et al., 2006a; Ku et al., 2008). These PcG-bound CGIs correspond to repressed promoters, and inhibition of transcription induces recruitment of PcG proteins to newly silenced CGIs (Riising et al., 2014). This supports a "chromatin sampling" model (Klose et al., 2013), which proposes that PcG proteins weakly interact with all potential binding sites but that their stable binding is blocked by active transcription or the presence of activating TFs. However, even after inhibition of transcription, a significant number of CGIs remains unbound by PcG proteins, and PcG proteins have been shown to be targeted to sites of active transcription, arguing for an alternative mechanism that complements chromatin sampling by PcG proteins. This might entail interaction of PcG proteins with TFs, such as E2F6, MGA/MAX, REST, SNAIL, RUNX1, JARID2, AEBP2, and YY1. Individually, these TFs induce recruitment only in specific circumstances and to a subset of their target sites. It is therefore possible that the PcG machinery is recruited to target genes via multiple interactions with diverse TFs, similar to the transcription machinery that can induce transcription via interactions with multiple enhancer-bound TFs.

Long non-coding RNAs (lncRNAs) also regulate recruitment of PcG complexes (Figure 3B). After the initial discovery that XIST plays a role in targeting PcG to the inactive X chromosome (Plath et al., 2003), additional work implicated other lncRNAs, like HOTAIR and KCNQ1, in targeting PcG proteins to *HOX* or imprinted gene loci (reviewed in Davidovich and Cech, 2015), although the direct role of lncRNAs in PcG recruitment remains disputed (Cerase et al., 2014; Portoso et al., 2017). In addition, short RNAs transcribed from repressed genes that form stem-loop structures resembling PRC2-binding sites in XIST can interact with PRC2 in *cis*, suggesting a general role for RNAs in PRC2 targeting (Kanhere et al., 2010). The complex relationship between RNAs and PRC2 is underscored by the observations that PRC2-RNA interactions can also prevent PRC2 recruitment to chromatin or inhibit its HMT activity (Beltran et al., 2016; Cifuentes-Rojas et al., 2014; Herzog et al., 2014). In summary, the current data suggest that a general RNA-PRC2 interaction counteracts silencing, whereas selective RNAs might favor PRC2-mediated silencing in the presence of specific cofactors.

A third factor that regulates PcG recruitment is chromatin marks deposited by other histone-modifying complexes or by PcG proteins themselves (Figure 3C). Whereas the PRC2-dependent mark H3K27me3 increases the affinity for Pc/CBX-containing PRC1 complexes to chromatin (Fischle et al., 2003; Min et al., 2003), recent evidence in mammals suggests the existence of a recruiting step upstream of this canonical hierarchy (Figure 2A). Tethering KDM2B induces the recruitment of ncPRC1 complexes in a PRC2-independent manner, which results in the deposition of H2AK119ub and promotes binding of PRC2 (Blackledge et al., 2014; Cooper et al., 2014). Moreover, a recent study showed that the ncPRC1.3/5 complex initiates recruitment of PRC1 and PRC2 to the inactive X chromosome via H2A ubiquitination, demonstrating a function of ncPRC1 and H2A ubiquitination for the initiation of Polycomb domains in a physiological context (Almeida et al., 2017). H2AK119ub-mediated recruitment of PRC2 might depend on AEBP2 and JARID2, which bind to H2AK119ub in vitro (Cooper et al., 2016; Kalb et al., 2014). Intriguingly, however, the E3 ligase activity of RING1B is not essential for PcG function in mouse (Illingworth et al., 2015), and H2A ubiquitination in flies seems to be largely dispensable for PRC2 recruitment (Kahn et al., 2016; Pengelly et al., 2015). Direct interactions between PRC1 and PRC2 complexes mediated by the proteins SCM or EED might support their co-binding to target genes, providing a potential explanation for why histone modifications are dispensable for PcG recruitment in many cases (Cao et al., 2014; Frey et al., 2016; Kang et al., 2015). The H3K36me3 modification, usually associated with active genes or DNA damage sites (Musselman et al., 2013), can also contribute to PcG targeting. The PRC2-associated factor Polycomb-like binds to H3K36me3 via its Tudor domain (Ballaré et al., 2012), facilitating the intrusion of PRC2 into active chromatin regions to promote silencing. Finally, multiple links have been established between H3K9 methylation and Polycomb. EED can bind H3K9me3 (Margueron et al., 2009), and PRC2 physically interacts with the H3K9 methyltransferases GLP/EHMT1 and G9A, whose loss results in reduced PRC2 binding to chromatin (Mozzetta et al., 2014).

In addition to H3K27me3, PRC2 catalyzes H3K27me1 and H3K27me2, which cover the large majority of the euchromatic genome (Ferrari et al., 2014; Lee et al., 2015). PRC2 is not detected in regions marked by H3K27me2, suggesting that its binding to these regions is transient. As PRC2 prefers lower methylated states (i.e., H3K27me0 or H3K27me1) over H3K27me2 for its methylation reaction (McCabe et al., 2012), this suggests a model in which PRC2 associates with most of the genomic chromatin through weak and transient interactions, which lead to mono- and dimethylation of H3K27. In contrast, additional components, including TFs, non-coding RNAs (ncRNAs), and specific chromatin-protein interactions, stabilize the chromatin targeting of PRC2 to a subset of regions where it can deposit H3K27me3 and facilitate cPRC1 targeting. Likewise, multiple targeting components might induce PRC1-specific chromatin binding in the absence of H3K27me3. The overarching picture is thus that PcG complexes have relatively poor DNA sequence specificity, but they can be specifically recruited to or evicted from selected regions by specific TFs, nascent RNAs, ncRNAs, and chromatin modifications.

Recruitment of TrxG Components

Compared to the wealth of data concerning PcG complexes, much less is known about the recruitment of TrxG complexes, although this is an equally important subject. In flies, TrxG complexes co-occupy a fraction of PREs with PcG complexes and are potentially recruited by a similar set of TFs implicated in PcG complex recruitment (Beisel et al., 2007). Whereas Trx is recruited independently of the activation status of its target gene, Ash1 and Brm are targeted to chromatin in an activation-dependent manner (Dejardin and Cavalli, 2004; Papp and Müller, 2006). In analogy to KDM2B-mediated recruitment of PRC1 to CGIs, sequence-specific DNA binding of COMPASS-like complexes to CGIs can be mediated by the CxxC domain of MLL1/2 (Ayton et al., 2004; Hu et al., 2017) (Figure 3A). COMPASS-like complexes can also be recruited to CGIs via their interaction with PAF1, which directly interacts with motifs flanking the CxxC domain of MLL1/2 (Muntean et al., 2010). Moreover, SET1/COMPASS complexes can be recruited to CGIs via their interaction with another CxxC-domain-containing protein, CFP1 (also known as CXXC1), that specifically associates with non-methylated CGIs (Thomson et al., 2010) (Figure 3A). Recently, the pioneering factor FOXA1 was shown to recruit MLL3 to enhancers, thereby mediating H3K4me1/2 (Jozwik et al., 2016). lncRNAs can also contribute to the recruitment of COMPASS to chromatin. *HOXA* transcript at the distal tip (HOTTIP) interacts with the COMPASS subunit WDR5 (Wang et al., 2011; Yang et al., 2014), resulting in COMPASS recruitment to the *HOXA* locus (Figure 3B). However, similar to the role of lncRNAs in PcG recruitment, the question of a more general role of lncRNAs in TrxG complex targeting to chromatin remains elusive. Finally, COMPASS can also be targeted to chromatin by preexisting histone marks (Figure 3C). MLL1 binds H3K4me3 via its PHD finger, thereby stabilizing its association with the *HOXA9* target gene locus (Milne et al., 2010). The COMPASS subunit WDR5 can contribute to chromatin binding by binding to H3K4 tails, irrespective of their methylation status. Another histone mark that contributes to targeting COMPASS to chromatin is the H3K36me2 mark deposited by ASH1L, which promotes binding of the epigenetic reader lens epithelium-derived growth factor (LEDGF) and MLL at key leukemia target genes. Conversely, overexpression of the demethylase KDM2A reduces H3K36me2, leading to the replacement of LEDGF/MLL and reduced expression of MLL target genes (Zhu et al., 2016). These data suggest that multifactorial contributions might stabilize the TrxG/COMPASS complex targeting to a subset of chromosomal regions, similar to the case of PcG complexes.

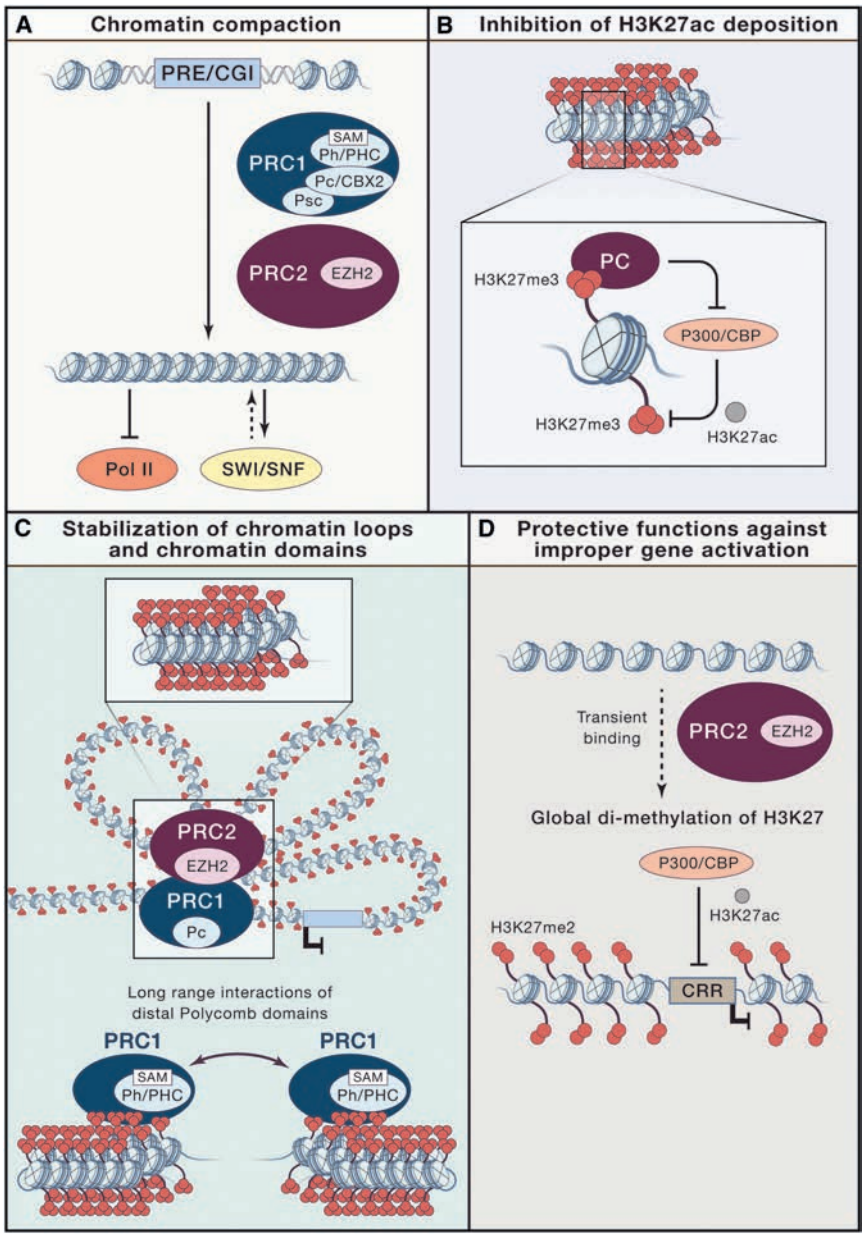
Mechanisms of PcG- and TrxG-Mediated Chromatin Regulation

PcG-mediated repression is considered to be the default state, with gene activation mediated by the counteracting activity of TrxG proteins (Klymenko and Müller, 2004). Intriguingly, both in flies and mammals, PcG and TrxG complexes extensively colocalize at chromatin, regardless of the activity state of the target gene (Beisel et al., 2007; Enderle et al., 2011; Papp and Müller, 2006). One key question is therefore how PcG and TrxG complexes exert their repressive or activating function and what tips the balance between PcG-mediated repression and TrxG-dependent gene activation.

One of the principal mechanisms of PcG-mediated repression is chromatin compaction, rendering chromatin inaccessible to the transcription machinery and inhibiting SWI/SNF-mediated chromatin remodeling (Shao et al., 1999) (Figure 4A). Importantly, different PRC1 variants might exert different levels of chromatin compaction that is mediated via a highly positively charged "compaction region" present in fly Psc (Grau et al., 2011) and vertebrate CBX2 bridging adjacent nucleosomes to compact chromatin (Lau et al., 2017). Moreover, PRC2 itself can lead to chromatin compaction (Margueron et al., 2008; Terranova et al., 2008). Together with the ability of PRC2 to bind H3K27 trimethylated nucleosomes (Margueron et al., 2009) and with the allosteric stimulation of catalysis by dense chromatin templates (Jiao and Liu, 2015; Yuan et al., 2012), this activity might lead to cooperation of PRC2 and PRC1 toward chromatin condensation and inhibition of transcription. The SWI/SNF complex, which was recently shown to be constantly involved in a dynamic competition with PRC1 thanks to a SMARCA4/BRG1-dependent PRC1 eviction activity (Kadoch et al., 2017; Stanton et al., 2017), might be key to switch this repressive chromatin toward activation. The outcome of this competition at individual loci might depend on the amount of PRC1 being recruited, which may lead to different net levels of PcG complexes on specific target genes and drive each locus into an open or closed chromatin state. This corroborates classic genetic studies showing that the SWI/SNF subunit Brm suppresses Polycomb mutations (Tamkun et al., 1992). Interestingly, beyond regulating Hox genes, the mammalian SWI/SNF complex is critically important in cell-fate decisions, such as neuronal differentiation and mitotic exit (Victor et al., 2014; Yoo et al., 2009, 2011). It will be interesting to see whether this role involves the activation of Polycomb target genes.

Polycomb can also act as a repressive agent by inhibiting the acetyltransferase activity of dCBP associated with regulatory regions to block H3K27 acetylation, thereby favoring H3K27 methylation (Tie et al., 2016) (Figure 4B). On the activating side, Trx associates with dCBP and is required for H3K27 acetylation to block PRC2-mediated methylation (Tie et al., 2009). Similarly, Trx-catalyzed H3K4me3 and Ash1-mediated methylation of H3K36 can inhibit PRC2 activity (Schmitges et al., 2011; Yuan et al., 2011) and Trx-dependent dimethylation on histone H3K4 marks PREs and contributes to the maintenance of activation (Rickels et al., 2016). The molecular trigger to switch PREs between silencing and activation might be mediated by non-coding transcription through PREs, affecting PRC2 activity (Herzog et al., 2014) rather than displacement of PcG proteins (Erokhin et al., 2015).

PRC1-mediated H2A ubiquitination was initially suggested to play an important role in PcG-mediated repression by interfering with multiple steps of the transcription process (Nakagawa et al., 2008; Stock et al., 2007; Zhou et al., 2008). However, more recent data have demonstrated that H2A ubiquitination is not required for PcG repression in flies and mammals (Illingworth et al., 2015; Pengelly et al., 2015), suggesting that PRC2 might be a key effector of PcG-mediated silencing. Indeed, flies carrying a point mutation in lysine 27 of histone H3 fail to repress PRC2-target genes, mimicking PRC2 mutant phenotypes and indicating that H3K27 is the PRC2 substrate relevant for PcG



Figures 4. PcG Proteins Use Multiple, Cooperative Mechanisms to Silence Their Target Genes

(A) PRC1 can induce chromatin compaction via a positively charged compacting region present in mammalian CBX2 or fly Psc. In addition, Pc/CBX (via H3K27me3-binding through its chromodomain) and Ph/PHC (via polymerization through its SAM domain) can contribute to chromatin compaction, which can interfere with SWI/SNF-mediated chromatin remodeling or PolII recruitment. PRC2 can also contribute to the compaction of nucleosomal arrays.

(B) Trimethylation of H3K27 (H3K27me3) can directly block acetylation of H3K27 (H3K27ac), which is involved in gene activation. Further, Pc/CBX can inhibit the acetyltransferase activity of CBP, therefore favoring methylation of H3K27.

(C) Distribution of the H3K27me3 mark over large genomic regions might stabilize chromatin-looping interactions between PcG-binding sites and gene regulatory regions, thereby contributing to stably locking genes in a repressed state. Oligomerization of the SAM domain of Ph/PHC is essential for PcG-mediated repression and can mediate long-range interactions between distal Polycomb domains.

(D) Dimethylation of H3K27 (H3K27me2) exerts protective functions by preventing acetylation of *cis* regulatory regions (CRRs), such as enhancers or promoters, thereby inhibiting their inappropriate activation.

repression (Pengelly et al., 2013). It seems that H3K27me3 stabilizes PcG-mediated repression but is not the initial cue (Kahn et al., 2016). This mark might also stabilize chromatin looping interactions of PRE-bound PcG complexes with neighboring gene regulatory elements, thereby contributing to locking target genes in a repressed state and promoting the formation of large repressive genomic domains (Figure 4C). Furthermore, the PRC1 subunit Ph cooperates with H3K27me3 to lock genes in a repressed state by mediating long-range interactions between distal PcG domains via oligomerization of its SAM domain, thereby establishing nuclear subcompartments enriched in PcG proteins (Wani et al., 2016). While PRC1 and PRC2 can and do cooperate at multiple target genes, the long-standing dogma that PcG repression requires both complexes has been challenged by evidence for an uncoupling of these complexes in gene regulation for at least a subset of PcG target genes that depend on PRC1, but not PRC2, for their silencing (Loubière et al., 2016; Schaaf et al., 2013). However, PRC1-dependent transcriptional repression is weaker at these genes than at canonical genes, suggesting that the cooperation between the PRC1 and PRC2 classes of complexes ensures more robust gene silencing.

Finally, the repressive function of PcG complexes might be mediated by the widespread H3K27 dimethylation, which can suppress pervasive chromatin opening and transcriptional activities by preventing H3K27 acetylation (Ferrari et al., 2014; Lee et al., 2015) (Figure 4D).

Higher-Order Genome Regulation

In metazoans, chromatin is organized in a hierarchical series of 3D architectures. At the lower level, nucleosome fibers fold into chromatin loops, which build topologically associating domains (TADs). TADs interact to form active and inactive chromosome compartments, which organize into chromosome territories (for reviews, see Bonev and Cavalli, 2016; Entrevan et al., 2016; Schwartz and Cavalli, 2017). PcG proteins display a punctate distribution in the cell nucleus, both in flies and in mammals (Buchenau et al., 1998; Saurin et al., 1998). These

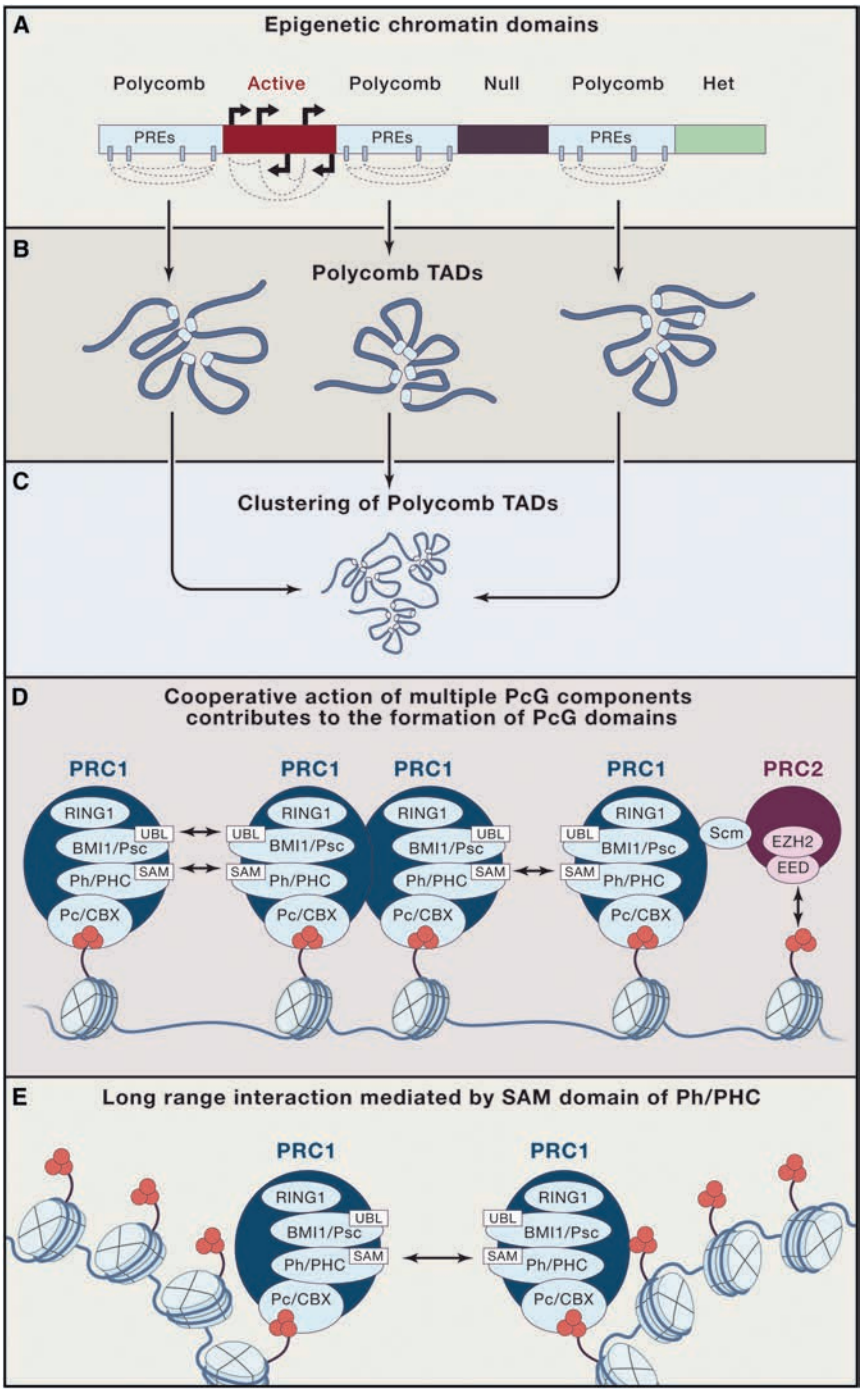


Figure 5. Roles of PcG Proteins in Regulating Higher-Order Chromosome Organization

(A) Chromatin domains are defined by a specific combination of chromatin marks. Intradomain chromatin contacts are much more frequent than contacts with regions outside domains.

(B) PcG components contribute to multi-looped higher-order structures that lead to the formation of repressive chromosomal domains or Polycomb TADs.

(C) Multiple Polycomb TADs can cluster, forming nuclear repressive compartments that stabilize silencing.

(D) EED and Pc/CBX can bind to H3K27me3, which might contribute to the propagation of PcG complexes in *cis* along the chromatin fiber. The PRC1 accessory protein Scm/SCM can contribute to the spreading of PcG complexes by acting as a molecular bridge connecting PRC1 and PRC2 via its SAM domain. In addition, the SAM domain of Ph/PHC and the UBL domain of PCGF4/BMI1 can contribute to PRC1 oligomerization, which might stabilize PcG complexes bound to their target sites. Similarly, *Drosophila* Psc has been show to bind to itself.

(E) Ph/PHC oligomerization mediated by the SAM domain can also facilitate binding of PcG complexes to extended chromatin regions by mediating long-range interactions between distal PRC1 binding sites.

PcG foci were shown to be the physical manifestation of PcG-mediated silencing (Grimaud et al., 2006a; Terranova et al., 2008), and recent developments have begun to unravel the molecular basis of PcG-mediated 3D genome regulation (Figure 5). At the level of oligonucleosomes and chromatin fibers, new technology has begun to pinpoint differences between Polycomb and other types of chromatin (Risca et al., 2017). In mammals, flies, and plants, several PcG components have been shown to be involved in regulatory chromatin loops (for a review, see Entrevan et al., 2016) as well as in the regulation of higher-order chromosome structure and function. Genome-wide mapping of a large set of DNA-binding proteins and histone marks revealed that the genome can be partitioned into epigenetic domains characterized by a specific combination of active or repressive chromatin marks (Filion et al., 2010; Kharchenko et al., 2011; Thurman et al., 2012) (Figure 5A). These states are distributed in chromosome domains, and the analysis of their 3D organization by Hi-C has shown that they form TADs, in which intradomain chromatin contacts have a much higher frequency than contacts with regions outside the domain (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014; Sexton et al., 2012). PcG components organize the 3D architecture of Polycomb-containing TADs (Figure 5B). Many studies focused on *HOX* gene clusters, which are regulated by PcG, TrxG, and insulator proteins in insects as well as in mammals. In flies, PcG- and TrxG-binding sites juxtapose insulator elements, and this proximity is important for *Hox* regulation (Schwartz and Pirrotta, 2007; Singh and Mishra, 2015). Within these domains, PcG binding sites establish preferential contacts, suggesting that silencing might involve the formation of a multi-looped structure (Lanzuolo et al., 2007). In mammals, repressed *HOX* clusters also

form a repressive domain spatially segregated from adjacent chromosome regions (Noordermeer et al., 2011). 3D contacts between independent PcG binding sites may support spreading of PcG complexes in *cis*, leading to the establishment of the highly condensed chromatin architecture described in these species. Local looping interactions and spreading in *cis* is likely mediated by the cooperative action of PRC2 (Eed) and PRC1 (Ph/PHC, Psc/BMI1, and Scm) subunits (Figure 5D) (Kundu et al., 2017; Wani et al., 2016). Mammalian *HOX* domains are progressively opened during activation, and 4C (circular chromosome conformation capture) studies showed that the active part of the cluster segregates away from the Polycomb-repressed domain (Noordermeer et al., 2011). The CTCF insulator protein regulates this process and is required for appropriate regulation of *HOX* genes (Narendra et al., 2015, 2016). These data thus suggest that regulation of 3D architecture of chromosome domains plays a critical role in specification of the body plan.

At a higher level of folding, Polycomb domains cluster to contact each other in the cell nucleus (Figure 5C). Again, this phenomenon is conserved in flies, plants, and mammalian systems (Bantignies et al., 2011; Rosa et al., 2013; Schoenfelder et al., 2015; Vieux-Rochas et al., 2015). Furthermore, Polycomb-mediated long-range interactions are dynamic. In flies, activation of a subset of the *Hox* genes prevents their interaction with their Polycomb-silenced counterparts (Bantignies et al., 2011), and in mouse ESCs, PRC2-dependent interactions are established during the transition from the naive to the primed state (Joshi et al., 2015). Both PRC2 and PRC1 have been shown to be required for long-range interactions of PcG-regulated regions in ESCs (Boettiger et al., 2016; Denholtz et al., 2013; Schoenfelder et al., 2015; Wani et al., 2016). However, PRC2 does not seem to be critical for establishing TADs (Nora et al., 2012), and its depletion has only limited effects on long-range interactions of Polycomb domains (Denholtz et al., 2013; Schoenfelder et al., 2015). In contrast, PRC1 plays a critical role, both in local chromatin condensation of Polycomb domains and in the establishment of their long-range interactions (Schoenfelder et al., 2015; Wani et al., 2016). In particular, oligomerization mediated by the SAM domain of the Ph/PHC1 subunit of cPRC1 (Figures 5D and 5E) is crucial for the condensation of individual Polycomb domains (Kundu et al., 2017) as well as for mediating long-range Polycomb domain interactions (Isono et al., 2013; Wani et al., 2016). Of note, not only Polycomb-silenced genes but also many active gene promoters interact in space with their enhancers and other promoters (Javierre et al., 2016; Schoenfelder et al., 2015). Whether COMPASS complexes are causally linked to 3D interactions, and whether this architecture affects function, are important questions for future research.

Equally important is to understand whether the 3D organization of PcG proteins plays a regulatory role or whether it is a consequence of the silencing process. In addition to a reciprocal stabilization of PRC1 and PRC2 complexes, PRC1 stabilizes the binding of the key recruiter protein Pho (Kahn et al., 2014; Schuettengruber et al., 2014), suggesting that, after initial PcG recruitment by DNA-binding factors, PcG complexes may form nuclear subcompartments that stabilize their binding as well as that of their own recruiters. This scenario suggests that 3D organization contributes to Polycomb-mediated silencing (Schuettengruber et al., 2014), yet how can one test this hypothesis? Mutations of PRC2 components such as EED, or of PRC1 components such as RING1 and Ph/PHC1, induce changes in gene expression and in 3D chromosome architecture, such that effects due to the loss of local PcG repressive components cannot be easily disentangled from those due to higher-order chromatin organization (Bantignies et al., 2011; Denholtz et al., 2013; Joshi et al., 2015; Schoenfelder et al., 2015; Wani et al., 2016). To tackle this, Wijchers et al. (2016) ectopically targeted EZH2 to naive loci containing lacO-binding sites, which led to local deposition of H3K27me3 and the establishment of new long-range contacts with other Polycomb domains, yet these changes had little impact on gene expression. In contrast, evidence for a direct role of 3D organization was obtained in *Drosophila* by deleting a DNA element in one of the *Hox* clusters (the BITHORAX complex). This mutation induced loss of contacts with the other *Hox* cluster (the ANTENNAPEDIA complex), which is located 10 Mb away along the same chromosome arm. In this configuration, genes in the ANTENNAPEDIA complex were upregulated, suggesting that 3D contacts contribute to the stability of Polycomb-dependent gene silencing (Bantignies et al., 2011). Similar genome engineering studies (i.e., manipulating the 3D neighborhood of the tested loci while preserving their *cis*-regulatory sequences) will be required to understand the relationship between nuclear organization and gene expression changes on a genome-wide scale.

The Role of PcG and TrxG Proteins in Somatic and Transgenerational Epigenetic Inheritance

PcG and TrxG proteins were originally studied for their ability to maintain the memory of *HOX* gene expression patterns throughout development (Cavalli and Paro, 1998; Grimaud et al., 2006b; Margueron and Reinberg, 2011; Poux et al., 2002; Schwartz and Pirrotta, 2007), but how powerful is this memory function, and what are its mechanisms? The two most challenging processes for chromatin inheritance during cell proliferation are DNA replication and mitosis/meiosis. PcG components were shown to remain bound to chromatin templates during replication, both in vitro (Francis et al., 2009; Lengsfeld et al., 2012; Lo et al., 2012) and in vivo (Hansen et al., 2008; Petruk et al., 2012). Although H3K27me3 is diluted during DNA replication, it can mediate short-term memory of repressed chromatin states (Coleman and Struhl, 2017), and post-replicative PRC2 activity is induced by binding to H3K27me3 on parental nucleosomes to restore original H3K27me3 levels (Alabert et al., 2015). However, sequence-specific targeting of PRC2 to PREs is required for efficient propagation of H3K27me3 levels over multiple cell divisions, indicating that the whole PcG machinery (including PRE sequences, chromatin-modifying activities, and their associated histone marks) acts as a epigenetic memory system that is needed for long-term memory of repressive chromatin states (Coleman and Struhl, 2017; Laprell et al., 2017). Likewise, Trx remains associated during DNA replication, and H3K4me3 is rapidly reconstituted after DNA replication (Alabert et al., 2015; Petruk et al., 2012), suggesting that Trx/COMPASS proteins use similar mechanisms to propagate active chromatin states.

Much less is known about the maintenance of PcG and TrxG components during mitosis. Quantitative in vivo analysis and genome-wide mapping studies in mitotic cells revealed the presence of Psc and Pc as well as the TrxG member Ash1 on mitotic chromatin (Follmer et al., 2012; Fonseca et al., 2012; Steffen et al., 2013). Likewise, both MLL and PcG components were detected on mammalian mitotic chromosomes, suggesting that at least a part of the PcG and TrxG machinery may bookmark mitotic chromosomes to inherit chromatin states. However, the mitotic chromosome distribution of both mammalian MLL components and fly Psc differs at least partly from that in interphase cells, and the mechanism of this targeting dynamic is still unknown (Blobel et al., 2009; Follmer et al., 2012).

Polycomb-mediated memory of chromatin states can last for long periods of time, as illustrated by the phenomenon of vernalization, in which exposure to prolonged cold accelerates flowering after a return to milder temperatures. In *Arabidopsis*, this phenomenon depends on PRC2-mediated silencing of the *FLC* locus, which progressively and quantitatively increases with longer periods of cold exposure (Angel et al., 2011; Bastow et al., 2004). In every generation, the chromatin state of the *FLC* locus is reset by the H3K27-specific demethylase ELF6, which prevents transgenerational inheritance of previous exposures to cold (Crevillén et al., 2014). Other observations suggest the possibility that inter- and transgenerational inheritance of chromatin states may occur under certain circumstances. Recent data suggest that intergenerational inheritance of maternally provided nucleosomes carrying H3K27me3 are propagated in early embryos to prevent inappropriate activation of lineage-specific enhancers during development (Zenk et al., 2017). PcG and TrxG proteins were also shown to induce transgenerational inheritance of alternative states of transgene expression in flies (Bantignies et al., 2003; Cavalli and Paro, 1998). In *C. elegans*, PRC2 proteins had been suggested to transmit inter- and transgenerational silencing of the X chromosome via both sperm and oocytes (Gaydos et al., 2014). In vertebrates, the mechanisms underlying transgenerational epigenetic inheritance are hotly debated (Blake and Watson, 2016; Heard and Martienssen, 2014; Hollick, 2017). This is mostly because specific pathways are responsible for resetting the state of the epigenome during germline development. DNA methylation is vastly reprogrammed, and most methylated DNA sites do not survive in the next generation (Smallwood and Kelsey, 2012). However, a small subset of loci can preserve their DNA methylation status (Surani, 2015). Improvement of chromatin-immunoprecipitation-sequencing (ChIP-seq) technologies allowed histone marks to be mapped in gametes and the early stages of mouse development. The data showed that both sperm and oocytes bear regions endowed with H3K4me3 and H3K27me3, which are extensively reprogrammed upon fertilization and during the early cell division cycles of the embryo (Dahl et al., 2016; Liu et al., 2016; Zhang et al., 2016; Zheng et al., 2016). However, a significant number of these regions are maintained during early embryogenesis (Zheng et al., 2016), and both PRC1 components and MLL2 are required in the oocyte for the correct specification of gene expression patterns in embryos of the subsequent generation (Andreu-Vieyra et al., 2010; Posfai et al., 2012). Likewise, the alteration of H3K4me3 levels by overexpression or inhibition of the H3K4 demethylase KDM5B induces severe early embryonic defects (Dahl et al., 2016; Liu et al., 2016; Zhang et al., 2016). Mouse and frog studies directly suggested the possibility of a role of PcG and TrxG proteins in transgenerational inheritance (Siklenka et al., 2015; Teperek et al., 2016). Finally, the analysis of a system in which the nuclear organization of a transgene was transiently perturbed demonstrated the PRC2-dependent transgenerational inheritance of different levels of H3K27me3 in *Drosophila.* Remarkably, inheritance could be transmitted through both the female and the male germline, suggesting that it is robust even to the partial replacement of histones by protamines during spermatogenesis (Ciabrelli et al., 2017). It will be important to define the extent to which this phenomenon can regulate endogenous genome function and its underlying molecular mechanisms, including the interplay with other components, such as ncRNAs, that have been linked to transgenerational inheritance (Sharma, 2017; Sharma et al., 2016).

Role in Stem Cell Renewal and Differentiation

While mutations of PcG proteins lead to homeotic transformations and other developmental defects in *Drosophila*, knockout of most PRC2 or PRC1 subunits in mice causes embryonic lethality at the gastrulation stage (Aloia et al., 2013). These genetic studies highlight a key role of PcG proteins in the specification and maintenance of cell fate rather than defects in the viability of early pluripotent cells (Montgomery et al., 2005). Indeed, genome-wide mapping of PRC1 and PRC2 subunits in pluripotent stem cells (Morey et al., 2012), more differentiated progenitors, and terminally differentiated cells (Morey et al., 2015) supports the key role of PcG proteins in cell identity and the maintenance of a proper differentiation program during development. In both embryonic and adult stem cells, a large number of PcG-bound regions are decorated by both the H3K27me3 repressive mark and the H3K4me3 activation mark (Shema et al., 2016). Those modifications, which are catalyzed by PRC2 and the MLL2 complex (Piunti and Shilatifard, 2016), respectively, contribute to set the promoter of genes implicated in cell-fate determination and development into a poised bivalent state. Such ''bivalent promoters'' are transcribed at very low levels and can be either activated or repressed, depending on developmental signals. How the correct balance between PcG and MLL2 occupancy is regulated at bivalent regions is still unclear, and the contributions of histone post-translational modifications (PTMs) in modulating promoter activity and chromatin organization of bivalent regions remain to be determined.

Bivalent regions and, more generally, the set of PcG-repressed genes differ depending on the state of the cells. In pluripotent ESCs, PcG complexes repress all lineage-specific genes, while genes necessary to maintain the proliferative and undifferentiated state of ESCs are continuously transcribed. Repression of specific lineage genes is relieved when ESCs are induced to differentiate. This occurs concomitantly with the silencing of pluripotency genes. Under this condition, PcG proteins still maintain the repression of alternative cell-fate genes, thus allowing cells to proceed to one specific cell type and preventing cells from de-differentiation or trans-differentiation. This exquisite regulation and the genome-wide redistribution of both PRC1 and PRC2 relies on the assembly and disassembly of PcG

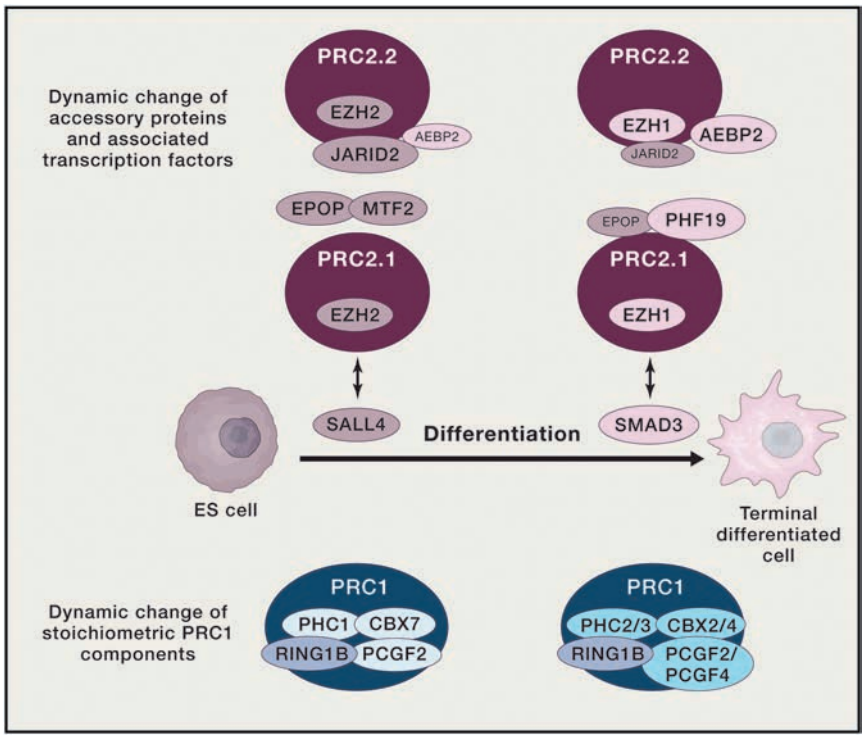


Figure 6. Dynamic Assembly of PRC1 and PRC2 during ES Cell Differentiation

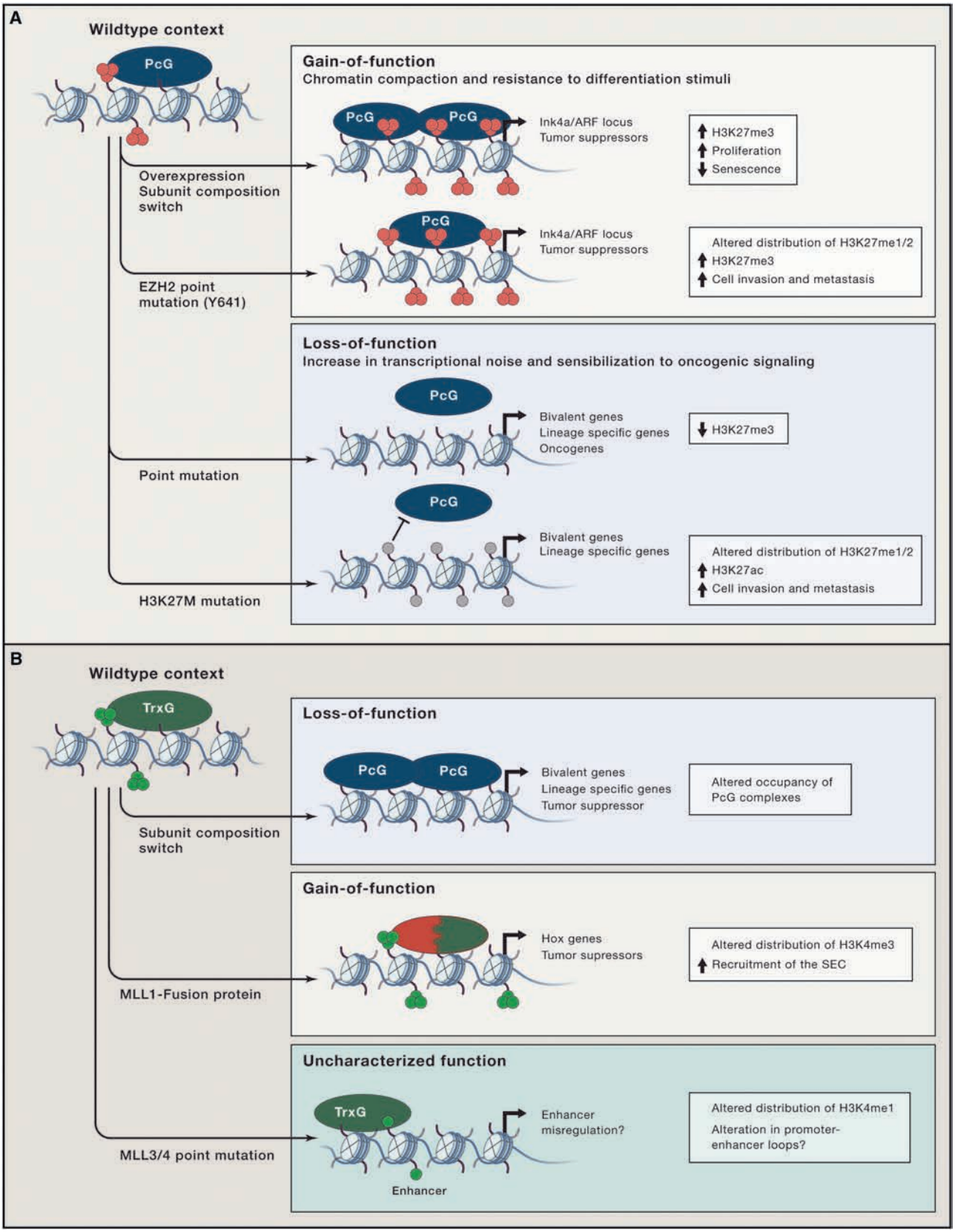
Different paralogs of PRC1 core components, such as PCGF or CBX proteins, are incorporated into PRC1 in ESCs or differentiated cells to potentially regulate PRC1 function. Similarly, different ancillary proteins assemble with PRC2 in ESCs or differentiated cells to modulate its function and localization during differentiation.

complexes for the proper balance between pluripotency and lineage commitment (Figure 6). In pluripotent ESCs, CBX7 is the most abundant CBX family member. To allow differentiation, CBX7 has to be replaced by other paralogous family members, such as CBX2, CBX4, and CBX8. Failure to replace CBX7 with other CBX family members causes hyper-proliferation of hematopoietic stem cells and thus leukemia (Klauke et al., 2013), while the function of CBX6 in gene regulation and in ESC differentiation is still enigmatic. Recently, several reports unveiled the specific role of PCGF family members in ESC maintenance (Endoh et al., 2017), mesodermal differentiation (Morey et al., 2015), and reprogramming (Zdzieblo et al., 2014). Thus, variation in assembly contributes to the dynamic targeting of PRC1 to different promoters (via CBX proteins) and provides specific biochemical functions that can regulate cell identity and cell-fate choices (via PCGF proteins). We know much less about how this occurs in adult stem cells or in differentiated cell types. In the past few years, several studies demonstrated an essential role for PRC1 and PRC2 in the patterning and homeostasis of the adult mouse intestine, skin, craniofacial structures, and hematopoietic systems (Chiacchiera et al., 2016; Ezhkova et al., 2009; Hidalgo et al., 2012; Minoux et al., 2017; Xie et al., 2014). Indeed, deletion of PcG proteins compromises the self-renewal capacity of adult stem cells, often triggering senescence and unscheduled activation of the differentiation program.

The situation is a bit simpler for the PRC2 core components, with an exchange of EZH2 with EZH1 during the switch between proliferation and differentiation. Although EZH1 can partially complement EZH2 functions in ESCs (Shen et al., 2008), biochemical data suggest that EZH2-PRC2 is enzymatically more active, while EZH1-PRC2 is more efficient in chromatin compaction assays (Margueron et al., 2008). However, most of the variation of the PRC2 complex relies on its dynamic interactions with accessory proteins (Figure 6); whereas EPOP, MTF2, PHF19, and JARID2 are more strongly enriched in undifferentiated cells, PHF1 and AEBP2 are preferentially associated with PRC2 core components in differentiated cells (Kloet et al., 2016; Oliviero et al., 2016). Clearly, these proteins can contribute to PRC2 targeting during cell-fate transitions via direct recognition of histone modifications (Ballaré et al., 2012) or by providing affinity for DNA sequences (Pasini et al., 2010). Indeed, depletion of PCLs or of JARID2 compromises the association of PRC2 with specific target genes as well as ESC differentiation. In contrast, EPOP, which is only present in the PCL-PRC2 complex, does not affect PRC2 occupancy (Beringer et al., 2016; Liefke et al., 2016). Together, these data suggest a scenario in which EPOP/PCLs and JARID2 associate with PRC2 in a mutually exclusive fashion (Grijzenhout et al., 2016; Hauri et al., 2016). Interestingly, EPOP also brings Elongins B/C into the PCL-PRC2 complex, opening the possibility of direct communication between PRC2 and the Pol II machinery. Since the expression pattern of PRC2-associated factors varies dramatically among different cell types, this adds a further layer of complexity to the assembly and function of the PRC2 complexes. To date, it is unclear how the expression of PRC1 and PRC2 core components and accessory proteins is regulated in different cell types or how the assembly of the different PcG variations is regulated. It is likely that post-translational modifications play an important role.

PcG and TrxG Proteins in Cancer

The role of PcG proteins in both cell identity and development is mainly achieved by the transcription control of pluripotency genes and linage-specific genes. Several years ago, van Lohuizen and coworkers reported that PcG can cooperate with c-MYC to generate mouse lymphomas via direct silencing of the *CDKN2a* locus (Jacobs et al., 1999a), thus suggesting that PcG proteins can play an oncogenic role by modulating cell proliferation and senescence. Later, the Chinnaiyan lab showed that elevated levels of EZH2 and the H3K27me3 mark in prostate tumors often correlates with a poor prognosis (Varambally et al., 2002). Along the same line, several subunits of PRC1 and PRC2 have been found to be overexpressed in both solid



(legend on next page)

cancers and leukemias and to be required for proliferation of cancer cells. This was then followed up by the identification of mutations in the demethylase UTX (KDM6A) in myeloma and renal cell carcinoma. Finally, a gain-of-function mono-allelic mutation in the catalytic pocket of EZH2 (Y641) has also been identified in lymphoma (Morin et al., 2010). The missense mutation renders the EZH2 Y641 incapable of either mono- or dimethylating H3K27, but it enhances the subsequent catalytic step (i.e., trimethylation to H3K27me3). Mechanistically, EZH2 expressed from the wild-type allele provides the enzymatic activity for the initial conversion of unmethylated H3K27 into H3K27me1/2, while mutated EZH2 activity leads to aberrantly high global levels of H3K27me3 in tumor samples. All these data suggest that hypermethylation at H3K27 acts as a driver in several human cancers (reviewed in Comet et al., 2016) and that it can initiate cell invasion and metastasis (Figure 7A). Based on these observations, several small-molecule inhibitors of EZH2 have been developed and are either being tested in clinical trials or already in use in clinical practice (Helin and Dhanak, 2013).

Unexpectedly, inactivating mutations in EZH2, EED, and SUZ12 have also been reported in myelodysplastic syndrome and malignant peripheral nerve sheath tumors (MPNST) (Lee et al., 2014), thus challenging the concept that PRC2 acts as an oncogene. Indeed, these loss-of-function mutations lead to reduced levels of H3K27me2/3 (Figure 7A). Likewise, a lysine-to-methionine mutation at position 27 of histone H3 (H3K27M) has been identified in diffuse intrinsic pontine glioma (DIPG). This mutation also results in reduced levels of H3K27me2/3 (Lewis et al., 2013). Although this was initially correlated to the capacity of H3K27me2/3 to strongly bind and sequester the PRC2 complex, recent data seem to support a model in which H3K27M excludes PRC2 binding, facilitating an aberrant accumulation of acetylated H3K27 (Figure 7A) on the other wild-type copies of histone H3 in heterotypic nucleosomes (Piunti et al., 2017). This has important clinical implications, since inhibition of proteins that bind to acetylated lysines, such as members of the bromodomain (BRD) family of proteins, offer new therapeutic avenues for DIPG. Moreover, recent crystal structural analyses of PcG complexes have provided important information about the mechanisms of histone modification catalysis and specificity. These studies propose a mechanism for the oncogenic effects of H3K27M, which explains how H3K27me3 binding to the PRC2 subunit EED enhances PRC2 activity and suggests possibilities for designing drugs that could interfere with PRC2 function (Brooun et al., 2016; Jiao and Liu, 2015; Justin et al., 2016; McGinty et al., 2014).

Should EZH2 inhibitors thus be considered for cancer treatments? In other words, do PcGs function as oncogenes, tumor suppressors, or both, depending on the cellular context? It is likely that they can be both, leaving us with a highly complex puzzle of understanding the PcG function in each cellular content. What is clear is that an unbalance in cellular levels of H3K27me3 can have devastating effects, compromising both cellular identity and homeostasis. PcG loss of function leads not only to the reactivation of PcG-silenced genes but also to increased transcriptional noise (Kar et al., 2017), which might sensitize the cell to oncogenic signaling. In contrast, PcG gain of function causes diffuse genome compaction, which is likely to be unresponsive to differentiation stimuli, thereby promoting uncontrolled growth.

A further interesting aspect to consider is that TrxG proteins are also often mutated in cancer (Figure 7B). Several SWI/SNF subunits, including SNF5/SMARCB1 and ARID1, are highly mutated in ovarian clear cell carcinomas and malignant rhabdoid tumors (Zinzalla, 2016). Mutations in the ATPase domain of SMARCA4 (also known as BRG1) are frequent in renal, ovarian, medulloblastoma, rhabdoid, colorectal, and lung tumors and lead to increased PcG occupancy on unmethylated CGI promoters, indicating a role for the SWI/SNF complex in displacing PcG complexes from chromatin (reviewed in Kadoch and Crabtree, 2015). COMPASS members are also heavily mutated in cancer. *MLL* genes were originally identified in a region that has frequently translocations in leukemia, and ELL elongation factor was the first reported fusion partner of MLL1 (Ziemin-van der Poel et al., 1991). Surprisingly, in the chimeric MLL-ELL protein, the catalytic SET domain of MLL is not retained, suggesting that altered levels of H3K4 methylation are not implicated in leukemogenesis. These initial observations were corroborated by the characterization of several other MLL fusion partners also implicated in transcription elongation, including AFF1, AF9, ELL, and ENL. Biochemical studies from the Shilatifard laboratory demonstrated that aberrant activation via the recruitment of the super elongation complex (SEC) is a common theme for all MLL translocations (Smith et al., 2011), thus exposing a potential Achilles' heel for treatment. In contrast, the other family members are often mutated rather than translocated in human tumors. *MLL4* missense mutations have been found in lymphomas and medulloblastomas, while *MLL3* mutations have been reported in bladder and kidney neoplasia (Piunti and Shilatifard, 2016). Given the dedicated role of MLL3 and MLL4 in methylating H3K4 at enhancers, it is intriguing to speculate that enhancer malfunction is the direct cause of these pathologies (Figure 7B). However, the underlying mechanisms are still unclear. Alterations in the chromatin loops between those enhancers affected by *MLL3/4* mutations with gene promoters are a likely possibility. Identifying direct

Figure 7. The Function of PcG and TrxG Proteins in Cancer

A summary of common mutations in PcG and TrxG subunits that affect gene expression and cellular functions.

(A) Gain-of-function mutations in PcG due to overexpression and changes in PcG complex composition or due to dominant-negative point mutation in the catalytic subunit (top) often leads to aberrant deposition of H3K27me3 at tumor suppressor loci. Loss-of-function mutations, such as missense mutations (bottom) or mutation in lysine 27 of histone H3, cause reactivation of PcG-repressed genes through depletion of H3K27me3 and accumulation of H3K27ac at lineage-specific genes and oncogenes.

(B) Similarly, deletion, mutation, and truncation of TrxG subunits results in reduced stability and/or activity of TrxG complexes, which in turn affects PcG genome-wide distribution (top). Gain of function, such as chromosome translocations, also are implicated in carcinogenesis processes (middle) through aberrant recruitment of the super elongation complex (SEC). Recently identified mutations in MLL3/4 (bottom panel) could possibly results in changes in enhancers' histone modifications and thus affect interactions with promoters.

target genes using Hi-C technology might shed some light on this.

The roles of PRC1 components in cancer are likely to reflect their functions in stem cells. CBX7 overexpression has been reported in human leukemias, prostate cancer, and ovarian clear cell carcinoma. Results obtained using animal models indicate that it acts as an oncogene by silencing the *INK4A/ARF* locus and preventing senescence and/or apoptosis. Conversely, no clear functional data have been reported so far for an oncogenic role of the other CBX paralogs, although correlative studies have shown their misexpression in several cancers. It is important to note that CBX family members have different affinity for H3K27me3, with CBX7 having the highest binding capability (Bernstein et al., 2006b). Other CBX proteins have been shown to recognize additional histone modifications as well as ncRNA. Understanding the exact mechanism of CBX chromatin recognition is a key challenge to properly understand the roles of this protein family in human cancer. Finally, specific inhibitors for each member could provide important tools to block cancer progression.

Similarly, BMI1 is highly expressed in hematopoietic stem cells but is replaced within the PRC1 complex by MEL18 during differentiation (Morey et al., 2015). Mechanistic insight provided by many studies suggests that BMI1 promotes cell proliferation by orchestrating the repression of *CDKN2*a and *PTEN* and by modulating the AKT pathway (Jacobs et al., 1999b), and its deletion in animal models confirms a key role in the initial steps of the carcinogenesis process. In contrast, MEL18 seems to act as a tumor suppressor by modulating the expression of several oncogenes, including *c-MYC* (Guo et al., 2007). Moreover, in several human tumors, *MEL18* is often downregulated, and its depletion in mouse models does not cause increased tumor formation. Since BMI1 and MEL18 have opposite effects on PRC1 activity (Cao et al., 2005), it is tempting to speculate that the switch of subunits might affect the deposition of the ubiquitin group and therefore gene expression programs. Whether this is a key aspect behind the oncogenic differences between BMI1 and MEL18 remains to be demonstrated, especially since the impact of H2Aub on gene regulation is still unclear.

Conclusions and Future Perspectives

Despite the boom in the PcG and TrxG research field, many open questions remain. Among the future developments, we suggest six major areas to watch. (1) PcG and TrxG members have now gained a reputation as important players in development, physiology, and cancer etiology. We also now have a reasonably good inventory of their target genes in a subset of cell types. However, we do not know how the PcG and TrxG complexes are targeted at specific subsets of genes in each cell type or which targets are critical for the regulation of specific functions in each cell type. While classical research showed, for instance, that *Hox* genes are crucial for regulating the anteroposterior body plan in flies, much more work will have to be devoted to understand critical PcG- and TrxG-regulated processes in the other cases. (2) PcG and TrxG proteins have an increasing evolutionary complexity with the gradual acquisition of more paralogs in more complex species, but what are the functions of individual paralogs? Changing the balance between paralogs in different protein complexes could restrict, expand, or even change the repertoire of target genes, and hence the regulatory outputs, in different cell types or in response to various cues. (3) Over the last decade, a clear link between PcG and TrxG proteins and 3D genome regulation has been established, and 3D contact networks have been shown to be established between active or PcG-silenced genes. However, these findings almost always described correlations. Are the architectural functions required, or are they a consequence of gene regulatory processes? A combination of analytical studies using microscopy and 3C-type technologies, as well as dedicated genome engineering studies, will be required to disentangle causes and consequences. (4) Recent studies have begun to decipher the molecular mechanisms that allow repressive or active chromatin states to be propagated through cell division. Furthermore, it is becoming clear that some PcG and TrxG components can convey transmission of inheritance not only throughout mitosis but also across multiple generations of an organism. On the other hand, additional work suggests that these components not only are able to carry out maintenance functions but also could be used dynamically to change gene regulatory states. It will be important to understand how these dynamic functions differ from memory functions and how transitions between labile, reprogrammable states and more stable, transmissible states might be regulated. Furthermore, in a specific context, these proteins can switch their function. For instance, PcG proteins may be involved in gene activation. They can associate with active genes (Brookes et al., 2012; Frangini et al., 2013; Kaneko et al., 2013; Mousavi et al., 2012; van den Boom et al., 2016) and can activate gene expression via PcG protein association with specific cofactors, positive regulation of PolII activity, recruitment of co-activators, and modulation of chromosome architecture (Arora et al., 2015; Frangini et al., 2013; Gao et al., 2014; Hauri et al., 2016; Lv et al., 2016; Morey et al., 2015). Moreover, a recent report demonstrated that a subset of *Drosophila* PREs can function as developmental enhancers in vivo, activating transcription in specific spatial domains (Erceg et al., 2017). Whether this dual functionality of PREs is conserved in mammals is unclear, and much remains to be done in this field. (5) The points above require a full, in-depth understanding of how PcG and TrxG functions are regulated, such as by post-translational modifications of individual components or by regulation of their gene expression. While multiple post-translational modifications of these components have been described (Gambetta and Müller, 2014; Gonzalez et al., 2014; Niessen et al., 2009) and subunit switching has also been described in various systems and cell types, the mechanistic regulation of these processes has not been investigated in detail. (6) Finally, new and unexpected discoveries will surely be made and will have to be taken into consideration, such as the role of Polycomb components in mitochondrial function (Liu et al., 2009) and the cytoplasm (Bodega et al., 2017; Chen et al., 2017). Future research is thus likely to deepen and expand the understanding the biological functions of epigenetic components in normal physiology and disease to provide the community with new prognostic and diagnostic tools and, finally, allow for the development of new therapeutic treatments for many cancers and other diseases.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.002.

ACKNOWLEDGMENTS

We thank all members of the Cavalli and Di Croce laboratories for helpful discussions. We thank Veronica A. Raker for editing the manuscript. Work in the Di Croce laboratory is supported by grants from the Spanish "Ministerio de Educación y Ciencia" (BFU2016-75008-P), AGAUR, and La Marato TV3. Work in the Cavalli laboratory is supported by grants from the CNRS, the FP7 European Network of Excellence EpiGeneSys, the European Union's Horizon 2020 research and innovation programme (grant 676556 MuG), the Agence Nationale de la Recherche (ANR-15-CE12-0006-01), the Fondation pour la Recherche Médicale (N. 144186), INSERM and ITMO Cancer (MM&TT), and the Laboratory of Excellence EpiGenMed.

REFERENCES

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. *29*, 585–590.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–1084.

Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. Development *140*, 2525–2534.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS Biol. *8*, e1000453.

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. Nature *476*, 105–108.

Arora, M., Packard, C.Z., Banerjee, T., and Parvin, J.D. (2015). RING1A and BMI1 bookmark active genes via ubiquitination of chromatin-associated proteins. Nucleic Acids Res. *44*, 2136–2144.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol. Cell. Biol. *24*, 10470–10478.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat. Struct. Mol. Biol. *19*, 1257–1265.

Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M., and Cavalli, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. Genes Dev. *17*, 2406–2420.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature *427*, 164–167.

Beisel, C., Buness, A., Roustan-Espinosa, I.M., Koch, B., Schmitt, S., Haas, S.A., Hild, M., Katsuyama, T., and Paro, R. (2007). Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proc. Natl. Acad. Sci. USA *104*, 16615–16620.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res. *26*, 896–907.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizán, P., Gutiérrez, A., Aranda, S., Payer, B., Wierer, M., and Di Croce, L. (2016). EPOP functionally links Elongin and Polycomb in pluripotent stem cells. Mol. Cell *64*, 645–658.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445–1459.

Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nat. Rev. Mol. Cell Biol. *16*, 643–649.

Blake, G.E., and Watson, E.D. (2016). Unravelling the complex mechanisms of transgenerational epigenetic inheritance. Curr. Opin. Chem. Biol. *33*, 101–107.

Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. Mol. Cell *36*, 970–983.

Bodega, B., Marasca, F., Ranzani, V., Cherubini, A., Della Valle, F., Neguembor, M.V., Wassef, M., Zippo, A., Lanzuolo, C., Pagani, M., and Orlando, V. (2017). A cytosolic Ezh1 isoform modulates a PRC2-Ezh1 epigenetic adaptive response in postmitotic cells. Nat. Struct. Mol. Biol. *24*, 444–452.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Brooun, A., Gajiwala, K.S., Deng, Y.L., Liu, W., Bolaños, B., Bingham, P., He, Y.A., Diehl, W., Grable, N., Kung, P.P., et al. (2016). Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. Nat. Commun. *7*, 11384.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. J. Cell Biol. *141*, 469–481.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol. Cell *20*, 845–854.

Cao, Q., Wang, X., Zhao, M., Yang, R., Malik, R., Qiao, Y., Poliakov, A., Yocum, A.K., Li, Y., Chen, W., et al. (2014). The central role of EED in the orchestration of polycomb group complexes. Nat. Commun. *5*, 3127.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505–518.

Cerase, A., Smeets, D., Tang, Y.A., Gdula, M., Kraus, F., Spivakov, M., Moindrot, B., Leleu, M., Tattermusch, A., Demmerle, J., et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc. Natl. Acad. Sci. USA *111*, 2235–2240.

Chen, R., Kong, P., Zhang, F., Shu, Y.N., Nie, X., Dong, L.H., Lin, Y.L., Xie, X.L., Zhao, L.L., Zhang, X.J., and Han, M. (2017). EZH2-mediated α-actin

methylation needs lncRNA TUG1, and promotes the cortex cytoskeleton formation in VSMCs. Gene *616*, 52–57.

Chiacchiera, F., Rossi, A., Jammula, S., Piunti, A., Scelfo, A., Ordóñez-Morán, P., Huelsken, J., Koseki, H., and Pasini, D. (2016). Polycomb complex PRC1 preserves intestinal stem cell identity by sustaining Wnt/β-catenin transcriptional activity. Cell Stem Cell *18*, 91–103.

Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuéreb, A., Klopp, C., Aravin, A., Paro, R., et al. (2017). Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat. Genet. *49*, 876–886.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, e00005.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. Mol. Cell *55*, 171–185.

Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science *356*, eaai8236.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nat. Rev. Cancer *16*, 803–810.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep. *7*, 1456–1470.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat. Commun. *7*, 13661.

Crevillén, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X., and Dean, C. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. Nature *515*, 587–590.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA *21*, 2007–2022.

Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., and King, N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. Dev. Biol. *357*, 73–82.

Dejardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. EMBO J. *23*, 857–868.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602–616.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

Dingwall, A.K., Beek, S.J., McCallum, C.M., Tamkun, J.W., Kalpana, G.V., Goff, S.P., and Scott, M.P. (1995). The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. Mol. Biol. Cell *6*, 777–791.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D., and Evans, G.A. (1992). A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. Nat. Genet. *2*, 113–118.

Dumesic, P.A., Homer, C.M., Moresco, J.J., Pack, L.R., Shanle, E.K., Coyle, S.M., Strahl, B.D., Fujimori, D.G., Yates, J.R., 3rd, and Madhani, H.D. (2015). Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell *160*, 204–218.

Eissenberg, J.C., Lee, M.G., Schneider, J., Ilvarsonn, A., Shiekhattar, R., and Shilatifard, A. (2007). The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. Nat. Struct. Mol. Biol. *14*, 344–346.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res. *21*, 216–226.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. eLife *6*, e27970.

Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of genome architecture and function by Polycomb proteins. Trends Cell Biol. *26*, 511–525.

Erceg, J., Pakozdi, T., Marco-Ferreres, R., Ghavi-Helm, Y., Girardot, C., Bracken, A.P., and Furlong, E.E. (2017). Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. Genes Dev. *31*, 590–602.

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. Proc. Natl. Acad. Sci. USA *112*, 14930–14935.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122–1135.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

Fauvarque, M.O., and Dura, J.M. (1993). Polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. Genes Dev. *7*, 1508–1520.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol. Cell *53*, 49–62.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Follmer, N.E., Wani, A.H., and Francis, N.J. (2012). A polycomb group protein is retained at specific sites on chromatin in mitosis. PLoS Genet. *8*, e1003135.

Fonseca, J.P., Steffen, P.A., Müller, S., Lu, J., Sawicka, A., Seiser, C., and Ringrose, L. (2012). In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. Genes Dev. *26*, 857–871.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110–122.

Frangini, A., Sjöberg, M., Roman-Trufero, M., Dharmalingam, G., Haberle, V., Bartke, T., Lenhard, B., Malumbres, M., Vidal, M., and Dillon, N. (2013). The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. Mol. Cell *51*, 647–661.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. Genes Dev. *30*, 1116–1127.

Gambetta, M.C., and Müller, J. (2014). O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. Dev. Cell *31*, 629–639.

Gambetta, M.C., Oktaba, K., and Müller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science *325*, 93–96.

Gao, Z., Lee, P., Stafford, J.M., von Schimmelmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. Nature *516*, 349–354.

Gaydos, L.J., Wang, W., and Strome, S. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science *345*, 1515–1518.

Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T., and Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. Mol. Cell *54*, 485–499.

Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev. *25*, 2210–2221.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development *143*, 2716–2723.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006a). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957–971.

Grimaud, C., Nègre, N., and Cavalli, G. (2006b). From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome Res. *14*, 363–375.

Guo, W.J., Datta, S., Band, V., and Dimri, G.P. (2007). Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. Mol. Biol. Cell *18*, 536–546.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A high-density map for navigating the human Polycomb complexome. Cell Rep. *17*, 583–595.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109.

Helin, K., and Dhanak, D. (2013). Chromatin proteins and modifications as drug targets. Nature *502*, 480–488.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., et al. (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. Nat. Genet. *46*, 973–981.

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nuñez, J., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*, 649–662.

Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. Cold Spring Harb. Perspect. Med. *6*, a026930.

Hollick, J.B. (2017). Paramutation and related phenomena in diverse species. Nat. Rev. Genet. *18*, 5–23.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Hu, D., Gao, X., Cao, K., Morgan, M.A., Mas, G., Smith, E.R., Volk, A.G., Bartom, E.T., Crispino, J.D., Di Croce, L., et al. (2017). Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. Mol. Cell *65*, 460–475.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. *29*, 1897–1902.

Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. Nature *306*, 591–593.

Ingham, P.W. (1985a). A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. J. Embryol. Exp. Morphol. *89*, 349–365.

Ingham, P.W. (1985b). Genetic control of the spatial pattern of selector gene expression in Drosophila. Cold Spring Harb. Symp. Quant. Biol. *50*, 201–208.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev. Cell *26*, 565–577.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999a). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature *397*, 164–168.

Jacobs, J.J., Scheijen, B., Voncken, J.W., Kieboom, K., Berns, A., and van Lohuizen, M. (1999b). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. Genes Dev. *13*, 2678–2690.

Jamieson, K., Rountree, M.R., Lewis, Z.A., Stajich, J.E., and Selker, E.U. (2013). Regional control of histone H3 lysine 27 methylation in Neurospora. Proc. Natl. Acad. Sci. USA *110*, 6027–6032.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell *167*, 1369–1384.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383.

Joshi, O., Wang, S.Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P.J., Habibi, E., Shaik, J., Saeed, S., Handoko, L., et al. (2015). Dynamic reorganization of extremely long-range promoter-promoter interactions between two states of pluripotency. Cell Stem Cell *17*, 748–757.

Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. Cell Rep. *17*, 2715–2723.

Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D., et al. (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat. Commun. *7*, 11316.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci. Adv. *1*, e1500447.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat. Genet. *49*, 213–222.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. PLoS Genet. *10*, e1004495.

Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb response elements. Nucleic Acids Res. *44*, 10132–10149.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat. Struct. Mol. Biol. *21*, 569–571.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. (2015). Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev. *29*, 1136–1150.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

Kar, G., Kim, J.K., Kolodziejczyk, A.A., Natarajan, K.N., Torlai Triglia, E., Mifsud, B., Elderkin, S., Marioni, J.C., Pombo, A., and Teichmann, S.A. (2017). Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat. Commun. *8*, 36.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. USA *85*, 8136–8140.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature *366*, 170–174.

Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat. Cell Biol. *15*, 353–362.

Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuijsen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat. Struct. Mol. Biol. *23*, 682–690.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling–an emerging perspective on targeting polycomb repressor proteins. PLoS Genet. *9*, e1003717.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5*, 373–377.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. Mol. Cell *65*, 432–446.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. *22*, 2799–2810.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat. Cell Biol. *9*, 1167–1174.

Laprell, F., Finkl, K., and Müller, J. (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. Science *356*, 85–88.

Lau, M.S., Schwartz, M.G., Kundu, S., Savol, A.J., Wang, P.I., Marr, S.K., Grau, D.J., Schorderet, P., Sadreyev, R.I., Tabin, C.J., and Kingston, R.E. (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science *355*, 1081–1084.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007a). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447–450.

Lee, N., Zhang, J., Klose, R.J., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2007b). The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. Nat. Struct. Mol. Biol. *14*, 341–343.

Lee, W., Teckie, S., Wiesner, T., Ran, L., Prieto Granada, C.N., Lin, M., Zhu, S., Cao, Z., Liang, Y., Sboner, A., et al. (2014). PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nat. Genet. *46*, 1227–1232.

Lee, H.G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. Genome Res. *25*, 1170–1181.

Lengsfeld, B.M., Berry, K.N., Ghosh, S., Takahashi, M., and Francis, N.J. (2012). A Polycomb complex remains bound through DNA replication in the absence of other eukaryotic proteins. Sci. Rep. *2*, 661.

Lewis, P.H. (1947). New mutants report. Drosoph. Inf. Serv. *21*, 69.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857–861.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP interacts with Elongin BC and USP7 to modulate the chromatin landscape. Mol. Cell *64*, 659–672.

Liu, J., Cao, L., Chen, J., Song, S., Lee, I.H., Quijano, C., Liu, H., Keyvanfar, K., Chen, H., Cao, L.Y., et al. (2009). Bmi1 regulates mitochondrial function and the DNA damage response pathway. Nature *459*, 387–392.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012). A bridging model for persistence of a polycomb group protein complex through DNA replication in vitro. Mol. Cell *46*, 784–796.

Loubière, V., Delest, A., Thomas, A., Bonev, B., Schuettengruber, B., Sati, S., Martinez, A.M., and Cavalli, G. (2016). Coordinate redeployment of PRC1 proteins suppresses tumor formation during Drosophila development. Nat. Genet. *48*, 1436–1442.

Lv, X., Han, Z., Chen, H., Yang, B., Yang, X., Xia, Y., Pan, C., Fu, L., Zhang, S., Han, H., et al. (2016). A positive role for polycomb in transcriptional regulation via H4K20me1. Cell Res. *26*, 529–542.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol. Cell *32*, 503–518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E., et al. (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). Proc. Natl. Acad. Sci. USA *109*, 2989–2994.

McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591–596.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G., and Allis,

C.D. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol. Cell *38*, 853–863.

Min, J., Zhang, Y., and Xu, R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. *17*, 1823–1828.

Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M.B., and Rijli, F.M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. Science *355*, eaal2913.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr. Biol. *15*, 942–947.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47–62.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E.P., Bruneau, B.G., and Di Croce, L. (2015). Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms. Cell Stem Cell *17*, 300–315.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat. Genet. *42*, 181–185.

Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol. Cell *45*, 255–262.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol. Cell *53*, 277–289.

Müller, J., and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. EMBO J. *10*, 3147–3155.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S., and Hess, J.L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell *17*, 609–621.

Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G., and Kutateladze, T.G. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. Nat. Commun. *4*, 2969.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M., and Ito, T. (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev. *22*, 37–49.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E.O., and Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev. *30*, 2657–2662.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J. *26*, 4078–4088.

Niessen, H.E., Demmers, J.A., and Voncken, J.W. (2009). Talking to chromatin: post-translational modulation of polycomb group function. Epigenetics Chromatin *2*, 10.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science *334*, 222–225.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Oliviero, G., Brien, G.L., Waston, A., Streubel, G., Jerman, E., Andrews, D., Doyle, B., Munawar, N., Wynne, K., Crean, J., et al. (2016). Dynamic protein interactions of the Polycomb repressive complex 2 during differentiation of pluripotent cells. Mol. Cell. Proteomics *15*, 3450–3460.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698–699.

Pengelly, A.R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes Dev. *29*, 1487–1492.

Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell *150*, 922–933.

Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science *352*, aad9780.

Piunti, A., Hashizume, R., Morgan, M.A., Bartom, E.T., Horbinski, C.M., Marshall, S.A., Rendleman, E.J., Ma, Q., Takahashi, Y.H., Woodfin, A.R., et al. (2017). Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. Nat. Med. *23*, 493–500.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. EMBO J. *36*, 981–994.

Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T.C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M., et al. (2012). Polycomb function during oogenesis is required for mouse embryonic development. Genes Dev. *26*, 920–932.

Poux, S., Horard, B., Sigrist, C.J., and Pirrotta, V. (2002). The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. Development *129*, 2483–2493.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rickels, R., Hu, D., Collings, C.K., Woodfin, A.R., Piunti, A., Mohan, M., Herz, H.M., Kvon, E., and Shilatifard, A. (2016). An evolutionary conserved epigenetic mark of Polycomb response elements implemented by Trx/MLL/COMPASS. Mol. Cell *63*, 318–328.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell *55*, 347–360.

Risca, V.I., Denny, S.K., Straight, A.F., and Greenleaf, W.J. (2017). Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. Nature *541*, 237–241.

Robinson, A.K., Leal, B.Z., Chadwell, L.V., Wang, R., Ilangovan, U., Kaur, Y., Junco, S.E., Schirf, V., Osmulski, P.A., Gaczynska, M., et al. (2012). The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. J. Biol. Chem. *287*, 8702–8713.

Rosa, S., De Lucia, F., Mylne, J.S., Zhu, D., Ohmido, N., Pendle, A., Kato, N., Shaw, P., and Dean, C. (2013). Physical clustering of FLC alleles during

Polycomb-mediated epigenetic silencing in vernalization. Genes Dev. *27*, 1845–1850.

Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. eLife *5*, e18591.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol. Cell. Biol. *28*, 2718–2731.

Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P., Otte, A.P., Sheer, D., and Freemont, P.S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J. Cell Biol. *142*, 887–898.

Schaaf, C.A., Misulovin, Z., Gause, M., Koenig, A., Gohara, D.W., Watson, A., and Dorsett, D. (2013). Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet. *9*, e1003560.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243–247.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. *7*, e13.

Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. Cell Rep. *9*, 219–233.

Schwartz, Y.B., and Cavalli, G. (2017). Three-dimensional genome organization and function in Drosophila. Genetics *205*, 5–24.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabidó, E., Gómez-Skarmeta, J.L., Di Croce, L., and Ruiz-Trillo, I. (2016). The dynamic regulatory genome of Capsaspora and the origin of animal multicellularity. Cell *165*, 1224–1237.

Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. Genes Dev. *21*, 537–551.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37–46.

Sharma, A. (2017). Transgenerational epigenetics: Integrating soma to germline communication with gametic inheritance. Mech. Ageing Dev. *163*, 15–22.

Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science *351*, 391–396.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O., and Bernstein, B.E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. Science *352*, 717–721.

Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol. Cell *32*, 491–502.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., Cohen, T., Xia, J., Suderman, M., Hallett, M., et al. (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science *350*, aab2006.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev. Biol. *158*, 131–144.

Singh, N.P., and Mishra, R.K. (2015). Specific combinations of boundary element and Polycomb response element are required for the regulation of the Hox genes in Drosophila melanogaster. Mech. Dev. *138*, 141–150.

Smallwood, S.A., and Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. Trends Genet. *28*, 33–42.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes Dev. *25*, 661–672.

Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M., Ku, W.L., Kadoch, C., Zhao, K., and Crabtree, G.R. (2017). Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat. Genet. *49*, 282–288.

Steffen, P.A., Fonseca, J.P., Gänger, C., Dworschak, E., Kockmann, T., Beisel, C., and Ringrose, L. (2013). Quantitative in vivo analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin. Nucleic Acids Res. *41*, 5235–5250.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. EMBO J. *4*, 3259–3264.

Surani, M.A. (2015). Human germline: a new research frontier. Stem Cell Reports *4*, 955–960.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell *68*, 561–572.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. USA *104*, 5521–5526.

Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G.E., Erkek, S., Kwon, T., Marcotte, E.M., Zegerman, P., Bradshaw, C.R., et al. (2016). Sperm is epigenetically programmed to regulate gene transcription in embryos. Genome Res. *26*, 1034–1046.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev. Cell *15*, 668–679.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. Proc. Natl. Acad. Sci. USA *113*, E744–E753.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat. Genet. *38*, 694–699.

van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W., Yuan, H., et al. (2016). Non-canonical PRC1.1 targets active genes independent of H3K27me3 and is essential for leukemogenesis. Cell Rep. *14*, 332–346.

van Lohuizen, M., Frasch, M., Wientjens, E., and Berns, A. (1991). Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2. Nature *353*, 353–355.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624–629.

Victor, M.B., Richner, M., Hermanstyne, T.O., Ransdell, J.L., Sobieski, C., Deng, P.Y., Klyachko, V.A., Nerbonne, J.M., and Yoo, A.S. (2014). Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. Neuron *84*, 311–323.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proc. Natl. Acad. Sci. USA *112*, 4672–4677.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637–646.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wani, A.H., Boettiger, A.N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R.I., Zhuang, X., Kingston, R.E., and Francis, N.J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat. Commun. *7*, 10291.

Wijchers, P.J., Krijger, P.H., Geeven, G., Zhu, Y., Denker, A., Verstegen, M.J., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., et al. (2016). Cause and consequence of tethering a SubTAD to different nuclear compartments. Mol. Cell *61*, 461–473.

Wilkinson, F., Pratt, H., and Atchison, M.L. (2010). PcG recruitment by the YY1 REPO domain can be mediated by Yaf2. J. Cell. Biochem. *109*, 478–486.

Xie, H., Xu, J., Hsu, J.H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S.H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell *14*, 68–80.

Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife *3*, e02046.

Yoo, A.S., Staahl, B.T., Chen, L., and Crabtree, G.R. (2009). MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. Nature *460*, 642–646.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. Nature *476*, 228–231.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971–975.

Zdzieblo, D., Li, X., Lin, Q., Zenke, M., Illich, D.J., Becker, M., and Müller, A.M. (2014). Pcgf6, a polycomb group protein, regulates mesodermal lineage differentiation in murine ESCs and functions in iPS reprogramming. Stem Cells *32*, 3112–3125.

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanović, O., and Iovino, N. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science *357*, 212–216.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting epigenetic memory by reprogramming of histone modifications in mammals. Mol. Cell *63*, 1066–1079.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol. Cell *29*, 69–80.

Zhu, L., Li, Q., Wong, S.H., Huang, M., Klein, B.J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., et al. (2016). ASH1L links histone H3 lysine 36 dimethylation to MLL leukemia. Cancer Discov. *6*, 770–783.

Ziemin-van der Poel, S., McCabe, N.R., Gill, H.J., Espinosa, R., 3rd, Patel, Y., Harden, A., Rubinelli, P., Smith, S.D., LeBeau, M.M., Rowley, J.D., et al. (1991). Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. Proc. Natl. Acad. Sci. USA *88*, 10735–10739.

Zinzalla, G. (2016). A new way forward in cancer drug discovery: inhibiting the SWI/SNF chromatin remodelling complex. ChemBioChem *17*, 677–682.

Cell

Reconstructing Prehistoric African Population Structure

Pontus Skoglund,[1,*] Jessica C. Thompson,[2] Mary E. Prendergast,[3] Alissa Mittnik,[4,5,33] Kendra Sirak,[2,6,33] Mateja Hajdinjak,[7,33] Tasneem Salie,[8,33] Nadin Rohland,[1] Swapan Mallick,[1,9] Alexander Peltzer,[4,10] Anja Heinze,[7] Iñigo Olalde,[1] Matthew Ferry,[1,11] Eadaoin Harney,[1,11,12] Megan Michel,[1,11] Kristin Stewardson,[1,11] Jessica I. Cerezo-Román,[13] Chrissy Chiumia,[14] Alison Crowther,[4,15] Elizabeth Gomani-Chindebvu,[14] Agness O. Gidna,[16] Katherine M. Grillo,[17] I. Taneli Helenius,[18] Garrett Hellenthal,[18] Richard Helm,[19] Mark Horton,[20] Saioa López,[18] Audax Z.P. Mabulla,[16] John Parkington,[21] Ceri Shipton,[22,23] Mark G. Thomas,[18] Ruth Tibesasa,[24] Menno Welling,[25,26] Vanessa M. Hayes,[27,28,29] Douglas J. Kennett,[30] Raj Ramesar,[8] Matthias Meyer,[7] Svante Pääbo,[7] Nick Patterson,[3,9] Alan G. Morris,[21] Nicole Boivin,[4] Ron Pinhasi,[6,31,32] Johannes Krause,[4,5,32] and David Reich[1,9,11,32,34,*]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Anthropology, Emory University, Atlanta, GA 30322, USA
[3]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA
[4]Max Planck Institute for the Science of Human History, Jena 07745, Germany
[5]Institute for Archeological Sciences, Eberhard-Karls-University, Tuebingen 72070, Germany
[6]School of Archaeology and Earth Institute, University College Dublin, Dublin 4, Ireland
[7]Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany
[8]Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
[9]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[10]Integrative Transcriptomics, Centre for Bioinformatics, University of Tuebingen, Tuebingen 72076, Germany
[11]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
[12]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[13]Department of Geography and Anthropology, California State Polytechnic University, Pomona, Pomona, CA 91768, USA
[14]Malawi Department of Museums and Monuments, Lilongwe 3, Malawi
[15]School of Social Science, The University of Queensland, Brisbane, Queensland 4072, Australia
[16]National Museums of Tanzania, Dar es Salaam, Tanzania
[17]Department of Archaeology and Anthropology, University of Wisconsin - La Crosse, La Crosse, WI 54601, USA
[18]Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
[19]Canterbury Archaeological Trust, Canterbury CT1 2LU, UK
[20]Department Archaeology and Anthropology, University of Bristol, Bristol BS8 1UU, UK
[21]Department of Archaeology, University of Cape Town, Cape Town 7700, South Africa

(Affiliations continued on next page)

SUMMARY

We assembled genome-wide data from 16 prehistoric Africans. We show that the anciently divergent lineage that comprises the primary ancestry of the southern African San had a wider distribution in the past, contributing approximately two-thirds of the ancestry of Malawi hunter-gatherers ~8,100–2,500 years ago and approximately one-third of the ancestry of Tanzanian hunter-gatherers ~1,400 years ago. We document how the spread of farmers from western Africa involved complete replacement of local hunter-gatherers in some regions, and we track the spread of herders by showing that the population of a ~3,100-year-old pastoralist from Tanzania contributed ancestry to people from northeastern to southern Africa, including a ~1,200-year-old southern African pastoralist. The deepest diversifications of African lineages were complex, involving either repeated gene flow among geographically disparate groups or a lineage more deeply diverging than that of the San contributing more to some western African populations than to others. We finally leverage ancient genomes to document episodes of natural selection in southern African populations.

INTRODUCTION

Africa harbors more genetic diversity than any other part of the world (Cann et al., 1987; Tishkoff et al., 2009). This is reflected both in a higher average number of differences among sub-Saharan African genomes than among non-African genomes (Cann et al., 1987; Ramachandran et al., 2005) and in the fact that the ancestry found outside of Africa is largely a subset of that within it (Tishkoff et al., 2009). Today, some of the earliest-branching African lineages are present only in populations with relatively small census sizes, including the southern African Khoe-San (see STAR Methods for terminology), central African rainforest hunter-gatherers, and Hadza of Tanzania (Gronau et al., 2011; Schlebusch et al., 2012; Veeramah et al., 2012). However, the population structure of Africa prior to the

[22]McDonald Institute for Archaeological Research, Cambridge CB2 3ER, UK
[23]British Institute in Eastern Africa, Nairobi 30710, Kenya
[24]Department of Anthropology and Archaeology, University of Pretoria, Pretoria 0083, South Africa
[25]African Studies Centre Leiden, Leiden University, Leiden 2300 RB, Netherlands
[26]African Heritage Ltd, Zomba, Malawi
[27]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[28]Central Clinical School, University of Sydney, Camperdown, NSW 2050, Australia
[29]School of Health Systems and Public Health, University of Pretoria, Gezina 0031, South Africa
[30]Department of Anthropology and Institutes for Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA
[31]Department of Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria
[32]Senior authors
[33]These authors contributed equally
[34]Lead Contact
*Correspondence: skoglund@genetics.med.harvard.edu (P.S.), reich@genetics.med.harvard.edu (D.R.)
http://dx.doi.org/10.1016/j.cell.2017.08.049

expansion of food producers (pastoralists and agriculturalists) remains unknown (Busby et al., 2016; Gurdasani et al., 2015; Patin et al., 2017). Bantu-speaking agriculturalists originating in western Africa are thought to have brought farming to eastern Africa by ~2,000 years BP (years before present, defined by convention as years before 1950 CE) and to southern Africa by ~1,500 BP, thereby spreading the largest single ancestry component to African genomes today (Russell et al., 2014; Tishkoff et al., 2009). Earlier migration(s), which brought ancestry related to the ancient Near East (Lazaridis et al., 2016; Pagani et al., 2012; Pickrell et al., 2014), brought herding to eastern Africa by ~4,000 BP (Marshall et al., 1984) and to southern Africa by ~2,000 BP (Sadr, 2015).

RESULTS

To reconstruct African population structure prior to the spread of food production, we generated new genome-wide data from 15 ancient sub-Saharan Africans (Table 1; Tables S1 and S2; STAR Methods). For three individuals from the western Cape of South Africa (~2,300–1,300 BP), we carried out direct shotgun sequencing to 0.7- to 2.0-fold coverage. For 12 individuals from eastern and south-central Africa, we used in-solution enrichment of ~1.2 million single nucleotide polymorphisms (SNPs). These included four individuals from the coastal region of Kenya and Tanzania (~1,400–400 BP), one from interior Tanzania (~3,100 BP), and seven from Malawi (ranging over ~8,100–2,500 BP) (Figure S1). All individuals had postmortem degradation characteristic of ancient DNA (Table 1), and we confirmed that key results are unlikely to be artifacts of contamination by restricting analysis to sequences with postmortem degradation (Skoglund et al., 2012; Skoglund et al., 2014a) (Figure S2). We merged the new ancient DNA data with previously reported shotgun sequence data from a ~4,500 BP Ethiopian highland individual (Llorente et al., 2015) and with SNP genotypes from 584 present-day African individuals from 59 diverse populations (including new data from 34 Malawi individuals; STAR Methods) (Lazaridis et al., 2014; Patterson et al., 2012), as well as 300 high-coverage genomes from 142 worldwide populations (Mallick et al., 2016). We also determined mitochondrial DNA and Y chromosome haplogroups for all newly reported samples (Tables S3 and S4).

An Ancient Cline of Southern and Eastern African Hunter-Gatherers

We used principal component analysis (PCA) (Patterson et al., 2006) and automated clustering (Alexander et al., 2009) to relate the 16 ancient individuals to present-day sub-Saharan Africans (Figures 1 and S3). Whereas the two individuals buried in ~2,000 BP hunter-gatherer contexts in South Africa share ancestry with southern African Khoe-San populations in the PCA, 11 of the 12 ancient individuals who lived in eastern and south-central Africa between ~8,100 and ~400 BP form a gradient of relatedness to the eastern African Hadza on the one hand and southern African Khoe-San on the other (Figure 1A). The genetic cline correlates to geography, running along a north-south axis with ancient individuals from Ethiopia (~4,500 BP), Kenya (~400 BP), Tanzania (both ~1,400 BP), and Malawi (~8,100–2,500 BP), showing increasing affinity to southern Africans (both ancient individuals and present-day Khoe-San). The seven individuals from Malawi show no clear heterogeneity, indicating a long-standing and distinctive population in ancient Malawi that persisted for at least ~5,000 years (the minimum span of our radiocarbon dates) but which no longer exists today.

We constructed a model where ancient and present-day African populations trace their ancestry to a putative set of nine ancestral populations. As proxies for these populations, three different ancient Near Eastern populations and six African populations were used that, according to our analyses, harbor substantial ancestry related to major lineages present in Africa today. The Mende from Sierra Leone are used in this model to represent a component of ancestry that exists in high proportions in western African populations, the ancient southern African genomes (South_Africa_2000BP) are used to represent the ancestry of southern Africa before agriculture, the Ethiopian individual (Ethiopia_4500BP) is used to represent northeastern African ancestry before agriculture, the Mbuti are used to represent central African rainforest hunter-gatherer ancestry, the individual from an eastern African pastoralist context (Tanzania_Luxmanda_3100BP) is used to represent an early pastoralist lineage from eastern Africa (see below; Figure 2), and the Dinka (from Sudan) are used to represent distinctive ancestry found in Nilotic speakers today. The ancient Near Eastern populations were representative of Anatolia, the Levant, and Iran,

Table 1. Summary of Ancient DNA from 15 Prehistoric Individuals Newly Reported in This Study

ID	Population label	Date (BP denotes calibrated years before present [defined as 1950], while "uncalibrated BP" denotes the radiocarbon age)*	Location	Lat.	Long.	Y chromosome haplogroup	mtDNA haplogroup	Damage rate at 5′ CpG dinucleotides	SNPs hit on 1.2M autosomal targets
I9028	South_Africa_2000BP	2241-1965 BP (2330 ± 25 uncalibrated BP, UGAMS-7255)	St. Helena, South Africa	−32.8	18.0	A1b1b2a	L0d2c1[$]	28%	731,098 (1.1X shotgun)
I9133	South_Africa_2000BP	2017-1748 BP (2000 ± 50 uncalibrated BP, Pta-5283)	Faraoskop Rock Shelter, South Africa	−32.0	18.5	A1b1b2a	L0d1b2b1b	33%	1,028,904 (2.3X shotgun)
I9134	South_Africa_1200BP	1282-1069 BP (1310 ± 50 uncalibrated BP, Pta-4373)	Kasteelberg, South Africa	−32.8	17.9	Female	L0d1a1a	21%	641,971 (0.8X shotgun)
I4427	Malawi_Fingira_6100BP	6175-5913 BP (5270 ± 25 uncalibrated BP, UCIAMS-186346)	Fingira, Malawi	−10.8	33.8	BT	L0d1b2b	37%	99,341
I4468	Malawi_Fingira_6100BP	6177-5923 BP (5290 ± 25 uncalibrated BP, UCIAMS-186347)	Fingira, Malawi	−10.8	33.8	BT	L0d1c	49%	30,257
I4421	Malawi_Chencherere_5200BP	5400-4800 BP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)	Chencherere, Malawi	−14.4	33.8	Female	L0k2	28%	59,470
I4422	Malawi_Chencherere_5200BP	5293-4979 BP (4525 ± 25 uncalibrated BP, UCIAMS-186348)	Chencherere, Malawi	−14.4	33.8	Female	L0k1	42%	9,355
I2966	Malawi_Hora_8100BP	10000-5000 BP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)	Hora, Malawi	−11.7	33.6	BT	L0k2 (PMDS > 3)	54%	610,605 (+0.26X shotgun)
I2967	Malawi_Hora_8100BP	8173-7957 BP (7230 ± 60 uncalibrated BP, PSUAMS-2535)	Hora, Malawi	−11.7	33.6	Female	L0a2	48%	65,686 (+0.03X shotgun)
I4426	Malawi_Fingira_2500BP	2676-2330 BP [2676-2343 BP (2425 ± 20 uncalibrated BP, PSUAMS-1734), 2483-2330 BP (2400 ± 20 uncalibrated BP, PSUAMS-1881)]	Fingira, Malawi	−10.8	33.8	Female	L0f	39%	635,427
I3726	Tanzania_Luxmanda_3100BP	3141-2890 BP (2925 ± 20 uncalibrated BP, ISGS-A3806)	Luxmanda, Tanzania	−4.3	35.3	Female	L2a1	65%	845,016
I0589	Tanzania_Zanzibar_1400BP	1370-1303 BP (1479 ± 23 uncalibrated BP, OxA-31427)	Kuumbi Cave, Zanzibar Island, Tanzania	−6.4	39.5	Female	L4b2a2c	22%	752,917
I1048	Tanzania_Pemba_1400BP	1421-1307 BP (1520 ± 30 uncalibrated BP, Beta-434912)	Makangale Cave, Pemba Island, Tanzania	−4.9	39.6	Female	L0a	42%	168,117
I2298	Tanzania_Pemba_600BP	639-544 BP (623 ± 20 uncalibrated BP, Wk-43308)	Makangale Cave, Pemba Island, Tanzania	−4.9	39.6	Female	L2a1a2	29%	695,242
I0595	Kenya_400BP	496-322 BP (388 ± 27 uncalibrated BP, OxA-30803)	Panga ya Saidi, Kenya	−3.7	39.7	E1b1b1b2	L4b2a2	30%	150,383

*Table S2 provides detailed information on the direct radiocarbon dating measurements.
[$]Consistent with previously published mtDNA sequence by Morris et al. (2014).

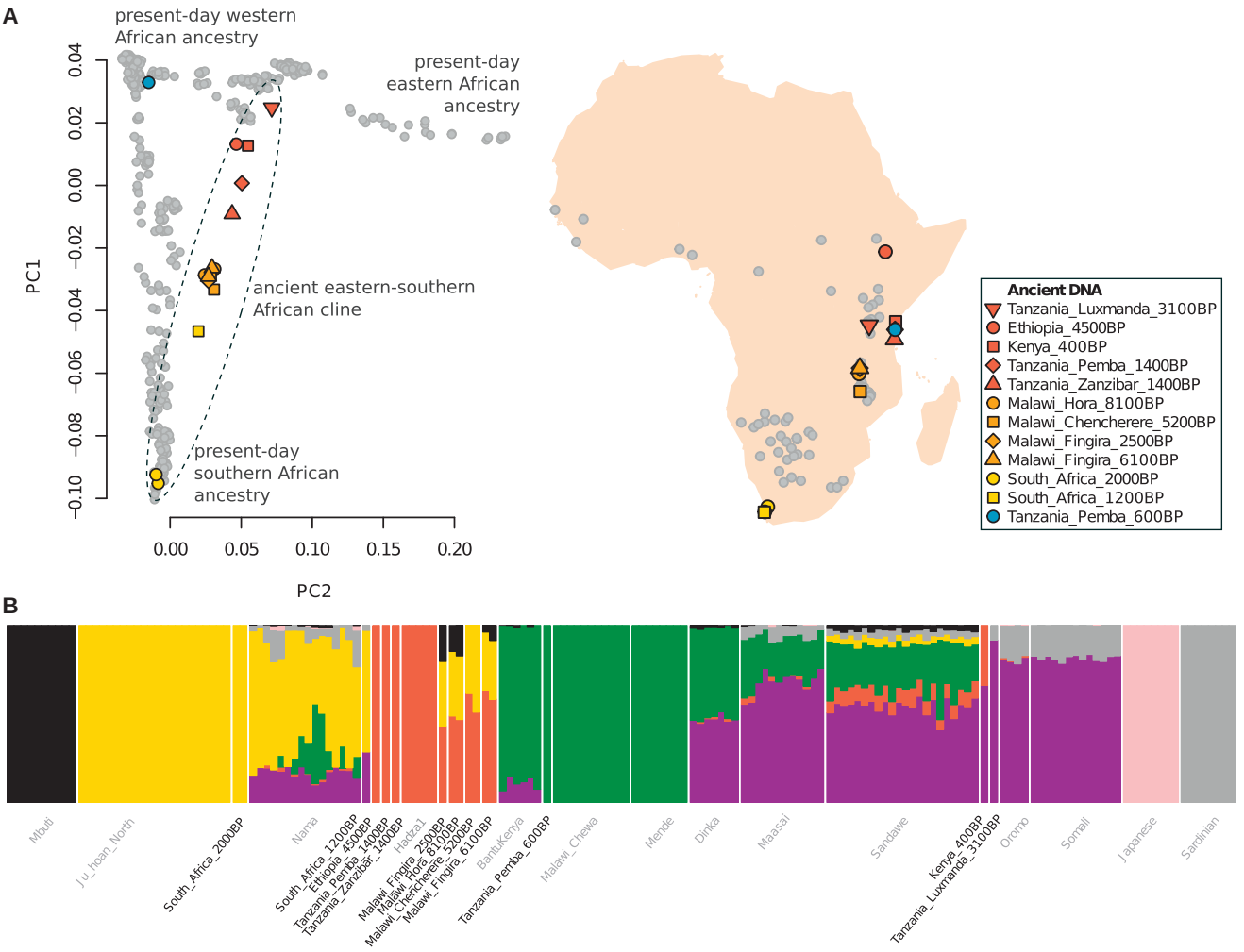


Figure 1. Overview of Ancient Genomes and African Population Structure

(A) Map of sampling locations in Africa and principal component analysis of all individuals. Present-day individuals are indicated with gray circles.

(B) Automated clustering of key ancient and present-day populations (for K = 7 cluster components). Present-day populations are labeled in gray.

respectively (Lazaridis et al., 2016; Mathieson et al., 2015). We used *qpAdm* (Haak et al., 2015), a generalization of f_4 symmetry statistics, to successively test 1-source, 2-source, or 3-source models and admixture proportions for all other ancient and present-day African populations, with a set of 10 non-African populations as outgroups (STAR Methods).

We find that ancestry closely related to the ancient southern Africans was present much farther north and east in the past than is apparent today. This ancient southern African ancestry comprises up to 91% of the ancestry of Khoe-San groups today (Table S5), and also 31% ± 3% of the ancestry of Tanzania_Zanzibar_1400BP, 60% ± 6% of the ancestry of Malawi_Fingira_6100BP, and 65% ± 3% of the ancestry of Malawi_Fingira_2500BP (Figure 2A). Notably, the Khoe-San-related ancestry in ancient individuals from Malawi and Tanzania is symmetrically related to the two previously identified lineages present in the San ($Z < 2$; Figure S2), estimated to have diverged at least 20,000 years ago (Mallick et al., 2016; Pickrell et al., 2012; Schlebusch et al., 2012), implying that this was an ancient divergent branch of this group that lived in eastern Africa at least until 1,400 BP. However, it was not present in all eastern Africans, as we do not detect it in the ~400-year-old individual from coastal Kenya nor in the present-day Hadza.

Displacement of Forager Populations in Eastern Africa

Both unsupervised clustering (Figure 1B) and formal ancestry estimation (Figure 2B) suggest that individuals from the Hadza group in Tanzania can be modeled as deriving all their ancestry from a lineage related deeply to ancient eastern Africans such as the Ethiopia_4500BP individual (Figure 3A; Table S5). However, this lineage appears to have contributed little ancestry to present-day Bantu speakers in eastern Africa, who instead trace their ancestry to a lineage related to present-day western Africans, with additional components related to the Nilotic-speaking Dinka and to the Tanzania_Luxmanda_3100BP

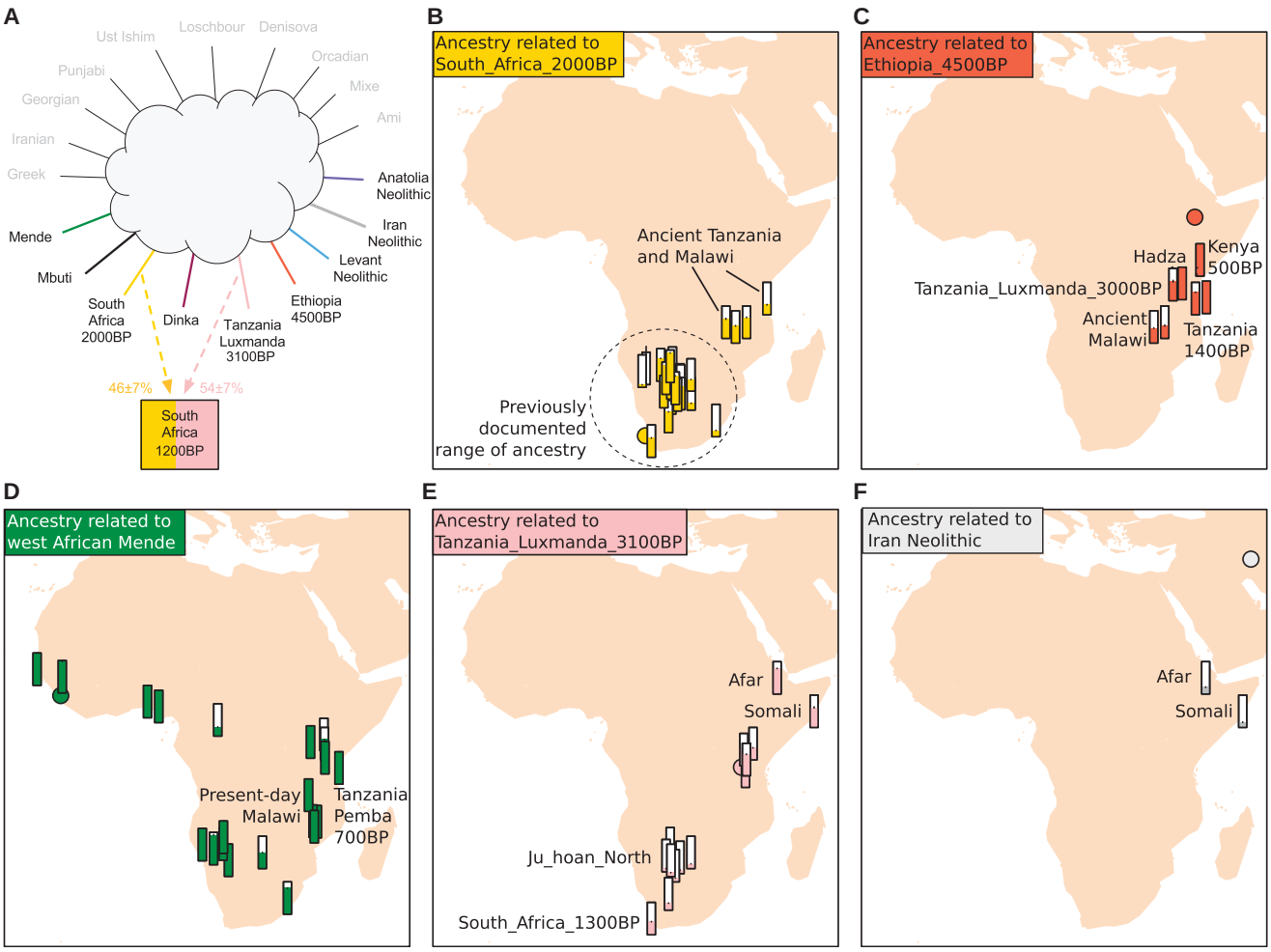


Figure 2. Ancestral Components in Eastern and Southern Africa

We show bar plots with the proportions inferred for the best model for each target population. We used a model that inferred the ancestry of each target population as 1-source, 2-source, or 3-source mixture of a set of potential source populations.

(A) We show an example of the inferred model for South_Africa_1200BP, an early pastoralist. A filled circle symbol in each panel indicates the geographic location of the sample that we use as a representative of the source population. We show five sources:

(B) South_Africa_2000BP, representing forager populations in southern Africa and a component of prehistoric Malawi and Tanzania that is no longer extant;

(C) Ethiopia_4500BP, which is today found in the Hadza but in the past was characteristic of eastern African hunter-gatherers;

(D) The Mende from Sierra Leone, which is related deeply to the western African ancestry that was spread with the Bantu expansion of agriculturalists;

(E) The Savanna Pastoral Neolithic sample Tanzania_Luxmanda_3100BP, a missing-link pastoralist population that almost certainly is responsible for having spread ancestry most closely related to the ancient Levant to southern Africa and which is also closely but not exclusively related to present-day Cushitic speakers;

(F) Ancestry more closely related to early Iranian farmers (Iran Neolithic) than what is found in Tanzania_Luxmanda_3100BP and which may have entered the Horn of Africa in later migrations.

pastoralist (see below; Figure 2). The Sandawe, another population that like the Hadza uses click consonants in their spoken language, are modeled as having ancestry similar to the Hadza but also admixture related to that of neighboring populations (Figure 3A; Table S5) consistent with previous findings (Henn et al., 2011; Tishkoff et al., 2009). Population replacement by incoming food producers appears to have been nearly complete in Malawi, where we detect little if any ancestry from the ancient individuals who lived ~8,100–2,500 BP. Instead, present-day Malawian individuals are consistent with deriving all their ancestry from the Bantu expansion of ultimate western African origin (Figure 3).

Among the ancient individuals analyzed here, only a ~600 BP individual from the Zanzibar archipelago has a genetic profile similar to present-day Bantu speakers (Figure 1). Notably, this individual has even more western-African-related ancestry than the present-day Bantu speakers we analyzed from Kenya, who also derive some of their ancestry from lineages related to Dinka and Tanzania_Luxmanda_3100BP (Figure 1B). Using linkage disequilibrium, we estimate that this admixture between western- and eastern-African-related lineages occurred an average of 800–400 years ago (STAR Methods). This suggests a scenario of genetic isolation between early farmers and

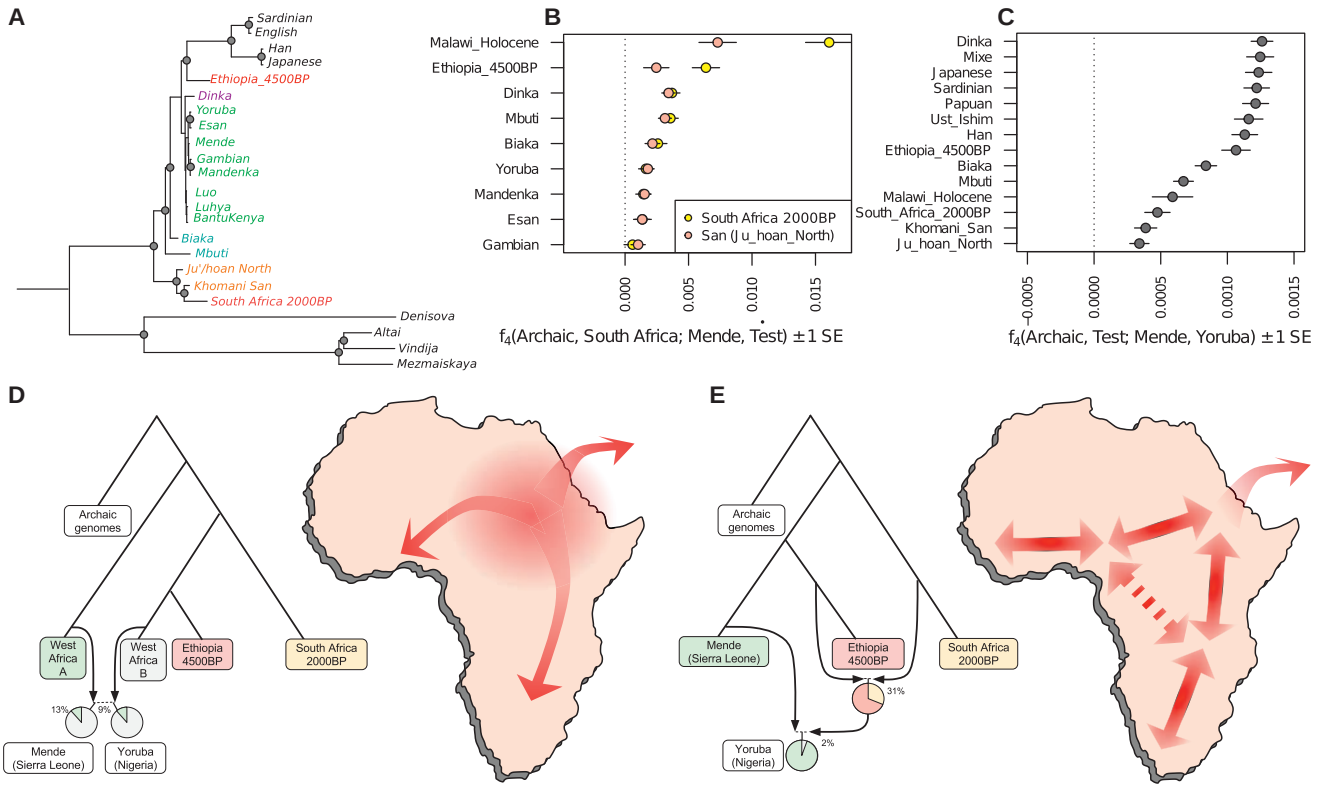


Figure 3. Mixture Events in the Deeper Population History of Continental African Lineages

(A) Maximum likelihood tree of genome sequences from present-day and ancient populations, excluding populations with evidence of asymmetrical allele sharing with non-Africans indicative of recent gene flow (Table S6). Nodes with bootstrap support >95% are indicated with a circle.

(B) A symmetry test of the hypothesis that ancient southern Africans are an outgroup lineage to other African populations, which can be rejected for most pairs.

(C) Asymmetry between western African Mende and Yoruba in the 1000 Genomes Project data is maximized in the Yoruba's excess affinity to eastern Africans and non-Africans but highly significant also for groups as distant as southern Africans.

(D) Admixture graph solution where Mende from Sierra Leone and Yoruba from Nigeria have ancestry from a basal western African lineage. The other source of western African ancestry is most closely related to eastern Africans and non-Africans (Figure S5D), which could be consistent with an expansion from eastern Africa. Note that the exact proportion "West Africa A" ancestry is not well constrained by the model, but the difference between Yoruba and Mende is highly significant (C).

(E) Admixture graph solution where the Yoruba have gene flow from a population related to both southern and eastern Africa, which could be consistent with a more complex pattern of isolation-by-distance on the continent.

previously established foragers during the initial phase of the Bantu expansion into eastern Africa (Crowther et al., 2017; Ribot et al., 2010), a barrier that broke down over time as mixture occurred. This parallels the patterns previously observed in genomic analyses of the Neolithic expansion into Europe (Haak et al., 2015; Skoglund et al., 2012) and the East Asian farming expansion into Remote Oceania (Skoglund et al., 2016). However, this process of delayed admixture did not always apply in Africa, as is evident in the absence of admixture from previously established hunter-gatherers in present-day Malawians.

Early Levantine Farmer-Related Admixture in an ~3,100-Year-Old Pastoralist from Tanzania

Western-Eurasian-related ancestry is pervasive in eastern Africa today (Pagani et al., 2012; Tishkoff et al., 2009), and the timing of this admixture has been estimated to be ~3,000 BP on average (Pickrell et al., 2014). We found that the ~3,100 BP individual (Tanzania_Luxmanda_3100BP), associated with a Savanna Pastoral Neolithic archeological tradition, could be modeled as having 38% ± 1% of her ancestry related to the nearly 10,000-year-old pre-pottery farmers of the Levant (Lazaridis et al., 2016), and we can exclude source populations related to early farmer populations in Iran and Anatolia. These results could be explained by migration into Africa from descendants of pre-pottery Levantine farmers or alternatively by a scenario in which both pre-pottery Levantine farmers and Tanzania_Luxmanda_3100BP descend from a common ancestral population that lived thousands of years earlier in Africa or the Near East. We fit the remaining approximately two-thirds of Tanzania_Luxmanda_3100BP as most closely related to the Ethiopia_4500BP ($p = 0.029$) or, allowing for three-way mixture, also from a source closely related to the Dinka ($p = 0.18$; the Levantine-related ancestry in this case was 39% ± 1%) (Table S4).

While these findings show that a Levant-Neolithic-related population made a critical contribution to the ancestry of present-day eastern Africans (Lazaridis et al., 2016), present-day Cushitic speakers such as the Somali cannot be fit simply as having Tanzania_Luxmanda_3100BP ancestry. The best fitting model for the Somali includes Tanzania_Luxmanda_3100BP ancestry, Dinka-related ancestry, and 16% ± 3% Iranian-Neolithic-related ancestry ($p = 0.015$). This suggests that ancestry related to the Iranian Neolithic appeared in eastern Africa after earlier gene flow related to Levant Neolithic populations, a scenario that is made more plausible by the genetic evidence of admixture of Iranian-Neolithic-related ancestry throughout the Levant by the time of the Bronze Age (Lazaridis et al., 2016) and in ancient Egypt by the Iron Age (Schuenemann et al., 2017).

Direct Evidence of Migration Bringing Pastoralism to Eastern and Southern Africa

In contrast to the Malawi and Zanzibar individuals, all three ancient southern Africans show affinities to the ancestry predominant in present-day Tuu speakers in the southern Kalahari more than to present-day Ju|'hoan speakers in the northern Kalahari (Figures S2B and S2C). However, the ~1,200 BP sample from the western Cape that is found in a pastoralist context has a specific similarity in clustering analyses to present-day Khoe-Khoe-speaking pastoralist populations such as the Nama (Figure 1B), and like them it has affinity to three groups: Khoe-San, western Eurasians, and eastern Africans. This supports the hypothesis that a non-Bantu-related population carried eastern African and Levantine ancestry to southern Africa by at least around 1,200 BP, providing direct evidence for claims previously made based on analysis of present-day populations (Pickrell et al., 2014).

We used our modeling framework to show that the South_Africa_1200BP pastoralist individual from the western Cape is consistent with being a mixture of just two streams of ancestry relative to non-southern African populations, with 40.3% ± 2.3% ancestry related to the Tanzania_Luxmanda_3100BP individual (54% ± 7% when restricting analysis to sequences with postmortem damage) and the remainder being related to the South_Africa_2000BP hunter-gatherers (Table S5). This supports the hypothesis that the Savanna Pastoral Neolithic archaeological tradition in eastern Africa is a plausible source for the spread of herding to southern Africa. Even the Ju|'hoan San group with the least genetic affinity to eastern Africans (Ju_hoan_North), have 9% ± 1% of their ancestry most closely related to Tanzania_Luxmanda_3100BP, consistent with previous findings that the ancestries of all present-day San and Khoe were affected by agro-pastoralist migrations in the last two millennia (Pickrell et al., 2014).

The Earliest Divergences among Modern Human Populations

Previous studies have suggested that the primary ancestry in the San is from a lineage that separated from all other lineages represented in modern humans today, before the latter separated from each other (Gronau et al., 2011; Veeramah et al., 2012). Such a model emerges when we automatically fit a tree without admixture to the data (Figure 3A), but we also find that a tree-like representation is a poor fit (Figure S4A), in the sense that ancient southern Africans who lived ~2,000 BP were not strictly an outgroup to extant lineages in other parts of sub-Saharan Africa. In particular, we find that ancient southern Africans, who have none of the eastern African admixture that is ubiquitous today, share significantly more alleles with present-day and ancient eastern Africans (including Dinka, Hadza, and Ethiopia_4500BP) than they do with present-day western Africans (Figure 3B; Table S6). Even within present-day western Africans, the genetic differences between Yoruba from Nigeria and the Mende from Sierra Leone are inconsistent with descent from a homogeneous ancestral population isolated from ancient southern Africans. The asymmetry between Yoruba and Mende is also observed with non-Africans but is no stronger than in eastern Africans (the most closely related Africans to the ancestral out-of-Africa population), and thus these signals are not driven by admixture from outside Africa and instead likely reflect demographic events entirely within Africa (Figure 3C; Table S6).

We carried out admixture graph modeling of the allele frequency correlations and found two parsimonious models that fit the data. The first posits that present-day western Africans harbor ancestry from a basal African lineage that contributed more to the Mende than it did to the Yoruba, with the other source of western African ancestry being related to eastern Africans and non-Africans (Figures 3D, S4, and S5; Table S7). The second model posits that long-range and long-standing gene flow has connected southern and eastern Africa to some groups in western Africa (e.g., the ancestors of the Yoruba) to a greater extent than to other groups in western Africa (e.g., the ancestors of the Mende) (Figure 3E) (Pleurdeau et al., 2012). The possible basal western African population lineage would represent the earliest known divergence of a modern human lineage that contributed a major proportion of ancestry to present-day humans. Such a lineage must have separated before the divergence of San ancestors, which is estimated to have begun on the order of 200–300 thousand years ago (Scally and Durbin, 2012). Such a model of basal western African ancestry might support the hypothesis that there has been ancient structure in the ancestry of present-day Africans, using a line of evidence independent from previous findings based on long haplotypes with deep divergences from other human haplotypes (Hammer et al., 2011; Lachance et al., 2012; Plagnol and Wall, 2006). One scenario consistent with this result could involve ancestry related to eastern Africans (and the out-of-Africa population) expanding into western Africa and mixing there with more basal lineages. Our genetic data do not support the theory that this putative basal lineage diverged prior to the ancestors of Neanderthals, since the African populations we analyze here are approximately symmetrically related to Neanderthals (Mallick et al., 2016; Prüfer et al., 2014).

A Selective Sweep Targeting a Taste Receptor Locus in Southern Africa

The availability of ancient African genomes provides an opportunity to search for genomic footprints of natural selection, manifested as regions of greater allele frequency differentiation between ancient and present-day populations than predicted by the genome-wide background. We compared the two ancient

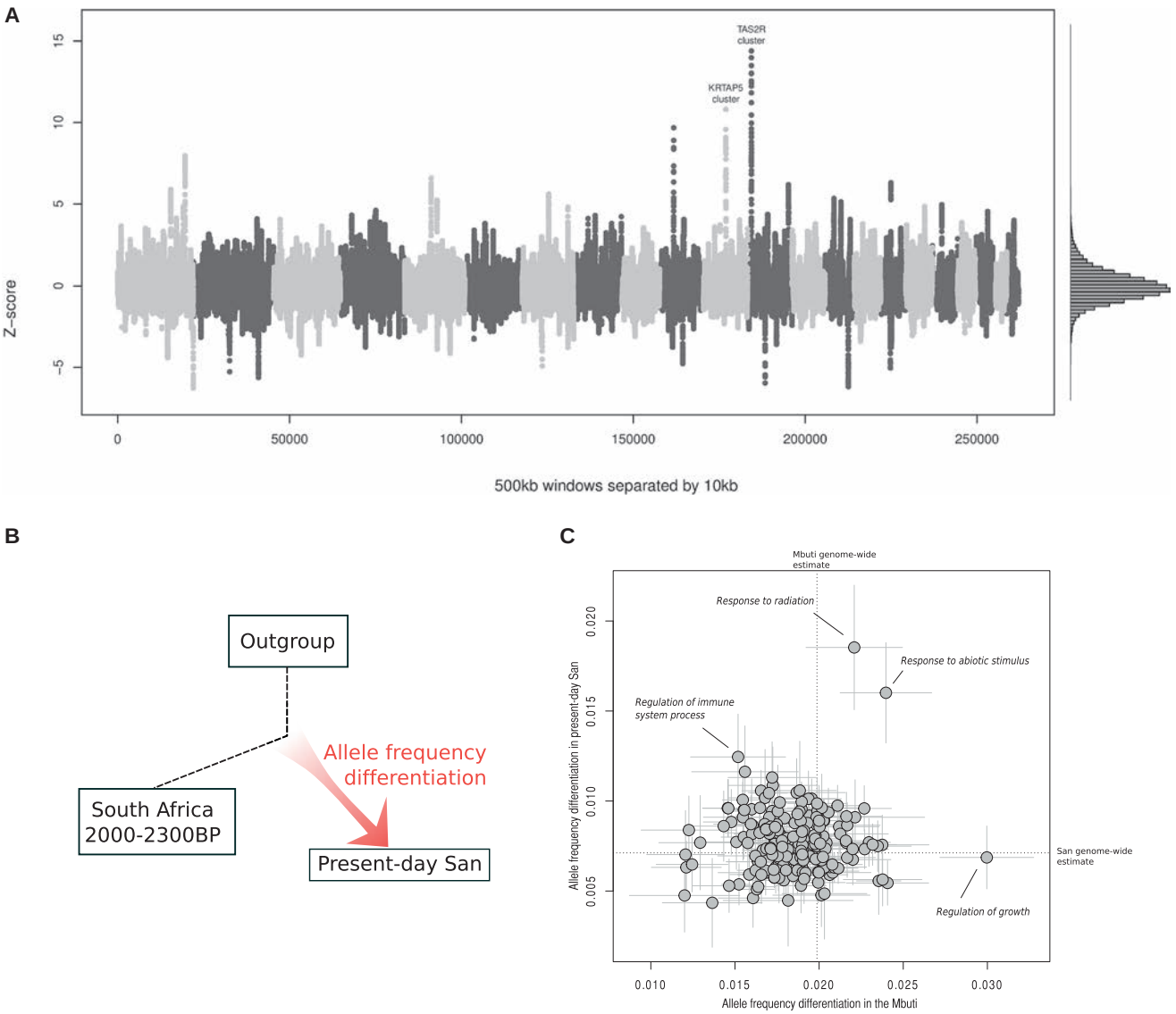


Figure 4. Ancient Genomes Provide Evidence of Natural Selection in Present-Day Southern African San Populations

We computed branch-specific allele frequency differentiation in six present-day high-coverage San genomes compared to a pool of two ~2,000 BP South African genomes as an outgroup using two approaches.

(A) We computed the statistic in windows of 500 kb separated by 10 kb. We also estimated genome-wide average and standard deviation of the statistic using windows separated by at least 5 Mb, and we transformed the genome-wide distribution of the sliding windows to be approximately normal (right). We observe outliers 15 standard deviations from the mean in a taste-receptor gene cluster on chromosome 12 and a secondary peak in the Keratin Associated Protein 4 gene cluster. The outgroup used was 4 Central African Mbuti genomes. See Table 2 for details on all major outlying regions.

(B) Illustration of the branch-specific allele frequency differentiation approach.

(C) y axis: We computed the statistic and block jackknife standard errors for 208 gene ontology categories with at least 50 genes each. The outgroup used was western Africans. x axis: As a control to confirm that outlier categories do not show larger magnitudes of allele frequency differentiation across populations, we replaced the present-day San with the central African Mbuti.

southern African ~2,000 BP shotgun sequence genomes to six present-day high-coverage San genomes with minimal recent mixture. The small number of ancient individuals does not permit inference of changing allele frequencies at single loci, so we performed a scan for high allele frequency differentiation in 500 kb windows with a step size of 10 kb. Using ~500 windows spaced at least 5 million base pairs apart as a null distribution, we found that the most differentiated locus was 15 standard deviations from the observed genome-wide mean and overlapped a cluster of eight taste-receptor genes on chromosome 12 (Figure 4A; Table 2). Taste-receptor genes have previously been identified as targets of natural selection in humans, as they modulate the ability to detect poisonous compounds in plants (Campbell et al., 2012).

Table 2. Top Five Candidate Regions Identified in Genome-wide Scan for Selective Sweeps in Present-Day San Populations in Southern Africa Compared to Ancient Genomes

Rank	Chrom.	start-end	f_3-statistic	Z-score	Genes in top 500 kb window in peak region
1	12	11,123,548-11,623,548	0.163	14.3	TAS2R43, PRH1-PRR4, TAS2R20, TAS2R50, TAS2R42, TAS2R46, TAS2R30, TAS2R31, PRB1, PRB2, PRB3, PRB4, LOC100129361, TAS2R19
2	11	71,208,258-71,708,258	0.125	10.8	LOC100129216, KRTAP5-7, DEFB108B, KRTAP5-8, NADSYN1, KRTAP5-9, FAM86C1, RNF121, ALG1L9P, LOC100133315, KRTAP5-10, KRTAP5-11
3	10	46,069,893-46,569,893	0.113	9.7	DQ577099, PTPN20B, PTPN20A, ZFAND4, AGAP4, DQ588224, FAM21C
4	1	224,960,062-225,460,062	0.095	8.0	DNAH14
5	5	82,375,629-82,875,629	0.08	6.6	VCAN, XRCC4

Polygenic Adaptation

Natural selection on phenotypic traits in humans is expected to only occasionally take the form of sweeps on a single locus, instead acting on multiple genes simultaneously to drive phenotypic adaptation (Coop et al., 2009). While a lack of genome-wide association studies in eastern and southern Africans has left the genetic basis of phenotypic traits far less well documented than it is for other populations, a variety of studies have linked broad functional classes of genes to phenotypic traits. To test for evidence of selection on specific functional categories of genes in present-day San since the divergence of the two ancient genomes from southern Africa (Figure 4B), we estimated allele frequency differentiation for 208 gene ontology categories with 50 or more genes in each, and we computed weighted block jackknife standard errors. The functional category that displays the most extreme allele frequency differentiation between present-day San and ancient southern Africans is "response to radiation" ($Z = 3.3$ compared to the genome-wide average). To control for the possibility that genes in this category show an inflated allele frequency differentiation in general, we computed the same statistic for the Mbuti central African rainforest hunter-gatherer group but found no evidence for selection affecting the response to radiation category (Figure 4C). Instead, the top category for the Mbuti is "response to growth," suggesting the possibility that the small stature of rainforest hunter-gatherer populations such as the Mbuti may be an acquired adaptation (although we have no ancient central African genome and thus no information about the time frame of selection). We speculate that the signal for selection in the response to radiation category in the San could be due to exposure to sunlight associated with the life of the ǂKhomani and Ju|'hoan North people in the Kalahari Basin, which has become a refuge for hunter-gatherer populations in the last millenia due to encroachment by pastoralist and agriculturalist groups (Morris, 2002).

DISCUSSION

This study, which multiplies by 16-fold the number of individuals with genome-wide ancient DNA data from sub-Saharan Africa, highlights the power of ancient African genomes to provide insights into prehistoric events that are difficult to discern based solely on analysis of present-day genomes. We reveal the presence of a hitherto unknown cline of geographically structured hunter-gatherer populations stretching from Ethiopia to South Africa, which we show existed prior to the great population transformations that occurred in the last few thousand years in association with the spread of herders and farmers. We also document deeper structure in western Africa, possibly predating the divergence of the ancestors of southern African hunter-gatherers from other population lineages. We finally provide case examples of how populations in eastern and southern Africa were transformed by the spread of food producers and show how the process gave rise to interactions with the previously established hunter-gatherers, with the outcomes ranging from no detectable mixture in present-day populations to substantial mixture. Our documentation of a radically different landscape of human populations before and after the spread of food producers highlights the difficulty of reconstructing the African past based solely on analysis of present-day populations and the importance of using ancient DNA to study deep African population history in an era in which technological improvements have now made this feasible. It is clear that ancient DNA studies with larger sample sizes and covering a broader chronological and geographic range have the potential to make major progress in improving our understanding of African prehistory.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Terminology
  - Panga ya Saidi Cave, Kilifi County, Kenya (n = 1)
  - Makangale Cave, Pemba, Tanzania (n = 2)
  - Kuumbi Cave, Zanzibar, Tanzania (n = 1)
  - Luxmanda, Babati District, Tanzania (n = 1)
  - Hora, Malawi (n = 2)
  - Fingira, Malawi (n = 3)
  - Chencherere II, Malawi (n = 2)
  - St. Helena Bay, South Africa (n = 1)
  - Faraoskop, South Africa (n = 1)
  - Kasteelberg, South Africa (n = 1)

- METHOD DETAILS
  - Direct AMS ^{14}C Bone Dates
  - Ancient DNA sample processing in Leipzig: St. Helena Bay sample
  - Ancient DNA sample processing in Tübingen: Faraoskop and Kasteelberg samples
  - Ancient DNA sample processing in Dublin: Malawi samples
  - Ancient DNA sample processing in Boston: Tanzania samples, Kenya samples, and Malawi sample powder
  - Shotgun genome sequencing
  - In-solution nuclear target enrichment
  - Genotyping and initial processing of 34 present-day individuals from Malawi
  - Data processing and preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Principal component analysis and ADMIXTURE clustering analyses
  - Symmetry statistics and admixture tests
  - Y chromosome and mitochondrial haplogroups
  - Ancestry model and estimates with *qpAdm*
  - Maximum likelihood tree model
  - Testing a tree-like model of African population history
  - Testing admixture graph models of African population history
  - Automated grafting of populations onto a skeleton admixture graph
  - Support for a single out-of-Africa founding population
  - Date of admixture between expanding agriculturalists and previously established foragers
  - Evidence for selective sweeps in the ancestry of present-day San
  - Evidence for polygenic selection
- DATA AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.049.

An audio PaperClip is available at http://dx.doi.org/10.1016/j.cell.2017.08.049#mmc8.

AUTHOR CONTRIBUTIONS

Conceptualization, P.S., J.C.T., M.E.P., V.M.H., R.R., A.G.M., N.B., R.P., J.K., and D.R.; Formal Analysis, P.S., S.M., A.P., and I.O.; Investigation, P.S., A.M., K.Si., M.Ha., T.S., N.R., A.H., M.F., E.H., M.Mi., K.St., and J.I.C.-R.; Resources, J.C.T., M.E.P., C.C., A.C., E.G.-C., A.O.G., K.M.G., I.T.H., G.H., R.H., M.Ho., S.L., A.Z.P.M., J.P., C.S., M.G.T., R.T., M.W., V.M.H., A.G.M., N.B., and N.P.; Data Curation, P.S., M.Ha., N.R., S.M., A.P., I.O., M.F., E.H., M.Me., K.St., D.J.K., N.P., and D.R.; Writing, P.S. and D.R.; Supervision, V.M.H., M.Me., S.P., N.P., N.N., N.B., R.P., J.K., and D.R.

ACKNOWLEDGMENTS

Permission to analyze the remains from Kenya and Tanzania was granted by the National Museums of Kenya; the Antiquities Division of the Ministry of Natural Resources and Tourism, Tanzania; and the Zanzibar Department of Museums and Antiquities. Permission to analyze the remains from Malawi was granted by the Malawi Department of Museums and Monuments and the University of Cape Town. We thank I. Lazaridis, M. Lipson, I. Mathieson, and S. Tishkoff for discussions and I. Kucukkalipci and K. Majander for laboratory support. P.S. was supported by the Wenner-Gren Foundation and the Swedish Research Council (VR grant 2014-453). J.C.T was supported by a grant from the Program for the Enhancement of Research at Emory University. J.K. and A.M. were supported by the DFG grant KR 4015/1-1 and the Max Planck Society. K.Si. was supported by NSF grant BCS-1613577. M.Ha., A.H., M.Me., and S.P. were supported by the Max Planck Society. A.G.M. and J.P. were supported by the National Research Foundation of South Africa. R.R. was supported by the South African Medical Research Council. N.B. was supported by ERC starting grant SEALINKS (206148), and R.P. was supported by ERC starting grant ADNABIOARC (263441). M.G.T. was supported by Wellcome Trust Senior Investigator Award (grant number 100719/Z/12/Z). D.R. was supported by NIH grant GM100233 and by NSF HOMINID BCS-1032255 and is a Howard Hughes Medical Institute investigator. The laboratory at Penn State was supported by the NSF Archaeometry program (BCS-1460369, D.J.K.).

Received: December 12, 2016
Revised: July 1, 2017
Accepted: August 29, 2017
Published: September 21, 2017

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am. J. Hum. Genet. *90*, 675–684.

Briggs, A.W., and Heyn, P. (2012). Preparation of next-generation sequencing libraries from damaged DNA. Methods Mol Biol. *840*, 143–154.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87.

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360.

Brown, T.A., Nelson, D.E., Vogel, J.S., and Southon, J.R. (1988). Improved collagen extraction by modified Longin method. Radiocarbon *30*, 171–177.

Busby, G.B.J., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M., et al.; Malaria Genomic Epidemiology Network (2016). Admixture into and within sub-Saharan Africa. eLife *5*, e15266.

Campbell, M.C., Ranciaro, A., Froment, A., Hirbo, J., Omar, S., Bodo, J.-M., Nyambo, T., Lema, G., Zinshteyn, D., and Drayna, D. (2012). Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. Mol. Biol. Evol. *29*, 1141–1153.

Cann, R.L., Stoneking, M., and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. Nature *325*, 31–36.

Chami, F. (2009). Zanzibar and the Swahili coast from c. 30,000 years ago (E&D Vision Pub.).

Chami, F., Khator, J., and Hamis Ali, A. (2009). The excavation of Mapangani cave, Pemba island, Zanzibar. In Studies in the African Past, F.A. Chami and C. Radimilahy, eds. (University of Dar es Salaam Press), pp. 74–79.

Clark, J. (1972). Prehistoric origins (London: The Early History of Malawi Longman), pp. 17–27.

Clark, J.D. (1956). Prehistory in Nyasaland. The Nyasaland Journal *9*, 92–119.

Clark, J.D. (1959). The prehistory of southern Africa (Penguin Books).

Clark, J.D. (1973). Archaeological investigation of a painted rock shelter at Mwana wa Chencherere, north of Dedza, Central Malaŵi, In July to September, 1972. Soc. Malawi J. *26*, 28–46.

Cole-King, P. (1973). Kukumba Mbiri Mu Malaŵi: A Summary of Archaeological Research to March, 1973 (Government Press).

Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., and Pritchard, J.K. (2009). The role of geography in human adaptation. PLoS Genet. *5*, e1000500.

Crader, D.C. (1984a). Faunal remains from Chencherere II rock shelter, Malawi. S. Afr. Archaeol. Bull. *39*, 37–52.

Crader, D.C. (1984b). Hunters in Iron Age Malaẇi: The Zooarchaeology of Chencherere Rockshelter (Lilongwe: Malaẇi Government Ministry of Education and Culture).

Crowther, A., Prendergast, M.E., Fuller, D.Q., and Boivin, N. (2017). Subsistence mosaics, forager-farmer interactions and the transition to food production in eastern Africa. Quat. Int. http://dx.doi.org/10.1016/j.quaint.2017.01.014.

Crowther, A., Veall, M.-A., Boivin, N., Horton, M., Kotarba-Morley, A., Fuller, D.Q., Fenn, T., Haji, O., and Matheson, C.D. (2015). Use of Zanzibar copal (*Hymenaea verrucosa* Gaertn.) as incense at Unguja Ukuu, Tanzania in the 7–8th century CE: chemical insights into trade and Indian Ocean interactions. J. Archaeol. Sci. *53*, 374–390.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

Dabney, J., and Meyer, M. (2012). Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques *52*, 87–94.

DeBusk, G.H. (1997). The distribution of pollen in the surface sediments of Lake Malawi, Africa, and the transport of pollen in large lakes. Rev. Palaeobot. Palynol. *97*, 123–153.

Fleisher, J., and LaViolette, A. (2013). The early Swahili trade village of Tumbe, Pemba Island, Tanzania, AD 600–950. Antiquity *87*, 1151–1168.

Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

Gansauge, M.-T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. *8*, 737–748.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., and Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. *43*, 1031–1034.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., and Stewardson, K. (2015). Massive migration from the steppe is a source for Indo-European languages in Europe. arXiv preprint arXiv:150202783.

Hammer, M.F., Woerner, A.E., Mendez, F.L., Watkins, J.C., and Wall, J.D. (2011). Genetic evidence for archaic admixture in Africa. Proc. Natl. Acad. Sci. USA *108*, 15123–15128.

Helm, R., Crowther, A., Shipton, C., Tengeza, A., Fuller, D., and Boivin, N. (2012). Exploring agriculture, interaction and trade on the eastern African littoral: preliminary results from Kenya. Azania *47*, 39–63.

Helm, R.M. (2000). Conflicting histories: the archaeology of the iron-working, farming communities in the central and southern coast region of Kenya (University of Bristol).

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc. Natl. Acad. Sci. USA *108*, 5154–5162.

Hogg, A.G., Hua, Q., Blackwell, P.G., Niu, M., Buck, C.E., Guilderson, T.P., Heaton, T.J., Palmer, J.G., Reimer, P.J., and Reimer, R.W. (2013). SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. Radiocarbon *55*, 1889–1903.

Jostins, L., Xu, Y., McCarthy, S., Ayub, Q., Durbin, R., Barrett, J., and Tyler-Smith, C. (2014). YFitter: Maximum likelihood assignment of Y chromosome haplogroups from low-coverage sequence data. arXiv preprint arXiv:14077988.

Juma, A. (2004). Unguja Ukuu on Zanzibar: An archaeological study of early urbanism (Uppsala University).

Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat Commun. *8*, 14115.

Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. In Ancient DNA, B. Shapiro and M. Hofreiter, eds. (Springer), pp. 197–228.

Kircher, M., Sawyer, S., and Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., and Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. Biotechniques *59*, 87–93.

Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., et al. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell *150*, 457–469.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science *350*, 820–822.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inference of admixture parameters in human populations using weighted linkage disequilibrium. Genetics *193*, 1233–1254.

Lohse, J.C., Madsen, D.B., Culleton, B.J., and Kennett, D.J. (2014). Isotope paleoecology of episodic mid-to-late Holocene bison population expansions in the Southern Plains, USA. Quat. Sci. Rev. *102*, 14–26.

Longin, R. (1971). New method of collagen extraction for radiocarbon dating. Nature *230*, 241–242.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

Manhire, A. (1993). A report on the excavations at Faraoskop Rock Shelter in the Graafwater district of the south-western Cape. Southern African Field Archaeology *2*, 3–23.

Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.

Marshall, F., Stewart, K., and Barthelme, J. (1984). Early domestic stock at Dongodien in northern Kenya. Azania *19*, 120–127.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., and Pääbo, S. (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature *505*, 403–406.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc *2010*, pdb.prot5448.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.

Miller, S.F. (1971). The age of Nachikufan industries in Zambia. S. Afr. Archaeol. Bull. *26*, 143–146.

Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet. *7*, e1001373.

Morris, A.G. (2002). Isolation and the origin of the Khoisan: Late Pleistocene and Early Holocene human evolution at the southern end of Africa. Hum. Evol. *17*, 231–240.

Morris, A.G., Heinze, A., Chan, E.K., Smith, A.B., and Hayes, V.M. (2014). First ancient mitochondrial human genome from a prepastoralist southern African. Genome Biol. Evol. *6*, 2647–2653.

Morris, A.G., and Ribot, I. (2006). Morphometric cranial identity of prehistoric Malawians in the light of sub-Saharan African diversity. Am. J. Phys. Anthropol. *130*, 10–25.

Odner, K. (1972). Excavations at Narosura, a Stone Bowl Site in the Southern Kenya Highlands. Azania *7*, 25–92.

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am. J. Hum. Genet. *91*, 83–96.

Parkington, J., and Dlamini, N. (2015). First People: Ancestors of the San (Creda Communications).

Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A., et al. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science *356*, 543–546.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

Phillipson, D.W. (1976). The Early Iron Age in eastern and southern Africa: a critical re-appraisal. Azania *11*, 1–23.

Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. Nat. Commun. *3*, 1143.

Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. USA *111*, 2632–2637.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.

Plagnol, V., and Wall, J.D. (2006). Possible ancestral structure in human populations. PLoS Genet. *2*, e105.

Pleurdeau, D., Imalwa, E., Détroit, F., Lesur, J., Veldman, A., Bahain, J.-J., and Marais, E. (2012). "Of sheep and men": earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia). PLoS ONE *7*, e40340.

Prendergast, M.E., Mabulla, A.Z.P., Grillo, K.M., Broderick, L.G., Seitsonen, O., Gidna, A.O., and Gifford-Gonzalez, D. (2013). Pastoral Neolithic sites on the southern Mbulu Plateau, Tanzania. Azania *48*, 498–520.

Prendergast, M.E., Rouby, H., Punnwong, P., Marchant, R., Crowther, A., Kourampas, N., Shipton, C., Walsh, M., Lambeck, K., and Boivin, N.L. (2016). Continental island formation and the archaeology of defaunation on Zanzibar, eastern Africa. PLoS ONE *11*, e0149565.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA *102*, 15942–15947.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.

Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon *55*, 1869–1887.

Renaud, G., Kircher, M., Stenzel, U., and Kelso, J. (2013). freelbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. Bioinformatics *29*, 1208–1209.

Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

Ribot, I., Morris, A.G., Sealy, J., and Maggs, T. (2010). Population history and economic change in the last 2000 years in KwaZulu-Natal, RSA. Southern African Humanities *22*, 89–112.

Robinson, K., and Sandelowsky, B. (1968). The Iron Age of northern Malawi: recent work. AZANIA. Journal of the British Institute in Eastern Africa *3*, 107–146.

Robinson, K.S.R. (1982). Iron Age of northern Malaẇi: an archaeological reconnaissance (Malaẇi Govt. Ministry of Education and Culture).

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130624.

Rohland, N., and Hofreiter, M. (2007). Comparison and optimization of ancient DNA extraction. Biotechniques *42*, 343–352.

Russell, T., Silva, F., and Steele, J. (2014). Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. PLoS ONE *9*, e87854.

Sadr, K. (2015). Livestock First Reached Southern Africa in Two Separate Events. PLoS ONE *10*, e0134215.

Sadr, K., Smith, A., Plug, I., Orton, J., and Mütti, B. (2003). Herders and foragers on Kasteelberg: interim report of excavations 1999-2002. S. Afr. Archaeol. Bull. *58*, 27–32.

Sandelowsky, B. (1972). Later Stone Age assemblages from Malawi and their technologies. PhD thesis (University of California, Berkeley).

Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. *13*, 745–753.

Schlebusch, C. (2010). Issues raised by use of ethnic-group names in genome study. Nature *464*, 487.

Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science *338*, 374–379.

Schuenemann, V.J., Peltzer, A., Welte, B., van Pelt, W.P., Molak, M., Wang, C.-C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. Nat Commun *8*, 15694.

Sealy, J.C., Patrick, M.K., Morris, A.G., and Alder, D. (1992). Diet and dental caries among later stone age inhabitants of the Cape Province, South Africa. Am. J. Phys. Anthropol. *88*, 123–134.

Shipton, C., Crowther, A., Kourampas, N., Prendergast, M.E., Horton, M., Douka, K., Schwenninger, J.-L., Faulkner, P., Quintana Morales, E.M., Langley, M.C., et al. (2016). Reinvestigation of Kuumbi Cave, Zanzibar, reveals Later Stone Age coastal habitation, early Holocene abandonment and Iron Age reoccupation. Azania *51*, 197–233.

Shriver, M.D., Kennedy, G.C., Parra, E.J., Lawson, H.A., Sonpar, V., Huang, J., Akey, J.M., and Jones, K.W. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum. Genomics *1*, 274–286.

Sinclair, P., Juma, A., and Chami, F. (2006). Excavations at Kuumbi Cave on Zanzibar 2005. In The African Archaeology Network: Research in Progress, J. Kinahan and J.H.A. Kinahan, eds. (Dar es Salaam University Press), pp. 95–106.

Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature *525*, 104–108.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014a). In Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal, R.G. Klein, ed. (Proceedings of the National Academy of Sciences), pp. 2229–2234.

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. Nature *538*, 510–513.

Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M., and Jakobsson, M. (2014b). Investigating population history using temporal genetic differentiation. Mol. Biol. Evol. *31*, 2516–2527.

Smith, A.B. (1992a). Kasteelberg. In Guide to Archaeological Sites in the South-western Cape, A.B. Smith and B. Mutti, eds. (Cape Town: South African Association of Archaeologists Conference), pp. 28–30.

Smith, A.B. (1992b). Pastoralism in Africa (Johannesburg: Witwatersrand University Press).

Smith, B.W. (1995). Rock art in south-central Africa: a study based on the pictographs of Dedza District, Malawi and Kasam District (Zambia: University of Cambridge).

Stafford, T.W., Hare, P.E., Currie, L., Jull, A.J.T., and Donahue, D.J. (1991). Accelerator radiocarbon dating at the molecular level. J. Archaeol. Sci. *18*, 35–72.

Stuiver, M., and Polach, H.A. (1977). Discussion: reporting of 14 C data. Radiocarbon *19*, 355–363.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.-M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035–1044.

Van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. *26*, 687–695.

van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.

Veeramah, K.R., Wegmann, D., Woerner, A., Mendez, F.L., Watkins, J.C., Destro-Bisol, G., Soodyall, H., Louie, L., and Hammer, M.F. (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol. Biol. Evol. *29*, 617–630.

Wang, Y., and Nielsen, R. (2012). Estimating population divergence time and phylogeny from single-nucleotide polymorphisms data with outgroup ascertainment bias. Mol. Ecol. *21*, 974–986.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–63.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75–78.

Zubieta, L.F. (2016). Learning through practise: Chewa women's roles and the use of rock art in passing on cultural knowledge. J. Anthropol. Archaeol. *43*, 13–28.

Cell

Article

The Kinetochore Receptor for the Cohesin Loading Complex

Stephen M. Hinshaw,[1,4] Vasso Makrantoni,[2,4] Stephen C. Harrison,[1,3,5,6,*] and Adèle L. Marston[2,5,*]
[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[2]The Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK
[3]Howard Hughes Medical Institute, Chevy Chase, MD, USA
[4]These authors contributed equally
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: harrison@crystal.harvard.edu (S.C.H.), adele.marston@ed.ac.uk (A.L.M.)
http://dx.doi.org/10.1016/j.cell.2017.08.017

SUMMARY

The ring-shaped cohesin complex brings together distant DNA domains to maintain, express, and segregate the genome. Establishing specific chromosomal linkages depends on cohesin recruitment to defined loci. One such locus is the budding yeast centromere, which is a paradigm for targeted cohesin loading. The kinetochore, a multiprotein complex that connects centromeres to microtubules, drives the recruitment of high levels of cohesin to link sister chromatids together. We have exploited this system to determine the mechanism of specific cohesin recruitment. We show that phosphorylation of the Ctf19 kinetochore protein by a conserved kinase, DDK, provides a binding site for the Scc2/4 cohesin loading complex, thereby directing cohesin loading to centromeres. A similar mechanism targets cohesin to chromosomes in vertebrates. These findings represent a complete molecular description of targeted cohesin loading, a phenomenon with wide-ranging importance in chromosome segregation and, in multicellular organisms, transcription regulation.

INTRODUCTION

Genome maintenance, expression, and transmission depend on organization of chromosomes into functional domains. At the heart of this organization, in both prokaryotes and eukaryotes, are protein complexes in the structural maintenance of chromosomes (SMC) family: cohesin, condensin, and SMC5/6 (Uhlmann, 2016). These complexes must be targeted to specific regions of the chromosome, including areas of damage, transcriptional units, and centromeres. Biological control of SMC complex activity thus relies on interactions that direct these complexes to particular chromosomal sites at appropriate times.

The best understood of the SMC complexes is cohesin, which links newly duplicated chromosomes in preparation for their segregation during cell division. At the core of the cohesin complex is a three-protein ring (Smc1, Smc3, and Scc1) (Losada and Hirano, 2005; Nasmyth, 2011; Onn et al., 2008). A separate and broadly conserved two-protein complex called Scc2/4 in yeast and NIPBL/Mau2 in humans loads cohesin onto chromosomes (Ciosk et al., 2000). While the yeast Scc2 and Scc4 genes were cloned initially due to their role in sister chromatid cohesion, the fly homolog of Scc2, Nipped-B, was first identified as a gene required for proper enhancer function (Rollins et al., 1999). Heterozygous mutations in NIPBL cause Cornelia de Lange syndrome (CdLS), a developmental disorder associated with aberrant transcription across the genome (Kawauchi et al., 2009; Krantz et al., 2004; Tonkin et al., 2004). The dual role of cohesin loading in chromosome segregation and transcription regulation (Fay et al., 2011) makes determining molecular pathways for targeted cohesin loading an issue of broad interest.

Genomic locations of cohesin enrichment exist in all eukaryotes, but the phenomenon is best understood in the context of the yeast centromere. Cohesin density peaks at or near centromeres and extends to nearby pericentromeric regions (Glynn et al., 2004; Lengronne et al., 2004). In mitosis, pericentromeric cohesin complexes enable spindle assembly checkpoint silencing by resisting the poleward forces exerted by kinetochore microtubules (Stern and Murray, 2001; Tanaka et al., 2000). Disabling this process causes high rates of chromosome missegregation and aneuploidy, both pervasive features of cancer (Holland and Cleveland, 2012). In meiosis, high pericentromeric cohesin occupancy serves at least three functions critical for avoiding aneuploidy (Nasmyth, 2001). First, at least in fission yeast, it prevents biorientation of sister kinetochores in meiosis I, thus facilitating reductional chromosome segregation (Sakuno et al., 2009; Watanabe and Nurse, 1999). Second, its selective preservation at centromeres throughout meiosis I is required for accurate chromosome segregation in meiosis II (Lee and Orr-Weaver, 2001). Third, it represses crossover recombination close to centromeres (Vincenten et al., 2015), which would otherwise lead to meiosis II nondisjunction (Rockmill et al., 2006).

Preferential cohesin loading at centromeres is a kinetochore-dependent process (Megee and Koshland, 1999; Tanaka et al., 2000; Weber et al., 2004). The yeast Ctf19 complex, the homolog of the human Constitutive Centromere Associated Network (CCAN), enables this function by recruiting the cohesin loader

 © 2017 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

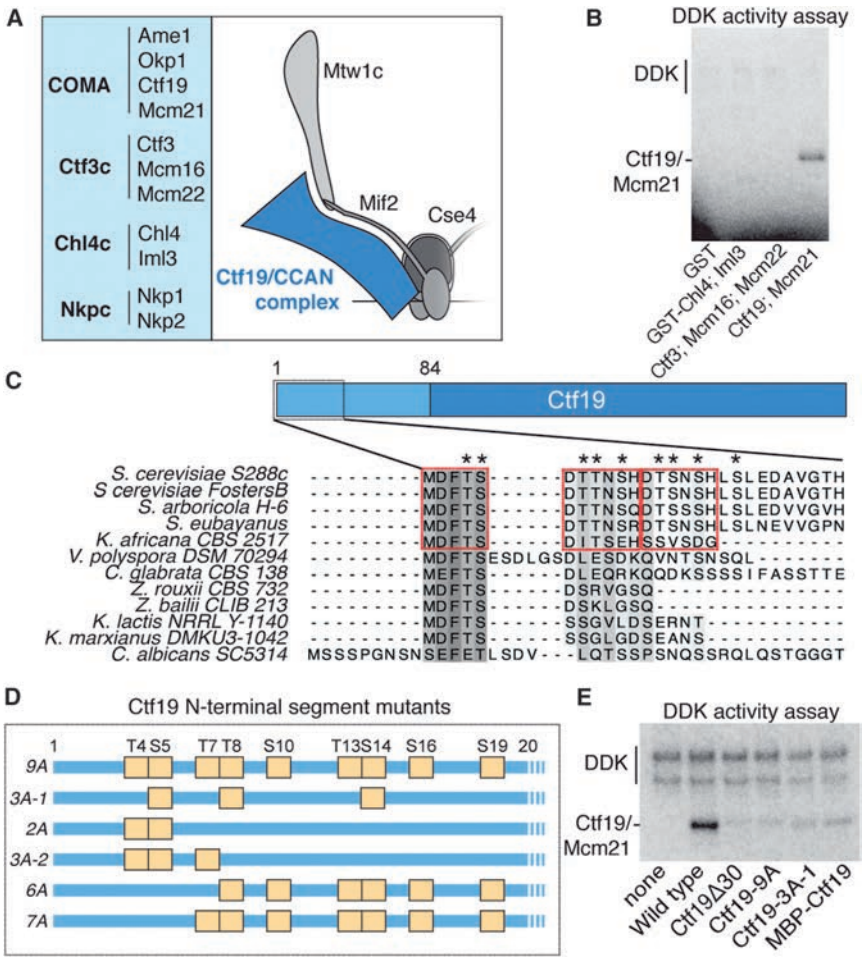


Figure 1. DDK Phosphorylates Ctf19

(A) Ctf19 subcomplexes and their members (left) listed next to a schematic of the kinetochore with the Ctf19 complex shown in blue (right).

(B) In vitro phosphorylation of the Ctf19-Mcm21 dimer by DDK. The indicated substrate proteins were incubated with purified DDK and γ-^{32}[P]-ATP. Reaction products were resolved by SDS-PAGE and visualized by autoradiography.

(C) Diagram of Ctf19 showing the fragment resolved crystallographically (dark blue, PDB: 3ZXU) and the unstructured N-terminal region (light blue). Sequence alignment shows the N-terminal fragment. Red boxes indicate candidate phosphorylation motifs. Asterisks mark residues mutated in the *ctf19-9A* allele.

(D) Ctf19 alleles used in this work. Yellow boxes indicate residues mutated to alanine.

(E) DDK phosphorylates residues in the Ctf19 N-terminal region. Kinase assay performed as in (B). Substrates are indicated below.

See also Figures S1–S3 and Table S1.

to centromeres (Eckert et al., 2007; Fernius and Marston, 2009; Fernius et al., 2013; Kogut et al., 2009; Ng et al., 2009). We have shown previously that Scc2 recruitment to centromeres requires a conserved surface on the Scc4 protein (Hinshaw et al., 2015). Mutation of amino acid residues at this surface abrogates Scc2/4 recruitment to centromeres, causes loss of cohesion specifically at centromeres, and leads to elevated rates of mini-chromosome missegregation. In the work reported here, we sought to identify the kinetochore factor mediating Scc2/4 recruitment, to describe the mode of regulation that restricts specific centromeric cohesin loading to a narrow window of the cell cycle, and to understand whether a similar recruitment process operates in other species.

One factor likely to play a broad role in regulating targeted cohesin loading is the Dbf4-dependent kinase (DDK). DDK is a cell-cycle-regulated kinase that is required for the preferential association of cohesin with centromeres in budding and fission yeast and with unreplicated chromatin in vertebrates (Bailis et al., 2003; Natsume et al., 2013; Takahashi et al., 2008). The essential function of DDK is to activate the Mcm2-7 helicase for DNA replication (Sheu and Stillman, 2010). In yeast, the Ctf19 complex recruits DDK to kinetochores in G1, where its activity enables subsequent Scc2 recruitment and robust centromere cohesion independently of its role in DNA replication (Natsume et al., 2013). In *Xenopus laevis* egg extracts, DDK activity is required for the association of Scc2/4 with replication origins (Takahashi et al., 2008). These findings suggest that DDK is part of a common mechanism for directing cohesin loading activity to defined chromosomal locations and for coupling this activity with the onset of DNA replication.

Using the budding yeast centromere as a model, we have determined the molecular pathway that drives targeted cohesin loading. We report here that DDK phosphorylates the Ctf19 protein. Phosphorylated Ctf19 interacts directly with the previously identified conserved surface of Scc4, explaining both the requirement for DDK in centromeric cohesion establishment and the nature of Scc4-mediated targeting of cohesin loading to centromeres. Our results thus provide a full description of the molecular events required to recruit cohesin to a defined chromosomal locus.

RESULTS

DDK Phosphorylates Ctf19-Mcm21

Several observations suggest that DDK phosphorylates a component of the Ctf19 complex (Figure 1A) and that this signal recruits Scc2 to centromeres. First, DDK localization at kinetochores immediately precedes Scc2 localization and subsequent cohesin loading (Natsume et al., 2013). Second, Scc4 has a conserved, positively charged, surface patch important for Scc2 localization, leading to the suggestion that this patch might be a phosphopeptide binding motif (Chao et al., 2015; Hinshaw et al., 2015). Third, the interaction between Scc2/4 and DDK is probably indirect in yeast, as it is not observable in a reconstituted pulldown system (our unpublished observations). To

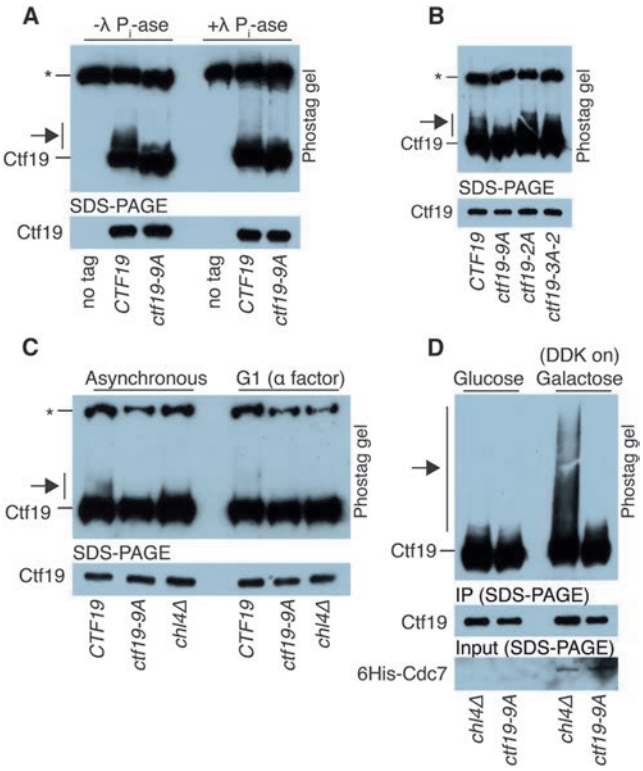


Figure 2. Observation of Ctf19 Phosphorylation In Vivo

(A) Ctf19-3FLAG, purified from exponentially growing cells expressing either wild-type *CTF19* (SMH372) or *ctf19-9A* (SMH378) was resolved on a phostag gel and detected by western blot (arrow, phosphorylated Ctf19; *, non-specific band).

(B) Ctf19 phosphorylation was assessed as in (A) for strains expressing *CTF19*, *ctf19-9A*, *ctf19-2A*, or *ctf19-3A-2* (SMH372, SMH378, SMH449, and SMH450, respectively).

(C) Ctf19 phosphorylation assessed as in (A) for strains expressing *CTF19*, *ctf19-9A*, or *CTF19* in a *chl4*Δ background (SMH372, SMH378, and SMH401, respectively). Cells were untreated or arrested in G1 as indicated (arrows, phosphorylated Ctf19; *, non-specific band).

(D) Strains bearing Ctf19-3FLAG and an inducible DDK expression cassette (Cdc7-6His, Dbf4 untagged) were not induced (left) or induced to express DDK (right) for 2 hr. Ctf19 phosphorylation was assessed in strains lacking *CHL4* (SMH429) or expressing *ctf19-9A* (SMH430).

determine whether DDK can bind proteins of the Ctf19 complex, we performed pulldowns using recombinant Ctf19 complex proteins and detected associated radiolabeled DDK produced by in vitro translation (Figure S1). We identified two candidate DDK interaction partners: the Ctf19-Mcm21 dimer and the Ctf3 complex. Although the Ctf19-Mcm21 dimer associates with the Okp1 and Ame1 proteins to make the COMA tetramer (CENP-P/O/Q/U in humans) (De Wulf et al., 2003; Okada et al., 2006), Okp1 and Ame1 were not required for DDK interaction. The Ctf3 complex is a trimeric assembly of Ctf3, Mcm16, and Mcm22 (CENP-I/H/K in humans) (Measday et al., 2002; Okada et al., 2006), and Ctf3 was the component required for DDK interaction.

The identification of Ctf19-Mcm21 and Ctf3 as DDK binding partners suggested that they could be the kinetochore substrates that generate the signal for recruitment of Scc2 to centromeres. We therefore purified DDK from yeast and performed kinase activity assays using Ctf19 complex proteins as substrates. Purified DDK phosphorylated Ctf19-Mcm21 but not the Ctf3 trimer or the Chl4-Iml3 dimer (Figure 1B). To confirm that phosphate transfer to Ctf19-Mcm21 depended on the catalytic activity of Cdc7 and not a co-purifying kinase, we purified an ATP analog (PP1)-sensitive mutant of DDK (Wan et al., 2006). Phosphate transfer to Ctf19 was inhibited by PP1 at concentrations (~25 μM) similar to those that inhibit kinase activity in vivo (Figure S2A) (Wan et al., 2006). These experiments confirm that Ctf19-Mcm21, but not Ctf3, is a DDK substrate in vitro.

DDK Phosphorylation Sites in the Ctf19 N-Terminal Region

We next sought to identify the Ctf19-Mcm21 residues targeted by DDK. A crystal structure of the Ctf19-Mcm21 dimer showed that N-terminal segments of both proteins are flexible (Schmitzberger and Harrison, 2012). We found that trypsin digestion of phosphorylated Ctf19-Mcm21 removed the phosphorylated residues under conditions that preferentially cleave the N-terminal tails from both proteins (Figure S2B), indicating that DDK phosphorylates residues in these unstructured regions.

Inspection of the N-terminal region of Ctf19 revealed three clusters of candidate phosphorylation sites (Figure 1C), and we found no candidates in the N-terminal region of Mcm21. DDK did not phosphorylate Ctf19-Mcm21 lacking the first 30 amino acids of Ctf19 (*ctf19-Δ30*), confirming that the DDK target site lies in this region. A Ctf19-Mcm21 complex in which Ctf19 was N-terminally tagged with MBP was also not phosphorylated in vitro, suggesting that residues very close to the N terminus of Ctf19 are important for phosphorylation by DDK. The serine and threonine residues near the Ctf19 N terminus are arranged in three near-repeat sequences, reminiscent of DDK target sites in Mcm4 and Rif1 (Hiraga et al., 2014; Sheu and Stillman, 2006). We mutated to alanine all candidate phosphate acceptors in the Ctf19 N-terminal region (*ctf19-9A*) or just a single candidate in each of the three near-repeat sequences (*ctf19-3A-1*) and tested for phosphate transfer to these complexes in vitro (Figures 1D and 1E). In all cases, phosphorylation was reduced to background levels. Mutation of just the first two candidate sites (T4, T5) to generate the *ctf19-2A* mutant (Figure 1D) did not abolish phosphorylation (Figure S2C). Finally, we also tested the effect of the Ctf3 complex on Ctf19 phosphorylation and found that it did not enhance phosphate transfer to Ctf19 in vitro (Figure S2D). We conclude that the Ctf19 N-terminal region is a specific substrate of DDK.

Determinants of Ctf19 Phosphorylation In Vivo

To test for phosphorylation in vivo, we purified Ctf19 from yeast and examined its mobility through a phostag-acrylamide matrix. We observed a slowly migrating species of wild-type Ctf19, which was susceptible to phosphatase treatment (Figure 2A, arrow). This Ctf19 mobility shift was evident in cells arrested in G1 (Figure 2C), S phase, and metaphase (Figures S3A and S3B). Ctf19 purified from *ctf19-9A* cells did not show this shift, indicating that the sites required for phosphorylation in vitro are also required for phosphorylation in vivo (Figure 2A). Mutating

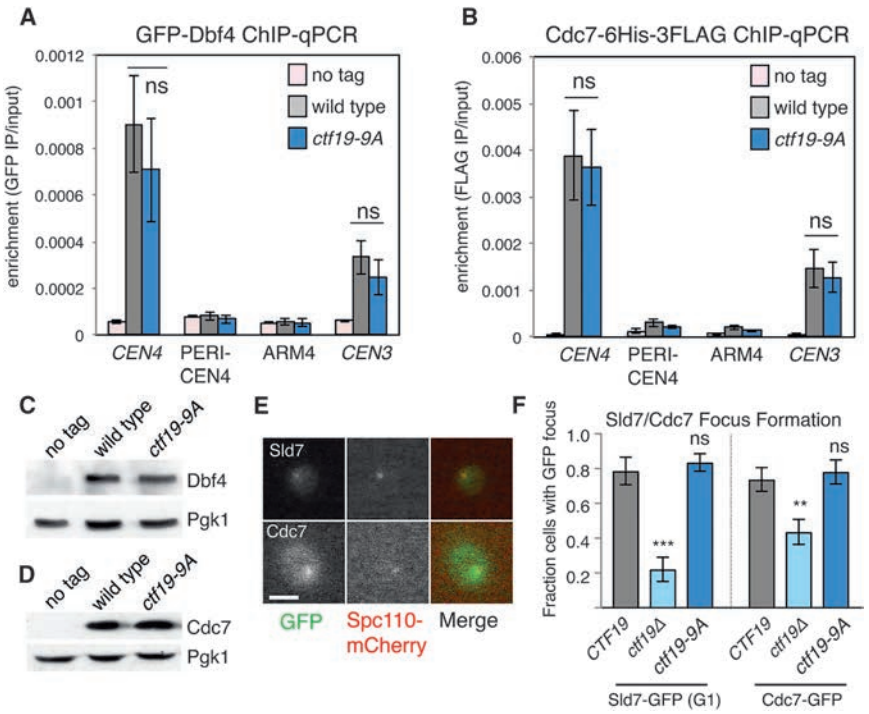


Figure 3. Ctf19 Phosphorylation Is Not Required for DDK Localization or Early Origin Activation

(A and B) DDK localization is not perturbed by *ctf19-9A*. Wild-type and *ctf19-9A* cells carrying *GFP-DBF4* (AM21871 and AM21872, respectively (A), or *CDC7-6His-3FLAG* (AM21110 and AM21234, respectively (B), together with an untagged control strain (AM1176) were arrested in G1 with alpha factor for 3 hr before harvesting for ChIP-qPCR (average of three biological repeats, error bars show SE; ns, not significant, Student's t test).

(C and D) Dbf4 or Cdc7 protein levels are unaltered in *ctf19-9A* strains. Representative anti-GFP and anti-HA immunoblots are shown (Pgk1, loading control).

(E and F) Live-cell imaging to assess Sld7 and Cdc7 localization. Strains expressing Spc110-mCherry to mark spindle pole bodies and either Sld7-GFP or Cdc7-GFP were examined for the presence of GFP foci in G1 (SMH477, SMH478, and SMH480 for Sld7-GFP; SMH473, SMH474, and SMH476 for Cdc7-GFP). Sld7-GFP-expressing strains were arrested in G1 with alpha factor, and Cdc7-GFP-expressing strains were observed during unperturbed growth. (E) Images of wild-type *CTF19* cells in G1 are shown for Sld7-GFP (top) and Cdc7-GFP (bottom; scale bar, 3 μm). (F) Fraction of cells with foci in the indicated strains (error bars, +/−SD for three experiments; $^{**}p < 0.01$; $^{***}p < 0.001$; ns, not significant; Student's t test versus the *CTF19* strain for the indicated GFP fusion protein, two tails, unequal variance).

See also Figure S4.

only the potential phosphorylation sites closest to the N terminus of Ctf19 (Figure 1D) did not eliminate the Ctf19 mobility shift (Figure 2B; Ctf19-2A; Ctf19-3A-2). Indeed, the Ctf19-2A protein with substitutions only at residues T4 and S5 was phosphorylated robustly in vivo as it was in vitro (Figures 2B and S2C).

To determine whether DDK localization is important for Ctf19 phosphorylation in vivo, we tested the effect of appending a C-terminal 9-myc tag to Dbf4, which perturbs DDK recruitment to the kinetochore in G1 (Natsume et al., 2013). Although the effect was subtle, we found decreased Ctf19 phosphorylation levels in this background (Figure S3C), supporting the idea that recruitment of DDK to kinetochores in G1 is important for robust Ctf19 phosphorylation. In the absence of Chl4, a Ctf19 complex protein required for DDK association with centromeres in G1 (Natsume et al., 2013), Ctf19 phosphorylation was only partly perturbed in asynchronous cultures but completely eliminated in G1-arrested cultures (Figures 2C and S3D). We conclude that the N-terminal region of Ctf19 is phosphorylated in vivo, and Chl4 determines the extent and timing of this phosphorylation.

If Chl4, which is not required for Ctf19 localization (Pot et al., 2003), determines Ctf19 phosphorylation levels by enabling DDK recruitment to kinetochores in G1, then overproduction of DDK should bypass this effect. To test this hypothesis, we overexpressed both DDK subunits from a galactose-inducible expression cassette and probed for Ctf19 phosphorylation (Figure 2D). Ctf19 phosphorylation was readily observable upon DDK overexpression in cells lacking Chl4 but not in the *ctf19-9A* strain. This result shows that Chl4 is dispensable for Ctf19 phosphorylation when DDK levels are not limiting.

In addition to the Ctf19 complex, robust centromere cohesion requires additional factors including Csm3 and Dcc1, members of the fork protection complex (FPC) and an alternative PCNA loading complex, respectively (Fernius and Marston, 2009; Mayer et al., 2004). Csm3 has a known role in DDK recruitment to the replisome in premeiotic S phase, suggesting a possible role in coordinating mitotic Ctf19 phosphorylation (Murakami and Keeney, 2014). We examined Ctf19 phosphorylation in *csm3*Δ and *dcc1*Δ strains and found it unperturbed in either background (Figure S3D), supporting the idea that these factors work in a pathway independent of cohesin loader recruitment (Fernius and Marston, 2009).

The Ctf19 N-Terminal Region Is Not Required to Recruit DDK or Replication Factors

Ctf19 is required for centromeric DDK recruitment and the subsequent early replication of centromeric DNA (Natsume et al., 2013), raising the possibility that the Ctf19 N-terminal region is not merely a substrate, but also a recruiter of DDK. To test whether the *ctf19-9A* allele interferes with proper DDK recruitment, we assessed DDK localization in cells arrested in G1 with alpha factor by performing chromatin immunoprecipitation (ChIP)-qPCR for the Dbf4 and Cdc7 subunits of the kinase complex (3GFP-Dbf4 and Cdc7-6His-3FLAG, respectively; Figures 3A–3D). We found that DDK localization to centromeres was independent of Ctf19 N-terminal region phosphorylation, a finding

confirmed by imaging Cdc7-GFP in live cells (Figures 3E and 3F). Furthermore, Sld7-GFP, a replication origin factor that depends on Ctf19 for its centromeric localization in G1 (Natsume et al., 2013), was also recruited normally to centromeres in the *ctf19-9A* background (Figures 3E and 3F). We conclude that Ctf19 phosphorylation is dispensable for DDK recruitment to centromeres and early origin activation.

Ctf19 Phosphorylation Is Required to Recruit Scc2 to Centromeres

Kinetochore-localized DDK and the Ctf19 complex enable cohesin loading at centromeres by recruiting Scc2/4. Our findings that the Ctf19 N-terminal region is a substrate of DDK and the Ctf3 complex is a DDK binding partner suggested that phosphorylation of Ctf19 might create a docking site for the Scc2/4 cohesin loader at the kinetochore (Figures 4A and 4B). To test this idea, we imaged cells expressing Scc2-GFP as they progressed through an unperturbed cell cycle (Figures 4C and 4D). Under these conditions, Scc2-GFP forms a kinetochore-associated focus as wild-type cells enter S phase (Fernius et al., 2013; Natsume et al., 2013). We found that Scc2-GFP did not make foci at kinetochores (identified by Mtw1-tdTomato in our experiments) in the *ctf19-9A* background, while kinetochore localization was readily observable in the wild-type background. This effect was specific for Ctf19 mutants, as Scc2-GFP localization was unperturbed in *csm3*Δ and *dcc1*Δ cells (Figure 4C).

We also used ChIP-qPCR to assess the localization of Scc2 (Figures 4E and 4G). Consistent with the imaging, we found that Scc2 occupancy at three test centromeres was reduced in the *ctf19-9A* strain to the levels seen at arm sites. Reduced Scc2 recruitment to centromeres is expected to result in decreased cohesin levels throughout the pericentromere. Indeed, ChIP-qPCR from the *ctf19-9A* mutant showed reduced levels of Scc1 at centromeric and pericentromeric sites but not at chromosomal arm sites (Figures 4F and 4H). Therefore, Ctf19 phosphorylation is required for Scc2 recruitment to centromeres and for cohesin enrichment in the pericentromere.

A Subset of Phosphorylated Ctf19 Residues Recruits Scc2 to Centromeres

Elevated conservation of the two phosphorylation sites closest to the N terminus of Ctf19 raised the possibility that these sites might be especially important for Scc2/4 recruitment to the kinetochore (Figure 1C), even though they are not the only sites phosphorylated in vivo (*ctf19*-2A; Figures 2B and S2C). To test this hypothesis, we examined Scc2-GFP localization in cells in which only the first three (T4A, S5A, T7A; *ctf19-3A-2*) or two (T4A, S5A; *ctf19-2A*) Ctf19 phosphate acceptor residues had been mutated. Like *ctf19-9A* cells, strains bearing either the *ctf19-3A-2* or the *ctf19-2A* allele did not efficiently localize Scc2 to centromeres as determined by fluorescence imaging (Scc2-GFP; Figure 4C) and by ChIP-qPCR (Figure 4E). Accordingly, cohesin recruitment to centromeres was diminished in the *ctf19-2A* strain, though less severely than in the *ctf19-9A* strain (Figure 4F). To determine whether phosphorylation of the residues closest to the Ctf19 N terminus is sufficient for Scc2/4 recruitment to kinetochores, we examined the effects of Ctf19 alleles in which all other Ctf19 phosphate acceptors had been mutated to alanine. The *ctf19-6A* strain, which is identical to *ctf19-9A* except that Ctf19 residues T4, S5, and T7 are unmutated, had only a mild defect in centromeric Scc2 occupancy (Figure 4C, E). Together, these findings show that phosphorylation of Ctf19 residues T4, S5, and T7 is both necessary and sufficient to drive Scc2/4 localization to kinetochores.

Phosphorylation of Ctf19 Is Required for Robust Centromere Cohesion

To determine whether failure to target Scc2 to centromeres upon mutation of the phosphorylation sites in Ctf19 results in defective centromere cohesion, we examined centromere separation in cells arrested in metaphase by Cdc20 depletion (Figures 5A–5C). In cells carrying the *ctf19-9A* allele, the distance between centromeric GFP foci was comparable to that in the *ctf19*Δ strain (Figure 5B). Alanine substitutions at a single phosphate acceptor site in each of the three N-terminal repeat-like sequences in Ctf19 (S5A, T8A, S14A; *ctf19-3A-1*) resulted in a partial centromere cohesion defect (Figure 5B).

Because phosphorylated sites near the N terminus of Ctf19 appear most critical for Scc2/4 recruitment, we examined the effect on centromere cohesion of mutating T4 and S5 to phospho-null (alanine; *ctf19-2A*) or phosphomimetic (aspartate; *ctf19-2D* and glutamate; *ctf19-2E*) residues (Figure 1D). While *ctf19-2A* resulted in a significant centromere cohesion defect, the phosphomimetic *ctf19-2D* and *ctf19-2E* mutants did not (Figure 5C), consistent with the idea that it is phosphorylation of Ctf19 residues near its N terminus that enables centromere cohesion.

Scc2 Recruitment by Artificially Tethered DDK Depends on Ctf19 N-Terminal Phosphorylation Sites

Our in vitro binding studies (Figure S1) indicated that, in addition to interacting with its substrate Ctf19, DDK also associates with Ctf3, which is not a substrate. ChIP-qPCR showed that the amounts of both Cdc7 and Dbf4 are strongly reduced at centromeres in G1-arrested *ctf3*Δ cells (Figures S4A–S4D). To test the hypothesis that the critical role of Ctf3 in centromeric cohesin loading is to enable Ctf19 phosphorylation by recruiting DDK to centromeres, we sought to artificially recruit DDK to centromeres in cells lacking Ctf3. We tagged Ctf19 with FKBP12 and Dbf4 with FRB, so that upon rapamycin addition, Dbf4-FRB would be tethered to Ctf19-FKBP12 at the kinetochore. Because kinetochore localization of Ctf19 does not depend on Ctf3 (Figures S4E and S4F) (Pot et al., 2003), forced Dbf4-Ctf19 association should bypass the normal requirement for Ctf3 in centromeric DDK recruitment. Consistent with a previous study (Natsume et al., 2013), tagging Dbf4 at its C terminus with either 13Myc or FRB diminished its association with centromeres in G1 cells (Figures S4G–S4J). However, centromeric Scc2 occupancy in metaphase cells producing Dbf4-FRB was comparable to that of wild-type cells, even in the absence of rapamycin (Figure 5F; *CTF19*, DMSO condition; compare with Figure 4E, wild-type). It is therefore likely that Dbf4 tagged at its C terminus with FRB retains the ability to access its Ctf19 substrate, at least during metaphase. Centromeric Scc2 localization in Dbf4-FRB cells in the absence of rapamycin nonetheless depended on Ctf3, while artificial tethering of Dbf4-FRB to Ctf19-FKBP12 by addition of

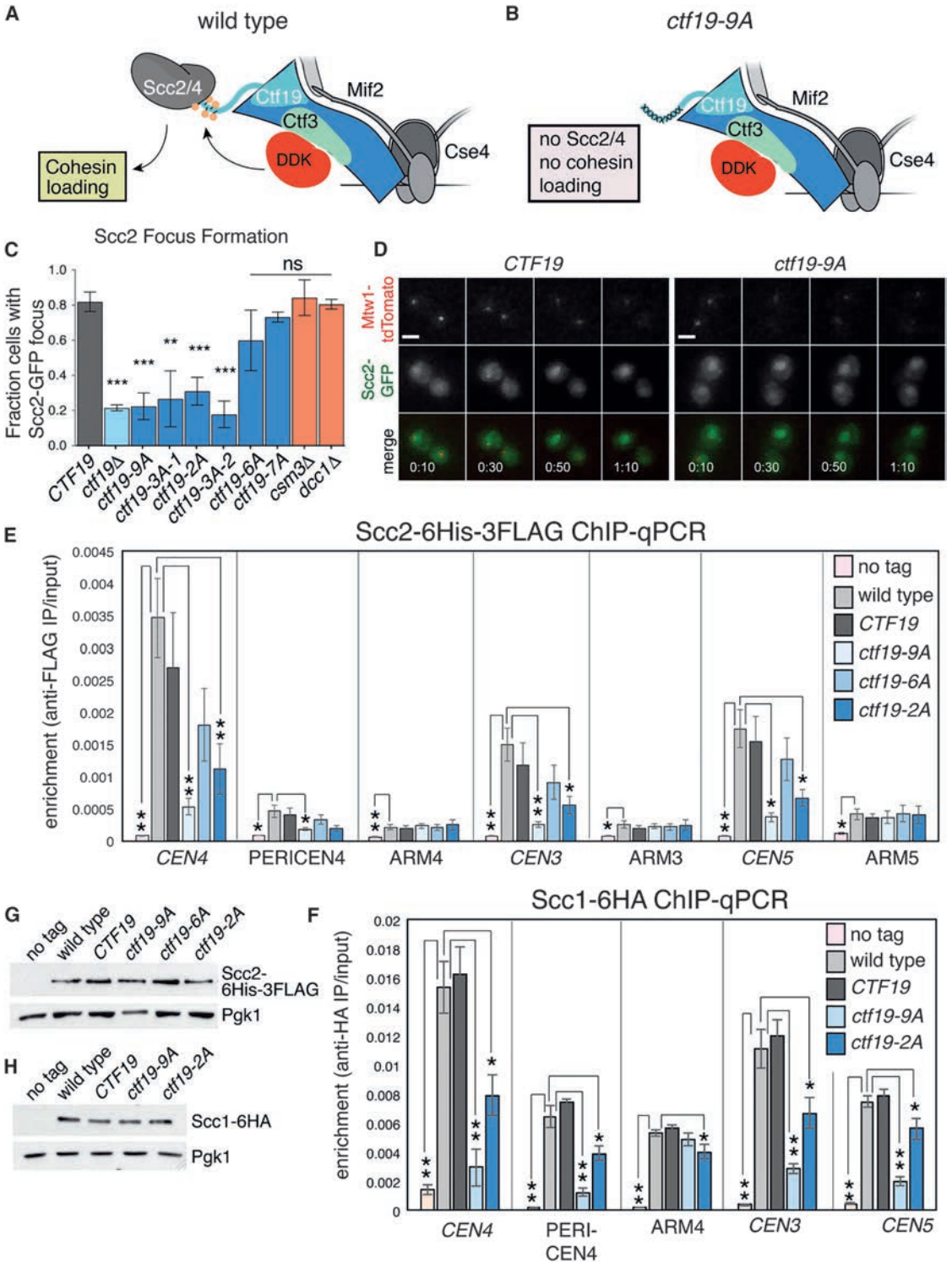


Figure 4. Ctf19 Phosphorylation Is Required for Scc2 Localization at Centromeres

(A and B) Model for the mechanism of Scc2/4 (gray) recruitment by the Ctf19 complex (blue) and DDK (red). Ctf3 is shown in green, and phosphate modifications are depicted as yellow circles. (A) Ctf3 recruits DDK, which phosphorylates Ctf19 to provide a docking site for Scc4. (B) Cells lacking the Ctf19 phosphorylation sites lack the Scc4 docking site, despite normal DDK recruitment.

(C and D) Ctf19 phosphorylation is required for efficient localization of Scc2 to centromeres. Strains of the indicated genotypes expressing Scc2-GFP and Mtw1-tdTomato (SMH353, SMH361, SMH354, SMH355, SMH451, SMH452, SMH459, SMH460, SMH416, and SMH417 listed in the order shown) were imaged during

(legend continued on next page)

rapamycin bypassed the Ctf3 requirement (Figure 5F). Thus, artificial tethering of Dbf4 to Ctf19 can substitute for Ctf3 in recruiting Scc2 to centromeres. Furthermore, mutation of phosphorylation sites near the Ctf19 N terminus completely (*ctf19-9A*) or partially (*ctf19-2A*) prevented tethered Dbf4 from restoring Scc2 association with centromeres in *ctf3*Δ cells (Figure 5F). We conclude that Ctf3 recruits DDK, which in turn phosphorylates Ctf19, thereby enabling Scc2 association with centromeres and cohesin loading.

Phosphorylated Ctf19 Binds Scc4

Scc4 is the localization module of the cohesin loading complex, but an interaction with kinetochore proteins has not yet been observed. To test the idea that phosphorylated Ctf19 is the centromeric receptor for Scc2/4, we attempted to reconstitute Scc2/4 targeting in vitro. Using a pulldown assay with recombinant $Scc2^{1–181}$-Scc4 as bait and Ctf19-Mcm21 as prey (Figure 6A), we found that Ctf19-Mcm21, when treated with DDK and ATP, associated with $Scc2^{1–181}$-Scc4, and this interaction required the N-terminal region of Ctf19. Therefore, Ctf19, when phosphorylated by DDK, is a direct binding partner of Scc2/4.

To determine whether the N-terminal region of Ctf19 is sufficient for Scc4 binding, we tested whether synthetic Ctf19 peptides interact with recombinant $Scc2^{1–181}$-Scc4. We found that $Ctf19^{1–6}$ phosphorylated at amino acid residues T4 and S5 bound $Scc2^{1–181}$-Scc4 in vitro, while an unphosphorylated peptide did not (Figures 6B and S5). The weak in vitro affinity (dissociation constant [Kd] ~23 μM) may reflect the absence of stabilizing factors present in vivo. To determine whether this interaction depended on the conserved surface of Scc4 required for kinetochore recruitment of Scc2 (Hinshaw et al., 2015), we performed identical experiments with a mutant version of the $Scc2^{1–181}$-Scc4 complex ($Scc2^{1–181}$-$Scc4^{m3}$; Figure 6B). This Scc4 mutant perturbs centromeric cohesin loading in vivo, and we found it deficient in its interaction with the phosphorylated Ctf19 peptide in vitro. We also tested longer Ctf19 peptides bearing at most one phosphate group for binding with $Scc2^{1–181}$-Scc4 (Figure S5). We found that unmodified $Ctf19^{1–11}$ did not interact with $Scc2^{1–181}$-Scc4. This observation is consistent both with the dependence of centromeric cohesin loading on DDK and with our inability to detect an interaction between Ctf19-Mcm21 and Scc2/4 using unmodified recombinant protein complexes expressed in bacteria (Figure 6A). Single phosphorylation at either T4 or S5 of Ctf19 resulted in detectable binding, with phosphorylation at S5 providing a higher apparent affinity (Figure S5). We conclude that the N-terminal segment of Ctf19, once it has been phosphorylated by DDK, specifically interacts with the conserved surface patch of Scc4. Furthermore, phosphorylated Ctf19 residues T4 and S5 each contribute to this interaction, at least in vitro. This finding is consistent with our observation that Ctf19 residues T4 and S5 (along with T7) are the necessary and sufficient DDK targets for centromeric Scc2/4 localization (Figures 4C and 4E). Our binding studies also explain the observation that mutation of Ctf19 serine 5 to alanine without mutation of threonine 4 leads to a partial but not complete cohesion defect (*ctf19-3A-1*, Figure 5B).

Structural Basis of Ctf19 Recognition by Scc4

To observe the interaction between phosphorylated Ctf19 and Scc4 directly, we determined the crystal structure of a phosphorylated Ctf19 N-terminal peptide (MDFpTpSD) bound to $Scc2^{1–181}$-Scc4 (Figure 6C). Inclusion of the Ctf19 peptide yielded crystals in a different space group than the one we had reported previously for unliganded $Scc2^{1–181}$-Scc4 (PDB: 4XDN), with protomers now packing so that the Scc4 conserved patch was far enough from crystallographic symmetry mates to allow occupancy by a small peptide. Iterative density modification with non-crystallographic symmetry averaging revealed extra density near the Scc4 conserved patch, allowing us to model key Ctf19 residues, including the conserved F3 and the phosphorylated T4 and S5 residues.

This crystal structure explains both the results of our binding studies using synthetic Ctf19 peptides and the behavior of the *ctf19-2A* and *ctf19-6A* alleles in vivo. The phosphorylated Ctf19 residues T4 and S5 make electrostatic contacts with lysine side chains protruding from the Scc4 surface required for centromere cohesion (Scc4 K331 and K267). Coordination of these phosphates traps Ctf19 F3 between two aromatic side chains deep within the Scc4 conserved pocket (Scc4 F328 and F263). The structure also explains why the *scc4-m3*, *scc4-m35*, and *scc4-m7* alleles abrogate Scc2 targeting to centromeres (Hinshaw et al., 2015). The conserved residues F324 and K331 are mutated in all three alleles. In addition to direct coordination of the phosphate at Ctf19 T4 by Scc4 K331, residue F324 stabilizes the position of Scc4 F328 and therefore aids in recognition of Ctf19 F3. Scc4 residues Y298 and K299 were mutated in the *scc4-m7* allele. Y298 helps coordinate the phosphate at Ctf19 S5, and K299 makes contact with the Ctf19 peptide backbone, explaining the contribution of these residues to the *scc4-m7* allele.

Conservation of Scc4-Mediated Scc2 Targeting in Vertebrates

The mechanism of Scc2/4 targeting we have described depends on a surface of Scc4 conserved from yeasts to vertebrates

unperturbed vegetative growth and scored for the presence of centromeric Scc2-GFP foci. Values are the averages of at least three independent experiments. At least 25 cells were scored per strain per experiment (error bars, +/−SD for the independent experiments; ***p < 0.001; **p < 0.01; ns, not significant; Student's t test versus CTF19 strain, two tails, unequal variance). (D) Representative images from experiments quantified in (C) (scale bar, 3 μm). Time after the beginning of image acquisition is shown.

(E and F) Scc2 and Scc1 localization require Ctf19 phosphorylation. Cells were arrested in metaphase following treatment with nocodazole and benomyl for 2.5 hr (AM1176, untagged control strain). Values are means from four independent experiments (error bars, SE; *p < 0.05; **p < 0.01; Student's t test for indicated comparisons, unpaired, two-tails). (E) ChIP-qPCR for Scc2 performed using wild-type (AM6006), *CTF19* (AM20613), *ctf19-9A* (AM20619), *ctf19-6A* (AM20622), and *ctf19-2A* (AM21908) cells, carrying *SCC2-6His-3FLAG*. (F) ChIP-qPCR for Scc1 performed using wild-type (AM1145), *CTF19* (AM20629), *ctf19-9A* (AM23456), and *ctf19-2A* (AM21958) cells, carrying *SCC1-6HA*.

(G and H) Immunoblots showing levels of Scc2-6His-3FLAG, Scc1-6HA, and Pgk1 (loading control) from a representative experiment.

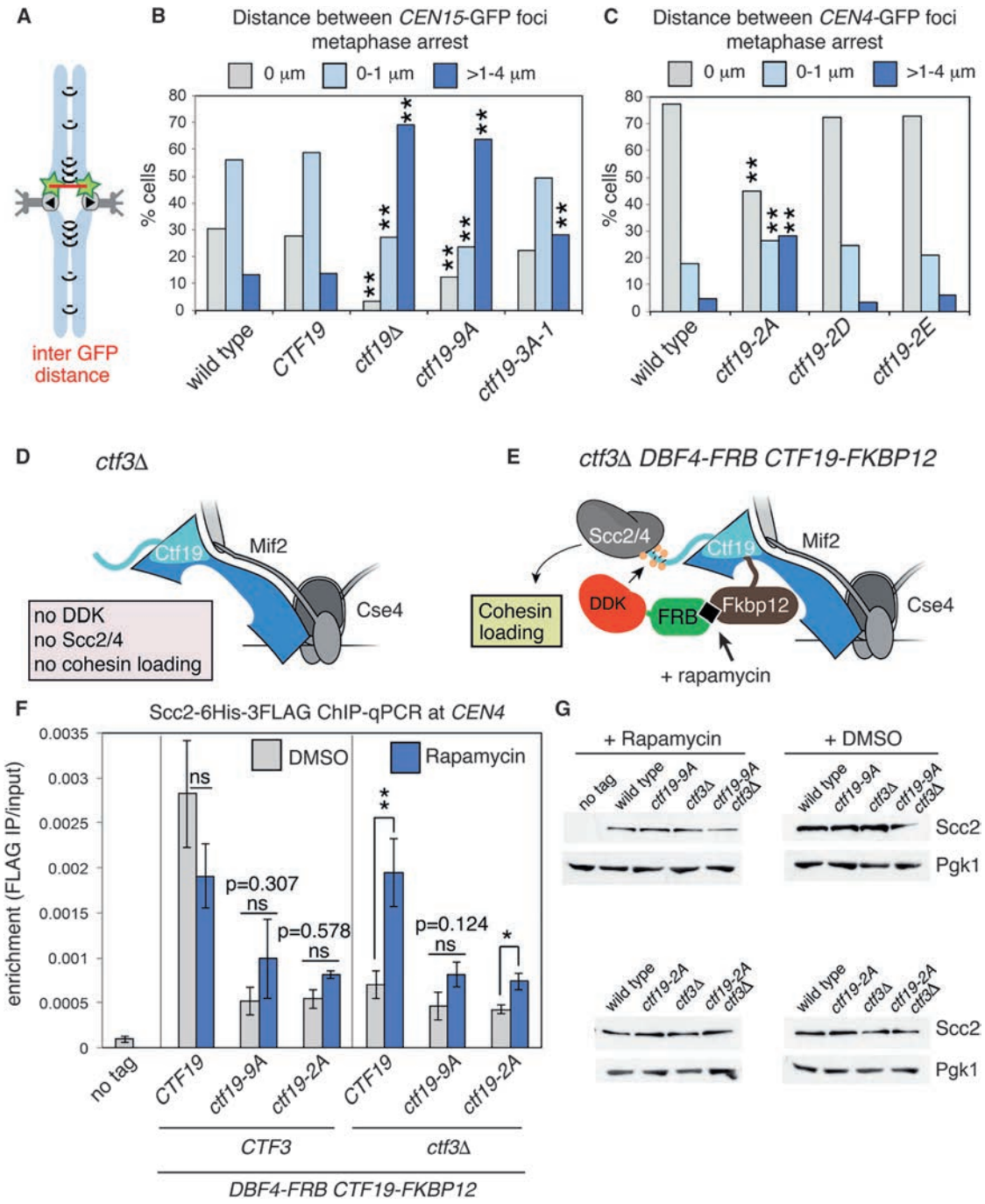


Figure 5. Ctf19 Phosphorylation Is Required for Centromeric Cohesion Establishment

(A) Schematic depicting sister centromere separation at metaphase. Distance between centromere-associated GFP foci (green stars) was measured (red line). (B and C) Sister centromeres are further apart in metaphase in *ctf19-9A* cells compared to wild-type cells. (B) Strains bearing a *lacO* array at *CEN15* (1.8 kb to the left of *CEN15*) and expressing LacI-GFP were imaged after arrest in metaphase by depletion of Cdc20 (*pMET-HA3-CDC20*) for 2h. Average centromere separation was determined for wild-type (SMH159), *CTF19* (SMH412), *ctf19*Δ (SMH395), *ctf19-9A* (SMH396), and *ctf19-3A-1* (SMH397) strains. Values were binned into categories (0 μm for no separation, 0–1 μm, and >1 μm–4 μm) and expressed as a percentage of all cells scored across three experiments. The fraction of cells in each bin was compared to the wild-type using Fishers exact test (**p < 0.0001). (C) Phospho-null but not phospho-mimetic mutations in phosphorylation sites closest to the Ctf19 N terminus cause a centromere cohesion defect. Wild-type (AM4643), *ctf19-2A* (AM22083), *ctf19-2D* (AM22481), and *ctf19-2E* (AM22484) strains, all carrying *+2.4 CEN4-GFP* (2.4 kb to the right of *CEN4*) were arrested in G1 before release into medium containing methionine to deplete *CDC20* for 2 hr before imaging and analyzed as in (B).

(D and E) Schematic showing the cohesin loading defect in *ctf3*Δ cells (D) and rescue of this defect by artificial tethering of Dbf4-FRB to the kinetochore (E).

(legend continued on next page)

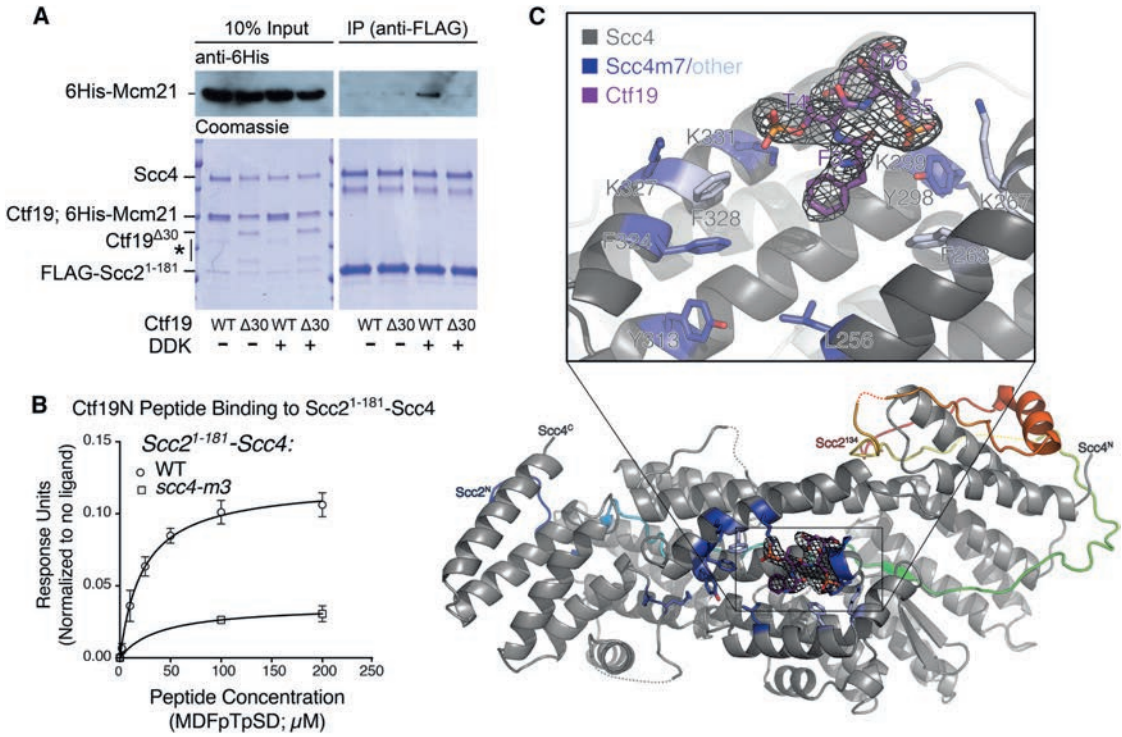


Figure 6. Scc4 Recognizes Phosphorylated Ctf19

(A) Reconstitution of DDK-dependent Scc4-Ctf19 interaction in vitro. Recombinant Ctf19-6His-Mcm21 complexes were phosphorylated with purified DDK and then used for pulldowns with recombinant FLAG-Scc2$^{1-181}$-Scc4. Proteins were immunopurified on anti-FLAG beads, resolved by SDS-PAGE, and detected by immunoblot against an 6His-Mcm21 (top) or by Coomassie stain (total protein, bottom; *, degradation product).

(B) Synthetic Ctf19$^{1-6}$ peptide (MDFpTpSD) was tested for binding to Scc2$^{1-181}$-Scc4^{WT} or Scc2$^{1-181}$-Scc4^{m3}. Mean average scaled response units are shown for three independent experiments for each data point.

(C) Structure of phosphorylated Ctf19$^{1-6}$ bound to Scc2$^{1-181}$-Scc4. An overview of the complex (bottom) is shown as a cartoon with Scc4 in gray and Scc2$^{1-181}$ in rainbow (violet, N terminus; red, C terminus). The Ctf19-Scc4 interaction is shown in detail above. Individual residues mutated in the *scc4-m7* allele are in purple. Other residues contributing to the interaction are in light blue. Ctf19 is purple with non-carbon atoms colored by element (red, oxygen; blue, nitrogen; yellow, phosphorous). Omit map density contoured to 0.8 sigma is shown for the Ctf19 peptide.

See also Figure S5 and Tables S2 and S3.

(Figure 7A). Ctf19 phosphorylation sites are broadly conserved in yeasts, but we found that vertebrate Ctf19 homologs, including *X. laevis* CENP-P, do not have candidate phosphorylation sites that align well with those required for Scc2 recruitment in yeast (Figure S6). DDK activity is nonetheless required for Scc2/4 to localize to chromatin in vertebrates (Takahashi et al., 2008). To test whether the Scc4 conserved patch is required for vertebrate Scc2/4 recruitment, we tested whether *X. laevis* Scc2$^{1-1024}$-Mau2^{Scc4} can associate with chromatin upon incubation with high speed egg supernatant (HSS; Figure 7B). We found that a version of Mau2^{Scc4} that mimics the yeast *scc4-m3* mutant (Mau2^{Scc4} Y307A, K310D, K331A; m3) did not associate with chromatin. We conclude that a mechanism analogous to the one we have described for Scc2/4 localization in yeast is also present in vertebrates.

DISCUSSION

We have presented a molecular explanation for targeted cohesin loading at centromeres. The following requirements for this pathway have been demonstrated previously: (1) kinetochore assembly and, more specifically, the Ctf19 complex (Eckert et al., 2007; Fernius and Marston, 2009; Ng et al., 2009), (2) the activity and centromeric localization of DDK (Natsume et al., 2013), (3) a conserved structural feature of the Scc4 protein (Hinshaw et al., 2015), and (4) expression of the Scc1 protein (Fernius

(F) Scc2 localization was assessed by ChIP-qPCR in the presence or absence of DDK tethering, *CTF3*, and Ctf19 phosphorylation sites. Wild-type (AM21613), *ctf19-9A* (AM21681), *ctf19-2A* (AM22360), *ctf3*Δ (AM21614), *ctf19-9A ctf3*Δ (AM21617), *ctf19-2A ctf3*Δ (AM22672), and control (AM21486) strains, all expressing *DBF4-FRB* and *CTF19-FKBP12*, were arrested in G1 for 3 hr before release into medium containing nocodazole, benomyl, and either DMSO (no DDK tethering) or rapamycin (DDK tethering; 1 μM) for 2.5 hr. Mean values of at least four independent experiments are shown. Error bars indicate SE (*$p < 0.05$; **$p < 0.01$; Student's t test, unpaired, two-tails; ns, not significant).

(G) Immunoblots showing relative levels of Scc2-6His-3FLAG and Pgk1 (loading control) in a representative experiment.

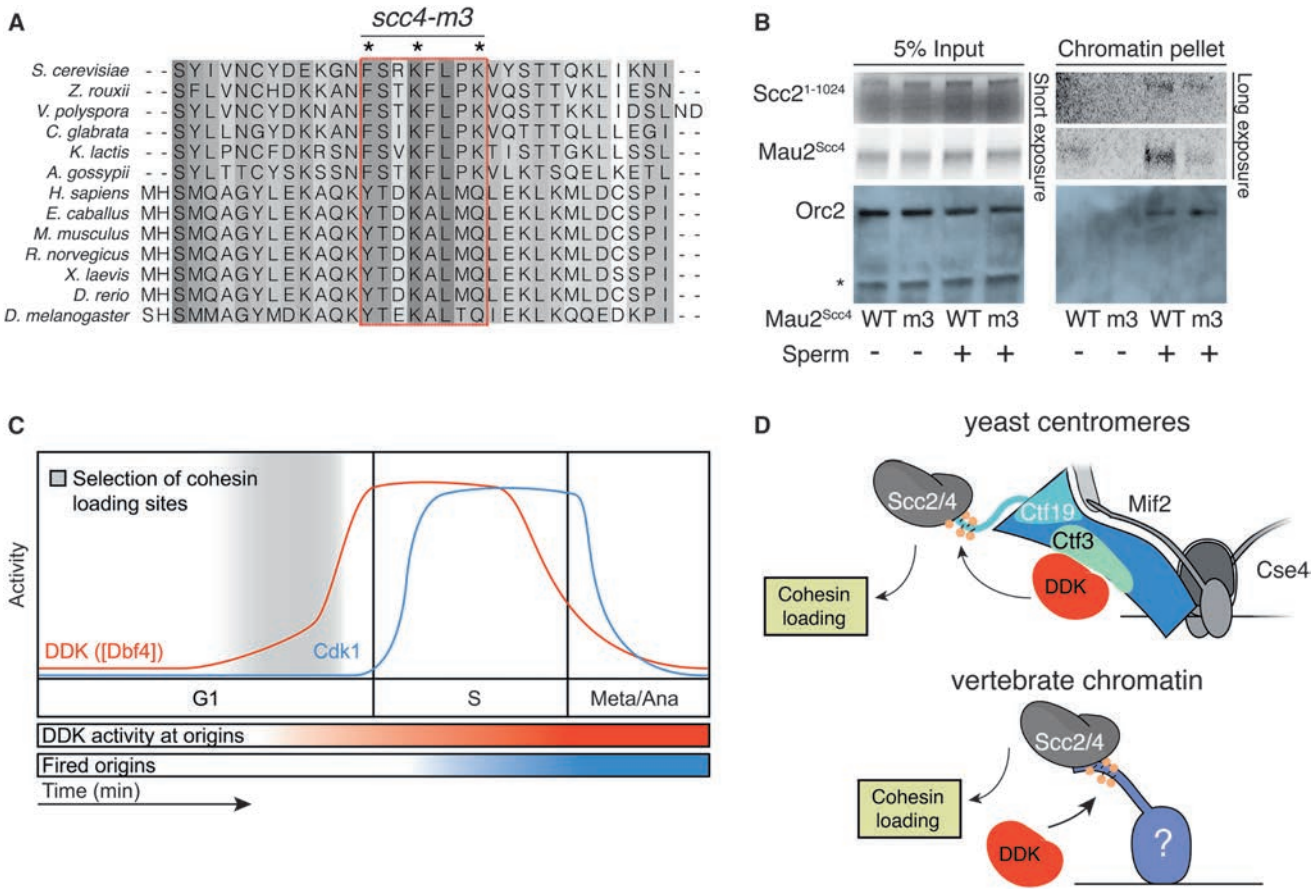


Figure 7. Scc4-Dependent Scc2/4 Localization in Vertebrates

(A) Multiple sequence alignment for Scc4 homologs in yeasts and animals with residues mutated in the *scc4-m3* allele shown.

(B) Chromatin association of *X. laevis* Scc2$^{1-1024}$-Scc4 depends on the Scc4 conserved patch. *X. laevis* high-speed supernatant (HSS) was incubated with sperm chromatin and recombinant Scc2$^{1-1024}$-Scc4. Chromatin-associated proteins were recovered by sucrose sedimentation and detected by autoradiography (Scc2$^{1-1024}$-Scc4) or western blot (Orc2; *non-specific band).

(C) Schematic showing idealized kinase activity during the cell cycle.

(D) Cartoon comparing DDK-dependent targeted cohesin loading at the yeast centromere with a similar process in vertebrates.

See also Figure S6.

et al., 2013). We have determined that phosphorylation of Ctf19 by DDK provides a binding site for Scc4 at kinetochores, explaining the first three of these four requirements.

These findings support a two-step process for cohesin recruitment to centromeres. First, the Ctf19 complex, through Ctf3, recruits DDK in G1, which phosphorylates the Ctf19 protein and possibly other targets. Second, phosphorylated Ctf19 recruits the cohesin loader. Centromere cohesion requires both steps, while early replication of centromeric DNA, requires only the first (Natsume et al., 2013). Our identification of separation of function mutations that do not affect DDK localization and activity but that do perturb cohesin localization to centromeres strongly supports the two-step model. Furthermore, deletion of factors important for DDK recruitment (*chl4*Δ or *ctf3*Δ) does not perturb Ctf19 localization at centromeres (Natsume et al., 2013; Pot et al., 2003), indicating that failure to phosphorylate Ctf19 is what prevents Scc2/4 recruitment. This division of labor between Ctf19 complex components—one acts as the DDK recruiter (Ctf3) and another acts as the DDK substrate (Ctf19)—was confirmed by our reconstitution of centromere-targeted cohesin loading in vivo by tethering DDK in the absence of the recruiter.

The first step in centromeric cohesin enrichment is specification of centromeres as preferred loading sites. Several lines of evidence have established that a Ctf19 complex-dependent pathway operates before S phase to achieve this (Figure 7C). First, DDK appears at centromeres before DNA replication (Natsume et al., 2013)(Figures 3, S4A, and S4B), and low levels of Ctf19 phosphorylation are apparent in G1 (Figures 2C and S5A). Second, overexpression of the limiting cohesin subunit, Scc1, drives Scc2 onto centromeres in G1 and depends on the Ctf19 complex to do so (Fernius et al., 2013). Third, vertebrate Scc2/4 associates with chromatin before DNA replication (Figure 7B) (Takahashi et al., 2008). The same is probably also true of yeast Scc2/4: (1) depletion of the DNA replication factor, Cdc6, does not prevent cohesin loading (Fernius et al., 2013), and (2) cells arrested in late G1 by expression of a dominant allele

of the Cdc28/Cdk1 inhibitor Sic1 have a cohesin distribution identical to that seen in metaphase (Lopez-Serra et al., 2013).

The second step in centromeric cohesin enrichment is the specific docking of Scc2/4 onto phosphorylated Ctf19 at kinetochores. The factors we have described here are sufficient for this activity in vitro (Figure 2A), but Scc1 expression is required in vivo. Scc1 production at cell-cycle entry completes the cohesin ring and enables its association with the Scc2/4 cohesin loader (Fernius et al., 2013). Therefore, cohesin potentiates the ability of Scc2/4 to associate with phosphorylated Ctf19 by a yet undetermined mechanism.

Cohesin remains on chromatin and is activated for cohesion during S phase (Rhodes et al., 2017; Uhlmann and Nasmyth, 1998). The two-step pathway we have described ensures that cohesin is efficiently loaded onto centromeres before they replicate, consistent with the observation that slowing DNA replication partially ameliorates chromosome segregation defects in Ctf19 complex mutants (Fernius and Marston, 2009). Moreover, our observation that phosphorylated Ctf19 accumulates in *chl4*Δ strains upon DDK overproduction suggests that Chl4 and Ctf3 amplify DDK's activity locally when DDK levels would otherwise be limiting. In this way, when Dbf4 levels begin to rise in late G1 (Jackson et al., 1993), the centromere is assured both of replicating early and of accumulating high levels of cohesin (Figure 7C).

We have previously proposed the existence of two modes of cohesin loading (Hinshaw et al., 2015). One mode operates specifically at centromeres and depends on the Scc4 conserved patch and the factors we have described here. A second mode results in cohesin loading genome-wide and is sufficient to confer viability. How Scc2/4 and cohesin are recruited to chromatin in the second mode is still unclear but may involve the RSC chromatin remodeling complex (Lopez-Serra et al., 2014).

How is Scc4 localization regulated in other chromosomal contexts? We propose an analogous process, in which phosphorylated chromosomal factors other than Ctf19 recruit Scc2/4. In support of this idea, the Scc4 conserved patch mutant (*scc4-m35*) displays meiotic cohesin localization defects on chromosome arms (Vincenten et al., 2015), hinting at the existence of meiosis-specific Scc4 binding partners. Furthermore, a DDK- and Scc4-dependent Scc2/4 localization pathway operates in *X. laevis* egg extracts (Takahashi et al., 2008), although the DDK substrate in this pathway is not known (Figure 7D). It is therefore likely that the mechanism we have described here reflects general principles in cohesin-mediated chromosome organization.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACTS FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Recombinant protein production
  - In vitro translation and pulldowns
  - DDK purification
  - DDK activity assay
  - Peptide binding experiments
  - Crystal preparation
  - Structure determination
  - Strain construction and growth conditions
  - Cohesion assay
  - Artificial tethering of Dbf4
  - Live-cell GFP fusion protein imaging
  - Chromatin immunoprecipitation
  - Western blotting
  - In vivo Ctf19 phosphorylation assay
  - Chromatin association in *X. laevis* HSS
  - Alignments and evolutionary conservation analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.017.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the project. All authors participated in writing the manuscript. S.M.H. and V.M. performed experiments. S.M.H., V.M., and A.L.M. generated reagents.

ACKNOWLEDGMENTS

We thank Terry Orr-Weaver, Johannes Walter, and Stephen Buratowski for helpful comments, the staff at the Nikon Imaging Center at Harvard Medical School for microscopy support, and Thomas Graham and James Dewar for vital instruction and reagents for the *X. laevis* experiments. The staff at APS/NE-CAT beamlines 24-IDC and 24-IDE helped with data collection. This work was supported by a Wellcome Senior Research Fellowship to A.L.M. (107827) and core funding for the Wellcome Centre for Cell Biology (203149). NE-CAT is supported by NIH grant P41 GM 103403. S.C.H. is an Investigator in the Howard Hughes Medical Institute.

Received: April 8, 2017
Revised: May 3, 2017
Accepted: August 9, 2017
Published: September 21, 2017

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. Biol Crystallogr. *66*, 213–221.

Bailis, J.M., Bernard, P., Antonelli, R., Allshire, R.C., and Forsburg, S.L. (2003). Hsk1-Dfp1 is required for heterochromatin-mediated cohesion at centromeres. Nat. Cell Biol. *5*, 1111–1116.

Chao, W.C., Murayama, Y., Muñoz, S., Costa, A., Uhlmann, F., and Singleton, M.R. (2015). Structural studies reveal the functional modularity of the Scc2-Scc4 cohesin loader. Cell Rep. *12*, 719–725.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol. Cell *5*, 243–254.

De Wulf, P., McAinsh, A.D., and Sorger, P.K. (2003). Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. Genes Dev. *17*, 2902–2921.

Eckert, C.A., Gravdahl, D.J., and Megee, P.C. (2007). The enhancement of pericentromeric cohesin association by conserved kinetochore components promotes high-fidelity chromosome segregation and is sensitive to microtubule-based tension. Genes Dev. *21*, 278–291.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Fay, A., Misulovin, Z., Li, J., Schaaf, C.A., Gause, M., Gilmour, D.S., and Dorsett, D. (2011). Cohesin selectively binds and regulates genes with paused RNA polymerase. Curr. Biol. *21*, 1624–1634.

Fernius, J., and Marston, A.L. (2009). Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. PLoS Genet. *5*, e1000629.

Fernius, J., Nerusheva, O.O., Galander, S., Alves, Fde.L., Rappsilber, J., and Marston, A.L. (2013). Cohesin-dependent association of scc2/4 with the centromere initiates pericentromeric cohesion establishment. Curr. Biol. *23*, 599–606.

Francis, L.I., Randell, J.C., Takara, T.J., Uchima, L., and Bell, S.P. (2009). Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev. *23*, 643–654.

Glynn, E.F., Megee, P.C., Yu, H.G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. PLoS Biol. *2*, E259.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Hinshaw, S.M., and Harrison, S.C. (2013). An Iml3-Chl4 heterodimer links the core centromere to factors required for accurate chromosome segregation. Cell Rep. *5*, 29–36.

Hinshaw, S.M., Makrantoni, V., Kerr, A., Marston, A.L., and Harrison, S.C. (2015). Structural evidence for Scc4-dependent localization of cohesin loading. eLife *4*, e06057.

Hiraga, S., Alvino, G.M., Chang, F., Lian, H.Y., Sridhar, A., Kubota, T., Brewer, B.J., Weinreich, M., Raghuraman, M.K., and Donaldson, A.D. (2014). Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. Genes Dev. *28*, 372–383.

Holland, A.J., and Cleveland, D.W. (2012). Losing balance: the origin and impact of aneuploidy in cancer. EMBO Rep. *13*, 501–514.

Jackson, A.L., Pahl, P.M., Harrison, K., Rosamond, J., and Sclafani, R.A. (1993). Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. Mol. Cell. Biol. *13*, 2899–2908.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kawauchi, S., Calof, A.L., Santos, R., Lopez-Burks, M.E., Young, C.M., Hoang, M.P., Chua, A., Lao, T., Lechner, M.S., Daniel, J.A., et al. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. PLoS Genet. *5*, e1000650.

Kogut, I., Wang, J., Guacci, V., Mistry, R.K., and Megee, P.C. (2009). The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. Genes Dev. *23*, 2345–2357.

Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A., et al. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat. Genet. *36*, 631–635.

Lee, J.Y., and Orr-Weaver, T.L. (2001). The molecular basis of sister-chromatid cohesion. Annu. Rev. Cell Dev. Biol. *17*, 753–777.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573–578.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics *23*, 127–128.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Lopez-Serra, L., Lengronne, A., Borges, V., Kelly, G., and Uhlmann, F. (2013). Budding yeast Wapl controls sister chromatid cohesion maintenance and chromosome condensation. Curr. Biol. *23*, 64–69.

Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A., and Uhlmann, F. (2014). The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat. Genet. *46*, 1147–1151.

Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev. *19*, 1269–1287.

Mayer, M.L., Pot, I., Chang, M., Xu, H., Aneliunas, V., Kwok, T., Newitt, R., Aebersold, R., Boone, C., Brown, G.W., and Hieter, P. (2004). Identification of protein complexes required for efficient sister chromatid cohesion. Mol. Biol. Cell *15*, 1736–1745.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

Measday, V., Hailey, D.W., Pot, I., Givan, S.A., Hyland, K.M., Cagney, G., Fields, S., Davis, T.N., and Hieter, P. (2002). Ctf3p, the Mis6 budding yeast homolog, interacts with Mcm22p and Mcm16p at the yeast outer kinetochore. Genes Dev. *16*, 101–113.

Megee, P.C., and Koshland, D. (1999). A functional assay for centromere-associated sister chromatid cohesion. Science *285*, 254–257.

Murakami, H., and Keeney, S. (2014). Temporospatial coordination of meiotic DNA replication and recombination via DDK recruitment to replisomes. Cell *158*, 861–873.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu. Rev. Genet. *35*, 673–745.

Nasmyth, K. (2011). Cohesin: a catenase with separate entry and exit gates? Nat. Cell Biol. *13*, 1170–1177.

Natsume, T., Müller, C.A., Katou, Y., Retkute, R., Gierliński, M., Araki, H., Blow, J.J., Shirahige, K., Nieduszynski, C.A., and Tanaka, T.U. (2013). Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. Mol. Cell *50*, 661–674.

Ng, T.M., Waples, W.G., Lavoie, B.D., and Biggins, S. (2009). Pericentromeric sister chromatid cohesion promotes kinetochore biorientation. Mol. Biol. Cell *20*, 3818–3827.

Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., 3rd, Desai, A., and Fukagawa, T. (2006). The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat. Cell Biol. *8*, 446–457.

Onn, I., Heidinger-Pauli, J.M., Guacci, V., Unal, E., and Koshland, D.E. (2008). Sister chromatid cohesion: a simple concept with a complex reality. Annu. Rev. Cell Dev. Biol. *24*, 105–129.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307.

Pot, I., Measday, V., Snydsman, B., Cagney, G., Fields, S., Davis, T.N., Muller, E.G., and Hieter, P. (2003). Chl4p and iml3p are two new members of the budding yeast outer kinetochore. Mol. Biol. Cell *14*, 460–476.

Rhodes, J.D.P., Haarhuis, J.H.I., Grimm, J.B., Rowland, B.D., Lavis, L.D., and Nasmyth, K.A. (2017). Cohesin can remain associated with chromosomes during DNA replication. bioRxiv. http://dx.doi.org/10.1101/124107.

Rockmill, B., Voelkel-Meiman, K., and Roeder, G.S. (2006). Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in Saccharomyces cerevisiae. Genetics *174*, 1745–1754.

Rollins, R.A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577–593.

Sakuno, T., Tada, K., and Watanabe, Y. (2009). Kinetochore geometry defined by cohesion within the centromere. Nature *458*, 852–858.

Schmitzberger, F., and Harrison, S.C. (2012). RWD domain: a recurring module in kinetochore architecture shown by a Ctf19-Mcm21 complex structure. EMBO Rep. *13*, 216–222.

Sheu, Y.J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol. Cell *24*, 101–113.

Sheu, Y.J., and Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature *463*, 113–117.

Stern, B.M., and Murray, A.W. (2001). Lack of tension at kinetochores activates the spindle checkpoint in budding yeast. Curr. Biol. *11*, 1462–1467.

Takahashi, T.S., Basu, A., Bermudez, V., Hurwitz, J., and Walter, J.C. (2008). Cdc7-Drf1 kinase links chromosome cohesion to the initiation of DNA replication in Xenopus egg extracts. Genes Dev. *22*, 1894–1905.

Tanaka, T., Fuchs, J., Loidl, J., and Nasmyth, K. (2000). Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. Nat. Cell Biol. *2*, 492–499.

Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat. Genet. *36*, 636–641.

Tropea, J.E., Cherry, S., and Waugh, D.S. (2009). Expression and purification of soluble His(6)-tagged TEV protease. Methods Mol. Biol. *498*, 297–307.

Uhlmann, F. (2016). SMC complexes: from DNA to chromosomes. Nat. Rev. Mol. Cell Biol. *17*, 399–412.

Uhlmann, F., and Nasmyth, K. (1998). Cohesion between sister chromatids must be established during DNA replication. Curr. Biol. *8*, 1095–1101.

Vincenten, N., Kuhl, L.M., Lam, I., Oke, A., Kerr, A.R., Hochwagen, A., Fung, J., Keeney, S., Vader, G., and Marston, A.L. (2015). The kinetochore prevents centromere-proximal crossover recombination during meiosis. eLife *4*, e10850.

Wan, L., Zhang, C., Shokat, K.M., and Hollingsworth, N.M. (2006). Chemical inactivation of cdc7 kinase in budding yeast results in a reversible arrest that allows efficient cell synchronization prior to meiotic recombination. Genetics *174*, 1767–1774.

Watanabe, Y., and Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature *400*, 461–464.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Weber, S.A., Gerton, J.L., Polancic, J.E., DeRisi, J.L., Koshland, D., and Megee, P.C. (2004). The kinetochore is an enhancer of pericentric cohesin binding. PLoS Biol. *2*, E260.

Cell

Dynamic Control of X Chromosome Conformation and Repression by a Histone H4K20 Demethylase

Katjuša Brejc,[1,3] Qian Bian,[1,3] Satoru Uzawa,[1] Bayly S. Wheeler,[1,4] Erika C. Anderson,[1] David S. King,[2] Philip J. Kranzusch,[1,5,6] Christine G. Preston,[1,7] and Barbara J. Meyer[1,8,*]

[1]Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720-3204, USA

[2]HHMI Mass Spectrometry Laboratory, University of California, Berkeley, Berkeley, California 94720-3204, USA

[3]These authors contributed equally

[4]Present address: Department of Biology, Rhodes College, Memphis, TN 38112, USA

[5]Present address: Department of Cancer Immunology and Virology at the Dana-Farber Cancer Institute, Boston, MA 02115, USA

[6]Present address: Department of Microbiology and Immunobiology at Harvard Medical School, Boston, MA 02115, USA

[7]Present address: Invitae, San Francisco, CA 94107, USA

[8]Lead Contact

*Correspondence: bjmeyer@berkeley.edu

http://dx.doi.org/10.1016/j.cell.2017.07.041

SUMMARY

Chromatin modification and higher-order chromosome structure play key roles in gene regulation, but their functional interplay in controlling gene expression is elusive. We have discovered the machinery and mechanism underlying the dynamic enrichment of histone modification H4K20me1 on hermaphrodite X chromosomes during *C. elegans* dosage compensation and demonstrated H4K20me1's pivotal role in regulating higher-order chromosome structure and X-chromosome-wide gene expression. The structure and the activity of the dosage compensation complex (DCC) subunit DPY-21 define a Jumonji demethylase subfamily that converts H4K20me2 to H4K20me1 in worms and mammals. Selective inactivation of demethylase activity eliminates H4K20me1 enrichment in somatic cells, elevates X-linked gene expression, reduces X chromosome compaction, and disrupts X chromosome conformation by diminishing the formation of topologically associating domains (TADs). Unexpectedly, DPY-21 also associates with autosomes of germ cells in a DCC-independent manner to enrich H4K20me1 and trigger chromosome compaction. Our findings demonstrate the direct link between chromatin modification and higher-order chromosome structure in long-range regulation of gene expression.

INTRODUCTION

Gene regulation in multicellular organisms entails diverse strategies that operate over different distances. Some mechanisms act locally on individual genes, whereas others act over long range to co-regulate expression of numerous genes within sub-chromosomal domains (e.g., imprinted gene clusters, *hox* clusters) (Andrey et al., 2013; Plasschaert and Bartolomei, 2014; Williamson et al., 2012), or across entire chromosomes (e.g., X chromosome dosage compensation) (da Rocha and Heard, 2017; Kuroda et al., 2016; Meyer, 2010; Strome et al., 2014). Post-translational modifications of histones and three-dimensional (3D) chromosome architecture are prime elements of long-range gene regulation, but their functional interplay in controlling gene expression has been elusive (Dekker and Mirny, 2016; Soshnev et al., 2016).

Here, we dissect X chromosome dosage compensation (DC) in *C. elegans* to determine the effect of chromatin modifications on higher-order chromosome structure during long-range gene regulation. DC is exemplary because it controls hundreds of genes simultaneously, it distinguishes X chromosomes from autosomes, and it discriminates between the sexes in modulating gene expression. The *C. elegans* dosage compensation complex (DCC), a ten-subunit complex with five condensin subunits, binds to both hermaphrodite X chromosomes to reduce gene expression by half, thereby equalizing it with that of the single male X (Meyer, 2010). Moreover, the DCC remodels the topology of hermaphrodite X chromosomes into a spatial conformation distinct from that of autosomes or male X chromosomes by forming topologically associating domains (TADs) (Crane et al., 2015). During that process, the chromatin modification H4K20me1 becomes enriched on the dosage-compensated X chromosomes (Kramer et al., 2015; Liu et al., 2011; Vielle et al., 2012; Wells et al., 2012). Neither the role of this modification in chromosome-wide gene control nor its effect on X chromosome structure has been determined.

H4K20me1 is also enriched on the inactive X of female mammals, revealing a common feature of diverse DC strategies (Kohlmaier et al., 2004). Its contribution to transcriptional silencing is unknown. In general, the role of H4K20me1 in gene regulation has remained a puzzle due to its context-dependent contribution to both gene activation and gene repression (Beck et al., 2012).

Although H4K20 methylation has been implicated in myriad nuclear functions including gene regulation, DNA replication and repair, mitotic chromosome condensation, and cell-cycle control, the mechanisms that regulate different H4K20me states

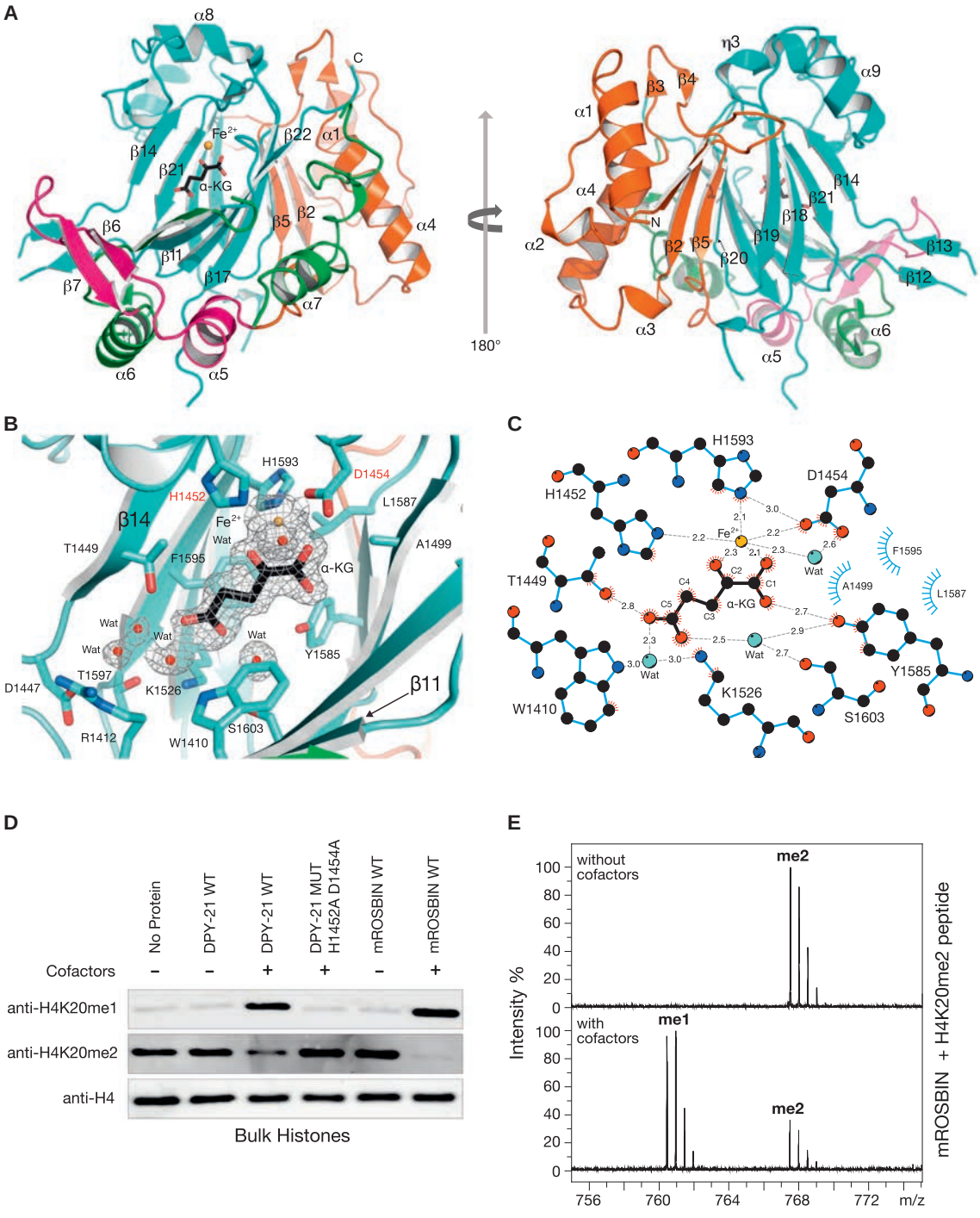


Figure 1. Structural and Functional Characterization of the DPY-21 JmjC Demethylase Domain

(A) DPY-21$^{1210–1617}$ structure in complex with α-KG (black) and Fe^{2+} (orange) showing JmjC (cyan), JmjN (orange), β-hairpin (pink), and mixed (green) domains.

(B) Active site of DPY-21$^{1210–1617}$ showing JmjC domain residues (cyan) in complex with α-KG (black), Fe^{2+} (orange), and water molecules (red). Facial triad residues H1452 and D1454 (red letters) were changed to alanines for in vitro and in vivo studies. The electron density, $2F_o$-F_c (mesh), contoured at 1.0 σ above the mean is shown for Fe^{2+}, α-KG, and water molecules.

(C) Schematic diagram of contacts made by α-KG with active-site residues and water molecules. Contacts include electrostatic (gray dashed lines in angstroms) and hydrophobic (blue and red eyelashes) interactions.

(legend continued on next page)

(H4K20me1/me2/me3) and transduce these states into properly executed nuclear functions are not well understood (Beck et al., 2012; Jørgensen et al., 2013; van Nuland and Gozani, 2016). In principle, H4K20me1 enrichment on *C. elegans* X chromosomes could occur by activating the methyltransferase (SET-1) that converts H4K20 to H4K20me1, by inhibiting the methyltransferase (SET-4) that converts H4K20me1 to H4K20me2/me3, by inhibiting the demethylase (JMJD-1.1/1.2) that converts H4K20me1 to H4K20, or by activating an unknown demethylase that converts H4K20me2 to H4K20me1. Although H4K20me2 is the predominant form of H4K20 in eukaryotic cells (Pesavento et al., 2008), only a neuron-specific H4K20me2 demethylase has been reported for any organism (Wang et al., 2015). No H4K20me2 demethylase has yet been identified that could regulate H4K20me1 levels during DC.

We pursued the machinery and mechanisms that catalyze H4K20me1 enrichment on X chromosomes and the role of H4K20me1 in higher-order chromosome structure and gene regulation. The key element controlling H4K20me1 levels is a DCC subunit. Apart from five condensin subunits (DPY-26, DPY-27, DPY-28, MIX-1, and CAPG-1), the DCC includes an XX-specific protein (SDC-2) that triggers the assembly of the DCC onto X chromosomes, two proteins (SDC-3 and DPY-30) that aid SDC-2 in recruiting the DCC to X, and two proteins (SDC-1 and DPY-21) required for full DCC activity, but not assembly (Meyer, 2010).

Through X-ray crystallography and biochemical assays, we found that DPY-21 defines a new subfamily of Jumonji C (JmjC) histone demethylases that converts H4K20me2 to H4K20me1 and is widely conserved from worms to mammals. Inactivation of JmjC activity in vivo abrogated H4K20me1 enrichment on X, reduced X chromosome compaction, disrupted X chromosome topology by weakening TAD boundaries, and elevated X-linked gene expression, demonstrating key roles for H4K20me1 in DC. Unexpectedly in germ cells, DPY-21 associates with autosomes, but not X chromosomes, in a DCC-independent manner to enrich H4K20me1 and facilitate chromosome compaction. Thus, DPY-21 is an adaptable chromatin regulator that is harnessed during development for distinct biological functions. In both somatic cells and germ cells, H4K20me1 enrichment modulates 3D chromosome architecture, demonstrating the direct impact of chromatin modification on establishing higher-order chromosome structure.

RESULTS

DPY-21 Is a JmjC Domain-Containing Protein

Although none of the ten DCC subunits had a demethylase domain identified by primary sequence analysis, structure prediction programs (STAR Methods) suggested homology between the C-terminal domain of DPY-21 and JmjC domain-containing lysine demethylases (KDMs), despite low (15%) sequence identity. JmjC KDMs are Fe^{2+} and α-ketoglutarate (α-KG)-dependent dioxygenases that demethylate lysines in histone and non-histone proteins (Markolovic et al., 2016). To investigate this potential connection, we determined a 1.8 Å crystal structure of the *C. elegans* DPY-$21^{1210–1617}$ fragment that encompasses the putative JmjC domain (Table S1; STAR Methods). Strikingly, comparisons to known structures using DALI (Holm and Rosenström, 2010) revealed DPY-21 to be a JmjC domain-containing protein most similar to the JmjC KDMs (Z score of 21.3–14.5; root-mean-square deviation of 2.9 Å–3.7 Å), strongly suggesting that DPY-21 indeed acts as a histone demethylase.

Similar to other JmjC KDMs, the DPY-$21^{1210–1617}$ structure includes a JmjC domain that is folded into a double-stranded β-helix (DSBH) and surrounded by a Jumonji N (JmjN) domain, a β-hairpin motif, and a mixed domain (Figure 1A) (Chen et al., 2006). Like in all α-KG dioxygenases, the DPY-21 DSBH core bears facial triad residues (H1452, D1454, and H1593) that chelate Fe^{2+} and form an active site (Figures 1B and 1C). The electron density map revealed an unexpected active site density that was identified as an α-KG molecule, although no α-KG was added during purification or crystallization. α-KG coordinates Fe^{2+} in a bidentate manner, and a water molecule completes the octahedral coordination of Fe^{2+} (Figures 1B and 1C). The α-KG is further stabilized by hydrogen bonding to side chains of T1449 and Y1585 and also by hydrogen bonding to side chains of W1410, K1526, and S1603 via water molecules. In addition, α-KG forms van der Waals contacts with A1499, L1587, and F1595 (Figures 1B and 1C).

The DPY-21 C-terminal domain is similar (33% sequence identity) to C-terminal domains of metazoan proteins named round spermatid basic proteins 1 (ROSBIN) (Yonker and Meyer, 2003) (Figure S1B). Mouse ROSBIN, an essential protein with minimal biochemical characterization, is expressed in embryos and male gonads (Koscielny et al., 2014; Takahashi et al., 2004). Structure-guided sequence comparison of DPY-21 with ROSBIN family members showed that all residues that interact with Fe^{2+} and α-KG are conserved (Figures S1A and S1B). Moreover, sequence conservation between these proteins (Figure S1A) extends to regions participating in histone peptide binding, as revealed by JmjC KDM structures (Markolovic et al., 2016). Extensive sequence conservation with DPY-21 suggests that ROSBINs may also be JmjC demethylases that share substrate specificity.

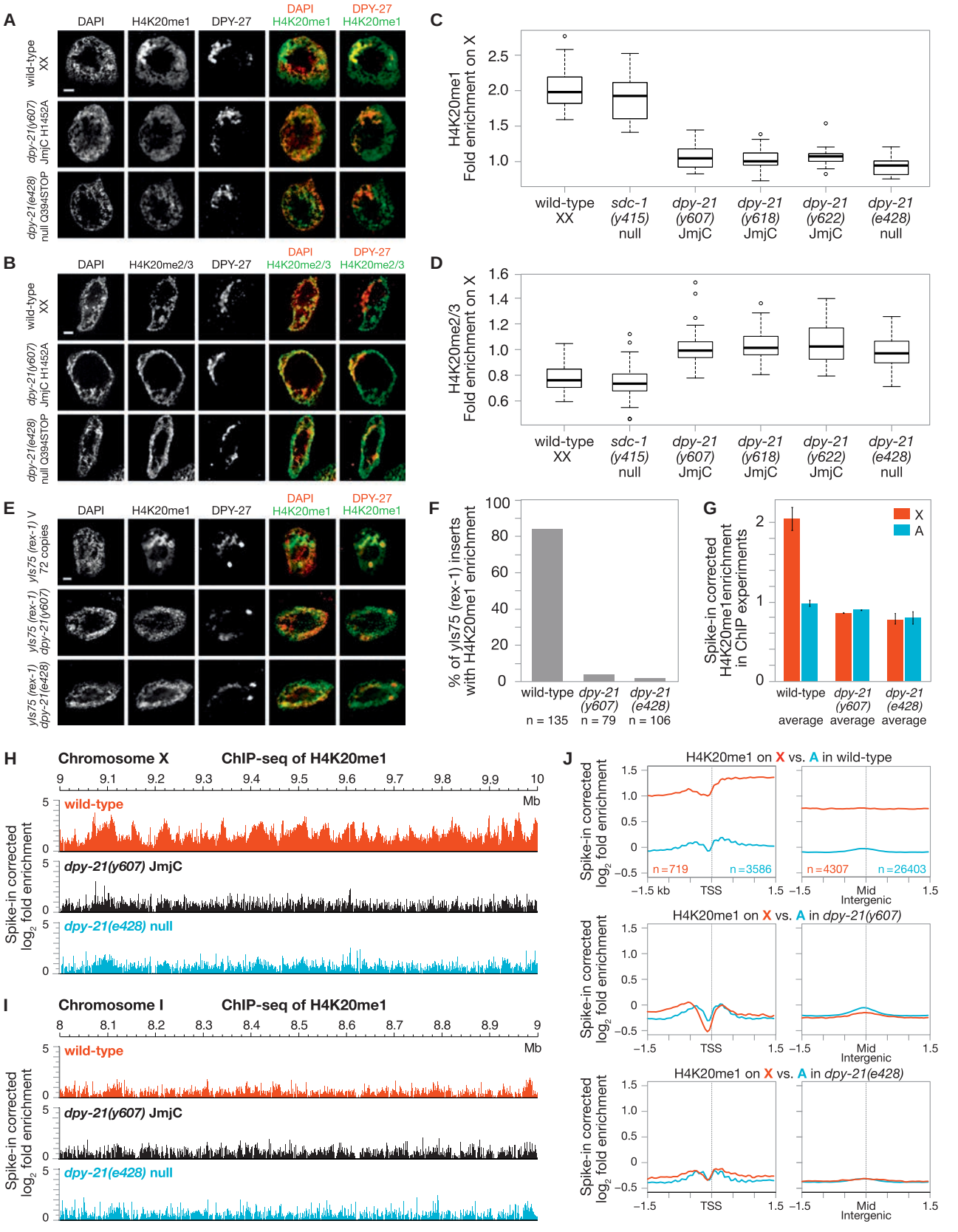
DPY-21 and mROSBIN Are the First Identified JmjC H4K20me2 Demethylases

We performed in vitro histone demethylase assays to determine whether DPY-21 is a bona fide histone demethylase and to define its substrate specificity. We incubated a set of purified DPY-21 fragments with bulk histones and assayed H3 and H4

(D) In vitro demethylase assay performed on bulk histones. DPY-$21^{1210-1641}$ and mROSBIN$^{350-795}$ specifically demethylate H4K20me2 only in the presence α-KG, $FeSO_4$, and ascorbate. Mutation of facial triad residues (H1452A D1454A) in DPY-$21^{1210-1641}$ abolished enzymatic activity.

(E) Mass spectrometry analysis of mROSBIN$^{350-795}$ activity showed demethylation of an H4K20me2 peptide. Upper and lower panels show spectra in the absence and presence of α-KG, $FeSO_4$, and ascorbate, respectively.

See Figures S1 and S2 and Table S1.



(legend on next page)

methylation marks by immunoblotting. DPY-21$^{1210-1641}$, which contains the core crystallized JmjC domain and the remaining 24 C-terminal residues, exhibited robust demethylase activity with histone H4 methyl substrates (Figure 1D), but not with histone H3 substrates (Figure S2A). DPY-21$^{1210-1641}$ demethylated H4K20me2 to H4K20me1 only in the presence of all cofactors (Figure 1D). Changes in facial triad residues H1452 and D1454 prevented demethylation of H4K20me2, confirming their critical role in demethylase activity. Together, these data establish DPY-21 as the first JmjC domain-containing demethylase with specificity for H4K20me2.

To assess the evolutionary conservation of this JmjC demethylase activity, we assayed a mouse ROSBIN fragment (mROSBIN$^{350-795}$) that includes the JmjC domain (Figure S1B). Similar to DPY-21$^{1210-1641}$, mROSBIN$^{350-795}$ displayed robust H4K20me2 demethylase activity only in the presence of cofactors (Figure 1D), demonstrating wide conservation of JmjC demethylase activity. mROSBIN$^{350-795}$ was also inactive against histone H3 methyl substrates (Figure S2A), suggesting specificity for H4.

The in vitro demethylase assay clearly demonstrated a reciprocal decrease in the level of H4K20me2 with an increase in H4K20me1, but the results for H4K20me3 were not compelling (Figure S2B). To further dissect substrate specificity, we assayed methylated H4K20 histone peptides. Mass spectrometry showed that mROSBIN$^{350-795}$ demethylated the H4K20me2 peptide (Figure 1E), but not the H4K20me3 peptide (data not shown), whereas DPY-21$^{1210-1641}$ was not active against either peptide (data not shown), plausibly because of differences in the C termini (Figure S1B). Collectively, our *in vitro* results demonstrate that DPY-21$^{1210-1641}$ and mROSBIN$^{350-795}$ are histone lysine demethylases, with a strong preference for H4K20me2 as the substrate.

DPY-21 Is the Founding Member of a New JmjC Subfamily, KDM9

Histone lysines are demethylated by two classes of demethylases that use different chemistry: flavin-dependent lysine-specific demethylases that belong to the KDM1 subfamily, and Fe^{2+} and α-KG dependent demethylases that belong to the JmjC family (Markolovic et al., 2016). JmjC demethylases are divided into subfamilies KDM2-KDM7 based on domain architecture and specificity toward histone methylation marks.

Very few demethylases of either class are known to act on methylated H4K20 substrates. For H4K20me2, only the neuron-specific demethylase LSD1n had been found prior to our work. LSD1n, an isoform of LSD1, is a flavin-dependent monoamine oxidase that demethylates both H4K20me2 and H4K20me1 (Wang et al., 2015). The only known JmjC H4K20 demethylases (PHF2 and PHF8) have specificity for H4K20me3 and H4K20me1, respectively (Liu et al., 2010; Qi et al., 2010; Stender et al., 2012) and belong to the KDM7 subfamily. Both PHF2 and PHF8 also have an N-terminal PHD domain required for H4K20 demethylase activity in vitro, which occurs only in the context of nucleosomes.

In contrast, DPY-21 and mROSBIN are unique compared to PHF2 and PHF8 not only because of their H4K20me2 substrate specificity but also their architecture. DPY-21 and mROSBIN lack PHD and other chromatin-interacting domains. They only require the JmjC domain for substrate recognition, allowing them to demethylate H4K20me2 in vitro on histones and histone peptides. Based on our structural and biochemical results, we propose a new subfamily of JmjC lysine demethyases, the KDM9 subfamily, founded by *Caenorhabditis* DPY-21 and the evolutionarily related ROSBIN proteins. These enzymes are the first JmjC demethylases shown to demethylate H4K20me2.

The DPY-21 JmjC Demethylase Acts In Vivo to Enrich H4K20me1 on X Chromosomes of Somatic Cells

To assess JmjC demethylase functions in vivo, we edited codons for catalytic residues in endogenous *dpy-21* loci to generate mutant proteins with the following substitutions in the facial triad: H1452A [allele, *dpy-21(y607 JmjC)*], D1454A [allele, *dpy-21(y618 JmjC)*], or H1452A D1454A [allele, *dpy-21(y622 JmjC)*].

In wild-type XX hermaphrodites (Figures 2A and 2C), but not XO males (Figure S3B), H4K20me1 is enriched on X chromosomes relative to autosomes (see also Vielle et al., 2012; Wells et al., 2012). Reciprocally, H4K20me2/me3 are depleted from X (Figures 2B and 2D) (Vielle et al., 2012). In all three engineered *dpy-21(JmjC)* mutants, sex-specific enrichment of H4K20me1

Figure 2. DPY-21 Demethylase Enriches H4K20me1 on X In Vivo

(A and B) Confocal images of gut nuclei from wild-type and *dpy-21* mutant adults show H4K20me1 (A) and H4K20me2/me3 (B) levels on X versus autosomes, with DPY-27 marking X.

(C and D) Boxplots summarize H4K20me1 (C) or H4K20me2/me3 (D) levels on X relative to their average levels on all chromosomes of wild-type and mutant nuclei. Boxes, middle 50% of ratios. Center bars, medians. X chromosomes of *dpy-21(null)* and *dpy-21(JmjC)* mutants have reduced H4K20me1 ($p < 10^{-14}$) and enhanced H4K20me2/me3 ($p < 10^{-9}$) relative to wild-type. H4K20me1 enrichment ($p = 0.1$) and H4K20me2/me3 depletion ($p = 0.2$) are not different in *sdc-1(null)* versus wild-type. p values, 2-sided Wilcoxon rank-sum test.

(E) DPY-21 tethering to ectopic *rex-1* DCC binding sites inserted on chromosome V caused JmjC-dependent H4K20me1 enrichment in gut nuclei.

(F) Quantification of DPY-27-marked *rex* insertions with HK20me1 enrichment from confocal images of strains in (E).

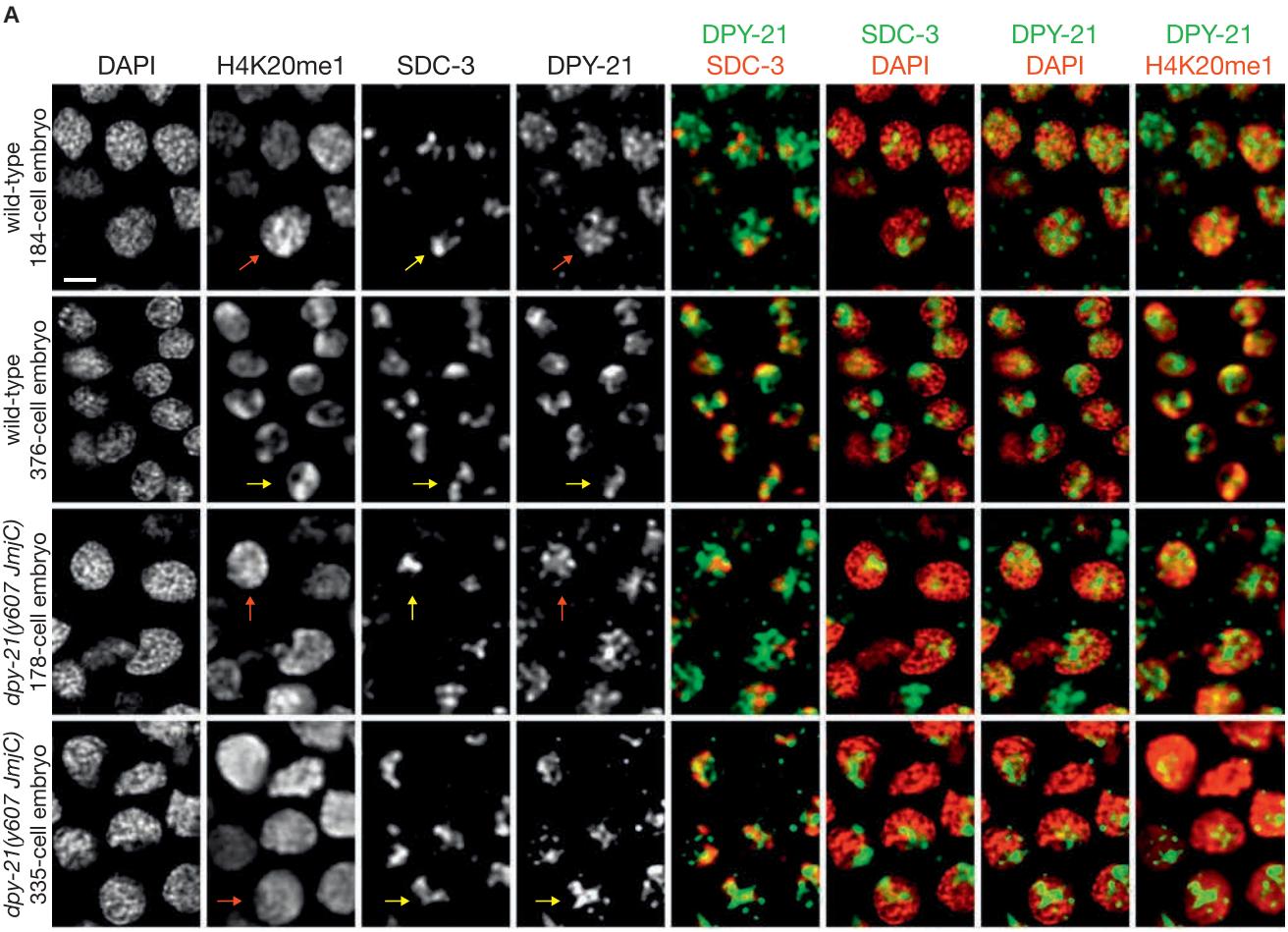
(G) Spike-in-corrected H4K20me1 enrichment on X (red) versus autosomes (blue) in ChIP-seq assays reveals a significant decrease in H4K20me1 on X in *dpy-21(JmjC)* or *dpy-21(null)* versus wild-type embryos. Average enrichment of two biological replicates is shown for each genotype with error bars indicating SD.

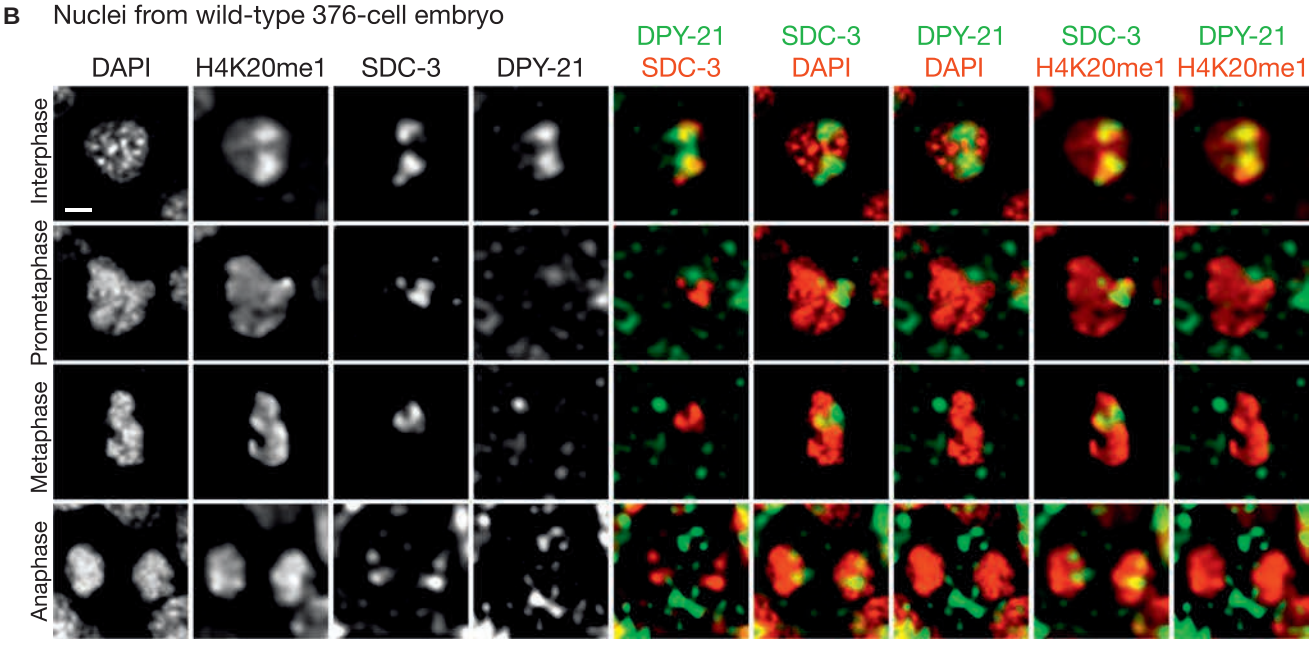
(H and I) ChIP-seq profiles show spike-in-corrected H4K20me1 enrichment in representative regions of chromosome (chr.) X (H) and chr. I (I) in wild-type and *dpy-21(JmjC)* or *dpy-21(null)* mutant embryos.

(J) Average spike-in-corrected H4K20me1 enrichment plotted across transcription start sites (TSS) defined by GRO-cap and GRO-seq (Kruesi et al., 2013) or across the middle of intergenic regions on X (red) and autosomes (blue). Similar decreases in H4K20me1 were found on all regions of X in *dpy-21* mutants versus wild-type. n, number of X or A regions in each set of related plots.

Scale bars in (A), (B), and (E), 2 μm.

See Figure S3.





(legend on next page)

was abolished from X chromosomes of XX adults (Figures 2A, 2C, and S3A) and embryos (Figures 3A and S3D), and H4K20me2/me3 levels were restored (Figures 2B, 2D, and S3C). Thus, H4K20me1 enrichment on X occurs by demethylation of H4K20me2 to H4K20me1.

To examine the genome-wide distribution of H4K20me1 at high resolution in wild-type and *dpy-21* mutant embryos, we performed chromatin immunoprecipitation sequencing (ChIP-seq) analysis using *C. tropicalis* embryos as a spike-in control to normalize read enrichment (Figures 2G–2J and S3G). Consistent with immunofluorescence experiments, H4K20me1 was enriched on X compared to autosomes in wild-type embryos. Enrichment was observed in both genic and intergenic regions (Figure 2J). Moreover, the H4K20me1 level on X was reduced by 2-fold in both *dpy-21(y607 JmjC)* and *dpy-21(e428 null)* mutants, making it comparable to that in autosomes (Figures 2G–2J and S3G). In contrast, autosomal levels of H4K20me1 were not significantly changed in *dpy-21(JmjC)* or *dpy-21(null)* mutants (Figures 2G–2J and S3G). Thus, DPY-21 functions in vivo as a chromosome-wide histone demethylase to enrich H4K20me1 on X.

Our findings predict that tethering of DPY-21 to an ectopic site on an autosome via its association with the DCC should cause JmjC-dependent H4K20me1 enrichment at that site. Indeed, inserting 72 copies of DCC binding site *rex-1* in a site on chromosome V was sufficient to induce site-specific enrichment of H4K20me1 that was abolished in *dpy-21(JmjC)* and *dpy-21(null)* mutants (Figures 2E and 2F). This result confirms that the DCC targets DPY-21 to chromosomes to perform demethylation.

DPY-21 Binding to X Chromosomes Coincides with Enrichment of H4K20me1 on X

H4K20me1 enrichment on X in XX embryos occurs long after initial recruitment of DCC subunits to X, which starts around the 30- to 40-cell stage of embryogenesis with the synthesis of SDC-2, the hermaphrodite-specific trigger of DCC binding to X. H4K20me1 enrichment was not evident before the 200-cell stage and became reliably detectable in most cells around the 300- to 350-cell stage (Figure 3A) (Vielle et al., 2012; Wells et al., 2012).

To assess the timing of DPY-21 binding to X relative to DCC binding and H4K20me1 enrichment, we performed immunofluorescence experiments using two different DPY-21-specific reagents. We used an antibody against the N-terminal region of DPY-21 to monitor DPY-21 in a wild-type strain and an antibody against 3× FLAG to monitor DPY-21 in a strain with *3× FLAG* inserted at the 5′ end of the endogenous *dpy-21* gene. The 3× FLAG tag had no adverse effects on DPY-21's DC functions (STAR Methods). In both cases, DPY-21 was not detected on X before the 200-cell stage and was only reliably found on X in most cells around the 300- to 350-cell stage, the same timing as X enrichment of H4K20me1, and long after binding of other DCC subunits (Figures 3A and S4A). The timing of DPY-21 binding is consistent with its demethylase activity causing H4K20me1 enrichment and our prior finding that DPY-21's association with the DCC is not as stable as that of other subunits (Yonker and Meyer, 2003).

Catalytically inactive DPY-21 proteins exhibited the same timing in X binding as active DPY-21 (Figures 3A, S4B, and S4C). Furthermore, western blots revealed no reduction in DPY-21 levels in *dpy-21(y607 JmjC)* versus wild-type embryos (Figures S3E and S3F). Hence, loss of H4K20me1 enrichment on X in JmjC mutants was not due to defective DPY-21 localization or protein degradation. Finally, mRNA levels of all other DCC subunits were not changed in JmjC mutants (GEO files), and X binding was not visibly affected (Figures 2A, 3A, S3A, S3D, S4B, and S4C), indicating that loss of H4K20me1 enrichment did not result from defective functioning of other DCC subunits.

Cell-Cycle-Dependent Activities of DPY-21

Unlike other DCC subunits, DPY-21's association with X is cell-cycle dependent. DPY-21 colocalizes with X during interphase but appears diffusely distributed in mitosis (Figures 3B and S4A). In contrast, other DCC subunits remain on X throughout the cell cycle. Coincident with this dynamic DPY-21 behavior, H4K20me1 is not enriched on X during mitosis (Figure 3B). Instead, all mitotic chromosomes have uniformly high levels of H4K20me1 (Figure 4A). In *dpy-21(JmjC)* mutants, the inactive protein exhibits the same dynamic behavior (Figures S4B and S4C), and H4K20me1 remains enriched on all mitotic chromosomes (Figure S4C), showing that DPY-21 does not control H4K20 methylation during mitosis. Thus, the selective enrichment of H4K20me1 on X is re-established by DPY-21 during interphase of each cell cycle. The factors that regulate this dynamic behavior are not known, but DPY-21 is phosphorylated at T565, S776, and T1329 (Zielinska et al., 2009), suggesting that phosphorylation might contribute to its regulation.

Mutations that Inactivate JmjC Activity Disrupt Dosage Compensation

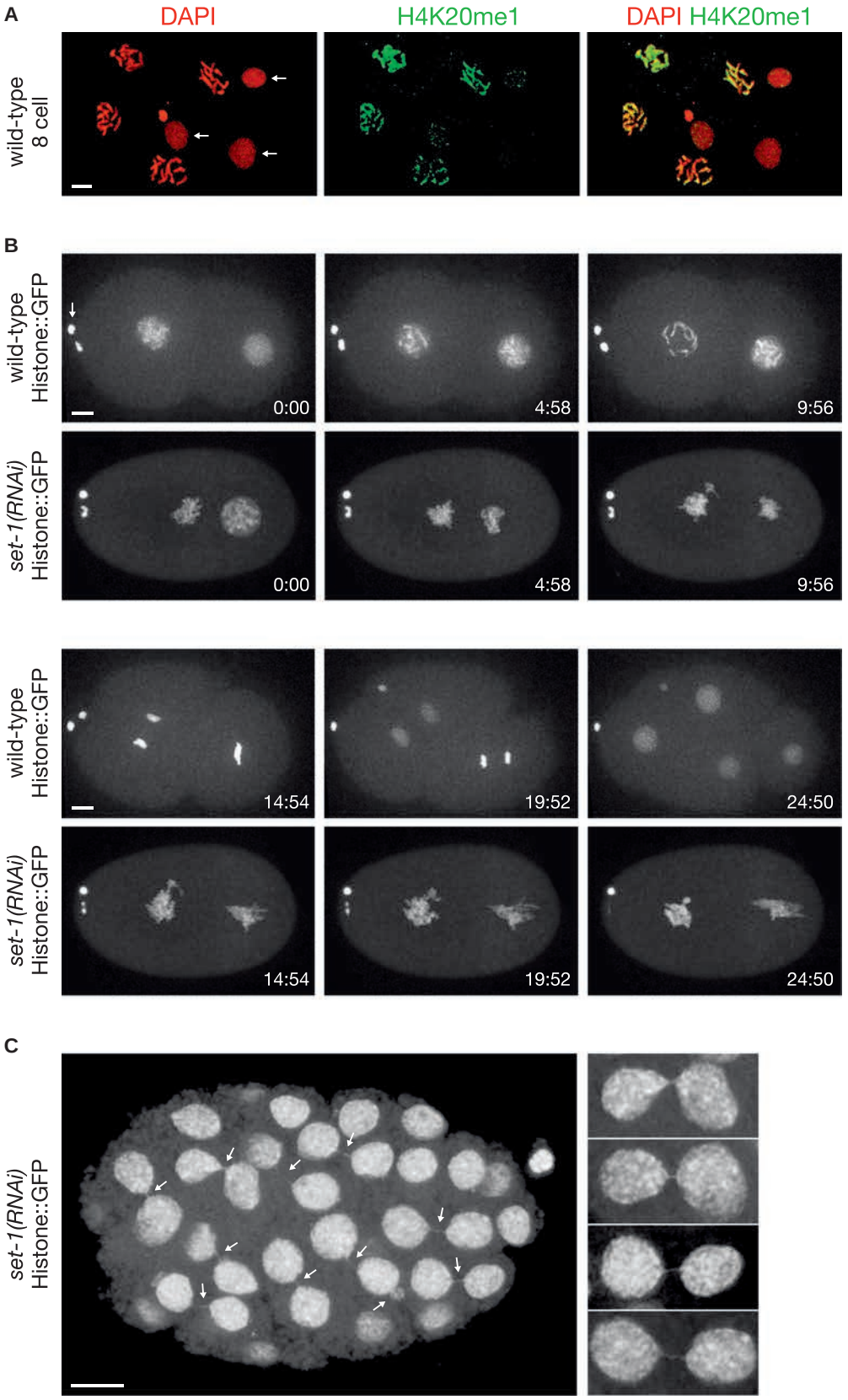
dpy-21(JmjC) mutations enabled us to selectively eliminate H4K20me1 enrichment on X and thereby assess whether H4K20me1 plays a direct role in X chromosome gene regulation or instead occurs as a secondary consequence of the DC process. Prior studies reduced H4K20me1 levels globally throughout the cell cycle via RNAi to the *set-1* methyltransferase gene or blocked its progression to H4K20me2/me3 via mutation

Figure 3. Recruitment of DPY-21 to X Chromosomes Is Cell-Cycle Dependent and Coincides with H4K20me1 Enrichment

(A) Nuclei from wild-type and *dpy-21(y607 JmjC)* XX embryos of different ages stained with DAPI and H4K20me1, SDC-3, and DPY-21 antibodies. DPY-21 localizes to X later in development than other DCC subunits, coincident with H4K20me1 enrichment on X. DPY-21 localized to X with similar timing in *dpy-21(JmjC)* and wild-type embryos. Yellow arrows, foci of SDC-3, DPY-21, or H4K20me1 on X. Red arrows, diffuse nuclear localization of H4K20me1 or DPY-21. Scale bar, 2 μm.

(B) Individual nuclei of a wild-type 376-cell embryo reveal cell-cycle-dependent association of DPY-21 on X during interphase, but not mitosis, unlike other DCC subunits. Scale bar, 1 μm.

See Figures S3 and S4.



(legend on next page)

of the *set-4* methyltransferase gene. Both conditions caused genome-wide changes in gene expression (Kramer et al., 2015; Vielle et al., 2012; Wells et al., 2012), thus complicating interpretations about H4K20me1's role.

In studies asking whether *set-1(RNAi)* enhances the XX-specific lethality caused by weak DC mutations, we (Table S2) and others (Vielle et al., 2012; Wells et al., 2012) found XX lethality to be enhanced. However, we also found that enhanced lethality is neither sex-specific nor specific to the disruption of DC (Table S2). The sex-independent effect is likely due to both genome-wide changes in mRNA levels (Figure 5C) (Kramer et al., 2015) and H4K20me1's role in mitosis. Supporting the latter, we found that strong depletion of SET-1 caused severe mitotic chromosome defects, resulting in arrested embryonic development (Figure 4B; Table S2), and weaker depletion caused chromosome bridges and milder chromosome segregation defects, permitting longer embryonic development (Figure 4C). These results compromise any conclusion that synergistic lethality is an indicator of H4K20me1's role in DC.

To assess the role of H4K20me1 in DC, we first asked whether *dpy-21(JmjC)* mutations suppress the XO-specific lethality caused by *xol-1* mutations, which inappropriately activate the DCC in males, causing reduced X expression and death (Figure 5A) (Miller et al., 1988; Rhind et al., 1995). If JmjC activity contributes to reducing X expression during DC, JmjC mutations should suppress the *xol-1* XO-specific lethality. Indeed, all three homozygous JmjC mutations suppressed the lethality (40%–56%), showing that JmjC activity, and hence H4K20me1 enrichment, is important for DC (Table S3A).

Because suppression by *dpy-21(JmjC)* mutations was less than that by a *dpy-21(null)* mutation (76%, $p < 0.01$, chi-square test) (Table S3A), we compared the viability of *dpy-21(JmjC/null)*; *xol-1* XO animals with *dpy-21(null/null); xol-1* XO animals under conditions that enabled us to detect all rescued XO animals regardles of their sex (Figure 5B; Table S4). The goal was to determine the relative contribution of the JmjC domain to DPY-21's overall functions in DC. We found that 62%–72% of JmjC/null XO animals were viable compared to 93% of null/null animals. These results confirmed the importance of DPY-21 demethylase activity in DC and suggested that DPY-21 has additional roles in DC. Supporting this idea, *dpy-21(null)* mutant hermaphrodites are less viable and have more developmental abnormalities than do *dpy-21(JmjC)* mutants (Table S3B).

The role of the DPY-21 demethylase in enriching H4K20me1 on X, and thereby reducing X expression, predicts that blocking H4K20me2/me3 production via a *set-4* mutation should prevent the rescue of *xol-1* XO males by *dpy-21(JmjC)* mutations. This prediction was met (Table S5), providing strong genetic evidence that production of H4K20me2/me3 on X is an intermediate step in the enrichment of H4K20me1 on X.

H4K20me1 Regulates X Chromosome Gene Expression

To assay more directly the effect of *dpy-21(JmjC)* mutations on X chromosome gene expression, we performed RNA sequencing (RNA-seq) on *dpy-21(JmjC)*, *dpy-21(null)*, and wild-type XX embryos and compared changes in gene expression on X and autosomes (Figure 5C). For all three *dpy-21(JmjC)* mutants, X chromosome gene expression was significantly elevated relative to that of all autosomes (Figure 5C). The relative elevation in X expression was greater in *dpy-21(null)* mutants, consistent with their more severe mutant phenotypes and greater suppression of *xol-1* XO lethality (Figure 5C). Thus, the DPY-21 demethylase, and hence X enrichment of H4K20me1, helps repress X-linked gene expression in vivo.

Malignant brain tumor repeat (MBT) proteins associate with H4K20me1 nucleosomes and repress gene expression (Blanchard et al., 2014; Trojer et al., 2007). Hence, we tested whether the MBT repeat proteins of *C. elegans* (Harrison et al., 2007) contribute to DC. Simultaneous knockout of the two MBT repeat genes *mbtr-1* and *lin-61* failed to suppress *xol-1* XO lethality (Table S6), making it unlikely that MBT repeat proteins play a significant role in X chromosome repression.

H4K20me1 Modulates X Chromosome Conformation in Somatic Cells

Cytological studies showed that DCC binding increases X chromosome compaction in interphase nuclei of somatic cells (Lau et al., 2014). Genome-wide chromosome conformation capture (Hi-C) studies showed the DCC remodels X chromosome structure by inducing topologically associating domains (TADs) using its highest-affinity binding sites (*rex* sites) to mediate long-range chromatin interactions (Crane et al., 2015). We asked whether the DPY-21 demethylase contributes to X compaction and to the DCC-driven remodeling of X chromosome topology.

To assess compaction, we determined the fraction of total chromosome volume occupied by X in gut nuclei of wild-type and *dpy-21(JmjC)* mutant XX adults (Figure 6A). Loss of demethylase activity caused a significant increase ($\sim$30%, $p < 10^{-6}$, 2-sided Wilcoxon rank-sum test) in X volume in all three *dpy-21(JmjC)* mutants (Figure 6A). An even greater increase in X volume ($\sim$55%, $p < 10^{-9}$) was found in *dpy-21(null)* mutants, suggesting additional roles for DPY-21 in X organization. In contrast, X volume was not significantly changed in *sdc-1(null)* mutants ($p = 0.2$), which have normal H4K20me1 enrichment on X, but

Figure 4. Depletion of Histone Methyltransferase SET-1 Causes Severe Mitotic Chromosome Segregation Defects

(A) Confocal images of a wild-type 8-cell embryo stained with DAPI and H4K20me1 antibody. H4K20me1 is significantly elevated in mitotic versus interphase (arrows) nuclei and is distributed across all chromosomes.

(B and C) Live images of *set-1(RNAi)* embryos show severe mitotic defects. (B) Defective mitotic chromosome segregation caused embryonic lethality. The time series of chromosome segregation starts with 2-cell wild-type or *set-1(RNAi)* embryos expressing histone::GFP. Both wild-type cells completed a round of division in 25 min. The *set-1(RNAi)* embryo arrested in mitosis and died. Arrow, two polar bodies. (C) Live image of an older *set-1(RNAi)* embryo that received less *set-1* RNA than in (B) shows mitotic segregation defects in nearly all cells, including chromatin bridges and micronuclei (arrows). Enlargements of chromatin bridges are shown.

Scale bars in (A)–(C), 5 μm.

See Table S2.

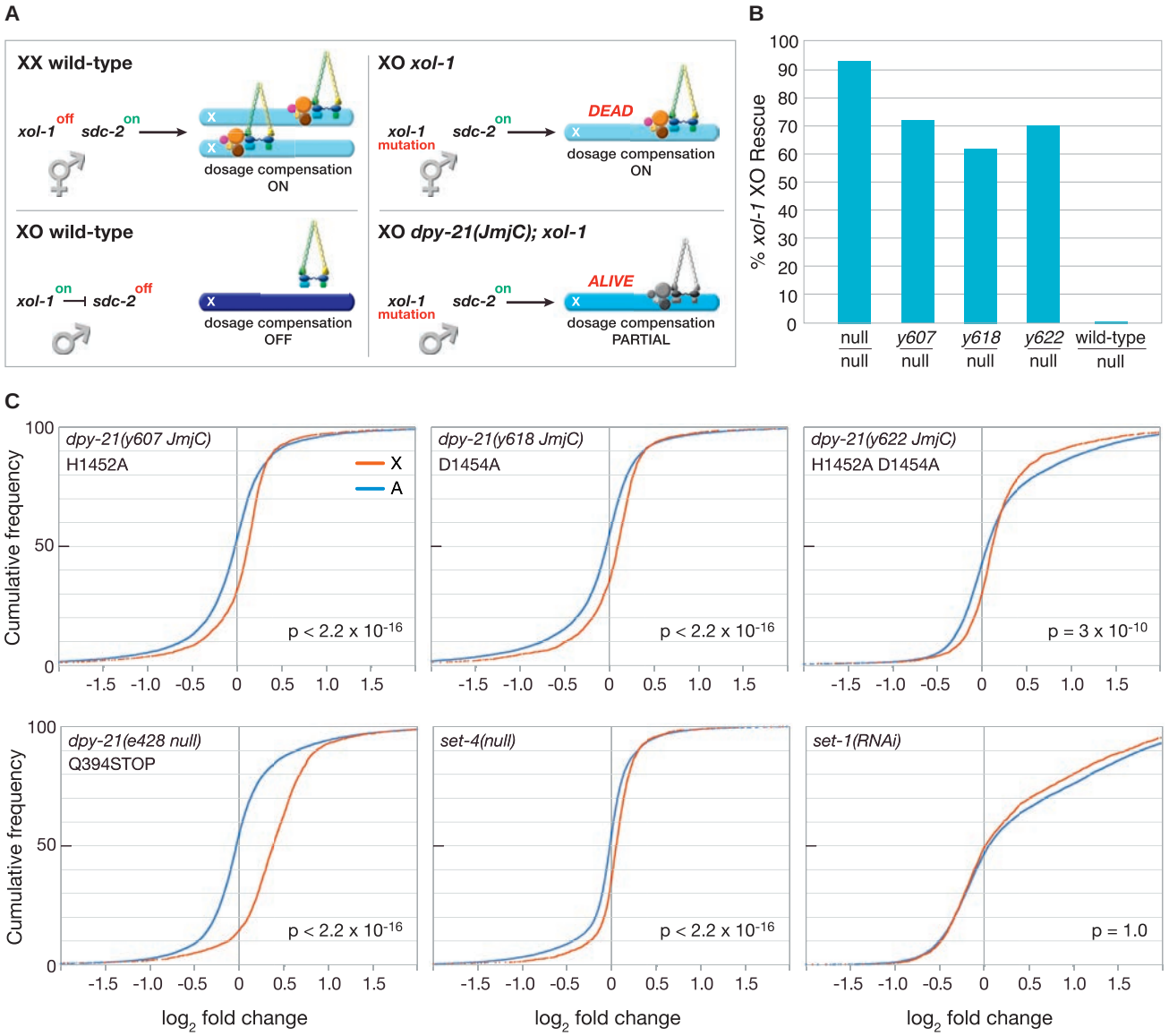


Figure 5. The DPY-21 Demethylase Regulates X Chromosome Gene Expression

(A) Cartoon of DCC regulation in different genotypes. In XX embryos, *xol-1* is repressed, allowing SDC-2 to trigger DCC assembly on X and reduce gene expression by half (light blue). In XO animals, *xol-1* is active and represses *sdc-2*, thereby blocking DCC binding and promoting full X expression (dark blue). In *xol-1* XO mutants, SDC-2 triggers DCC binding to X, causing reduced X expression (light blue) and death. Most *dpy-21(JmjC); xol-1* XO animals are viable because *dpy-21(JmjC)* mutations reduce activity of the X-bound DCC (gray) and partially restore X expression (medium blue) (Table S3A).

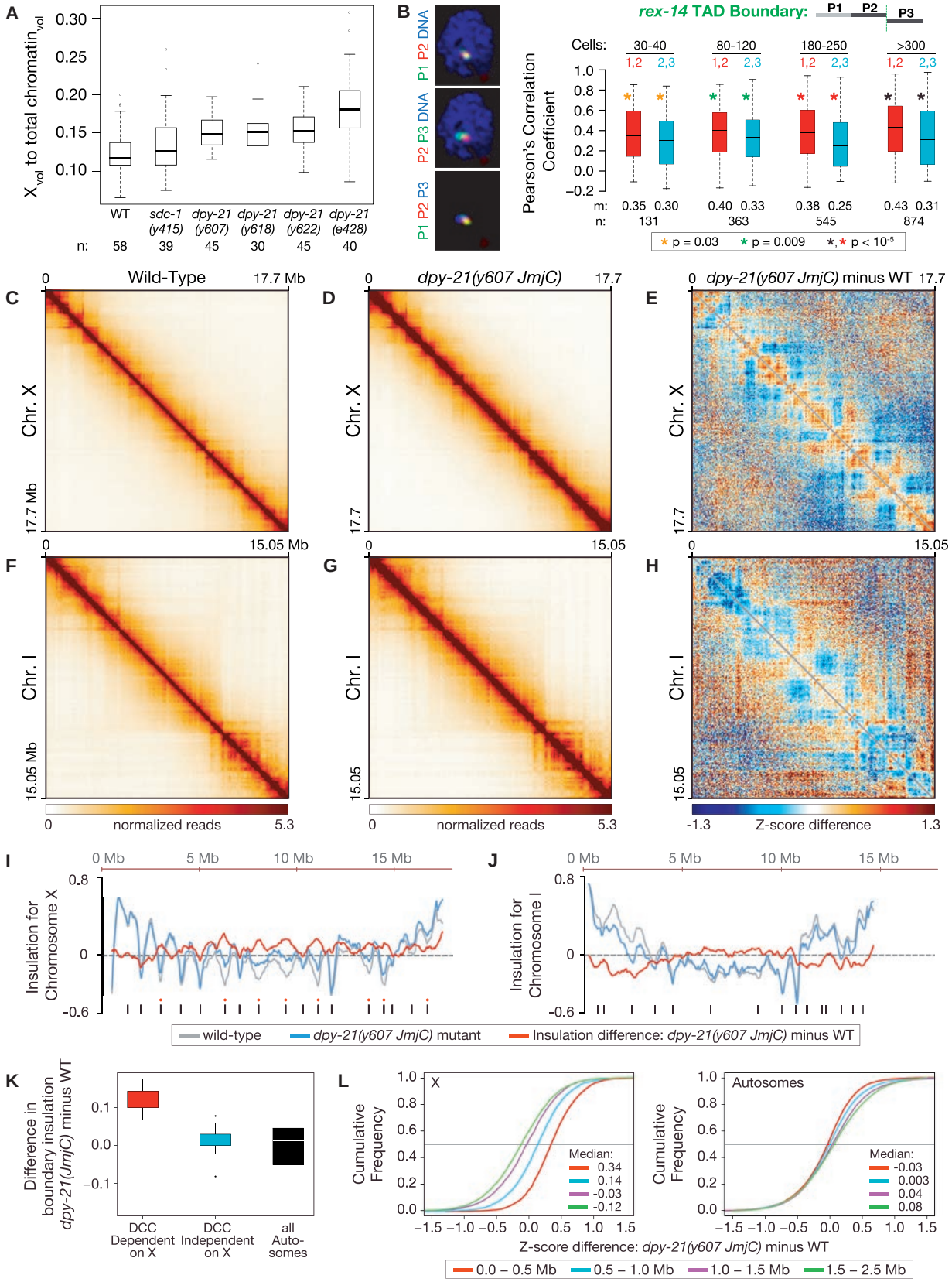
(B) Rescue of *xol-1* XO lethality by *dpy-21(JmjC)/dpy-21(null)* heterozygous mutations shows that *JmjC* mutations disrupt DC less severely than a *dpy-21(null)* mutation (Table S4).

(C) Cumulative plots show the distribution of expression changes for genes on X (red) and autosomes (blue) in mutant versus wild-type embryos as assayed by RNA-seq. x axis represents the $\log_2$ fold change in expression. X chromosomes have elevated expression compared to autosomes in all mutants, except for *set-1(RNAi).* p values, 1-sided Wilcoxon rank-sum tests.

See Tables S3, S4, S5, and S6.

modest DC defects like *dpy-21(JmjC)* mutants. Thus, H4K20me1 plays a direct role in X chromosome compaction, and loss of compaction is not a byproduct of DC defects.

To assess the impact of H4K20me1 on the remodeling of X topology, we first examined the relative timing of TAD boundary formation and DPY-21 binding. In embryos of different ages, we imaged pairs of fluorescent probes that tile adjacent 500 kb regions within a TAD or flanking a DCC-dependent TAD boundary (Crane et al., 2015) (Figure 6B). In 80-120-cell embryos, less overlap occurred between probes flanking the *rex-14* TAD boundary than between probes within the TAD, indicating boundary formation had begun. The boundary was well formed



(legend on next page)

in 180- to 250-cell embryos (Figure 6B), before DPY-21 was reliably detected on X (Figure 3). Thus, demethylase activity is not essential to initiate formation of DCC-dependent TAD boundaries.

To determine the effect of H4K20me1 enrichment on X chromosome topology at higher resolution, we performed in situ Hi-C (Figures 6C–6H and S5A–S5C; Table S7) (Crane et al., 2015; Rao et al., 2014). X chromosomes were organized into TADs in both wild-type and JmjC mutant embryos (Figures 6C and 6D), but TADs were less distinct in JmjC mutants (Figures 6D and 6E).

To quantify changes in TAD boundaries, we calulated insulation profiles for X and autosomes (Figures 6I and 6J). An insulation score reflects the aggregate of interactions across each interval, and local minima in insulation profiles denote TAD boundaries. On X chromosomes of both wild-type and JmjC mutant embryos, TAD boundaries formed at the same 17 locations, consistent with TAD formation starting prior to DPY-21 binding. However, the difference in X insulation profiles showed an increase in insulation, and thus a reduction in boundary strength, at all eight DCC-dependent TAD boundaries (Figures 6I and 6K) in JmjC mutants. In contrast, the nine DCC-independent TAD boundaries on X and the boundaries on autosomes were not significantly weakened. These fluorescence in situ hybridization (FISH) and Hi-C results suggest a two-tier model for TAD formation during DC: the DCC initiates TAD formation at specific locations prior to DPY-21 recruitment. DPY-21 then binds to X and enriches H4K20me1, which strengthens TAD boundaries.

To further define the role of H4K20me1 in X chromosome remodeling, we compared genomic-distance-normalized *Z* scores between wild-type and JmjC mutant embryos (Figures 6E, 6H, and S5A). We found distance-dependent changes in interactions across the entire X (Figures 6E and 6L, left). In JmjC mutants, interactions between sites within 1 Mb increased significantly ($p < 10^{-15}$, 2-sided KS test), while interactions between sites more than 1 Mb apart decreased significantly ($p < 10^{-15}$). Notably, long-range interactions among the DCC-dependent TAD boundaries and the 25 highest-affinity *rex* sites key for TAD boundary formation were significantly reduced (Figure S5D), consistent with reduced TAD boundary strength. In contrast, interactions on autosomes exhibited modest changes in the opposite direction (Figures 6H and 6L, right, and S5A), consistent with X chromosomes having a different pattern of distance-dependent chromatin interaction decay than autosomes in JmjC mutants (Figure S5E). These findings show that H4K20me1 promotes long-range interactions across X that enhance the integrity of TADs.

We verified the role of demethylase activity in X chromosome conformation and compaction using FISH to examine the spatial separation between pairs of genomic loci in wild-type and JmjC mutant embryos. Based on Hi-C data (Crane et al., 2015), we chose several pairs of *rex* and non-*rex* sites that were separated by more than 1 Mb and had a comparable decrease in interaction frequency following DCC depletion. FISH utilized 30- to 40-kb probes on embryos with more than 300 cells.

In *dpy-21(JmjC)* mutants, we found a significant increase in the spatial separation of all three pairs of non-*rex* sites (Figure S6A), consistent with the decrease in Hi-C interaction frequencies and in chromosome compaction. Equivalently spaced sites on chromosome I showed no change in spatial separation (Figure S6C). However, the spatial separation of *rex* sites was not changed by either the *dpy-21(JmjC)* or *dpy-21(null)* mutation, but was increased by the *sdc-2(null)* mutation (Figure S6B). The discrepancy between FISH and Hi-C at these *rex* sites likely reflects a difference in sensitivity of the two methods for detecting changes in chromosome structure (Giorgetti and Heard, 2016).

Our combined results show that H4K20me1 enrichment contributes significantly to X chromosome remodeling during DC. At least two different mechanisms contribute to X chromosome structure: a demethylase-independent mechanism that initiates TAD formation and a demethylase-dependent mechanism that compacts X chromosomes and strengthens TADs.

The DPY-21 Demethylase Selectively Enriches H4K20me1 and Compacts Autosomes in Germ Cells via a DCC-Independent Mechanism

H4K20me1 exhibits a dynamic pattern of chromatin enrichment in germ cells (Vielle et al., 2012) (Figures 7A and S7A–S7C),

Figure 6. DPY-21 Demethylase Activity Modulates Spatial Organization of X Chromosomes

(A) Boxplots showing ratios of X volume relative to the total volume of chromatin in gut nuclei of different genotypes reveal that demethylase activity promotes X compaction. n, number of nuclei assayed.

(B) TAD boundary formation occurs prior to DPY-21 binding to X. Boxplots show the distribution of Pearson's correlation coefficients between FISH probes within (red) or across (blue) the *rex-14* TAD boundary in embryos of different cell number. Boxes, middle 50% of coefficients. Center bars, median (m) coefficients. n, total number of nuclei. Asterisks of the same color specify datasets compared using 1-sided Wilcoxon rank-sum test.

(C, D, F, and G) Heatmaps of Hi-C data binned at 50-kb resolution show chromatin interaction frequencies on chr. X or chr. I in wild-type and *dpy-21(JmjC)* mutant embryos.

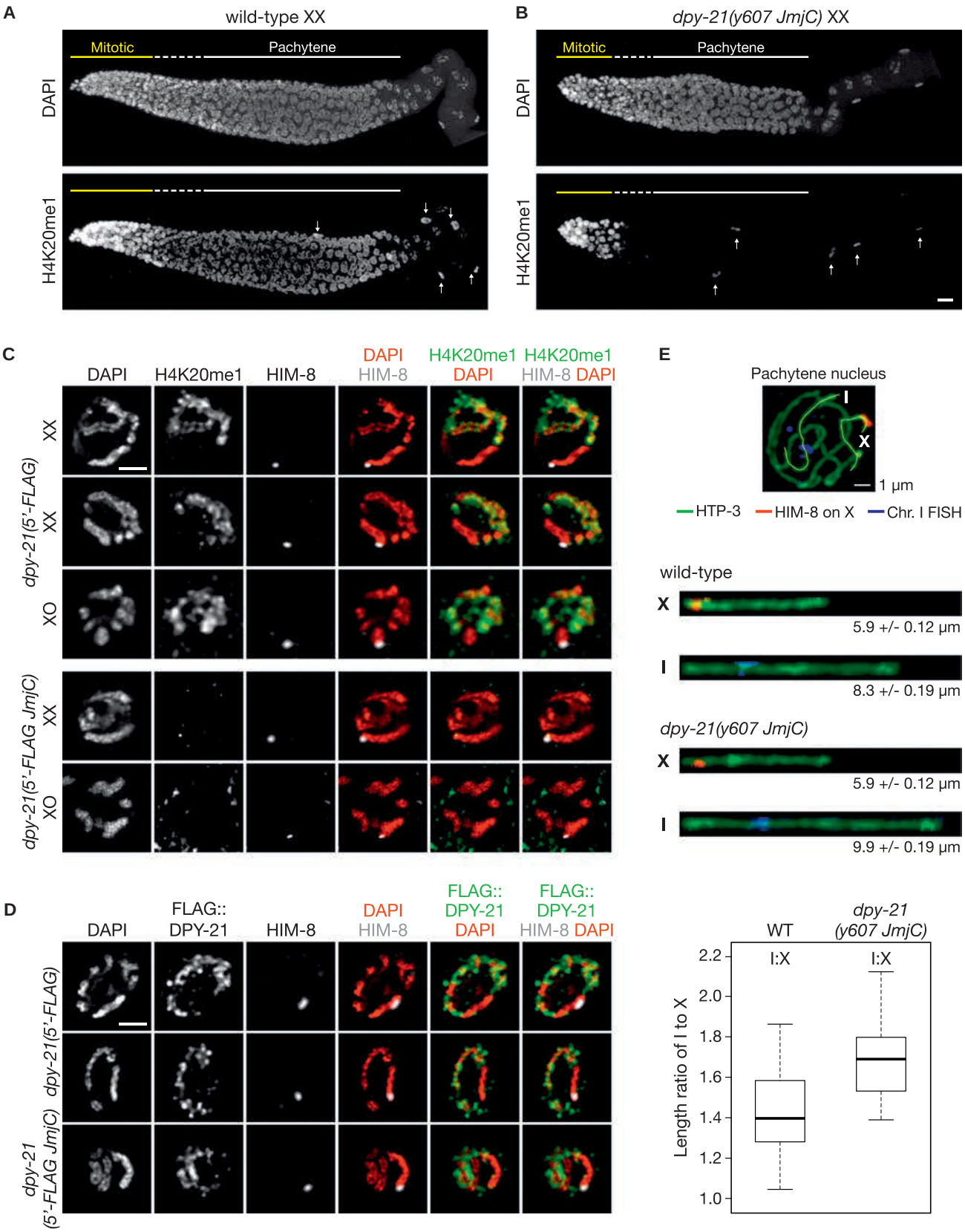
(E and H) *Z* score difference heatmaps at 50-kb resolution show chromatin interactions that increase (orange-red) and decrease (blue) on X (E) or I (H) in *dpy-21(JmjC)* mutant versus wild-type embryos.

(I and J) Insulation plots for chr. X (I) or chr. I (J) of wild-type (gray) or *dpy-21(JmjC)* mutant (blue) embryos and insulation difference plots (red). Black bars, location of TAD boundaries in wild-type embryos. Red dots, DCC-dependent boundaries greatly diminished or eliminated following DCC depletion (Crane et al., 2015).

(K) Boxplots compare insulation changes for different classes of TAD boundaries. In *dpy-21(JmjC)* mutants, DCC-dependent TAD boundaries on X (red) have greater insulation changes than do DCC-independent boundaries on X (blue) ($p < 10^{-3}$) or autosomal boundaries (black) ($p < 10^{-4}$). p values, Wilcoxon ranked-sum test.

(L) Cumulative distributions of *Z* score differences between *dpy-21(JmjC)* mutant and wild-type embryos for Hi-C interactions at specified distances on X (left) or all autosomes (right) showed an increase in short-range interactions on X (<1 Mb apart) and a decrease in long-range interactions on X (>1 Mb) in JmjC mutants. Medians of *Z* score differences for each distance range are shown.

See Figures S5 and S6 and Table S7.



(legend on next page)

prompting the question of whether DPY-21 also regulates H4K20me1 in the gonad, where the DCC does not function. H4K20me1 levels are high on all mitotic chromosomes and become reduced as nuclei enter meiosis (Figure 7A). By early pachytene, H4K20me1 is enriched only on autosomes and is absent from X chromosomes of both XX and XO germ cells (Figures 7C, S7B, and S7C). Autosomal enrichment persists through mid-late pachytene.

To determine the function of DPY-21 JmjC activity in germ cells, we examined H4K20me1 localization in *dpy-21(JmjC)* XX and XO mutant gonads. H4K20me1 levels remained high on mitotic chromosomes of mutant gonads (Figures 7B and S7A), but H4K20me1 was absent from all chromosomes of pachytene nuclei (Figures 7B and 7C). Thus, demethylase activity is essential for the autosomal enrichment of H4K20me1 during germ cell meiosis but is not required for H4K20me1 enrichment during germ cell mitosis.

These results lead to three predictions about DPY-21 localization in the germline: DPY-21 should localize to all autosomes of pachytene nuclei but not to X chromosomes of pachytene nuclei or to any mitotic chromosomes. All three predictions were met, indicating that DPY-21 is a direct regulator of H4K20me1 during germ cell meiosis (Figures 7D and S7A–S7C). Because the DCC does not assemble or function in germlines of XX or XO animals, DPY-21 must operate via a DCC-independent mechanism. DPY-21's selective localization to autosomes during meiosis must occur through different partners that either recruit it to autosomes or repel it from X.

The association of DPY-21 with autosomes in germ cells raised the question of whether H4K20me1 enrichment plays a general role in regulating chromosome compaction across cell types (Figure 7E). A change in meiotic chromosome compaction should alter the length of the chromosome axis. We measured and compared axis lengths of chromosomes X and I in pachytene nuclei of wild-type and *dpy-21(JmjC)* mutant gonads by tracing the axes in three dimensions and computationally straightening them. Chromosome X axis length was not different in *dpy-21(JmjC)* versus wild-type nuclei, as predicted by the lack of DPY-21 on X, but chromosome I axis length was extended in the mutant. The ratio of I:X lengths increased significantly from 1.4 in wild-type nuclei to 1.7 in *dpy-21(JmjC)* nuclei ($p = 1.7 \times 10^{-8}$). Thus, higher-order structure of different chromosomes can be regulated in somatic cells versus germ cells by selective localization of DPY-21, and hence H4K20me1 enrichment. DPY-21 is an adaptable chromatin regulator that controls distinct biological functions in development.

DISCUSSION

The function of chromatin modification in establishing higher-order chromosome structure during gene regulation has been elusive. Here, we have determined the mechanism underlying the dynamic control of histone modification H4K20me1 throughout the cell cycle and demonstrated the H4K20me1-driven remodeling of X chromosome architecture and repression of X chromosome gene expression.

Our structural and biochemical analysis of DCC subunit DPY-21 revealed a novel, widely conserved subfamily of JmjC demethylases (KDM9) that catalyzes the demethylation of H4K20me2 to H4K20me1 in worms and mammals. DPY-21 is the first known H4K20me2 demethylase with pan-cellular activity. DPY-21 lacks a recognizable DNA binding domain to confer target specificity, but it is recruited to diverse genomic locations by different partners to catalyze H4K20me2 demethylation. In somatic cells, the DCC, a condensin complex, targets the demethylase to X chromosomes. DCC binding and spreading distributes demethylase activity, enhancing H4K20me1 chromosome-wide.

Thus, the DCC acts as an "eraser" of a prominent histone post-translational modification, promoting H4K20me1 enrichment on X by activating an H4K20me2 demethylase. DPY-21's role in chromatin modification further illustrates how dosage compensation evolved by co-opting conserved machinery used in other biological processes for the new task of fine-tuning X chromosome expression. Moreover, our results suggest that other condensin and SMC complexes could act as scaffolds to recruit proteins with chromosome-modifying activities.

Chromatin modifications can recruit specialized proteins to regulate the structure of nucleosome arrays and gene expression. In flies and mammals, malignant brain tumor repeat (MBT) proteins associate with nucleosomes enriched in H4K20me1 or H4K20me2, compact the chromatin fiber, and help repress transcription (Blanchard et al., 2014; Trojer et al., 2007). Knockout of worm MBT proteins failed to disrupt DC, implying that if H4K20me1 binding proteins modulate

Figure 7. DPY-21 Demethylase Acts Independently of the DCC in Germ Cells to Enrich H4K20me1 and Compact Autosomes

(A and B) Maximum intensity projections of wild-type (A) and *dpy-21(y607 JmjC)* mutant (B) XX gonads stained with DAPI and H4K20me1 antibodies. Demethylase inactivation does not reduce H4K20me1 levels in mitotic nuclei but prevents H4K20me1 accumulation in meiotic nuclei. Arrows, somatic sheath cells. Scale bar, 10 μm.

(C) Pachytene nuclei from *dpy-21(5′-FLAG)* and *dpy-21(5′-FLAG JmjC)* XX and XO gonads stained with antibodies to H4K20me1 and X-specific marker HIM-8. H4K20me1 is selectively enriched on autosomes in pachytene nuclei of both sexes in the *dpy-21(5′-FLAG)* strain but absent from X (top). H4K20me1 is absent from all pachytene chromosomes of *dpy-21(JmjC)* mutant XX and XO gonads (bottom).

(D) Pachytene nuclei from *dpy-21(5′-FLAG)* and *dpy-21(5′-FLAG JmjC)* XX gonads stained with FLAG and HIM-8 antibodies. DPY-21 localizes to autosomes but not X, consistent with its demethylation of only autosomes.

(E) DPY-21 demethylase is required for full compaction of autosomes in germ cells. (Top) Pachytene nucleus from a wild-type gonad stained with a FISH probe to chr. I (blue) and antibodies to HIM-8 (red) and axis protein HTP-3 (green). The 3D traces of X and I (yellow) from wild-type and *dpy-21(JmjC)* gonads were used to straighten the chromosomes. (Middle) Computationally straightened X and I with average axis lengths and SEMs from 34 wild-type and 44 *dpy-21(JmjC)* mutant nuclei. (Bottom) Boxplots show that the ratio of I:X axis lengths is significantly lower in wild-type versus *dpy-21(JmjC)* pachytene nuclei ($p = 1.7 \times 10^{-8}$, 2-sided Wilcoxon rank-sum test). Scale bars in (C) and (D), 2 μm.

See Figure S7.

X chromosome conformation and gene repression, they will define a new class of H4K20me1-specific histone "readers." Alternatively, H4K20me1 may control nucleosome or chromatin folding directly or antagonize other chromatin modifying activities.

While histone modifications can alter inter-nucleosomal interactions that govern the compaction state of a chromatin fiber (Francis et al., 2004; Kalashnikova et al., 2013; Lu et al., 2008), the potential impact of histone modifications on higher-order chromosome organization beyond chromatin-fiber compaction is not well understood. Only recently has super-resolution imaging shown that chromatin domains enriched in H3K27me3 and polycomb proteins adopt unique folded states (Boettiger et al., 2016). Once the polycomb repressive complex 1 is recruited to H3K27me3-enriched chromatin, it generates chromatin domains 20–140 kb in size that are distinct from previously defined TADs (Kundu et al., 2017). Our results show that H4K20me1 controls chromosome structure over a larger length scale. H4K20me1 modulates the formation of TADs ($\geq$1 Mb) by enhancing long-range DNA interactions across X.

Our combined data support a two-tier model for the DCC-dependent remodeling of X chromosome topology. In the first stage, the condensin DCC initiates TAD boundary formation via a demethylase-independent mechanism that promotes long-range interactions between the highest affinity DCC binding sites (*rex* sites) prior to DPY-21's assembly onto X. In the second stage, DPY-21 demethylase catalyzes enrichment of H4K20me1 on X, thereby enhancing long-range DNA interactions across X. X chromosome compaction is generally increased, and TAD boundaries are strengthened by the elevation of long-range *rex* interactions critical for TAD formation.

Unexpectedly, higher-order chromosome structure is also regulated in germ cells by DPY-21. In meiotic nuclei of both sexes, DPY-21 is localized to autosomes via a DCC-independent mechanism. The targeted H4K20me1 enrichment then condenses the lengths of autosomal axes, providing broad evidence for H4K20me1's regulation of higher-order chromosome structure. Axis expansion can alter crossover frequency and distribution (Mets and Meyer, 2009), raising the possibility that autosomal H4K20me1 might have consequences for crossover recombination as well as gene expression.

DPY-21's DCC-independent activity is relevant for recent findings that DPY-21 modulates nematode growth and metabolism through the TORC2 pathway (Webster et al., 2013) and entry into the quiescent dauer state through the insulin-signaling pathway (Delaney et al., 2017; Dumas et al., 2013). Evidence in both cases suggests at least partial DCC-independence for DPY-21's functions. Determining the effect of JmjC-specific mutations on both pathways will be key for distinguishing the contribution of demethylase activity from general DCC functions. Moreover, because metabolism is altered in both biological contexts and α-KG is a necessary DPY-21 co-factor and a product of the Krebs cycle, the worm's metabolic state, rather than the DCC, could influence DPY-21's functions (Chin et al., 2014).

Relevance of our studies for mammalian development is underscored by the enrichment of H4K20me1 on the inactive female X chromosome (Kohlmaier et al., 2004). Mechanisms underlying the H4K20me1 enrichment and its role in chromosome silencing are not known. The long non-coding RNA *XIST*, the trigger of mammalian X inactivation, induces accumulation of H4K20me1 on X, but X inactivation per se is not required for H4K20me1 deposition, suggesting that H4K20me1 might contribute to establishing inactivation. Consistent with this idea, knockout of the H4K20me1 methyltransferase causes decondensation of X (Oda et al., 2009). Our discoveries offer new directions for unraveling the regulation and function of H4K20me1 in X chromosome inactivation and other long-range mechanisms of gene control.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - *C. elegans* and *C. tropicalis* Strains
  - *E. coli* Strains
- METHOD DETAILS
  - Naming of the DPY-21-ROSBIN Subfamily of JmjC Demethylases
  - Structure Prediction
  - Expression and Purification of DPY-21$^{1210–1617}$
  - Crystallization and Structure Determination
  - Immunoblotting-Based Demethylase Assay
  - Dot Blot Analysis of Antibody Specificity Against H4K20me3
  - Mass Spectrometry-Based Demethylase Assay
  - Genome Editing
  - Quantification of Progeny from Matings on Feeding RNAi
  - Generating Males for Mating RNAi Experiments
  - *sdc-2* RNAi
  - *set-1* dsRNA for RNAi Injection
  - Immunofluorescence
  - FISH
  - Microscopy and Image Analysis
  - Live Imaging of Embryos
  - Growth of *set-1*(RNAi) and Control RNAi Nematodes for RNA-seq
  - RNA-seq Library Preparation
  - Gene Expression Analysis
  - Western Blot Analysis
  - ChIP-seq and Spike-in Analysis
  - In situ Hi-C
  - Hi-C Data Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.07.041.

AUTHOR CONTRIBUTIONS

K.B., Q.B., S.U., E.C.A., P.J.K., B.S.W., D.S.K., and C.G.P. conducted experiments and analyzed data. K.B., Q.B., P.J.K., and E.C.A. edited the manuscript. B.J.M. guided the study, analyzed data, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank D. Fujimori, D. Minor, T. Cline, E. Ralston, and M. Marletta for general discussions; J. Dekker for Hi-C discussions; D. Stalford for figures; T. Cline, J. Berger, B. Farboud, and N. Fuda for manuscript comments; and J. Doudna for sharing equipment and expertise. We used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by an NIH Instrumentation Grant (S10 OD018174). X-ray data were collected at the Lawrence Berkeley National Laboratory Advanced Light Source (Beamline 8.3.1). Huygens software was licensed to the CNR Biological Imaging Facility at UC Berkeley. Research was supported, in part, by Miller Institute funds and NIGMS grants (R01 GM030702 to B.J.M.; F32 GM100647 to B.S.W.). B.J.M. is an investigator of the Howard Hughes Medical Institute.

Received: March 22, 2017
Revised: June 25, 2017
Accepted: July 25, 2017
Published: August 31, 2017

SUPPORTING CITATIONS

The following reference appears in the Supplemental Information: Tan et al. (2007).

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167.

Beck, D.B., Oda, H., Shen, S.S., and Reinberg, D. (2012). PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. Genes Dev. *26*, 325–337.

Blanchard, D.P., Georlette, D., Antoszewski, L., and Botchan, M.R. (2014). Chromatin reader L(3)mbt requires the Myb-MuvB/DREAM transcriptional regulatory complex for chromosomal recruitment. Proc. Natl. Acad. Sci. USA *111*, E4234–E4243.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Chen, H., Hughes, D.D., Chan, T.A., Sedat, J.W., and Agard, D.A. (1996). IVE (Image Visualization Environment): a software platform for all three-dimensional microscopy applications. J. Struct. Biol. *116*, 56–60.

Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T.G., Simpson, M., Mao, Q., Pan, C.H., Dai, S., et al. (2006). Structural insights into histone demethylation by JMJD2 family members. Cell *125*, 691–702.

Chin, R.M., Fu, X., Pai, M.Y., Vergnes, L., Hwang, H., Deng, G., Diep, S., Lomenick, B., Meli, V.S., Monsalve, G.C., et al. (2014). The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. Nature *510*, 397–401.

Chuang, P.T., Albertson, D.G., and Meyer, B.J. (1994). DPY-27:a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. Cell *79*, 459–474.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240–244.

da Rocha, S.T., and Heard, E. (2017). Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. Nat. Struct. Mol. Biol. *24*, 197–204.

Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. Cell *164*, 1110–1121.

Delaney, C.E., Chen, A.T., Graniel, J.V., Dumas, K.J., and Hu, P.J. (2017). A histone H4 lysine 20 methyltransferase couples environmental cues to sensory neuron control of developmental plasticity. Development *144*, 1273–1282.

Dumas, K.J., Delaney, C.E., Flibotte, S., Moerman, D.G., Csankovszki, G., and Hu, P.J. (2013). Unexpected role for dosage compensation in the control of dauer arrest, insulin-like signaling, and FoxO transcription factor activity in *Caenorhabditis elegans*. Genetics *194*, 619–629.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. *67*, 282–292.

Farboud, B., and Meyer, B.J. (2015). Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. Genetics *199*, 959–971.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. Genome Biol. *17*, 215.

Harrison, M.M., Lu, X., and Horvitz, H.R. (2007). LIN-61, one of two *Caenorhabditis elegans* malignant-brain-tumor-repeat-containing proteins, acts with the DRM and NuRD-like protein complexes in vulval development but not in certain other biological processes. Genetics *176*, 255–271.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Jørgensen, S., Schotta, G., and Sørensen, C.S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. Nucleic Acids Res. *41*, 2797–2806.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. J. R. Soc. Interface *10*, 20121022.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature *421*, 231–237.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. *32*, W526–W531.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. *2*, E171.

Koscielny, G., Yaikhom, G., Iyer, V., Meehan, T.F., Morgan, H., Atienza-Herrero, J., Blake, A., Chen, C.K., Easty, R., Di Fenza, A., et al. (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic Acids Res. *42*, D802–D809.

Kramer, M., Kranz, A.L., Su, A., Winterkorn, L.H., Albritton, S.E., and Ercan, S. (2015). Developmental dynamics of X-chromosome dosage compensation by the DCC and H4K20me1 in *C. elegans*. PLoS Genet. *11*, e1005698.

Kruesi, W.S., Core, L.J., Waters, C.T., Lis, J.T., and Meyer, B.J. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. eLife *2*, e00808.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. Mol. Cell *65*, 432–446.e5.

Kuroda, M.I., Hilfiker, A., and Lucchesi, J.C. (2016). Dosage Compensation in *Drosophila*-a Model for the Coordinate Regulation of Transcription. Genetics *204*, 435–450.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. *51*, 2778–2786.

Lau, A.C., Nabeshima, K., and Csankovszki, G. (2014). The *C. elegans* dosage compensation complex mediates interphase X chromosome compaction. Epigenetics Chromatin *7*, 31.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, W., Tanasa, B., Tyurina, O.V., Zhou, T.Y., Gassmann, R., Liu, W.T., Ohgi, K.A., Benner, C., Garcia-Bassets, I., Aggarwal, A.K., et al. (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature *466*, 508–512.

Liu, T., Rechtsteiner, A., Egelhofer, T.A., Vielle, A., Latorre, I., Cheung, M.S., Ercan, S., Ikegami, K., Jensen, M., Kolasinska-Zwierz, P., et al. (2011). Broad chromosomal domains of histone modification patterns in *C. elegans*. Genome Res. *21*, 227–236.

Lo, T.W., Pickle, C.S., Lin, S., Ralston, E.J., Gurling, M., Schartner, C.M., Bian, Q., Doudna, J.A., and Meyer, B.J. (2013). Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics *195*, 331–348.

Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat. Struct. Mol. Biol. *15*, 1122–1124.

Markolovic, S., Leissing, T.M., Chowdhury, R., Wilkins, S.E., Lu, X., and Schofield, C.J. (2016). Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. Curr. Opin. Struct. Biol. *41*, 62–72.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Mets, D.G., and Meyer, B.J. (2009). Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. Cell *139*, 73–86.

Meyer, B.J. (2010). Targeting X chromosomes for repression. Curr. Opin. Genet. Dev. *20*, 179–189.

Miller, L.M., Plenefisch, J.D., Casson, L.P., and Meyer, B.J. (1988). *xol-1*: a gene that controls the male modes of both sex determination and X chromosome dosage compensation in *C. elegans*. Cell *55*, 167–183.

Oda, H., Okamoto, I., Murphy, N., Chu, J., Price, S.M., Shen, M.M., Torres-Padilla, M.E., Heard, E., and Reinberg, D. (2009). Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. Mol. Cell. Biol. *29*, 2278–2295.

Paix, A., Folkmann, A., Rasoloson, D., and Seydoux, G. (2015). High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. Genetics *201*, 47–54.

Pesavento, J.J., Yang, H., Kelleher, N.L., and Mizzen, C.A. (2008). Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol. Cell. Biol. *28*, 468–486.

Plasschaert, R.N., and Bartolomei, M.S. (2014). Genomic imprinting in development, growth, behavior and stem cells. Development *141*, 1805–1813.

Qi, H.H., Sarkissian, M., Hu, G.Q., Wang, Z., Bhattacharjee, A., Gordon, D.B., Gonzales, M., Lan, F., Ongusaha, P.P., Huarte, M., et al. (2010). Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. Nature *466*, 503–507.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rhind, N.R., Miller, L.M., Kopczynski, J.B., and Meyer, B.J. (1995). *xol-1* acts as an early switch in the *C. elegans* male/hermaphrodite decision. Cell *80*, 71–82.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater than the sum of parts: complexity of the dynamic epigenome. Mol. Cell *62*, 681–694.

Stender, J.D., Pascual, G., Liu, W., Kaikkonen, M.U., Do, K., Spann, N.J., Boutros, M., Perrimon, N., Rosenfeld, M.G., and Glass, C.K. (2012). Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. Mol. Cell *48*, 28–38.

Strome, S., Kelly, W.G., Ercan, S., and Lieb, J.D. (2014). Regulation of the X chromosomes in *Caenorhabditis elegans*. Cold Spring Harb. Perspect. Biol. *6*, a018366–a018366.

Takahashi, T., Tanaka, H., Iguchi, N., Kitamura, K., Chen, Y., Maekawa, M., Nishimura, H., Ohta, H., Miyagawa, Y., Matsumiya, K., et al. (2004). Rosbin: a novel homeobox-like protein gene expressed exclusively in round spermatids. Biol. Reprod. *70*, 1485–1492.

Tan, F.J., Fire, A.Z., and Hill, R.B. (2007). Regulation of apoptosis by *C. elegans* CED-9 in the absence of the C-terminal transmembrane domain. Cell Death Differ. *14*, 1925–1935.

Terwilliger, T.C. (1999). Reciprocal-space solvent flattening. Acta Crystallogr. D Biol. Crystallogr. *55*, 1863–1871.

Trojer, P., Li, G., Sims, R.J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S.D., et al. (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell *129*, 915–928.

van Nuland, R., and Gozani, O. (2016). Histone H4 lysine 20 (H4K20) methylation, expanding the signaling potential of the proteome one methyl moiety at a time. Mol. Cell. Proteomics *15*, 755–764.

Vielle, A., Lang, J., Dong, Y., Ercan, S., Kotwaliwale, C., Rechtsteiner, A., Appert, A., Chen, Q.B., Dose, A., Egelhofer, T., et al. (2012). H4K20me1 contributes to downregulation of X-linked genes for *C. elegans* dosage compensation. PLoS Genet. *8*, e1002933.

Wang, J., Telese, F., Tan, Y., Li, W., Jin, C., He, X., Basnet, H., Ma, Q., Merkurjev, D., Zhu, X., et al. (2015). LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. Nat. Neurosci. *18*, 1256–1264.

Webster, C.M., Wu, L., Douglas, D., and Soukas, A.A. (2013). A non-canonical role for the *C. elegans* dosage compensation complex in growth and metabolic regulation downstream of TOR complex 2. Development *140*, 3601–3612.

Wells, M.B., Snyder, M.J., Custer, L.M., and Csankovszki, G. (2012). *Caenorhabditis elegans* dosage compensation regulates histone H4 chromatin state on X chromosomes. Mol. Cell. Biol. *32*, 1710–1719.

Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., and Bickmore, W.A. (2012). Anterior-posterior differences in HoxD chromatin topology in limb development. Development *139*, 3157–3167.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

Yonker, S.A., and Meyer, B.J. (2003). Recruitment of *C. elegans* dosage compensation proteins for gene-specific versus chromosome-wide repression. Development *130*, 6519–6532.

Zielinska, D.F., Gnad, F., Jedrusik-Bode, M., Wiśniewski, J.R., and Mann, M. (2009). *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. J. Proteome Res. *8*, 4039–4049.

Article

Cell

Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate

Takeshi Isoda,[1] Amanda J. Moore,[1] Zhaoren He,[1] Vivek Chandra,[1] Masatoshi Aida,[1] Matthew Denholtz,[1] Jan Piet van Hamburg,[1] Kathleen M. Fisch,[2] Aaron N. Chang,[2] Shawn P. Fahl,[3] David L. Wiest,[3] and Cornelis Murre[1,4,*]

[1]Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA

[2]Center for Computational Biology & Bioinformatics, Institute for Genomic Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

[3]Blood Cell Development and Function, Fox Chase Cancer Center, 333 Cottman Avenue, PA, Philadelphia, PA 19111, USA

[4]Lead Contact

*Correspondence: cmurre@ucsd.edu

http://dx.doi.org/10.1016/j.cell.2017.09.001

SUMMARY

It is now established that Bcl11b specifies T cell fate. Here, we show that in developing T cells, the Bcl11b enhancer repositioned from the lamina to the nuclear interior. Our search for factors that relocalized the Bcl11b enhancer identified a non-coding RNA named ThymoD (thymocyte differentiation factor). ThymoD-deficient mice displayed a block at the onset of T cell development and developed lymphoid malignancies. We found that ThymoD transcription promoted demethylation at CTCF bound sites and activated cohesin-dependent looping to reposition the Bcl11b enhancer from the lamina to the nuclear interior and to juxtapose the Bcl11b enhancer and promoter into a single-loop domain. These large-scale changes in nuclear architecture were associated with the deposition of activating epigenetic marks across the loop domain, plausibly facilitating phase separation. These data indicate how, during developmental progression and tumor suppression, non-coding transcription orchestrates chromatin folding and compartmentalization to direct with high precision enhancer-promoter communication.

INTRODUCTION

The differentiation of T cells is orchestrated in the thymus. Upon exposure to Delta-Notch signaling, early T cell progenitors (ETPs) differentiate into multipotent DN2a cells, which in turn develop into committed DN2b cells. DN2b cells subsequently progress into DN3a cells in which TCRβ VDJ rearrangement is initiated. Once a productive TCRβ chain has been assembled, DN3b cells expand and differentiate into $CD4^+CD8^+$ double-positive (DP) thymocytes. In the DP compartment, thymocytes die by either neglect or negative selection or persist through positive selection to differentiate into CD4 single-positive (CD4SP) or CD8SP cells (Klein et al., 2014; Naito et al., 2011).

The developmental progression of T cells is regulated by the combined activities of an ensemble of transcriptional regulators. T-lineage development is initiated by the E-proteins that activate the expression of genes encoding components involved in Notch signaling (Bain and Murre, 1998; Ikawa et al., 2006; Miyazaki et al., 2017). Once instructed to respond to Notch signaling, T cell progenitors activate the expression of Bcl11b, GATA-3, and TCF1 (Yui and Rothenberg, 2014). Specifically, Bcl11b expression is initiated at the DN2a cell stage to promote developmental progression to the DN2b cell stage. At the DN2b cell stage, Bcl11b expression is further elevated and, in concert with E2A, activates a T-lineage-specific program of gene expression and suppresses the expression of genes associated with alternative cell fates (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010a; Longabaugh et al., 2017). The activation of Bcl11b expression in DN2 cells involves Notch signaling, GATA-3, TCF1, and RUNX1 that bind to an enhancer, named Major Peak, located in the Bcl11b intergenic locus control region (Guo et al., 2008; Weber et al., 2011; García-Ojeda et al., 2013; Li et al., 2013). Recent elegant studies indicated that full activation of Bcl11b expression in developing T cell progenitors requires a rate-limiting transition from an inactive to an active chromatin state (Kueh et al., 2016).

Here, we have examined how Bcl11b expression is activated to establish T cell fate and suppress the development of lymphoid malignancies. We found that, in developing T cell progenitors, the Bcl11b locus control region, containing a well-characterized enhancer, repositioned from the lamina to the nuclear interior. The repositioning of the Bcl11b enhancer was orchestrated by a non-coding RNA, named ThymoD (thymocyte differentiation factor). ThymoD transcription promoted demethylation at sites associated with CTCF occupancy across the transcribed region and activated cohesin-dependent looping, plausibly involving loop extrusion, to bring the Bcl11b promoter and enhancer into a single loop domain. These results are consistent with a model in which non-coding transcription dictates enhancer-promoter communication at multiple levels: (1) demethylation of CpG residues across the ThymoD transcribed region to permit CTCF occupancy, (2) recruitment of the cohesin complex to the transcribed region to activate cohesin-dependent looping, (3) loop extrusion to juxtapose with great precision the enhancer and promoter into a

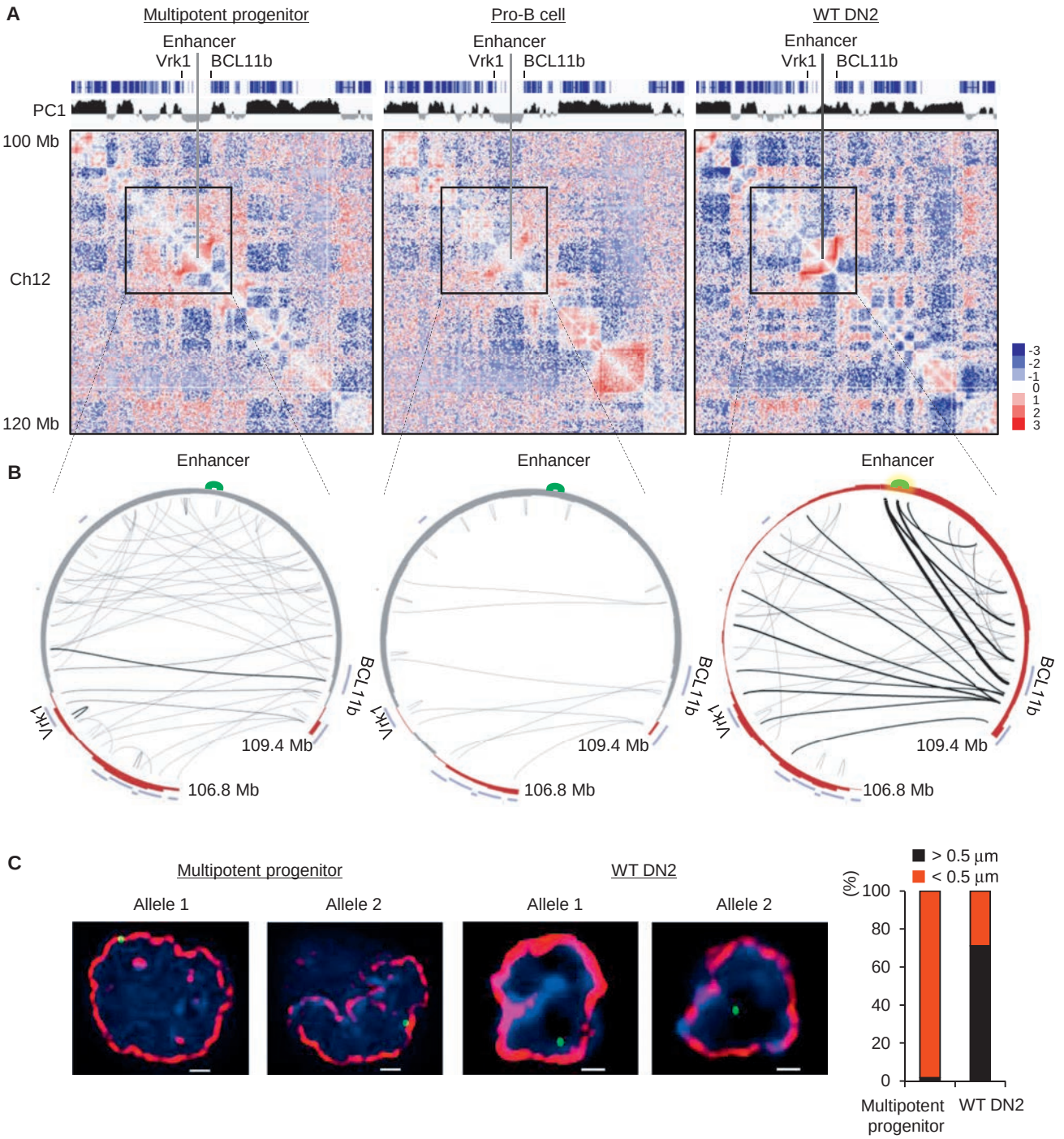


Figure 1. Large-Scale Changes in Nuclear Architecture in Developing T Cell Progenitors

(A) Normalized genome-wide contact matrices for multipotent progenitors, pro-B cells, and DN2 cells of a genomic region located on chromosome 12 are displayed. Top panel shows coding elements (blue) associated with chromosome 12. Positive PC1 values are indicated in black, whereas negative PC1 values are displayed in gray. Heatmap indicates the ratio of observed versus expected frequencies for genomic interactions as revealed by HiC. Numerical and statistical details are described in STAR Methods.

(B) Circos diagrams representing genomic interactions across the Bcl11b intergenic region in multipotent progenitors, pro-B cells, and DN2 cells. Indicated are the Vrk1 and Bcl11b coding regions, as well as the Bcl11b intergenic locus control region. The enhancer is shown in green. Thickness of connecting lines reflects p values associated with the indicated interactions. Positive PC1 values are indicated in red, whereas negative PC1 values are shown in gray. Numerical and statistical details are described in STAR Methods.

(legend continued on next page)

single-loop domain, (4) repositioning the enhancer from a heterochromatic to an euchromatic environment, and (5) permitting the deposition of activating epigenetic marks across the loop domain to facilitate phase separation.

RESULTS

The Bcl11b Locus Control Region Repositions in Developing T Cell Progenitors

In previous studies, we demonstrated that in multipotent progenitors, the Bcl11b intergenic region was associated with the transcriptionally repressive compartment (Lin et al., 2012). To explore the possibility that the Bcl11b locus repositioned in developing thymocytes, adult hematopoietic progenitors were isolated from the bone marrow and cultured on an OP9-DL1 stromal layer in the presence of IL7 and FLT3L (Schmitt and Zúñiga-Pflücker, 2002). Cells were harvested upon reaching the DN2 cell stage and examined for nuclear architecture using HiC (Table S1) (Lieberman-Aiden et al., 2009; Rao et al., 2014). To identify genomic regions that repositioned in developing T cell progenitors, we segregated the DN2 genome into transcriptionally permissive (A) and transcriptionally repressive (B) compartments using principal component analysis. We found that during the transition from the multipotent progenitor to the committed DN2 cell stage, a wide spectrum of genomic regions switched compartments (data not shown). Conspicuous among the genomic regions that repositioned was an intergenic region containing the Bcl11b enhancer (Figure 1A) (Li et al., 2013). Specifically, we found that the Bcl11b intergenic region repositioned from the transcriptionally repressive compartment B to the transcriptionally permissive compartment A (Figure 1A). Notably, the repositioning of the Bcl11b intergenic region was associated with *de novo* genomic interactions involving the Bcl11b enhancer and promoter regions (Figure 1B). To validate these findings, formaldehyde fixed cells were hybridized with fluorescently labeled fosmid probes that span the Bcl11b enhancer. As predicted by the HiC analysis, we found that in a large majority of multipotent progenitor cells, the Bcl11b intergenic region was sequestered at the nuclear envelope, whereas in DN2 cells, it was predominantly localized in the nuclear interior (Figure 1C). These data indicate that in developing thymocytes, the Bcl11b intergenic region harboring the enhancer repositions from the lamina to the nuclear interior to direct the Bcl11b enhancer to the Bcl11b promoter.

Non-coding RNA Transcription Initiated within the Bcl11b Intergenic Region Activates Bcl11b Transcription

Previous studies have suggested that intergenic transcription is associated with the repositioning of genomic regions from heterochromatic to euchromatic genomic regions (Schmitt et al., 2005). As a first approach to identify such transcripts in developing T cell progenitors, we employed assay for transposase-accessible chromatin sequencing (ATAC-seq) and analyzed RNA sequencing (RNA-seq) reads to detect nucleosome depletion and non-coding transcripts across the Bcl11b intergenic region (Table S1) (Buenrostro et al., 2013; Bossen et al., 2015). This analysis revealed a cluster of ATAC-sensitive sites centered on a previously identified enhancer known to regulate Bcl11b expression (Li et al., 2013) (Figure 2A). Four ATAC-sensitive regions associated with this region were identified as CR1, CR2, CR3, and CR4. CR1 corresponds to the Bcl11b enhancer previously named Major Peak (Li et al., 2013). The roles of CR2, CR3, and CR4 were unknown. CR1, CR2, CR3, and CR4 were depleted of nucleosomes in DN2 cells, but not in pro-B cells, indicating lineage specificity (Figure S1A). We next examined this region for developmentally regulated non-coding transcripts whose expression preceded or overlapped with Bcl11b expression (Zhang et al., 2012). Two distinct non-coding RNAs, previously designated as GM16084, initiated from either CR1 (Major Peak) or CR2 (Figure S1B). We will hereafter refer to GM16084 as ThymoD. At the DN1 cell stage, ThymoD transcription was initiated from CR2 immediately prior to the induction of Bcl11b expression (Figure S1B; data not shown). In the DN2 and DN3 cell stages, ThymoD transcription was initiated from both CR1 and CR2 (Figure S1B).

As a first approach to examine potential roles for ThymoD in modulating Bcl11b expression, we inserted a poly(A) site upstream of either CR1 or CR2. As a recipient cell line, we used SCIDadh, representing the DN3 cell stage (Carleton et al., 1999). Upon electroporation of the repair template and gRNAs targeting non-conserved genomic regions either upstream of CR1 or CR2, clones were isolated, expanded, and examined for proper poly(A) site insertion. Clones that harbored the poly(A) insertion on both alleles, flanking either CR1 or CR2, were expanded and examined for ThymoD and Bcl11b expression using real-time PCR. We found that insertion of a poly(A) site upstream of CR1 did not interfere with either ThymoD or Bcl11b transcription (Figure S1C and S1D). In contrast, when inserted upstream of CR2, ThymoD, as well as Bcl11b transcript abundance, significantly declined (Figure S1D). Thus, the insertion of a poly(A) site upstream of CR2, but not CR1, interfered with ThymoD and Bcl11b expression.

Generation of ThymoD p(A)/p(A) Mice

To examine the role of ThymoD in thymocyte development, we generated mice that carried a poly(A) cassette inserted immediately upstream of CR2 (Figure 2A). gRNAs targeting this region in conjunction with the repair template and Cas9 mRNA were injected into C57BL/6 zygotes (Figure S2A) (Yang et al., 2013). Mice that carried at least three poly(A) cassettes in tandem in the ThymoD locus were viable and generated offspring at the expected Mendelian ratios (data not shown). To validate for proper insertion, PCR primers were generated that were located

(C) The Bcl11b intergenic region repositions during the developmental progression from multipotent progenitors to DN2 cells from the nuclear lamina to the nuclear interior. Images represent 3D-FISH analysis using a fosmid probe corresponding to the Bcl11b locus control region. The nuclear envelope was visualized using antibodies directed against the lamina and is indicated in red. The two left images show localization for the Bcl11b enhancer at the nuclear lamina in multipotent progenitors. Right images indicate location of the Bcl11b enhancer in the nuclear interior of DN2 cells. DAPI staining is shown in blue. Bar graph shows fraction of spatial distances (< 500nm) separating the Bcl11b super-enhancer from the nuclear lamina in multipotent progenitor cells (n = 175) and WT DN2 cells (n = 135). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm.

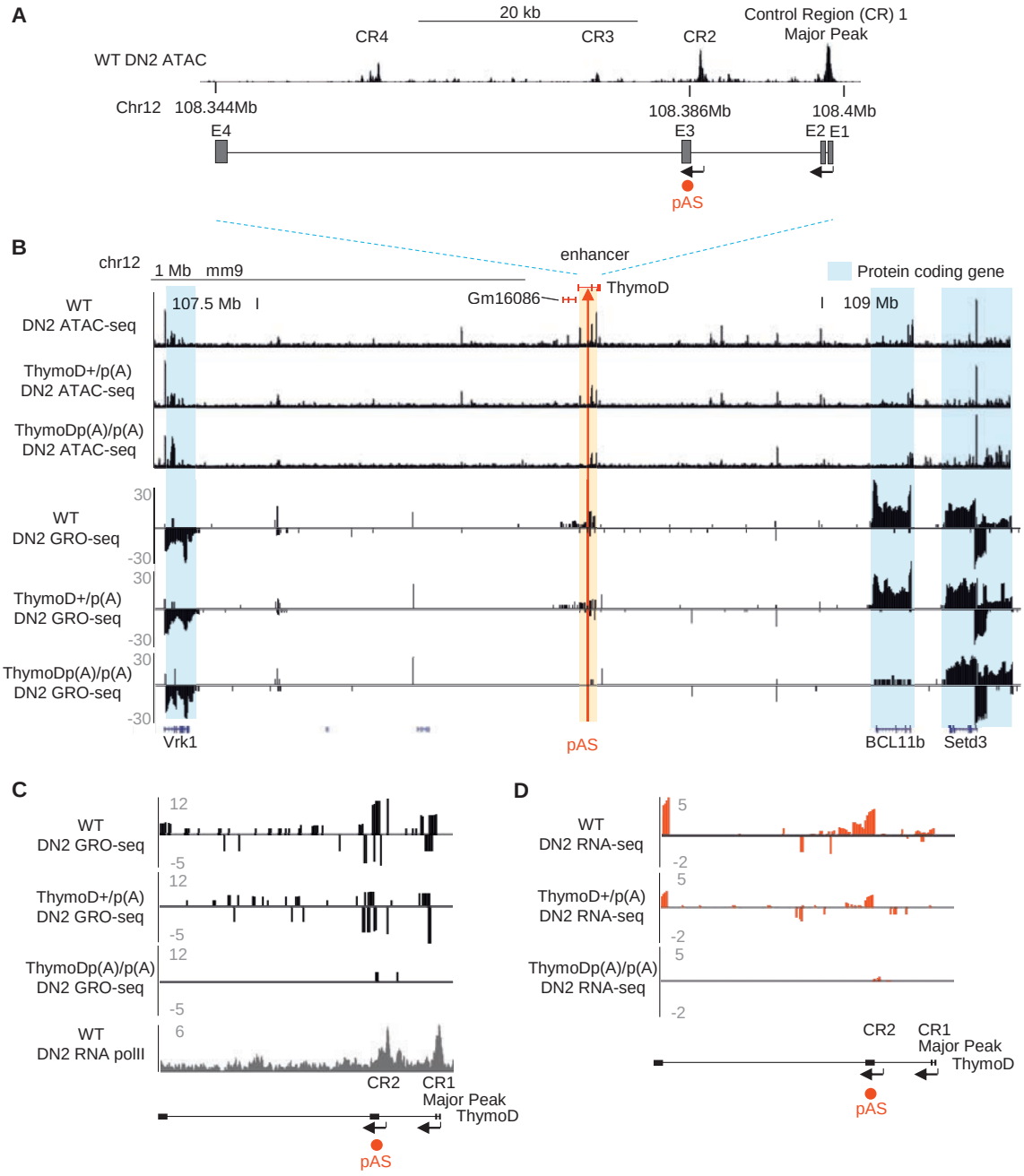


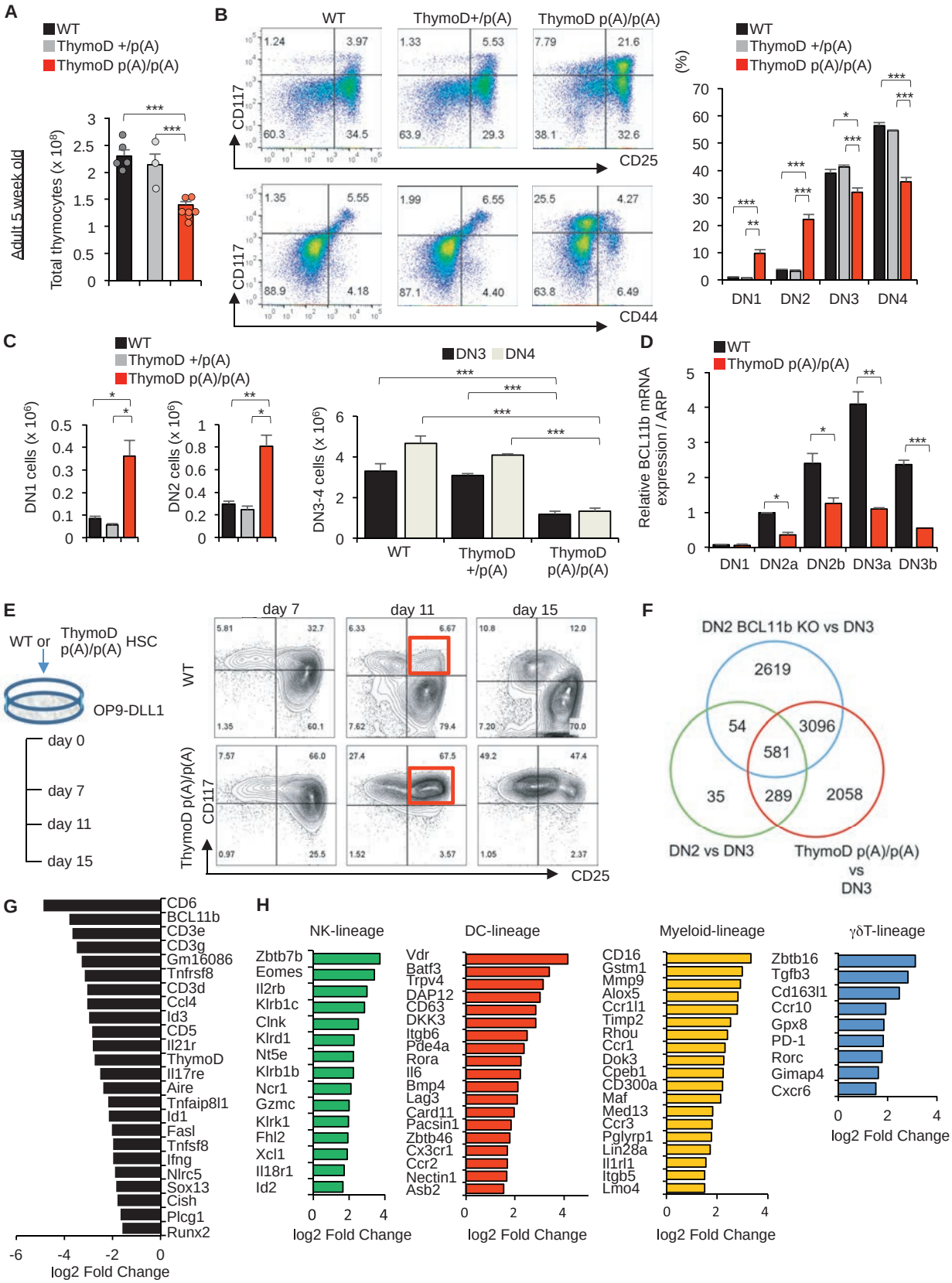
Figure 2. Non-coding RNA ThymoD Activates Bcl11b Expression across Vast Genomic Distances

(A) Schematic diagram depicting the ThymoD region. Upper panel shows ATAC-seq reads for wild-type DN2 cells across the Bcl11b locus control region. CR1 (major peak) indicates the Bcl11b enhancer. CR2, CR3, and CR4 refer to genomic regions depleted for nucleosomes in DN2 cells. Positions of pAS insertions are indicated in red.

(B) ThymoD transcription activates with high precision and selectivity Bcl11b expression. ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells were examined for chromatin accessibility and nascent transcription using ATAC-seq and GRO-seq. ThymoD region is indicated. pAS insertion is shown by red arrow. Enhancer, also named Major Peak or CR1, refers to a previously identified Bcl11b enhancer (Li et al., 2013).

(C and D) pAS insertion abolishes ThymoD transcription. Both GRO-seq and RNA-seq tracks are shown for wild-type, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells.

adjacent and within the targeting construct (Figure S2A). Fragment size of amplified products was as expected (Figure S2B). Sequenced PCR products verified proper insertion of the p(A) sites (Figure S2B; data not shown). DNA sequencing revealed that the putative regulatory elements CR1, CR2, CR3, and CR4 were not mutated by insertion of the poly(A) cassette (data



(legend on next page)

not shown). To validate that the insertion of poly(A) sites interfered with ThymoD transcription, RNA was isolated from ThymoD +/+, ThymoD +/p(A), and ThymoD p(A)/p(A) thymi. We found that ThymoD transcript levels were severely reduced in thymi isolated from ThymoD p(A)/p(A) mice beyond the p(A) insertion site (Figure S2C; data not shown). To examine how ThymoD transcripts were localized in nuclei derived from DN2 cells, we performed RNA-fluorescence *in situ* hybridization (RNA-FISH). ThymoD transcripts were localized at low abundance in wild-type DN2 cells (data not shown).

To determine how p(A) insertion affected the transcription of ThymoD and Bcl11b, we performed global run-on sequencing (GRO-seq). We found that Bcl11b transcripts, as well as transcripts associated with nearby genes, including Vrk1 and Setd3, were readily detectable in wild-type DN2 cells (Figure 2B). To define the transcriptional start sites associated with ThymoD, RNA polymerase II occupancy was analyzed in DN2 cells. Both CR1 and CR2 were enriched for RNA polymerase II occupancy (Figure 2C; lowest panel). Using 3′-RACE (rapid amplification of cDNA ends), we confirmed that CR1 and CR2 act as promoter elements (data not shown). ThymoD sense and anti-sense transcripts were readily detected in wild-type DN2 cells, albeit at relatively low abundance (Figure 2B). However, ThymoD sense and anti-sense transcript levels were severely decreased in ThymoD p(A)/p(A) DN2 cells, indicating that ThymoD transcription is essential to generate a transcriptionally permissive environment (Figures 2C and 2D). Notably, Bcl11b, but not Setd3 and Vrk1, transcript levels were sharply reduced in ThymoD p(A)/p(A) DN2 cells as measured by RNA-seq and GRO-seq (Figure 2B; lowest panel). We conclude that ThymoD transcription activates Bcl11b expression across vast genomic distances with great precision and specificity.

ThymoD Specifies T Cell Fate

To determine whether ThymoD expression specifies T cell fate, thymi were isolated from ThymoD +/+, ThymoD +/p(A), and ThymoD p(A)/p(A) mice and analyzed for developmental abnormalities. We found that thymocyte cellularity in ThymoD p(A)/p(A) mice was reduced 2-fold when compared to thymi derived from wild-type mice (Figure 3A). The proportion and number of the DN1 and DN2 compartments were increased in thymi isolated from ThymoD p(A)/p(A) mice as compared to wild-type mice, whereas the DN3 and D4 compartments were reduced (Figures 3B and 3C). ThymoD p(A)/p(A) thymi also displayed abnormal ratios of DP and SP cells (Figure S3A). The $\gamma\delta$ T cell compartment was reduced in ThymoD p(A)p(A) thymi as compared to wild-type thymi (Figure S3B). We also observed increases in the percentages of $NK1.1^+$, as well as $NK1.1^+B220^+$ cells, in thymi isolated from ThymoD p(A)/p(A) mice (Figures S3C and S3D). Aged ThymoD p(A)/p(A) mice showed more pronounced defects that varied and frequently were accompanied by skewed DN ratios (Figure S3E). Ultimately, the majority of aged ThymoD p(A)/p(A) mice succumbed from leukemia and lymphoma (Figures S3F and S3G). The lymphomas and leukemias that developed in ThymoD p(A)/p(A) mice were largely composed of $CD4^+CD8^+$ double-positive $TCR\beta^+$ T-lineage cells (data not shown). Since the developmental defects and the development of lymphoid malignancies observed in ThymoD p(A)/p(A) mice closely resembled those observed in radiation-induced lymphoma in Bcl11b heterozygous mice, DN1-3 cells were examined for Bcl11b expression (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010b; Kamimura et al., 2007; Gutierrez et al., 2011). Indeed, Bcl11b expression in DN2-DN3 cells sorted from ThymoD p(A)/p(A) mice was significantly reduced, as was observed upon insertion of a poly(A) cassette into the ThymoD locus of the SCIDadh cell line (Figure 3D). To further validate the notion that ThymoD and Bcl11b expression were linked, DN2 cells were generated *in vitro* using OP9-DL1 cells. We found that hematopoietic progenitors derived from ThymoD p(A)/p(A) mice failed to develop beyond the DN2 stage upon culture on OP9-DL1 monolayers, consistent with the developmental arrest observed in ThymoD p(A)/p(A) thymi *in vivo* (Figure 3E). Likewise, when cultured on OP9-DL1 cells, ThymoD p(A)p(A) progenitors displayed defects in T (Figure S4A). To examine Bcl11b expression levels in cultured DN2 cells derived from ThymoD p(A)/p(A) mice, RNA was extracted and analyzed by real-time PCR. As was observed in progenitors derived from ThymoD p(A)/p(A) mice, Bcl11b levels were significantly decreased in ThymoD

Figure 3. ThymoD Specifies T Cell Fate

(A) Reduced cellularity in thymi derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice. Bar graphs show thymocyte cell numbers in 5-week-old ThymoD+/+, ThymoD+/−, and ThymoD−/− mice.

(B) Thymocyte development is partially blocked at the DN2 cell stage in ThymoD p(A)/p(A) mice. Thymocytes derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice were stained for the expression of CD25 versus CD117 and CD44 versus CD117. Cells were gated on the DN compartment. Right panels show bar graphs that display percentages of DN1, DN2, DN3, and DN4 cells in thymi derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice.

(C) Left panels show increased numbers of DN1 and DN2 cells in thymi derived from ThymoD p(A)/p(A) mice when compared to thymi isolated from ThymoD+/+ and ThymoD+/p(A) mice. Right panels show decreased numbers of DN3 and DN4 cells in thymi derived from ThymoD p(A)/p(A) mice when compared to thymi isolated from ThymoD+/+ and ThymoD+/p(A) mice.

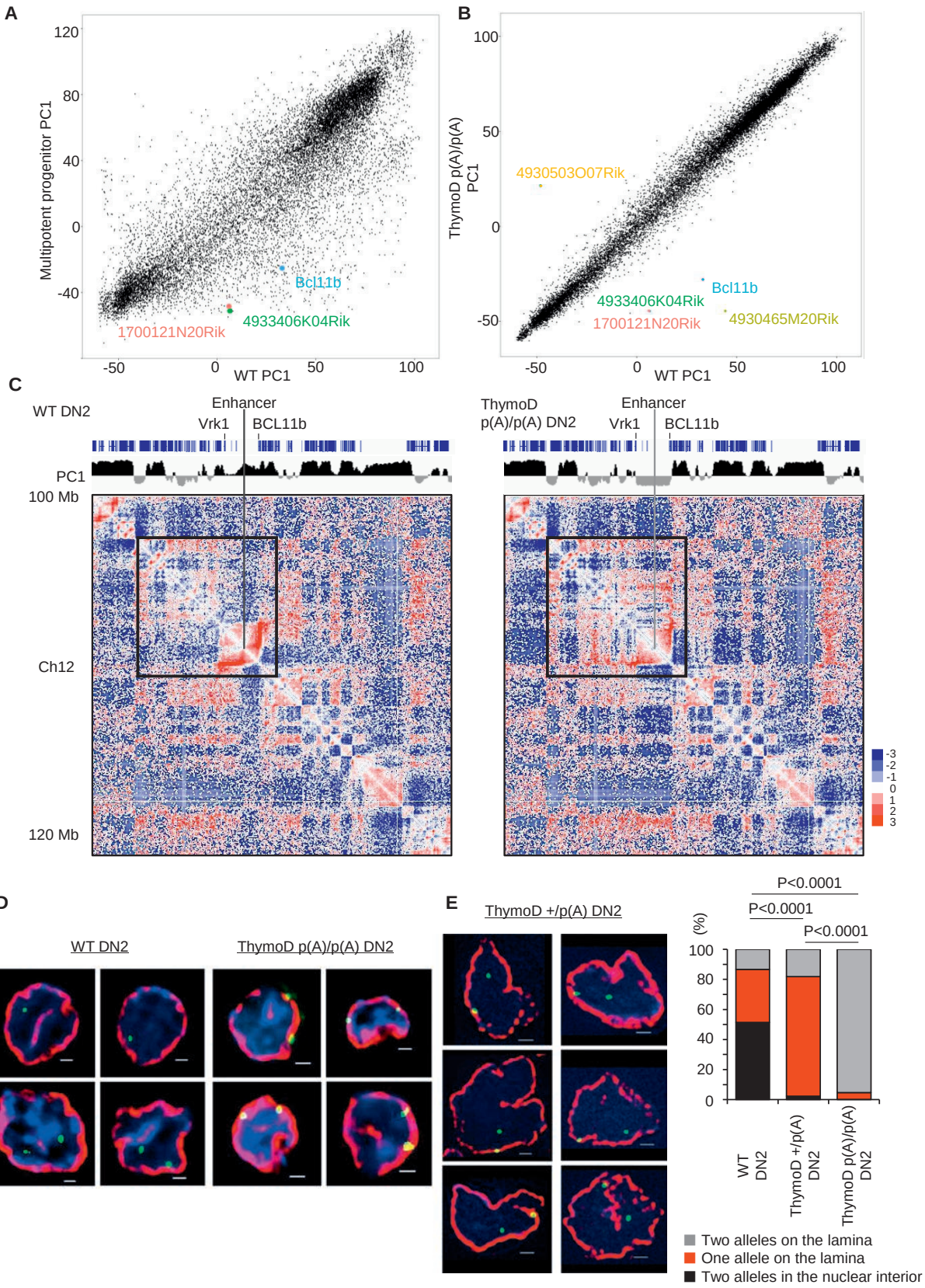
(D) Bar graph indicating relative levels of Bcl11b expression in the DN1-DN3b cells derived from 5-week-old ThymoD+/+ and ThymoD−/− thymi. Results show the mean ± SD (n = 3).

(E) Developmental arrest at the DN2 cell stage in *in vitro* differentiated ThymoD p(A)/p(A) using OP9-DL1 stromal cell cultures. Representative FACS plots for DN compartments are shown.

(F) ThymoD and Bcl11b share a common set of target genes in DN2 cells. Venn diagram is shown displaying differentially expressed genes for indicated genotypes.

(G) ThymoD and Bcl11b are linked in a common pathway to induce a T-lineage specific program of gene expression. Data were obtained from three independent RNA-seq experiments. Table displays a selected group of target genes differentially expressed in sorted ThymoD+/+ and ThymoD p(A)/p(A) DN2 cells. Diagram reveals decreased levels of Bcl11b, Id3, CD3e, and Sox13 expression.

(H) ThymoD acts to suppress the expression of genes associated with the developmental progression of alternative cell lineages. Data were obtained from three independent RNA-seq experiments.



(legend on next page)

p(A)/p(A) DN2 cells cultured *in vitro* (Figure S4B). To determine the extent of overlap between the transcription signatures between ThymoD p(A)/p(A) and Bcl11b-deficient DN2 cells, RNA-seq profiles were compared (Longabaugh et al., 2017). We found that many of the differentially expressed genes in ThymoD p(A)/p(A) were modulated during the DN2 to DN3 transition, and the vast majority overlapped with those differentially expressed in Bcl11b-deficient DN2 cells (Figure 3F). This set of differentially expressed genes was closely linked with specification and commitment of adaptive and innate immune cells (Figures 3G, 3H, and S4C). Prominent among the regulators that characterize adaptive and innate immune cell development were Id2, CD3ε, Runx2, Sox13, Il2rb, Zbtb7b, Fcgr3, Ikzf3, c-kit, Notch3, and Dntt (Figure 3H) (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010a; Rothenberg et al., 2016). These data directly link ThymoD and Bcl11b into a common pathway to specify T cell fate.

ThymoD Expression Releases the Bcl11b Enhancer from the Nuclear Lamina

Given that the Bcl11b enhancer repositions away from the lamina in developing T cell progenitors and that ThymoD modulates Bcl11b expression, we considered the possibility that ThymoD transcription acts to release the Bcl11b locus from the lamina to the nuclear interior. To address this possibility, we performed HiC analysis on wild-type and ThymoD p(A)/p(A) DN2 cells. Specifically, adult hematopoietic progenitors were isolated from wild-type and ThymoD p(A)/p(A) hematopoietic progenitors, cultured on OP9-DL1 cells, isolated using cell sorting, formaldehyde fixed, and analyzed using HiC. From the HiC reads, we generated contact maps and used principal component analysis to segregate the DN2 genomes into A versus B compartments. Upon comparing PC1 values derived from multipotent progenitors versus wild-type DN2 cells, we found a large ensemble of genomic regions (2260) that switched compartments during the transition from multipotent progenitors to committed DN2 cells (Figure 4A and Table S1). In contrast, a mere 30 genomic regions were associated in different compartments upon comparing ThymoD+/+ and ThymoD p(A)/p(A) DN2 cells (Figure 4B and Table S1). Conspicuous among these was the Bcl11b intergenic region (Figure 4B). The finding that a rather restricted number of genomic regions flipped PC1 values raised the question as how this compares to changes in compartmentalization during the DN2 to DN3 transition. To explore this possibility, we performed HiC on isolated RAG-deficient thymocytes derived from fetal liver cells and cultured on OP9-DL1 cells. Contact maps were generated from DN3 HiC reads and directly compared to those observed for DN2 cells. We found that a total of sixty genomic regions switched compartments during the DN2 to DN3 transition (Tables S2 and S3). We next compared changes in the PC1 distribution across chromosomes for wild-type DN2 versus ThymoD p(A)/p(A) DN2, wild-type DN2 versus wild-type DN3, and wild-type DN2 versus multipotent progenitors. The correlation coefficient of interactions for each locus for the paired cell types was calculated and plotted as the mean correlation coefficient for all loci across each chromosome. We found that most of the differences in PC1 values in wild-type versus ThymoD p(A)/p(A) DN2 cells were restricted to chromosome 12, where the Bcl11 locus is located (Figure S4D). These data are consistent with the notion that ThymoD acts locally to orchestrate compartmentalization. As expected, the changes in compartmentalization in wild-type DN2 versus ThymoD p(A)/p(A) DN2 cells were also closely associated with differences in long-range genomic interactions that span the Bcl11b locus control region (Figure 4C). In wild-type DN2 cells, we observed an elaborate pattern of looping involving the Bcl11b locus control region and the Bcl11b coding region (Figure 4C). In contrast, ThymoD p(A)/p(A) DN2 cells were severely depleted for such interactions. Rather, we observed genomic interactions spanning vast genomic distances and primarily involving neighboring heterochromatic regions (Figure 4C). To validate the changes in compartmentalization, we performed immune-3D-FISH using antibodies directed against the lamina and fluorescently labeled probes corresponding to the Bcl11b intergenic region. As expected, we found that in wild-type DN2 cells, the Bcl11b intergenic region was predominantly localized at the nuclear interior, whereas in ThymoD p(A)/p(A) DN2 cells, the Bcl11b locus was retained for the large majority of cells at or near the nuclear lamina (Figure 4D). Collectively, these observations indicate that ThymoD expression is essential to release the Bcl11b intergenic region from the nuclear lamina.

Nascent ThymoD Transcription Acts in *Cis* to Release the Bcl11b Locus Control Region from the Lamina

While the data described above indicate that ThymoD transcription is essential to release the Bcl11b intergenic region from the nuclear lamina, it remained to be determined whether ThymoD transcription acts in *cis* or in *trans*. To address this

Figure 4. ThymoD Acts in *Cis* to Reposition the Bcl11b Enhancer from the Lamina to the Nuclear Interior

(A) Large ensemble of genomic regions reposition during the transition from multipotent progenitors to differentiated DN2 cells. PC1 values derived from HiC reads are plotted for multipotent progenitors versus differentiated DN2 cells.

(B) ThymoD expression is essential to reposition the Bcl11b locus control region in developing T cell progenitors. Comparison of PC1 values derived from HiC reads generated from ThymoD+/+ versus ThymoD p(A)/p(A) DN2 cells are shown.

(C) ThymoD expression regulates Bcl11b enhancer and promoter compartmentalization. Contact maps derived from HiC reads derived from wild-type and ThymoD p(A)/p(A) DN2 cells are shown. Positive PC1 values are indicated in black, whereas negative PC1 values are displayed in gray. Heatmap indicates the ratio of observed versus expected frequencies for genomic interactions, as revealed by HiC. Numerical and statistical details are described in STAR Methods.

(D) ThymoD expression is essential to release the Bcl11b super-enhancer from the nuclear lamina. Images represent 3D-FISH analysis using a fosmid probe containing the Bcl11b super-enhancer. The nuclear envelope was visualized using antibodies directed against the lamina and is indicated in red. Green signal indicate Bcl11b super-enhancer. Blue reflects DAPI staining.

(E) ThymoD acts in *cis* to release the Bcl11b intergenic region from the lamina. Representative images were taken from ThymoD+/p(A) DN2 cells. Bar graph shows localization of Bcl11b in ThymoD p(A)/p(A) (n = 226), ThymoD+/p(A) (n = 176), and ThymoD p(A)/p(A) DN2 cells (n = 85). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm. p values are indicated.

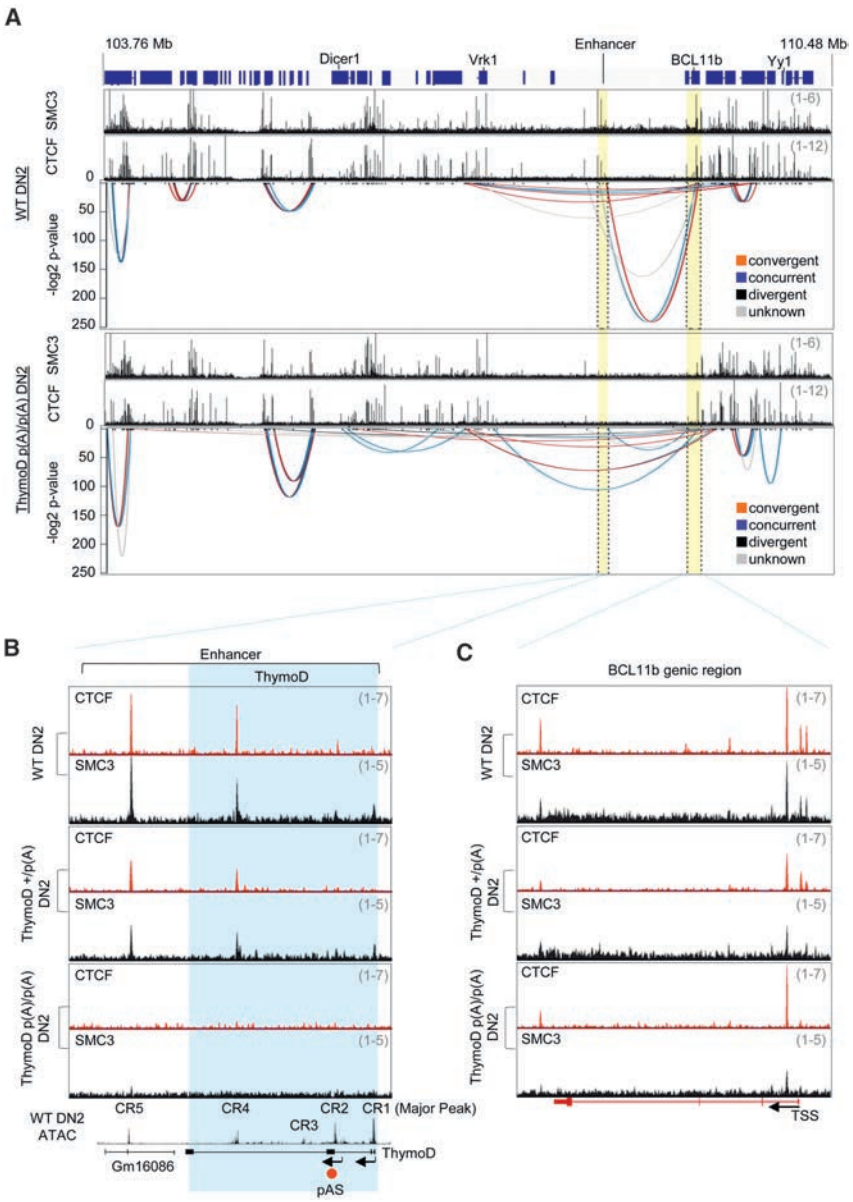


Figure 5. ThymoD Transcription Juxtapose the Bcl11b Super-Enhancer and Promoter Regions into a Single-Loop Domain

(A) ThymoD transcription orchestrates cohesin-dependent looping to bring the Bcl11b promoter and enhancer region into a single-loop domain. Genomic interactions derived from HiC reads across the Bcl11 intergenic and genic regions are shown for wild-type and ThymoD p(A)/p(A) DN2 cells.

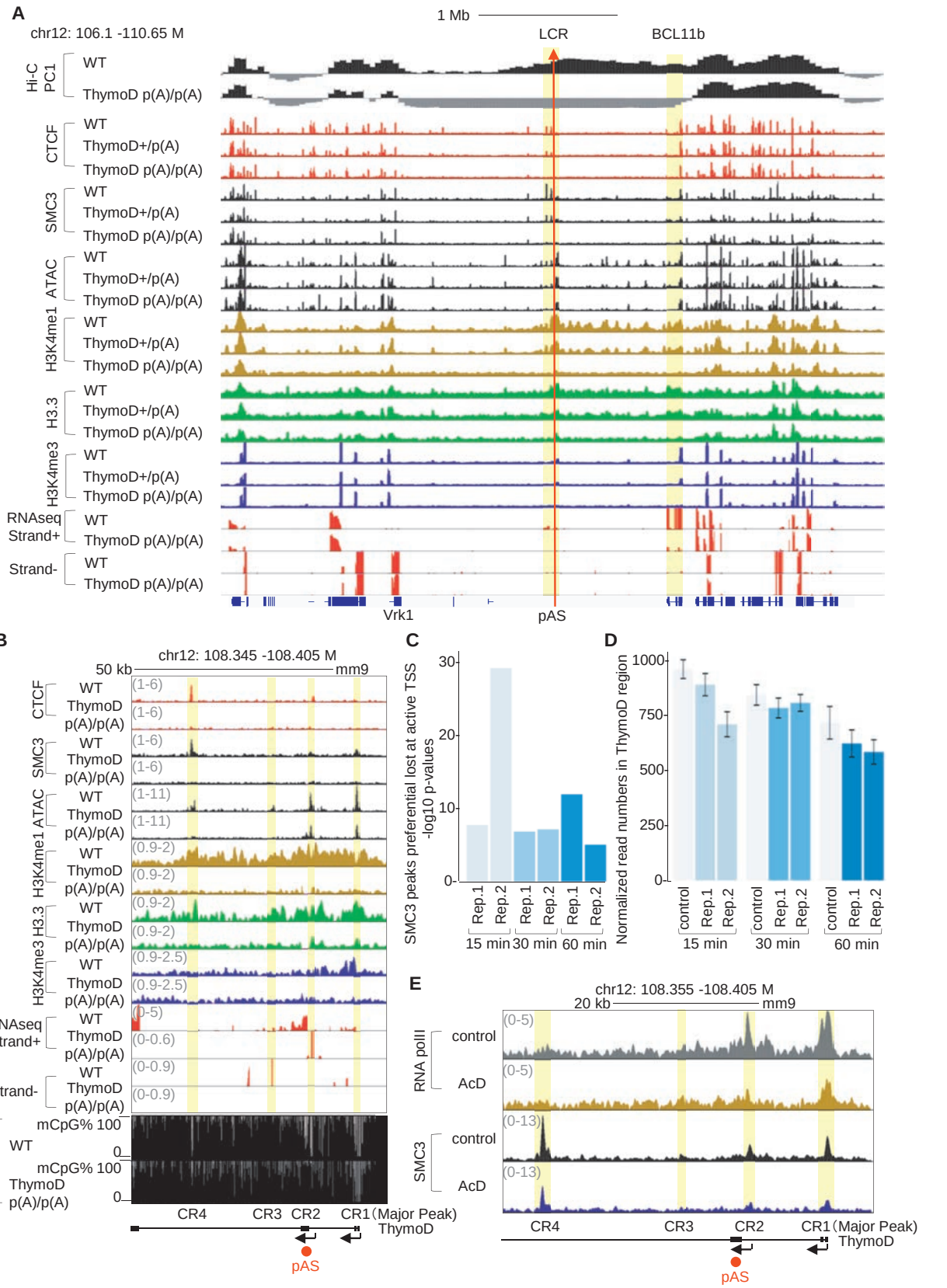
(B) ThymoD p(A)/p(A) DN2 cells are severely depleted for CTCF and SMC3 occupancy across the ThymoD locus. ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells were generated from adult hematopoietic progenitors that were cultured in the presence of OP9DL1 and cytokines. Tracks display CTCF and SMC3 occupancy across a genomic region containing the Bcl11b super-enhancer in ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells. Note that recent mapping data (mm10) indicates that ThymoD and Gm16086 span a single transcription unit. CR1, also named Major Peak, refers to a previously identified Bcl11b enhancer (Li et al., 2013). ChIP-seq reads were corrected for input reads.

(C) ThymoD p(A)/p(A) DN2 cells are severely depleted of CTCF and SMC3 occupancy across the Bcl11b genic region. Tracks display CTCF and SMC3 occupancy across the Bcl11b genic region in ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells. CTCF and SMC3 peaks were called using HOMER and marked as red and black tags, respectively. Note that ChIP-seq reads were corrected for input reads.

question, ThymoD+/+ and heterozygous ThymoD+/p(A) DN2 cells were examined for attachment of the Bcl11b intergenic region to the nuclear lamina using immune-3D-FISH. We found that in the vast majority of DN2 +/p(A) cells, a single allele of the Bcl11b super-enhancer was sequestered at the nuclear lamina, whereas the other allele was positioned in the nuclear interior (Figure 4E). Thus, these data indicate that ThymoD expression acts in *cis* to modulate the nuclear location of the Bcl11b locus control region. To test whether forced ThymoD or ThymoD2 expression in ThymoD +/p(A) DN2 cells has the ability to release the ThymoD p(A) allele in *trans*, hematopoietic progenitor cells isolated from ThymoD +/p(A) bone marrow were transduced with virus expressing vector alone, ThymoD (exons 1–3), and ThymoD2 (exons 1, 2, and 4) and were examined for the nuclear positioning of the Bcl11b locus control region (data not shown). We found that the ThymoD p(A) allele remained at the nuclear lamina regardless of ectopic ThymoD or ThymoD2 expression (data not shown). Taken together, these findings demonstrate that (1) nascent ThymoD transcription is required to release the Bcl11b locus control region from the nuclear lamina and (2) ThymoD transcripts generated at wild-type alleles do not have the ability to act in *trans* and release the other Bcl11b allele from the lamina. We conclude that in developing T cell progenitors ThymoD expression acts in *cis* to release the Bcl11b locus control region from the nuclear lamina.

ThymoD-Induced Structural Changes in Chromatin Topology Foster Bcl11b Enhancer-Promoter Communication

To determine how ThymoD transcription modulates local chromatin folding, we examined HiC contact reads derived from wild-type and ThymoD p(A)/p(A) DN2 for genomic interactions across the Bcl11b intergenic and genic regions (Figure 5A). In wild-type DN2 cells, the most significant interactions across a 7 Mb genomic region involved the Bcl11b enhancer and promoter (Figure 5A). Strikingly, genomic interactions associated



(legend on next page)

with the enhancer and promoter were greatly diminished in DN2 cells derived from ThymoD p(A)/p(A) mice (Figure 5A). Rather, we found that ThymoD p(A)/p(A) DN2 cells were highly enriched for loops involving the Bcl11b genic and promoter regions with heterochromatin (Figures 4C and 5A). Thus, ThymoD transcription promotes a pattern of genomic interactions that greatly enriches genomic encounters involving the Bcl11b promoter and enhancer.

To determine how genomic interactions were affected by ThymoD expression, we examined wild-type and ThymoD p(A)/p(A) DN2 cells for CTCF and cohesin (SMC3) occupancy (Figure 5B). We found that in wild-type DN2 cells, CTCF and SMC3 bound to multiple sites that spread across the Bcl11b intergenic and genic regions (Figures 5B and 5C). As expected, the majority of CTCF bound sites overlapped with SMC3 occupancy (Figure 5A). Importantly, both CTCF and SMC3 occupancy across the Bcl11b intergenic region were severely affected in ThymoD p(A)/p(A) DN2 cells (Figure 5B). CTCF and SMC3 occupancy was also affected at the Bcl11b promoter (Figures 5C and S5A).

It could be argued that ThymoD p(A)/p(A) DN2 cells were developmentally distinct from that of wild-type DN2 cells and that the abnormalities observed in CTCF and SMC3 occupancy were caused by a developmental arrest. To address this possibility, we examined CTCF and SMC3 occupancy in ThymoD+/p(A) mice, in which thymocyte development was not perturbed. We found that in ThymoD +/p(A) DN2 cells, CTCF and SMC3 occupancy was reduced 2-fold across the Bcl11b intergenic region, suggesting a direct link between ThymoD transcription and CTCF and SMC3 occupancy (Figure 5B).

As a first approach to determine the mechanism by which ThymoD transcription promotes CTCF and SMC3 occupancy, we examined the ThymoD transcribed region for nucleosome depletion, as well as for the deposition of epigenetic marks (Figures 6A and 6B). We found that ThymoD transcription affected nucleosome depletion and the deposition of epigenetic marks associated with active transcription and the deposition of H3.3 across the Bcl11b promoter, but not genomic regions that flank the Bcl11b regulatory elements (Figures 6A, 6B, and S5A).

The absence of CTCF occupancy in ThymoD p(A)/p(A) DN2 cells raised the question as to why CTCF failed to bind its cognate sites across the ThymoD transcribed region. Since recent studies have demonstrated that non-coding transcription is closely associated with CpG hypomethylation, we considered the possibility that transcription across the ThymoD region acts to recruit members of the Tet protein family (Benner et al., 2015). Hence, we examined DNA isolated from wild-type and ThymoD p(A)/p(A) DN2 cells for cytosine DNA methylation using genome-wide bisulfite sequencing (Figure 6B; bottom tracks). Notably, several CpG residues associated with CTCF binding sites across the ThymoD transcribed region were hypermethylated in ThymoD p(A)/p(A) DN2 cells (Figure 6B; upper and bottom tracks). Taken together, these data indicate that ThymoD transcription is essential to (1) demethylate CpG residues closely associated with CTCF bound sites, (2) recruit members of the cohesin complex to the transcribed region, and (3) reposition the Bcl11b enhancer from a repressive to a transcriptionally permissive neighborhood.

Nascent ThymoD Transcription Is Essential to Recruit Cohesin to the Bcl11b Locus Control Region

The data described above indicate that ThymoD transcription is associated with modulating chromatin topology to bring the Bcl11b enhancer and promoter elements within close spatial proximity. To determine whether nascent ThymoD transcription is required to recruit members of the cohesin complex to the Bcl11b locus control region, wild-type DN2 cells were incubated in the presence of actinomycin D for 15, 30, and 60 min. Treated cells were formaldehyde fixed and immunoprecipitated with antibodies directed against RNA polymerase II and SMC3 followed by genome-wide sequencing. Consistent with previous observations, we found that SMC3 occupancy was significantly depleted at sites associated with RNA polymerase II occupancy (Figure 6C and Table S1) (Izumi et al., 2015; Bhardwaj et al., 2016). We next evaluated whether in DN2 cells nascent ThymoD transcription directly recruits SMC3. We found that DN2 cells treated with actinomycin D showed a significant reduction of SMC3 occupancy across the transcribed region (Figures 6D and S5B). Notably, a brief 15 min incubation with actinomycin D significantly affected SCM3 occupancy (Figure 6E). We conclude that the nascent

Figure 6. ThymoD Transcription Promotes the Deposition of Histone Variant H3.3, Activating Histone Marks and Hypomethyation across the Bcl11b Intergenic Region

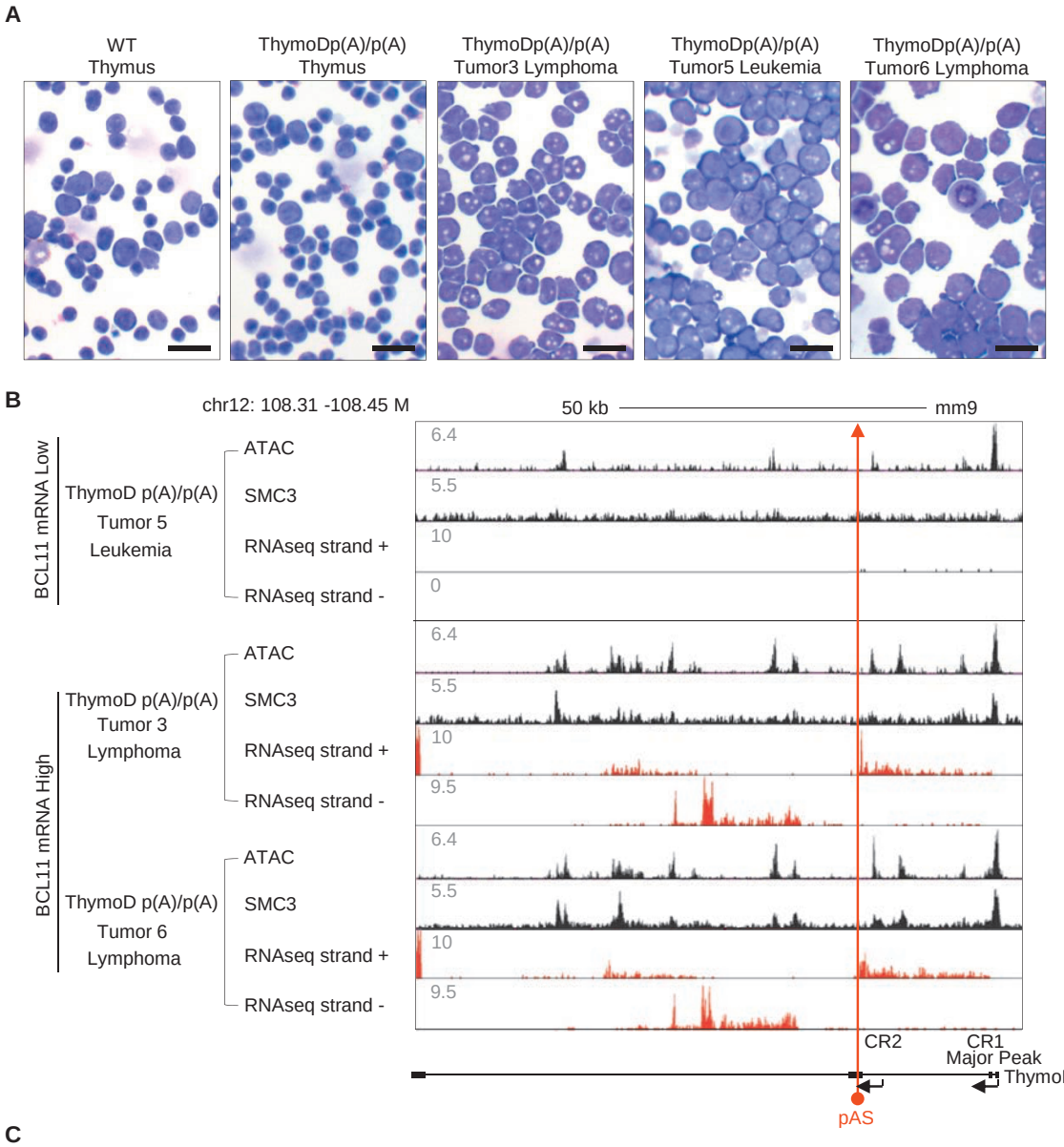
(A) ThymoD expression promotes nucleosome depletion and deposition of epigenetic marks, chromatin accessibility, and H3.3 across the Bcl11b intergenic region. Tracks indicate PC1 values derived from HiC reads, ATAC reads, and CHiP-seq reads for CTCF, SMC3, H3K4me1, H3K4me3, and H3.3 for wild-type and ThymoD p(A)/p(A) DN2 cells. Bottom panels indicated RNA-seq reads across the Bcl11b locus. pAS insertion site is indicated by red arrow. Note that ChIP-seq reads were corrected for input reads.

(B) ThymoD expression modulates chromatin accessibility and the deposition of epigenetic marks across the ThymoD locus. Tracks indicate ATAC reads and ChIP-seq reads for CTCF, SMC3, H3K4me1, H3K4me3, and H3.3 for wild-type and ThymoD p(A)/p(A) DN2 cells. RNA-seq reads across the ThymoD locus for wild-type and ThymoD p(A)/p(A) DN2 cells are indicated. Lower panels show CpG DNA hyper methylation across the CR2 and CR4 regions in ThymoD p(A)/p(A) DN2 cells.

(C) Cohesin occupancy in DN2 cells is closely associated with active transcription start sites. DN2 cells were incubated in the presence of actinomycin D for the indicated time points, formaldehyde fixed, and analyzed for SMC3 occupancy using ChIP-seq. Active transcription start sites were identified using RNA polymerase II ChIP-seq. Plotted is the –log 10 of p values of the Fisher test of the contingency table (Table S1). The analysis indicates preferential loss of cohesin occupancy at active TSS sites.

(D) SMC3 loading across the ThymoD transcribed region requires nascent transcription. SMC reads were determined across the ThymoD locus (chr12: 108.345–108.405 Mb) in DN2 cells treated in the absence or presence of actinomycin D for indicated times and normalized to the total read number. The mean and standard deviation for each experiment were derived from 20 independent samples using boot strapping for all reads across the genome.

(E) Nascent RNA transcription targets SMC3 to the ThymoD locus. DN2 cells were incubated in the absence or presence of actinomycin D for 15 min and analyzed for SMC3 occupancy using ChIP-seq. Representative track shows a reduction of SMC3 occupancy across the ThymoD locus.



(legend on next page)

ThymoD transcription acts in *cis* to directly promote SMC3 loading across the transcribed region.

Compartmentalization and Cohesin-Dependent Looping Is Reversible in ThymoD-Deficient Lymphomas and Leukemias

As described above, ThymoD p(A)/p(A) mice readily develop lymphoma or leukemia. To determine whether the Bcl11b intergenic remained attached to the nuclear lamina in ThymoD p(A)/p(A) lymphomas, long-term cultures were established from the tumors. Tumors isolated from aged ThymoD p(A)/p(A) mice displayed a higher fraction of blasting cells that frequently adopted irregular nuclear morphologies and readily expanded in tissue culture (Figure 7A). Long-term cultures were established either in the absence or presence of OP9-DL1 stromal cells. From these expanding cultures, we established independent cell lines representing either lymphomas or leukemias. To determine whether the expression of ThymoD and Bcl11b remained silent in the lymphomas and leukemias, the tumors were examined for ThymoD and Bcl11b expression using RNA-seq (Figure S6). As expected, ThymoD transcription was blocked at the poly(A) insertion in all of the tumors (Figure S6). Intriguingly, three of the tumors expressed anti-sense transcription at relatively high levels in three of the six tumors (Figure S6). Notably, lymphomas and leukemias that expressed high levels of anti-sense transcription also abundantly expressed Bcl11b (Figure S6). These data indicate that a subset of lymphoid malignancies that develop in ThymoD p(A)/p(A) mice display coordinate ThymoD anti-sense and Bcl11b transcription, whereas in tumors that failed to induce anti-sense ThymoD transcription, Bcl11b expression remained suppressed.

To examine whether the activation of ThymoD and Bcl11b expression in a subset of lymphoma cell lines was associated with chromatin accessibility and SMC3 occupancy across the ThymoD region, three lymphomas (3, 5, and 6) were examined using ATAC-seq and SMC3 ChIP-seq (Figure 7B). As predicted by the analyses described above, lymphomas that were associated with coordinate ThymoD and Bcl11b expression showed widespread ATAC reads across the ThymoD locus and were enriched for SMC3 occupancy (Figure 7B).

To determine whether the induction of ThymoD and Bcl11b expression, as well as the selective increase in SCM3 occupancy, was associated with changes in nuclear location, tumors 3 and 5 were examined by 3D-FISH for compartmentalization. We found that the majority of cells derived from tumor 3, which coordinately expressed high levels of ThymoD and Bcl11b, exhibited localization of the Bcl11b intergenic region that was localized away from the nuclear lamina (Figure 7C). In contrast, the localization of ThymoD in tumor 5, which expresses low levels of ThymoD and Bcl11b, was near or on the nuclear lamina (Figure 7C). These data indicate that the insertion of a poly(A) cassette in the ThymoD locus does not irreversibly attach the ThymoD locus to the nuclear lamina. Instead, ThymoD transcription, either sense or anti-sense, promotes SMC3 occupancy and cohesin-dependent looping, changes in compartmentalization, and the activation of Bcl11b expression to specify T cell fate (Figure S7).

DISCUSSION

Specifying Immune Cell Fate by Non-coding Transcription

While it is now well established that genomic regions reposition during developmental progression, it is not understood how such changes in nuclear location are regulated. Here, we provide insight into the mechanism that underpins the repositioning of regulatory elements. We identify a non-coding RNA, named ThymoD, which acts to reposition the Bcl11 enhancer away from the lamina to the nuclear interior and sequester the Bcl11b promoter and enhancer region into a single-loop domain. How does ThymoD accomplish the repositioning of the Bcl11b enhancer? The phenotype revealed in ThymoD+/p(A) thymocytes strongly suggests that ThymoD acts in *cis* on nearby gene expression, consistent with previous observations that nascent transcription, rather than biochemical activities, associated with a subset of non-coding RNAs executes their function (Melé and Rinn, 2016; Engreitz et al., 2016). Here, we found that the vast majority of DN2+/p(A) cells display one Bcl11b allele located at the nuclear lamina and one in the nuclear interior. Thus, ThymoD acts in *cis* to release the Bcl11b enhancer from the lamina and to active Bcl11b gene expression. How does ThymoD perform this task? We propose a stepwise mechanism. The induction of Bcl11b expression is initiated and maintained in early T cell progenitors by activating a yet-to-be-identified enhancer associated with the ThymoD locus. Next, we suggest that the ThymoD transcribed region directs dioxygenases, Tet proteins, to specific residues that are associated with CTCF occupancy. This explains why, in the absence of ThymoD expression, CTCF occupancy is perturbed across the transcribed region, consistent with previous observations indicating

Figure 7. Compartmentalization and Cohesin-Dependent Looping Is Reversible in ThymoD-Deficient Lymphomas

(A) Development of leukemias and lymphomas in ThymoD p(A)/p(A) mice. Tumors isolated from aged ThymoD p(A)/p(A) mice displayed a higher fraction of blasting cells that frequently adopted irregular nuclear morphologies and readily expanded in tissue culture. DN2 cells derived from wild-type and ThymoD p(A)/p(A) DN2 cells, as well as several lymphomas and leukemias, are shown. Black bar represents 40 μm.

(B) ThymoD expression, either sense or anti-sense transcription, promotes SMC3 occupancy and cohesin occupancy, chromatin accessibility, and the activation of Bcl11b expression. pAS insertion is indicated by red arrow.

(C) ThymoD transcription is closely associated with compartmentalization. Wild-type and ThymoD p(A)/pA) DN2 cells and ThymoD p(A)/pA) lymphomas were examined for nuclear localization of the Bcl11b intergenic region. The two left images show localization for the Bcl11b super-enhancer at the nuclear lamina in multipotent progenitors. Upper images indicate location of the Bcl11b enhancer in the nuclear interior of ThymoD p(A)/p(A) tumor 3. The bottom images show localization of the Bcl11b enhancer at the nuclear lamina in tumor 5. DAPI staining is shown in blue. Bar graph shows fraction of spatial distances (< 500nm) separating the Bcl11b super-enhancer from the nuclear lamina in multipotent progenitor cells (n = 175) and WT DN2 cells (n = 135). ThymoD p(A)/p(A) DN2 cells (n = 120), ThymoD p(A)/p(A) tumor3 (n = 206), ThymoD p(A)/p(A) tumor5 (n = 207). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm.

that nascent transcription is closely associated with hypomethylated DNA (Benner et al., 2015). Previous observations have also indicated that mice deficient for *Tet2* and *Tet3* expression displayed defects in NKT cell development and lymphoproliferative disease and also exhibited significantly lower abundance of *Bcl11b* expression (Tsagaratou et al., 2017). Although still to be proven, it is conceivable that the absence of *Tet2* and *Tet3* expression in NKT cells leads to hypermethylated CpG residues across the ThymoD transcribed region and in a failure to efficiently activate Bcl11b expression, plausibly resulting in lympho-proliferative disease and defects in NKT cell development. How the Tet proteins target specific CpG residues spanning the transcribed region is an important question that needs to be addressed but may involve distinct secondary DNA structures that are induced either by sense or anti-sense transcription. Regardless of the precise mechanism, these data link non-coding transcription and Tet targeting into a common pathway.

ThymoD transcription was associated with cohesin occupancy consistent with previous studies (Izumi et al., 2015; Bhardwaj et al., 2016; Busslinger et al., 2017). Recruitment of the cohesin complex, in turn, activates an elaborate and coherent pattern of looping, likely involving loop extrusion, to bring the enhancer and promoter region in a singular loop domain (Nasmyth, 2001; Sanborn et al., 2015; Fudenberg et al., 2016). Juxtaposing paired enhancer and promoter regions, separated by vast genomic distances, would yield greatly reduced first-passage times (Lucas et al., 2014). We note that loop extrusion would permit the repositioning of paired enhancers and promoters with great precision even when separated by vast genomic distances. Finally, loop extrusion would sequester the loop domain away from the transcriptionally repressive environment to the euchromatic compartment. We suggest that, once localized in the transcriptionally permissive compartment, transcriptional regulators associated with T-lineage commitment, including GATA-3, TCF1, and RUNX1, would bind enhancer elements across the loop domain to orchestrate the deposition of activating histone marks, as well as the histone variant H3.3, facilitating enhancer-promoter communication that may involve phase separation (Hnisz et al., 2017). Such a sequential order of gene activation is consistent with the notion that ThymoD transcription (DN1 stage) precedes the expression of non-coding RNAs initiated from the enhancer regions (DN2a stage). In sum, we propose a model in which nascent ThymoD transcription dictates enhancer-promoter communication at multiple steps that precede gene activation: (1) demethylation at CpG residues closely associated with CTCF occupancy across the ThymoD genomic region, (2) activation of cohesin-dependent looping to juxtapose the enhancer and promoter into a single-loop domain and to relocalize the enhancer from a heterochromatic to an euchromatic environment, and (3) deposition of activating epigenetic marks across the loop domain once positioned in the euchromatic compartment to facilitate interactions of paired regulatory elements by phase separation.

Mechanisms that Orchestrate Immune Cell Development

Previous studies have established that a subset of genomic regions reposition in the nucleus during the transition from the multipotent progenitor cell stage to committed B lineage cells (Lin et al., 2012; Miyazaki et al., 2014; van Steensel and Belmont, 2017). Switching the nuclear location of genomic regions is often associated with changes in gene expression (Lin et al., 2012; Kind et al., 2015). Prominent among the genomic regions that switch nuclear location in B cell progenitors are genes encoding for transcriptional regulators that specify B cell fate, including the EBF1 and FOXO1 loci (Lin et al., 2012; Mansson et al., 2012). Both loci are associated with the heterochromatic compartment in multipotent progenitors but reposition to the transcriptionally permissive compartment upon developing into committed B lineage cells (Lin et al., 2012). Here, we demonstrate that in developing T cell progenitors, a genomic region associated with the Bcl11b locus that harbors critical regulatory elements repositions from the nuclear lamina to the transcriptionally permissive nuclear interior. These findings point to a common mechanism that specifies B and T cell fate. In developing B cell progenitors, the EBF1 locus switches from the nuclear lamina to the transcriptionally permissive compartment, leading to the induction of EBF1 expression to establish B cell identity. In contrast, in T cell progenitors, the Bcl11b enhancer repositions from the lamina to the nuclear interior to activate Bcl11b expression. Why has such a complex mechanism, involving repression, nuclear repositioning, and activation, evolved to promote adaptive immune cell fate? We suggest that anchoring regulatory elements to the nuclear lamina permits efficient transcriptional repression across a large genomic region preventing stochastic and premature activation of a lymphoid-specific program of gene expression in hematopoietic progenitors. Only upon reaching a distinct developmental stage are progenitor cells instructed to release locus control regions or super enhancers from the nuclear lamina to the transcriptionally active compartment. In B cell progenitors, a repositioned EBF1 locus control region in the transcriptionally permissive compartment would associate with B-lineage-specific transcription factors such as E2A and FOXO1 to induce a B-lineage-specific program of gene expression (Mansson et al., 2012). In T cell progenitors, the Bcl11b locus control region being positioned in the transcriptionally permissive compartment would be able to respond to Notch signaling and transcriptional regulators that include GATA3, TCF1, and RUNX1 to activate a T-lineage-specific transcription signature (Kueh et al., 2016). Thus, specification of adaptive immune cell fate in early lymphoid progenitors is orchestrated by a common mechanism that involves the repositioning of regulatory regions from the lamina to the euchromatic compartment.

Non-coding Transcription and Tumor Suppression

The activation of anti-sense ThymoD transcription associated with a subset of lymphomas that were derived from ThymoD p(A)/p(A) mice was initially puzzling. How is anti-sense ThymoD transcription induced in a subset of lymphomas? Murine T cell lymphomas are frequently associated with mutations that activate the Notch signaling pathway, which activates Bcl11b expression (Weng et al., 2004; Yui and Rothenberg, 2014). Although a detailed characterization of cooperating oncogenic mutations in the ThymoD p(A)/p(A) lymphomas will be required, we speculate that the aberrant activation of the Notch signaling cascade leads to the induction of high levels of anti-sense

ThymoD transcription, ultimately leading to release of the Bcl11b intergenic region from the lamina. Since both sense and antisense ThymoD transcription are associated with changes in compartmentalization and chromatin folding, we suggest that ThymoD transcription, rather than its primary nucleotide sequence, is the critical determinant that dictates the specificity of Bcl11b enhancer-promoter communication.

Previous studies have demonstrated that somatic mutation of an intergenic region may generate an oncogenic superenhancer (Mansour et al., 2014). Our observations indicate that non-coding transcription may act as a tumor suppressor. Given that non-coding transcription is widespread, we suggest that, likewise, non-coding transcription is a common mechanism to activate the expression of nearby genes that encode for proteins with tumor suppressive function.

Conclusion

It is now well established that the genomes of animal and plant kingdoms are organized into euchromatic and heterochromatic regions. Numerous studies have also identified genomic regions that switch nuclear location during developmental progression. The underlying mechanism that directs changes in compartmentalization, however, remained to be revealed. Here, we present data indicating that non-coding transcription dictates remodeling of local chromatin structure, possibly involving loop extrusion, to reposition regulatory elements from the heterochromatic to the euchromatic compartment. We suggest that local remodeling of chromatin topology by non-coding transcription-induced loop extrusion is a universal mechanism that permits genomic regions to readily switch compartments.

Enhancers are often separated from promoter regions by vast genomic distances. How enhancer elements select their cognate promoter regions with high precision and specificity remains to be revealed (Farley et al., 2015). We suggest that developmental and lineage-specific patterns of non-coding transcription sequester, by loop extrusion, enhancer and promoter elements into a single-loop domain. Non-coding transcription-induced loop extrusion explains how enhancers and promoters, segregated from each other by vast genomic distances and positioned in separate compartments, find each other with great specificity during developmental progression.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - SCIDadh cell line culture
  - Mice
  - T cell culture
  - ThymoD p(A)/p(A) tumor culture
- METHOD DETAILS
  - Generating SCIDadh pA/pA cell line
  - sgRNA template sequence at upstream of ThymoD CR1
  - sgRNA template sequence at exon3 of thymoD (upstream of CR2)
  - Generating ThymoD pA/pA mice
  - Primer for genotyping
  - Flow cytometry
  - Wright Giemsa staining
  - Real-time PCR
  - Primer for qPCR
  - 3D-FISH
  - In Situ Hi-C
  - ATAC-Seq
  - ChIP-Seq
  - RNA-Seq
  - GRO-seq
  - Whole genome bisulfite sequencing (WGBS)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Filtering and Normalization of HiC Reads
  - Analysis for ATAC-Seq
  - ChIP-Seq
  - Analysis of CTCF direction and significant interaction
  - Analysis of SMC3 Occupancy in DN2 Cells Exposed to Actinomycin D
  - RNA-Seq
  - Analysis of GRO-seq
  - Analysis of WGBS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information includes seven figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.09.001.

AUTHORS CONTRIBUTIONS

T.I. and C.M. designed experiments. T.I. performed the majority of the experiments. A.J.M. performed HiC. V.C. identified ThymoD. Z.H., M.D., K.M.F., and A.N.C. performed bioinformatics. J.P.v.H. characterized ThymoD isoforms. S.P.F. and D.L.W. provided cells. T.I. and C.M. wrote the manuscript. C.M. conceived and supervised the study.

ACKNOWLEDGMENTS

We thank Alex Bortnick for editing the manuscript. This study was supported by funding from the Center for Computational Biology & Bioinformatics (UL1TRR001442), the California Institute for Regenerative Medicine (RB5-07025), and the NIH (AI102853, AI00880, AI082850, and AI09599) to C.M. D.L.W. was supported by the NIH (AI102853). T.I. was supported by the Uehara Memorial Foundation.

Received: January 30, 2017
Revised: June 24, 2017
Accepted: August 30, 2017
Published: September 21, 2017

REFERENCES

Bain, G., and Murre, C. (1998). The role of E-proteins in B- and T-lymphocyte development. Semin. Immunol. *10*, 143–153.

Benner, C., Isoda, T., and Murre, C. (2015). New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. Proc. Natl. Acad. Sci. USA *112*, 12776–12781.

Bhardwaj, S., Schlackow, M., Rabajdova, M., and Gullerova, M. (2016). Transcription facilitates sister chromatid cohesion on chromosomal arms. Nucleic Acids Res. *44*, 6676–6692.

Bossen, C., Murre, C.S., Chang, A.N., Mansson, R., Rodewald, H.R., and Murre, C. (2015). The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. Nat. Immunol. *16*, 775–784.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., and Peters, J.M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature *544*, 503–507.

Carleton, M., Ruetsch, N.R., Berger, M.A., Rhodes, M., Kaptik, S., and Wiest, D.L. (1999). Signals transduced by CD3epsilon, but not by surface pre-TCR complexes, are able to induce maturation of an early thymic lymphoma in vitro. J. Immunol. *163*, 2576–2585.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

Farley, E.K., Olson, K.M., and Levine, M.S. (2015). Regulatory principles governing tissue specificity of developmental enhancers. Cold Spring Harb. Symp. Quant. Biol. *80*, 27–32.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. Cell Rep. *15*, 2038–2049.

García-Ojeda, M.E., Klein Wolterink, R.G., Lemaître, F., Richard-Le, Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T cell specification by repressing B cell potential in pro-T cells in mice. Blood *121*, 1749–1759.

Guo, Y., Maillard, I., Chakraborti, S., Rothenberg, E.V., and Speck, N.A. (2008). Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T-cell specification. Blood *112*, 480–492.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Gutierrez, A., Kentsis, A., Sanda, T., Holmfeldt, L., Chen, S.C., Zhang, J., Protopopov, A., Chin, L., Dahlberg, S.E., Neuberg, D.S., et al. (2011). The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood *118*, 4169–4173.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23.

Ikawa, T., Kawamoto, H., Goldrath, A.W., and Murre, C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. J. Exp. Med. *203*, 1329–1342.

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science *329*, 93–96.

Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., et al. (2015). Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat. Genet. *47*, 338–344.

Kamimura, K., Ohi, H., Kubota, T., Okazuka, K., Yoshikai, Y., Wakabayashi, Y., Aoyagi, Y., Mishima, Y., and Kominami, R. (2007). Haploinsufficiency of Bcl11b for suppression of lymphomagenesis and thymocyte development. Biochem. Biophys. Res. Commun. *355*, 538–542.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat. Rev. Immunol. *14*, 377–391.

Kueh, H.Y., Yui, M.A., Ng, K.K., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., and Rothenberg, E.V. (2016). Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. Nat. Immunol. *17*, 956–965.

Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329*, 89–93.

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.C., Lu, D., Campos, L., Goulding, D., Ng, B.L., et al. (2010b). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science *329*, 85–89.

Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. Blood *122*, 902–911.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. *13*, 1196–1204.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, P., Li, P., and Burke, S. (2010). Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. Immunol. Rev. *238*, 138–149.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., and Rothenberg, E.V. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proc. Natl. Acad. Sci. USA *114*, 5800–5807.

Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell *158*, 339–352.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a non-coding intergenic element. Science *346*, 1373–1377.

Mansson, R., Welinder, E., Åhsberg, J., Lin, Y.C., Benner, C., Glass, C.K., Lucas, J.S., Sigvardsson, M., and Murre, C. (2012). Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. Proc. Natl. Acad. Sci. USA *109*, 21028–21033.

Melé, M., and Rinn, J.L. (2016). "Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription. Mol. Cell *62*, 657–664.

Miyazaki, K., Miyazaki, M., and Murre, C. (2014). The establishment of B versus T cell identity. Trends Immunol. *35*, 205–210.

Miyazaki, M., Miyazaki, K., Chen, K., Jin, Y., Turner, J., Moore, A.J., Saito, R., Yoshida, K., Ogawa, S., Rodewald, H.R., et al. (2017). The E-Id protein axis specifies adaptive lymphoid cell identity and suppressed thymic innate lymphoid cell development. Immunity *46*, 818–834.e4.

Naito, T., Tanaka, H., Naoe, Y., and Taniuchi, I. (2011). Transcriptional control of T-cell development. Int. Immunol. *23*, 661–668.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu. Rev. Genet. *35*, 673–745.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rothenberg, E.V., Ungerbäck, J., and Champhekar, A. (2016). Forging T-lymphocyte identity: Intersecting networks of transcriptional control. Adv. Immunol. *129*, 109–174.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, 6456–6465.

Schmitt, T.M., and Zúñiga-Pflücker, J.C. (2002). Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. Immunity *17*, 749–756.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctinos with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tsagaratou, A., González-Avalos, E., Rautio, S., Scott-Browne, J.P., Togher, S., Pastor, W.A., Rothenberg, E.V., Chavez, L., Lähdesmäki, H., and Rao, A. (2017). TET proteins regulate the lineage specification and TCR-mediated expansion of iNKT cells. Nat. Immunol. *18*, 45–53.

van Steensel, B., and Belmont, A.S. (2017). Lamina-associated domains: Links with chromosome architecture, heterochromatin, and gene repression. Cell *169*, 780–791.

Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature *476*, 63–68.

Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., 4th, Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., and Aster, J.C. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science *306*, 269–271.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. *14*, 529–545.

Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467–482.

Cell

Article

Structure of an Intron Lariat Spliceosome from *Saccharomyces cerevisiae*

Ruixue Wan,[1,4] Chuangye Yan,[1,4] Rui Bai,[1,4] Jianlin Lei,[1,2] and Yigong Shi[1,3,5,*]

[1]Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China

[2]Technology Center for Protein Sciences, Ministry of Education Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

[3]Institute of Biology, Westlake Institute for Advanced Study, Westlake University, Shilongshan Road No. 18, Xihu District, Hangzhou 310064, Zhejiang Province, China

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: shi-lab@tsinghua.edu.cn

http://dx.doi.org/10.1016/j.cell.2017.08.029

SUMMARY

The disassembly of the intron lariat spliceosome (ILS) marks the end of a splicing cycle. Here we report a cryoelectron microscopy structure of the ILS complex from *Saccharomyces cerevisiae* at an average resolution of 3.5 Å. The intron lariat remains bound in the spliceosome whereas the ligated exon is already dissociated. The step II splicing factors Prp17 and Prp18, along with Cwc21 and Cwc22 that stabilize the 5′ exon binding to loop I of U5 small nuclear RNA (snRNA), have been released from the active site assembly. The DEAH family ATPase/helicase Prp43 binds Syf1 at the periphery of the spliceosome, with its RNA-binding site close to the 3′ end of U6 snRNA. The C-terminal domain of Ntr1/Spp382 associates with the GTPase Snu114, and Ntr2 is anchored to Prp8 while interacting with the superhelical domain of Ntr1. These structural features suggest a plausible mechanism for the disassembly of the ILS complex.

INTRODUCTION

The removal of an intron requires two sequential S_N2-type transesterification reactions. The step I reaction generates a free 5′ exon and an intron lariat-3′ exon intermediate, and step II results in the ligation of the two flanking exons and production of a free intron lariat (Grabowski et al., 1984; Padgett et al., 1984; Ruskin et al., 1984). Pre-mRNA splicing is executed by the spliceosome, an exceptionally dynamic supramolecular ribonucleoprotein complex (Will and Lührmann, 2011). During each splicing cycle, the assembled spliceosome assumes at least seven consecutive states: pre-catalytic (B), activated (B^{act}), step I catalytically activated (B*), catalytic step I (C), step II catalytically activated (C*), post-catalytic (P), and the intron lariat spliceosome (ILS). The step I and II reactions proceed in the B* and C* complexes, resulting in the C and P complexes, respectively. The B-to-B^{act}, B^{act}-to-B*, and C-to-C* transitions are driven by the RNA-dependent ATPases/helicases Brr2, Prp2, and Prp16, respectively (Cordin et al., 2012; Jankowsky, 2011). The ligated exon in the P complex is released by the DEAH-family ATPase/helicase Prp22 (Company et al., 1991; Schwer, 2008; Schwer and Gross, 1998; Wagner et al., 1998), resulting in the ILS complex. The intron lariat in the ILS complex is released through the action of the ATPase/helicase Prp43 (Arenas and Abelson, 1997; Fourmann et al., 2013; Martin et al., 2002; Tsai et al., 2005).

The first atomic structure of an intact spliceosome, determined for the endogenous ILS complex from *S. pombe* at 3.6 Å resolution, reveals the spatial configuration of the splicing active site and the intricate organization of 37 proteins and four RNA elements (Hang et al., 2015; Yan et al., 2015). The spliceosome is proven to be a protein-directed metalloribozyme, with the catalytic magnesium (Mg^{2+}) ions coordinated by conserved nucleotides in U6 small nuclear RNA (snRNA) (Fica et al., 2013; Gordon et al., 2000; Hang et al., 2015; Sontheimer et al., 1997; Steitz and Steitz, 1993; Yan et al., 2015; Yean et al., 2000). The intramolecular stem loop (ISL) of U6 snRNA, loop I of U5 snRNA, helix I of the U2/U6 duplex, and the associated metal ions constitute the splicing active site, and they are accommodated by a positively charged cavity in the central spliceosomal component Prp8 (Hang et al., 2015; Yan et al., 2015). Remarkably, nearly all key features initially observed in the *S. pombe* ILS complex are conserved in the *S. cerevisiae* B^{act} (Rauhut et al., 2016; Yan et al., 2016), C (Galej et al., 2016; Wan et al., 2016), and C* (Fica et al., 2017; Yan et al., 2017) complexes. These features include the active site configuration, the conformations of at least 20 core components, and the overall organization of the spliceosome (Shi, 2017).

Disassembly of the ILS complex marks the end of a splicing cycle. The disassembly of the ILS complex and other spliceosomal intermediates requires the DEAH family ATPase/helicase Prp43 and/or the Ntr complex (Boon et al., 2006; Chen et al., 2013a; Koodathingal et al., 2010; Mayas et al., 2010; Pandit et al., 2006; Tanaka et al., 2007; Tsai et al., 2005, 2007). The Ntr complex in *S. cerevisiae* comprises Prp43 and two associated proteins Ntr1/Spp382 and Ntr2. In the presence of ATP or other NTP, the purified Ntr complex efficiently disassembled

CrossMark

the ILS complex into the free intron lariat, U6 snRNA, U2 small nuclear ribonucleoprotein (snRNP), U5 snRNP, and other discrete components in an in vitro assay (Fourmann et al., 2013). Notably, Prp43 fused to the Gly-rich motif (G-patch motif) of Ntr1/Spp382 disassembled the ILS complex as efficiently as the entire Ntr complex (Fourmann et al., 2016). Although the ATPase/helicase Brr2 and its activating protein Snu114 were thought to play an important role in the disassembly of the ILS complex (Hahn and Beggs, 2010; Small et al., 2006; Tsai et al., 2007), recent biochemical analyses strongly suggest otherwise (Fourmann et al., 2013).

In the cryoelectron microscopy (cryo-EM) structure of the *S. pombe* ILS complex (Hang et al., 2015; Yan et al., 2015), all three components of the Ntr complex are ostensibly absent. Consequently, the structural mechanism for the disassembly of the ILS complex remains largely unknown. In addition, some key features of the *S. pombe* ILS complex may be different from those of the *S. cerevisiae* ILS complex. For example, Cwf19 interacts with Spp42 and the intron lariat close to the active site of the *S. pombe* ILS complex (Hang et al., 2015; Yan et al., 2015); Cwf19 has close homologs in mammals, but it only shares sequence homology with the C-terminal CwfJ domains of the *S. cerevisiae* protein Drn1 (Garrey et al., 2014).

In this paper, we report the cryo-EM structure of the *S. cerevisiae* ILS complex at an average resolution of 3.5 Å, which allows identification of the disassembly proteins Prp43, Ntr1/Spp382, and Ntr2 and the splicing factor Cwc23. Analysis of the structural features reveals important insights into the disassembly mechanism of the ILS complex. This structure, together with those reported earlier and that of the B complex (Plaschka et al., 2017), offers a comprehensive picture on pre-mRNA splicing in *S. cerevisiae*.

RESULTS

Overall Structure

Using an affinity-tagged approach (Yan et al., 2016), a population of distinct endogenous spliceosomal complexes was purified from *S. cerevisiae*. The final dataset of 761,767 particles gave rise to cryo-EM structures of the B^{act} complex at 3.5-Å (Yan et al., 2016), the C complex at 3.4-Å (Wan et al., 2016), and the C* complex at 4.0-Å resolution (Yan et al., 2017). Using a similar 3D classification strategy (Yan et al., 2016), we picked 163,051 particles from this dataset that may represent the ILS complex. These particles only allowed structure determination of the ILS complex at a moderate resolution. To improve the data quality, we prepared a new batch of the spliceosomal complex using affinity-tagged Yju2 (Figure S1). A total of 4,375 micrographs were collected on the new cryo-EM sample, generating 139,470 particles for the ILS complex (Figure S2A). These particles were combined with the earlier ILS particles. Following 3D classifications, a subset of 150,363 particles yielded a reconstruction of the ILS complex at 3.5 Å. An additional step of 3D classification further improved the EM density map (Figure S2A).

The final cryo-EM structure of the *S. cerevisiae* ILS complex, derived from 92,108 particles, displayed an average resolution of 3.5 Å on the basis of the Fourier shell correlation (FSC) value of 0.143 (Figures S2B–S2E and S3; Tables S1 and S2). The local resolutions of the EM density maps reached 2.9–3.5 Å in the center of the ILS complex (Figure S2C; Table S2), which allowed convenient atomic modeling of the spliceosome. The final atomic structure of the *S. cerevisiae* ILS complex contained 36 spliceosomal proteins, three snRNA molecules (U2, U5, and U6), and an intron lariat, which amount to 9,329 amino acids and 344 nt (Figure 1A; Tables S1 and S2). Similar to other spliceosomal complexes (Bertram et al., 2017; Fica et al., 2017; Galej et al., 2016; Hang et al., 2015; Plaschka et al., 2017; Rauhut et al., 2016; Wan et al., 2016; Yan et al., 2015, 2016, 2017; Zhang et al., 2017), the ILS complex adopted a highly asymmetric architecture (Figure 1A).

The overall structure and organizational features of the *S. cerevisiae* ILS complex were very similar to those of the *S. pombe* ILS complex (Figure 1B). These similarities included the conformation of the protein components in the spliceosomal core and the RNA map, which refers to the spatial placement of and interactions among the snRNAs and pre-mRNA. Details of the similarities are discussed subsequently. Importantly, however, four new proteins were identified in the structure of the *S. cerevisiae* ILS complex (Figure S3). The *S. cerevisiae* ILS complex, but not the *S. pombe* complex, contained all three components of the Ntr complex: the ATPase/helicase Prp43 and its associated proteins Ntr1/Spp382 (hereafter referred to as Ntr1) and Ntr2. In addition, the *S. cerevisiae* complex, but not the *S. pombe* complex, contained the splicing factor Cwc23. The *S. pombe* complex, on the other hand, contained four proteins: Cwf11, Cwf19, Cwf17, and Cyp1. Except Cwf19, the other three proteins lack homologs in *S. cerevisiae*. Cwf19 is a paralog of the C-terminal CwfJ domains of the *S. cerevisiae* protein Drn1 (Garrey et al., 2014).

In the center of the *S. cerevisiae* ILS complex, Prp8, Snu114, three Nineteen Complex (NTC) proteins (Cef1, Clf1, and Syf2), and six NTC-related (NTR) components (Cwc2, Cwc15, Bud31, Ecm2, Prp45, and Prp46) interacted with one another and with the RNA elements at the active site (Figure 1C). These features are similar to those in the C* complex (Fica et al., 2017; Yan et al., 2017). However, the exon sequences were dissociated from loop I of U5 snRNA; the splicing factors Cwc21, Cwc22, Prp17, and Prp18 were displaced from the active site of the ILS complex (Figure 1C). In addition, the ATPase/helicase Prp22 that interacted with the Linker and RT Fingers/Palm of Prp8 in the C* complex (Fica et al., 2017; Yan et al., 2017) was now absent in the ILS complex.

The RNA Map and the Active Site

The three snRNA molecules and the intron lariat are unambiguously identified by the EM density maps. Aided by knowledge of the C* complex (Fica et al., 2017; Yan et al., 2017), we modeled 82, 117, and 101 nt for U2, U5, and U6 snRNAs, respectively (Figure 2A; Table S2). We also built 36 nt for the intron lariat. Similar to the C* complex (Fica et al., 2017; Yan et al., 2017), the 5′ and 3′ ends of the three snRNAs were positioned away from the active site, and the lariat junction was located away from the catalytic metal ions by approximately 20 Å (Figures 2A and 2B). Despite dissociation of the ligated exon, the splicing active site remained largely intact and was nearly identical to that in the C* complex (Fica et al., 2017; Yan et al., 2017) (Figures 2B and 2C). Different

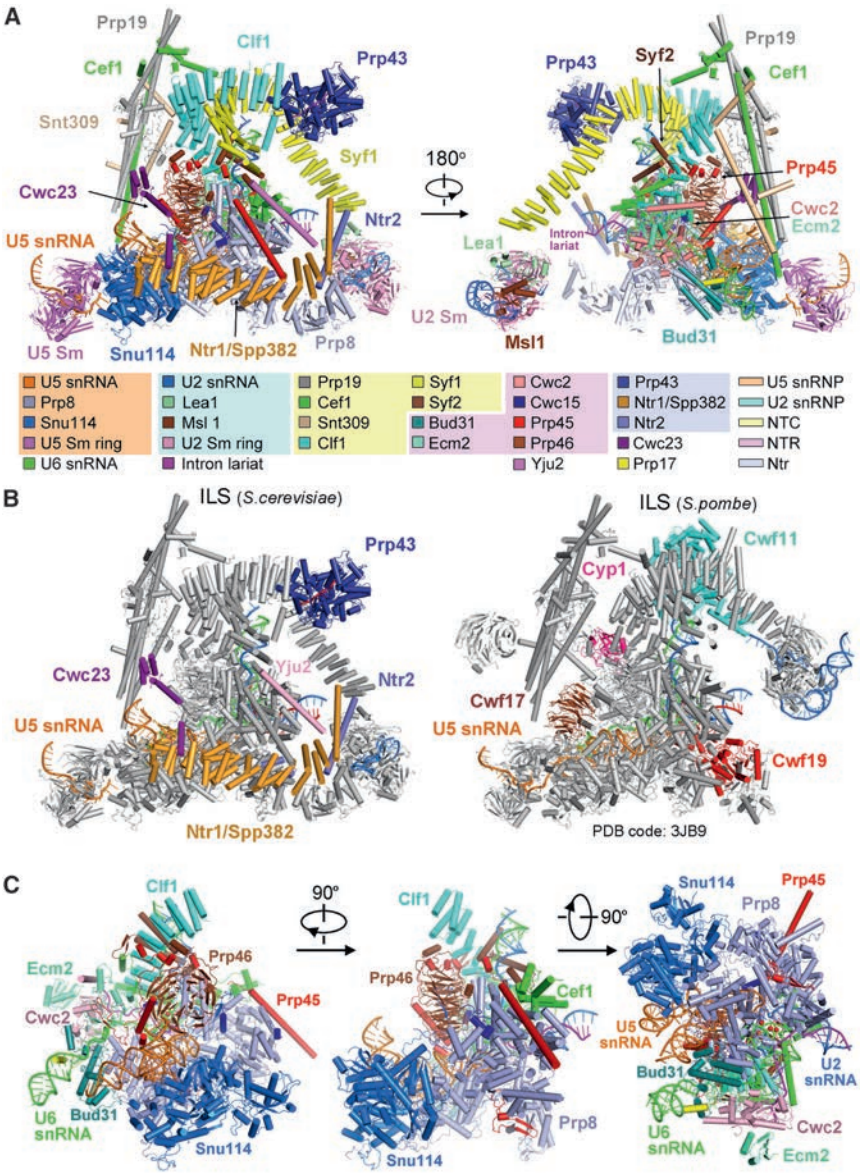


Figure 1. Structure of an Intron Lariat Spliceosome (ILS) from *Saccharomyces cerevisiae* (*S. cerevisiae*)

(A) Overall structure of the *S. cerevisiae* ILS complex at an average resolution of 3.5 Å. The front and back views are shown. The individual spliceosomal components are color coded and shown below the structure. The structure displayed here includes 36 proteins, three snRNAs (U2, U5, and U6), and an intron lariat. The ligated exon has been dissociated from the ILS complex.

(B) Comparison of the overall structure between the two *S. cerevisiae* and *S. pombe* ILS complexes. Compared to the *S. pombe* ILS complex (Hang et al., 2015; Yan et al., 2015) (right panel), the *S. cerevisiae* complex (left panel) contains four more proteins that coordinate with each other to disassemble the spliceosome. These proteins are Prp43, Ntr1/Spp382, Ntr2, and Cwc23. But the *S. pombe* ILS complex also contains four proteins that are absent in the *S. cerevisiae* complex: Cwf11, Cyp1, Cwf17, and Cwf19.

(C) Protein components around the active site. The conformation of the RNA elements at the active site is supported by Prp8, Snu114, three Nineteen Complex (NTC) proteins (Cef1, Clf1, and Syf2), and six NTC-related (NTR) components (Cwc2, Cwc15, Bud31, Ecm2, Prp45, and Prp46). Three mutually perpendicular views are shown. All structural images were generated in PyMOL (DeLano, 2002).

from that in the C* complex (Fica et al., 2017; Yan et al., 2017), the loop I of U5 snRNA was unengaged in the ILS complex.

Comparison of the RNA elements between the C* and ILS complexes revealed nearly identical conformations for U5 and U6 snRNAs and nucleotides 1–30 of U2 snRNA (Figures 2C and 2D). Importantly, the location and conformation of the intron lariat remained largely unchanged between the C* and ILS complexes (Figure 2C), which strongly suggests a similar location and conformation of the intron lariat in the P complex. The conformation of U5 loop I and the positions of associated Mg^{2+} ions remained indistinguishable between the C* and ILS complexes (Figure S4). As previously noted (Yan et al., 2017), however, the bulk of U2 snRNA sequences were translocated by distances of approximately 30–80 Å from the C* to the ILS complex (Figure 2D). The movement of U2 snRNA was likely triggered by the action of the ATPase/helicase Prp22, which is thought to dissociate the ligated exon from loop I of U5 snRNA by pulling the 3′ end sequences of the exon (McPheeters et al., 2000; Schwer, 2008; Schwer and Gross, 1998; Semlow et al., 2016).

The ATPase/Helicase Prp43

All three components of the Ntr complex have been identified in the *S. cerevisiae* ILS complex (Figure 3A; Figure S3). At the periphery of the spliceosome, a characteristic lobe of the EM density is closely associated with the middle portion of the superhelical protein Syf1. Prp43 is likely the only remaining DEAH family ATPase/helicase in the ILS complex; the crystal structure of the *S. cerevisiae* Prp43 bound to an RNA element (PDB: 5I8Q) can be nicely docked into this density, which only required minor adjustment of the two RecA domains (Figure 3B). Prp43 sequentially consists of six domains: N domain, RecA1, RecA2, Winged-helix domain (WH), Ratchet-like domain, and an oligonucleotide/oligosaccharide-binding fold (OB). All six domains are identified in the EM density map. Following assignment of the secondary structural elements, some weak EM density remained in the RNA-binding tunnel of Prp43, suggesting the presence of RNA (Figure S3M). Therefore, Prp43 in our structure likely represents the substrate-bound, nucleotide-bound state. Because of the relatively low resolution

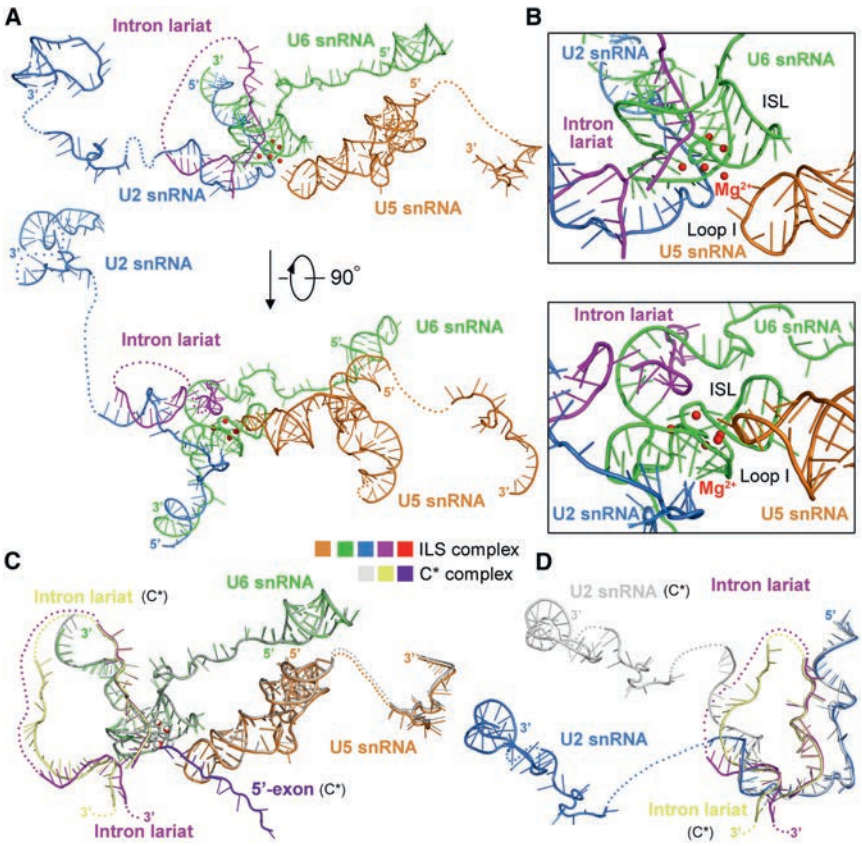


Figure 2. The RNA Map and the Splicing Catalytic Center of the *S. cerevisiae* ILS Complex

(A) The overall RNA map. The U2, U5, and U6 snRNAs are colored blue, orange, and green, respectively. The intron lariat is highlighted in magenta and the metal ions are shown in red spheres. The disordered nucleotide sequences are represented by dotted lines. Two perpendicular views are shown.

(B) Structure of the splicing catalytic center of the ILS complex. The two perpendicular views shown here correspond to those in (A). The catalytic center comprises the intramolecular stem loop (ISL) of U6 snRNA, the helix I of the U2/U6 duplex, the loop I of U5 snRNA, and the associated Mg^{2+} ions. Notably, the intron lariat is already translocated away from the active site center.

(C) The U5 and U6 snRNAs and the intron lariat are structurally very similar between the C* (Yan et al., 2017) and ILS complexes from *S. cerevisiae*. Shown here is a structural overlay of the RNA maps of the C* and ILS complexes.

(D) The 30 consecutive nucleotides at the 5′ end of U2 snRNA remain structurally very similar between the C* and ILS complexes. In contrast, the sequences downstream of nucleotide 30 are highly mobile and adopt very different locations between the two complexes.

of the EM density around Prp43, we were unable to conclusively identify the bound nucleotide.

Compared to the *C. thermophilum* Prp43 bound to an RNA element (Tauchert et al., 2017), the two RecA domains in our structure moved farther apart from each other (Figure S5A), generating a pronounced gap (He et al., 2017) (Figure 3B). Structural comparison of Prp43 in three different conformational states revealed additional movements of the RecA domains that appeared to regulate RNA loading and/or translocation (Figure S5B). The RNA-binding tunnel on the RecA domains was closed in the ADP-bound Prp43, but it remained open in the ADP-BeF_3-bound Prp43 (Tauchert et al., 2017; Walbott et al., 2010), presumably because ADP-BeF_3 leads to the activation of Prp43. The RNA-binding tunnel was closed again in the ADP-BeF_3-bound Prp43 that was already loaded with an RNA element (Tauchert et al., 2017); this Prp43 was presumably ready to exert its helicase activity and translocate the single-stranded RNA element.

In all cases, the static portion of Prp43 was the C-lobe, which contained the WH, Ratchet-like, and OB domains, and the mobile portion was the N-lobe, which comprised the N domain, RecA1, and RecA2 (Figure S5B). This arrangement is perfectly suited for the translocation of the RNA element, whose phosphate backbone is predominantly bound by the two RecA domains. The movement of RecA1 relative to RecA2 is presumably powered by ATP binding and hydrolysis, which generates the pulling force of the DEAH family helicase activity (Semlow et al., 2016).

Ntr1 Is Anchored on Snu114

The 708-residue protein Ntr1 sequentially consisted of three distinct domains: a G-patch motif at the N terminus, a central superhelical domain, and a folded β sheet domain at the C terminus (CTD) (Figure 3C). The vast majority of the Ntr1 sequences were fitted into the EM density map due to good connectivity; the amino acid sequences of the CTD were fully assigned in our structure. In contrast to Prp43, Ntr1 spanned a long distance of approximately 150 Å and linked several key proteins that may play a role in the disassembly of the ILS complex (Figure 3A). The G-patch motif, conserved among eukaryotic RNA-processing proteins, plays an important role in pre-mRNA splicing through stimulation of the DEAH-box helicase activity (Aravind and Koonin, 1999; Silverman et al., 2004). The G-patch motif in Ntr1, which presumably interacts with Prp43 (Christian et al., 2014; Tanaka et al., 2007), remained to be identified in the EM density map due to the limited local resolution. The extended superhelical domain loosely contacted Ntr2, Prp45, and the step I splicing factor Yju2 (Figure 3D).

Most notably, the CTD was anchored to the cleft between domains III and IV of Snu114 (Figure 3E). The interaction between the Ntr1 CTD and Snu114 was further stabilized by an α helix of Cwc23 that directly associated with the CTD on top. Compared to other parts of Ntr1, the CTD appeared to play a major role in the association of Ntr1 to the ILS complex. Compared to the C* complex (Yan et al., 2017), domain IV of Snu114 and the RT Fingers/Palm domain of Prp8 in the ILS complex were translocated by about 4–10 Å relative to the well-aligned domain III of Snu114 (Figure 3F). Because the position

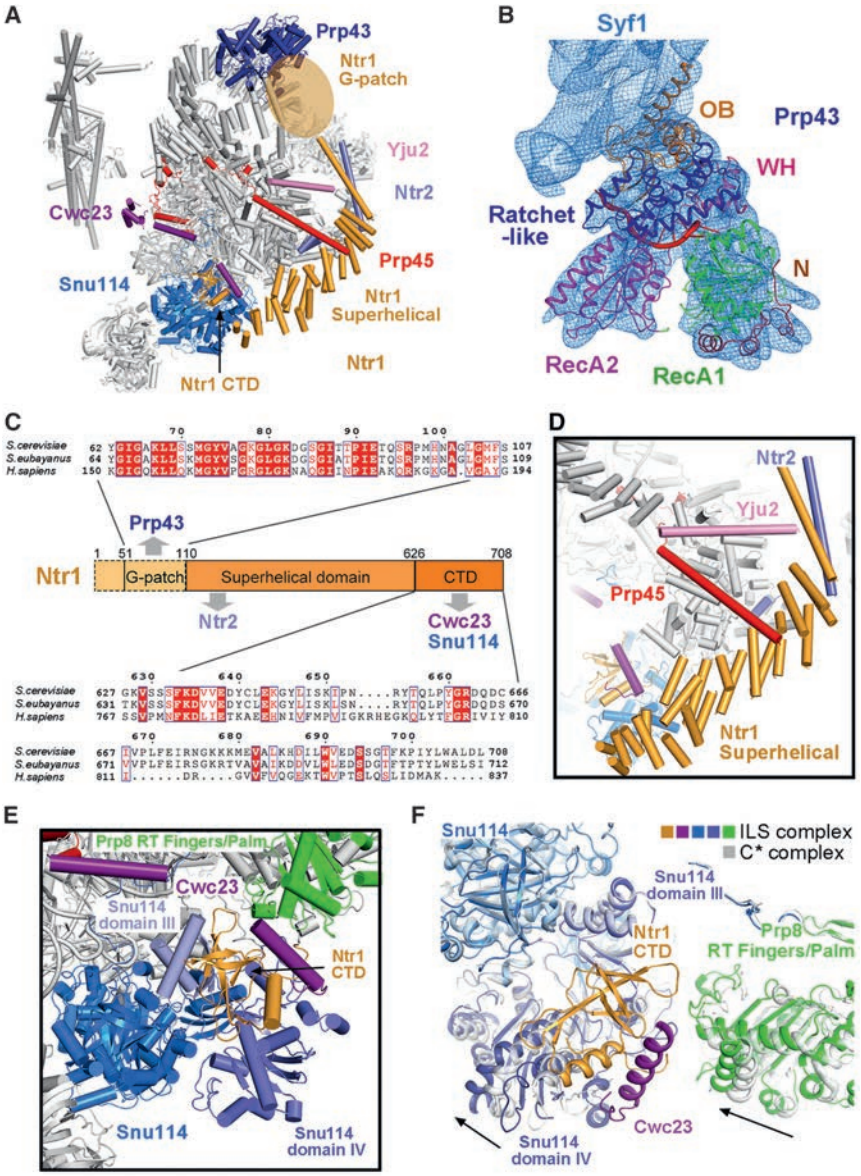


Figure 3. Structure of the ATPase/Helicase Prp43 and Its Cofactor Ntr1

(A) An overall view of the Ntr complex in the ILS complex. The G-patch motif at the N terminus of Ntr1, which remains largely unidentified in the current structure, interacts with Prp43 (colored dark blue). The superhelical domain of Ntr1 binds Ntr2 (light blue) and Prp45 (red). The CTD of Ntr1 associates with Cwc23 (purple) and Snu114 (blue).

(B) Overall structure of the ATPase/helicase Prp43. The crystal structure of Prp43 bound to an RNA element (PDB: 5I8Q) was fitted into the local masked EM density map (blue mesh), with minor rigid-body adjustment made in COOT (Emsley and Cowtan, 2004). The six domains of Prp43 are color coded as follows: N domain (N, brown), RecA1 (green), RecA2 (magenta), Winged-helix (WH, hot pink), Ratchet-like (blue), and oligopeptide/oligo-saccharide-binding fold (OB, orange).

(C) Sequence alignment of the G-patch motif and the CTD of Ntr1. Ntr1 sequentially consists of three well-defined domains: a G-patch motif at the N terminus, an extended superhelical domain, and a CTD. The amino acid sequences of Ntr1 from two yeast species (*S. cerevisiae* and *S. eubayanus*) and *Homo sapiens* are aligned for the G-patch and the CTD. The invariant residues (shaded red) and the conserved residues (colored red) are boxed.

(D) A close-up view on the superhelical domain of Ntr1. Notably, portions of Ntr1, Ntr2, Prp45, and Yju2 interact with one another in this region.

(E) Ntr1 is anchored to the ILS complex through the interactions of its CTD with Snu114. The CTD of Ntr1 closely interacts with both domain III and domain IV of Snu114. An α helix from Cwc23 binds atop the CTD, stabilizing its interaction with Snu114.

(F) Compared to the C* complex (Yan et al., 2017), domain IV of Snu114 and the RT Fingers/Palm domain of Prp8 undergo a lateral translation of about 4–10 Å. Because domain III of Snu114 remains unchanged, this movement is likely coupled to the binding of Ntr1.

of domain III remained unchanged between the C* and ILS complexes, this translocation was likely triggered by the binding of Ntr1.

Ntr2 Is Anchored on Prp8

Although only a small portion of the Ntr2 sequences were assigned in our structure, they appeared to play a key role by anchoring Ntr2 to the ILS complex. These sequences formed two consecutive α helices and a short intervening loop (hereafter referred to as the helix-loop-helix motif), which functioned like a tweezer to closely interact with the Linker domain of Prp8 (Figures 4A and 4B). Notably, some of the highly conserved residues of Ntr2 exemplified by Trp242 were located at the interface and directly contacted residues of the Linker domain (Figure 4A). This finding is consistent with the biochemical observation that Ntr2 can be independently recruited to Prp8 (Boon et al., 2006). Intriguingly, the helix-loop-helix motif of Ntr2 in the ILS complex occupied the same general location as elements of Prp45 and Hsh155 in the B^{act} complex (Yan et al., 2016) (Figure 4C, left panel). This analysis suggests that it is unlikely for Ntr2 to be recruited into the B^{act} complex. A fusion protein between Prp43 and the G-patch motif of Ntr1, but not the Ntr complex, was able to disassemble the B^{act} complex (Fourmann et al., 2016). This observation can be explained by the scenario whereby the recruitment of Ntr2 is a pre-requisite for efficient recruitment of Ntr1, which then activates Prp43 through its G-patch motif. Similar to the B^{act} complex, the potential binding site for the helix-loop-helix motif of Ntr2 in the C* complex was partially blocked by the ATPase/helicase Prp22 (Yan et al., 2017; Zhang et al., 2017) (Figure 4C, right panel). In contrast, this site in the C complex remained unoccupied (Galej et al., 2016; Wan et al., 2016) (Figure 4C, middle panel). This analysis

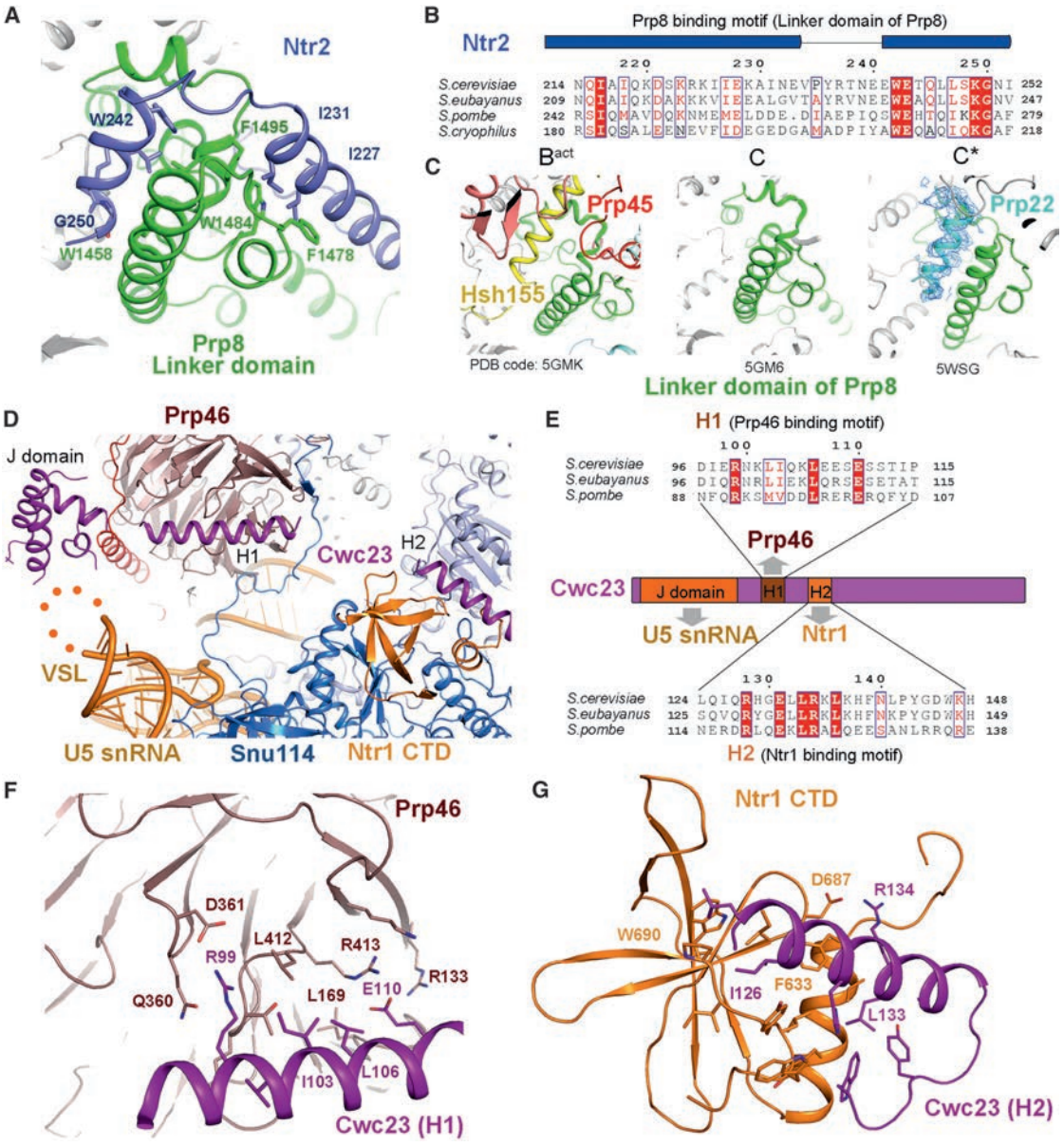


Figure 4. The Disassembly Protein Ntr2 and the Splicing Factor Cwc23

(A) Ntr2 is anchored to the ILS complex through the interactions between the Prp8-binding motif of Ntr2 and the Linker domain of Prp8. The Prp8-binding motif of Ntr2 comprises two consecutive α helices and an intervening sequence, constituting a helix-loop-helix motif. Key residues at the interface are labeled.

(B) Sequence alignment of the helix-loop-helix motif of Ntr2 from four yeast species. The invariant residues are shaded red and the conserved residues are boxed. These residues directly interact with Prp8 at the interface.

(C) Ntr2 may be responsible for the selection of specific spliceosomal complexes for disassembly. Ntr2 can potentially associate with the C complex, but not the B^{act} or C* complexes. In contrast to the C complex (middle panel) where the Ntr2-binding site of Prp8 remains free, it is occupied by elements of Prp45 and Hsh155 in the B^{act} complex (left panel) and by a short N-terminal helix of Prp22 in the C* complex (right panel). The EM density map for the short helix of Prp22 in the yeast C* complex is shown here (EMDB: EMD-6684; Yan et al., 2017), and it was modeled after the human C* complex (PDB: 5XJC; Zhang et al., 2017).

(D) A close-up view on the structure of the splicing factor Cwc23. Only the N-terminal half of Cwc23 is identified in our structure. It sequentially consists of an N-terminal J domain and two α helices H1 and H2. The J domain likely interacts with the variable stem loop (VSL) of U5 snRNA. The H1 helix binds the WD40 repeats of Prp46, whereas the H2 helix stacks against the CTD of Ntr1.

(E) Sequence alignment of the H1 and H2 helices of Cwc23 from three yeast species. The invariant residues are shaded red and the conserved residues are boxed. These residues play an important role at the interfaces with Prp46 and Ntr1, as shown in (F) and (G), respectively.

(F) A close-up view on the interactions between the H1 helix of Cwc23 and the WD40 repeats of Prp46.

(G) A close-up view on the interactions between the H2 helix of Cwc23 and the CTD of Ntr1.

suggests that the Ntr complex has the potential to disassemble the C complex. In addition, biochemical studies have demonstrated that the Ntr complex can be recruited to mediate the disassembly of the spliceosomal intermediates, linking spliceosome disassembly to proofreading (Chen et al., 2013a; Koodathingal et al., 2010; Mayas et al., 2010).

The Splicing Factor Cwc23

We were also able to model several discrete sequence elements of the splicing factor Cwc23, which might be important for pre-mRNA splicing but remains poorly characterized (Pandit et al., 2009). The N-terminal half of Cwc23 contained a DnaJ/Hsp40 motif (J domain) followed by two extended α helices H1 and H2 (Figures 4D and 4E). The J domain, which appeared to play a role in spliceosome disassembly, but not for pre-mRNA splicing (Sahi et al., 2010), was located in the vicinity of the disordered variable stem loop (VSL) of U5 snRNA. The two helices H1 and H2 made specific interactions with the WD40 domain of Prp46 and the CTD of Ntr1, respectively (Figures 4F and 4G). Importantly, at both interfaces, the highly conserved residues of Cwc23 were directly involved in the inter-molecular interactions with specific residues from Prp46 and Ntr1. Deletion of these two helices crippled pre-mRNA splicing (Pandit et al., 2009). The interactions between Cwc23 and the CTD of Ntr1 further stabilized the recruitment of Ntr1 to the ILS complex through its interactions with Snu114.

Comparison between the C* and ILS Complexes

Although the overall structures of the C* and the ILS complexes are similar, there are important compositional and conformational differences. In the C* complex (Fica et al., 2017; Yan et al., 2017), the distinct active site conformation was maintained in part by four splicing factors: Cwc21, Cwc22, Prp17, and Prp18 (Figure 5A, left panel). In particular, an extended sequence from Cwc21 recognized the 5′ exon, and this interaction was stabilized by the Switch loop of Prp8. The WD40 domain of Prp17 was positioned between the intron/U6 and intron/U2 duplexes, whereas Prp18 interacted with the RNaseH-like domain of Prp8. In the ILS complex (Figure 5A, right panel), the exon sequences were dissociated from U5 snRNA through the action of Prp22, which likely caused the dissociation of Cwc21 and Cwc22 and a 180-degree flipping of the Switch loop (Figure S6A). The ATPase/helicase Prp22 and the WD40 domain of Prp17 and Prp18 in the C* complex were also absent in the ILS complex. An N-terminal fragment of Prp17 remained associated with the 5′ end sequences of U6 snRNA. Instead, the Ntr complex, along with the splicing factor Cwc23, was present in the ILS complex.

Structural overlay revealed steric clash between Ntr1/Ntr2 of the ILS complex and Prp22 of the C* complex (Figure 5B). Thus, the dissociation of Prp22 likely preceded the recruitment of the Ntr complex into the spliceosome. Compared to the C* complex (Fica et al., 2017; Yan et al., 2017), the endonuclease-like domain and the Linker domain were separated 5–7 Å away from the N domain of Prp8 in the ILS complex (Figure 5C). In particular, an extended loop in the Linker domain, which stabilized the bound 5′ exon in the C* complex, was translocated by approximately 6 Å in the ILS complex, leaving a gap between the Linker and the N domain of Prp8 (Figure 5D). This conformational change was likely coupled to the dissociation of the 5′ exon.

Despite these pronounced structural rearrangements, the bulk of Prp8—both the N domain and the Core—remained essentially unchanged between the C* and ILS complexes (Figure S6A, left panel). In contrast, the RNaseH-like domain of Prp8, which is known to associate with a mobile RNA element (Wan et al., 2016), underwent a marked positional movement. The β finger of the RNaseH-like domain, which directly interacted with the lariat junction in the C* complex (Fica et al., 2017; Yan et al., 2017), was positioned away from the lariat junction by about 30–40 Å and pointed away from the active site. The dissociation of the step II splicing factor Prp18 was likely a consequence of the translocation of the RNaseH-like domain. These structural changes may have perpetuated to Prp17, causing its dissociation from the intron/U2 and intron/U6 duplexes.

Comparison between the *S. cerevisiae* and *S. pombe* ILS Complexes

Advent of the ILS structure from *S. cerevisiae* allows detailed comparison with that from *S. pombe* (Figure 1B) (Hang et al., 2015; Yan et al., 2015). The core components, including the snRNAs and conserved protein components, adopted nearly identical conformations in these two ILS complexes. Both the N domain and the Core of Prp8 (including the Switch loop) were nearly indistinguishable between these two complexes (Figure S6A, right panel). The ISL and the locations of the associated Mg^{2+} ions in the *S. cerevisiae* complex were also very similar to those in the *S. pombe* complex, except that one extra structural metal ion was identified in the *S. cerevisiae* ILS complex (Figures S6B and S6C).

The protein components at the periphery, along with the associated RNA elements, had similar structures but different spatial positions in the two ILS complexes (Figure 1B). For example, the 3′ end sequences of U2 snRNA, together with Msl1, Lea1, and the heptameric Sm ring, were positioned 40–50 Å apart between these two complexes. Although the bulk of Prp8 remained structurally superimposable, the RNaseH-like domain adopted two different orientations relative to Prp8 Core in the two ILS complexes (Figure S6A, right panel).

The most prominent difference between the two ILS complexes was the presence of Prp43, Ntr1, Ntr2, and Cwc23 in the *S. cerevisiae* complex, but not in the *S. pombe* complex (Figure 1B). Notably, among these four *S. cerevisiae* proteins, the homolog of Ntr1 is yet to be identified in *S. pombe*. Structural overlay between the two ILS complexes revealed two distinct conformations of the superhelical protein Syf1/Cwf3: one specific for binding to Prp43 in the *S. cerevisiae* ILS complex and the other specific for interacting with Cwf11 in the *S. pombe* complex (Figure 6A). This analysis suggests that Cwf11 must be dissociated prior to the recruitment of Prp43 into the spliceosome. Similarly, Cwf19 in the *S. pombe* ILS complex sterically clashed with the helix-loop-helix motif of Ntr2 in the *S. cerevisiae* complex (Figures 6B and 6C), suggesting dissociation of Cwf19 prior to the recruitment of Ntr2. Taken together, these analyses strongly argue that the *S. pombe* complex and the *S. cerevisiae* complex may represent two different states of the spliceosome on its way to disassembly, with the

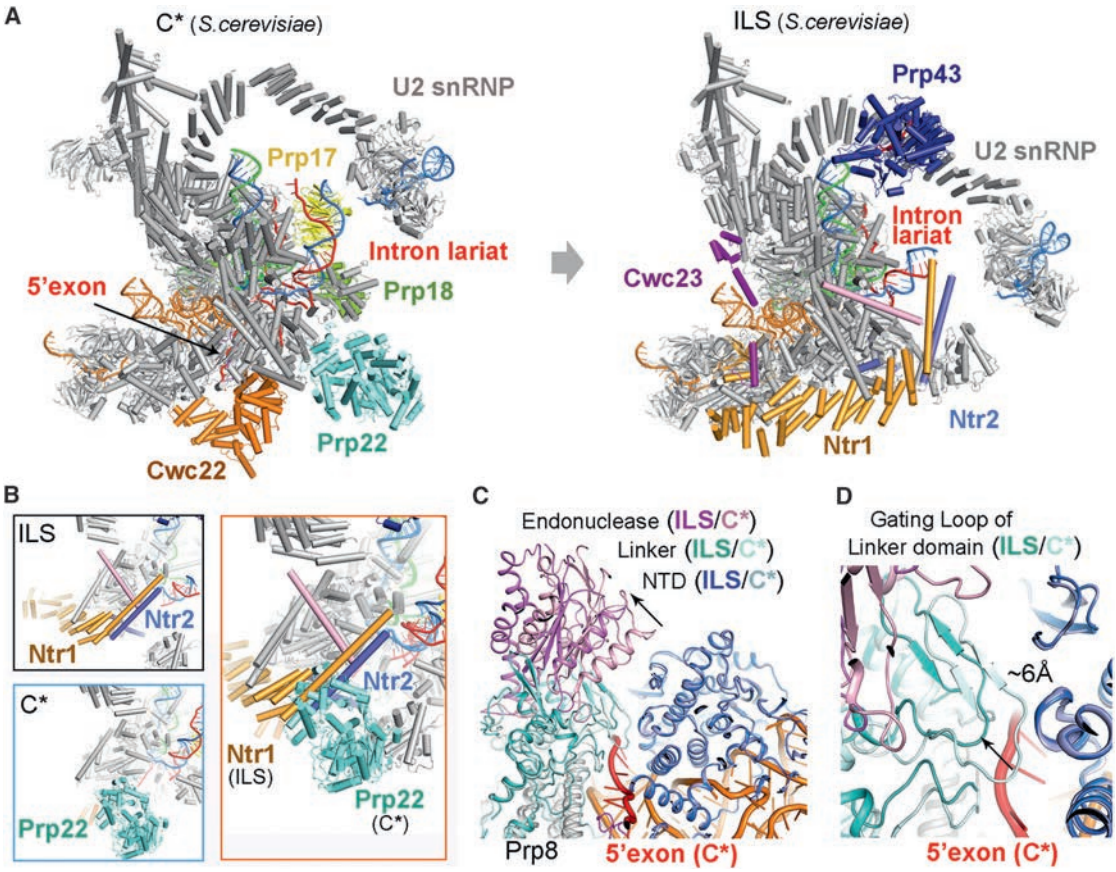


Figure 5. Comparison between the C* and the ILS Complexes from *S. cerevisiae*

(A) Comparison of the overall structure between the C* (Yan et al., 2017) and the ILS complexes. Compared to the C* complex (left), the ILS complex (right) no longer contains the exon; the splicing factors Cwc21, Cwc22, and Prp18; or the ATPase/helicase Prp22. Instead, four new proteins important for disassembly of the spliceosome—the ATPase/helicase Prp43, Ntr1, Ntr2, and Cwc23—have been loaded into the ILS complex.

(B) Recruitment of the disassembly factors into the ILS complex requires the departure of the ATPase/helicase Prp22. With the core components of the C* and ILS complexes aligned, Prp22 of the C* complex partially overlaps with Ntr1 and Ntr2 of the ILS complex (right panel). The local structure in the C* and ILS complexes is shown in the left panels.

(C) Major conformational rearrangement around the binding site of the 5′ exon between the C* (Yan et al., 2017) and ILS complexes. With the core components of the two complexes aligned, the endonuclease-like domain of Prp8 in the ILS complex (magenta) is shifted by about 10 Å relative to that in the C* complex (light magenta). This movement widens the cleft under which the 5′ exon is bound in the C* complex. The Linker domain of Prp8 undergoes a similar movement to allow exit of the 5′ exon. Details of the Linker domain movement are shown in (D).

(D) An extended loop of the Prp8 Linker domain (named the Gating loop) is shifted by about 6 Å between the C* and the ILS complexes.

S. cerevisiae ILS complex representing a very late spliceosome just prior to disassembly.

DISCUSSION

The ATPase/helicase Prp43, located at the periphery of the ILS complex (Figure 7A), is thought to pull the RNA sequences in the disassembly of the spliceosome (Arenas and Abelson, 1997; Fourmann et al., 2016; Martin et al., 2002). Consistent with previous biochemical studies (Bohnsack et al., 2009; Fourmann et al., 2016), the location of Prp43 suggests two distinct possibilities: pulling the intron lariat sequences or the 3′ end sequences of U6 snRNA. In the first scenario, the intron sequences between the 5′ splice site (5′SS) and the branch site (BS) would loop through Prp43 (Figure 7B; Figure S7A). Driven by ATP binding and hydrolysis, Prp43 would pull the single-stranded intron lariat sequences, which would result in the unwinding of the intron/U2 and intron/U6 duplexes and consequent dissociation of the NTC and NTR components around the catalytic center. This scenario is supported by results of direct cross-linking experiments (Fourmann et al., 2016). This mechanism requires a relatively long intron sequence of at least 70 nt between the 5′SS and the BS, which is easily met for most introns in *S. cerevisiae* (Bon et al., 2003; Kupfer et al., 2004; Spingola et al., 1999).

In the second scenario, the sequences at the 3′ end of U6 snRNA would be grabbed by Prp43 (Figure 7C; Figure S7B). Notably, the 5′ end of the bound RNA element within Prp43 is close to the 3′ end of U6 snRNA. The distance between nucleotide 99 of the 112-nt U6 snRNA and the nucleotide at the 5′ end of

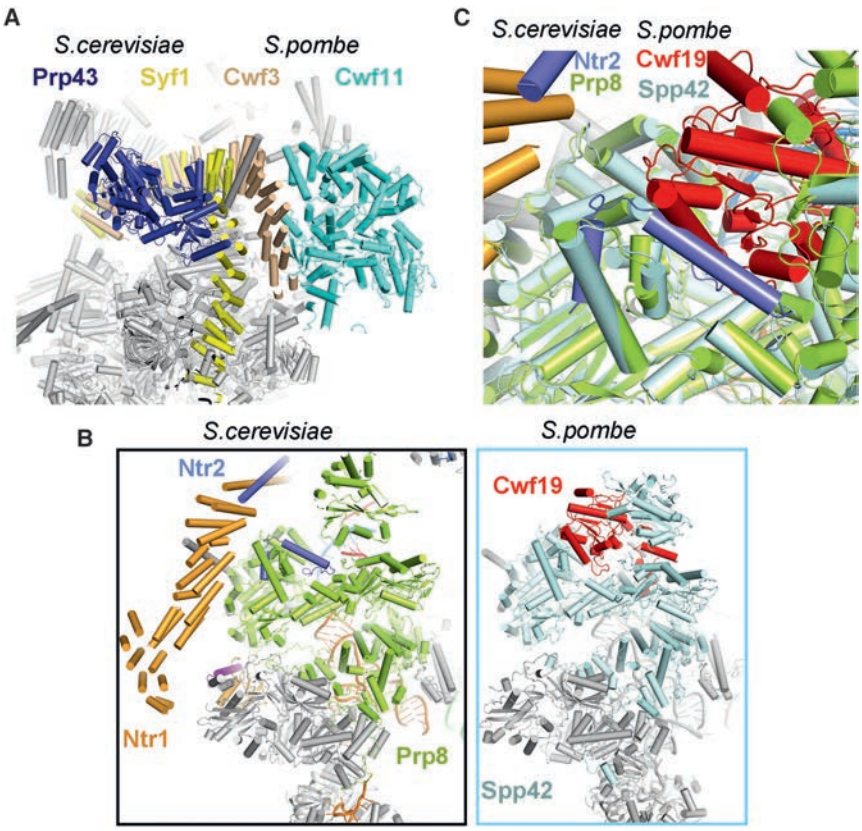


Figure 6. Comparison between the *S. cerevisiae* and *S. pombe* ILS Complexes

(A) Superposition of the two ILS complexes reveals incompatible binding conformation around *S. cerevisiae* Syf1 (*S. pombe* Cwf3). With the core components of the two ILS complexes aligned, Syf1 of *S. cerevisiae* exhibits a positional shift of up to 30 Å compared to Cwf3 of *S. pombe*. The conformational differences between Syf1 and Cwf3 only allow Prp43 to bind Syf1 and Cwf11 to interact with Cwf3, but not the other way around or simultaneous binding.

(B) Comparison of the two ILS complexes around *S. cerevisiae* Prp8 (*S. pombe* Spp42). Cwf19 closely interacts with Spp42 (right panel); the corresponding region in Prp8 is partially occupied by Ntr2 (left panel).

(C) Superposition of *S. cerevisiae* Prp8 with *S. pombe* Spp42 reveals steric clash between Ntr2 (light purple) and Cwf19 (red).

the bound 8-mer RNA element is only 25 Å (Figure S7B), which can be covered by four consecutive nucleotides. Thus, it is possible for Prp43 to bind the 3′ end sequences of U6 snRNA and pull in the 3′-to-5′ direction. Pulling the single-stranded 3′ end sequences of U6 snRNA would unwind the U2/U6 and intron/U6 duplexes and dissociate the affiliated protein components, resulting in the disassembly of the spliceosome. Notably, these two scenarios are not mutually exclusive of each other, and it is possible that both mechanisms are used in cells, depending on the length of the introns.

The recruitment of Ntr2 into the spliceosome likely relies on the interaction between the helix-loop-helix motif of Ntr2 and the Linker domain of Prp8. Similarly, the CTD of Ntr1 anchors Ntr1 to the ILS complex through interactions with Snu114, which are further strengthened by the splicing factor Cwc23. Compared to Ntr1 and Ntr2, Prp43 through its OB domain only appears to weakly associate with the superhelical protein Syf1. Ntr1 and Ntr2 are known to stably interact with each other (Tsai et al., 2007). Together, these structural findings suggest a working mechanism for the disassembly of the ILS complex. In this mechanism, both the recruitment of Prp43 to the ILS complex and the helicase activity of Prp43 are greatly facilitated by the G-patch motif of Ntr1. However, the recruitment of Ntr1 to the ILS complex is linked to the interaction with Ntr2 and the availability of the binding surface of Snu114. In this regard, Ntr2 and Ntr1 collectively control the specificity for the disassembly of the spliceosome. The recruitment of Ntr1 into the ILS complex is greatly strengthened by the presence of Ntr2, and only Ntr1-activated Prp43 has the ability to disassemble the spliceosome.

In addition to the Ntr complex, the ATPase/helicase Brr2 and its activating protein Snu114 may also play a role in the disassembly of the ILS complex (Small et al., 2006); but, this role is likely to be considerably less important than that of the Ntr complex (Fourmann et al., 2013, 2016). In the ILS structure, Prp43 is located more than 120 Å away from Snu114 (Figure 7D). Brr2, on the other hand, remains to be identified in our structure; its potential role in spliceosome disassembly is yet to be further investigated. Snu114, on the other hand, may play a role in the disassembly of the ILS complex through its interactions with Ntr1. Therefore, the dissociation of the ILS complex likely begins with the recruitment of Ntr2 and Ntr1, followed by Prp43 (Figure 7E). The association of the G-patch motif of Ntr1 with Prp43 triggers its NTPase activity, resulting in the disassembly of the ILS complex into free U6 snRNA, 20S U2 snRNP, 18S U5 snRNP, and individual NTC and NTR proteins (Fourmann et al., 2013) (Figure 7E).

The P complex, which contains the ligated exon stably associated with loop I of U5 snRNA, is preceded by the C* complex and followed by the ILS complex. The P-to-ILS transition is driven by the ATPase/helicase Prp22 and involves the dissociation of the ligated exon and several associated splicing factors; in contrast, the C*-to-P transition proceeds naturally and without the need of ATP hydrolysis. Therefore, the structural difference between the C* and the P complexes should be minimal: the covalent linkage between the 3′ end nucleotide of the intron lariat and the 5′-phosphate of the 3′ exon is replaced by that between the 3′ end nucleotide of the 5′ exon and the 5′-phosphate of the 3′ exon. Such a difference is likely to be accommodated by local structural adjustment. This analysis suggests that the structural differences between the C* and

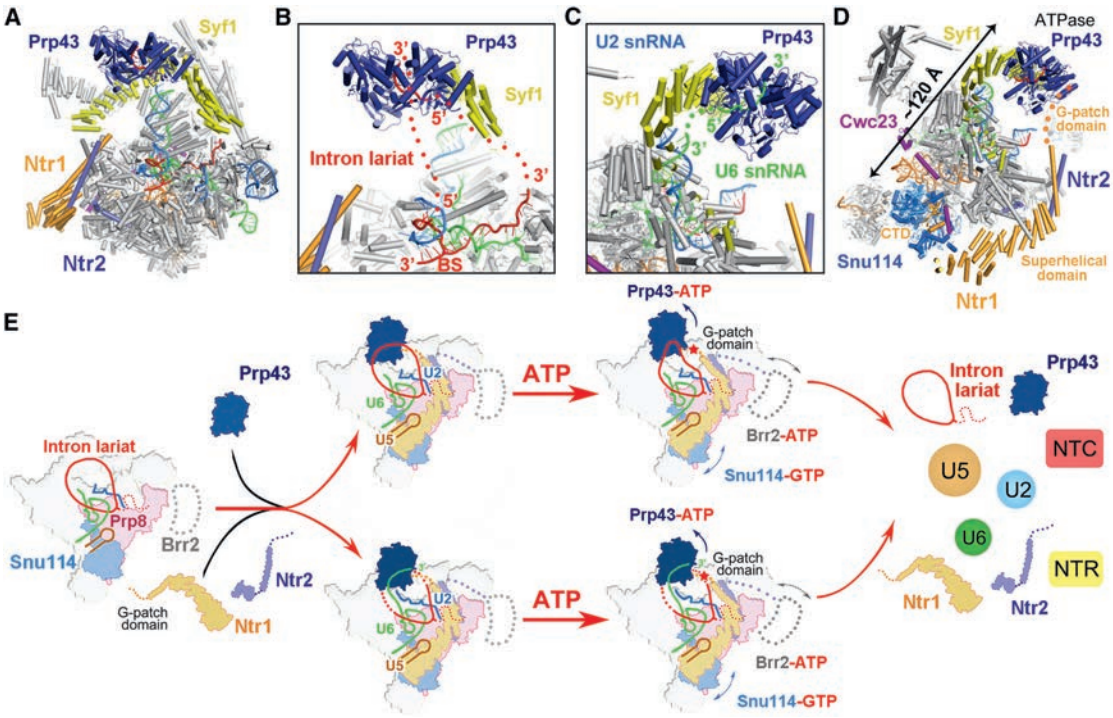


Figure 7. Prp43 May Disassemble the ILS Complex by Pulling the Intron Sequences or the 3′ End Sequences of U6 snRNA

(A) Prp43 binds to the middle portion of the superhelical protein Syf1.

(B) Prp43 may disassemble the ILS complex by pulling the intron sequences. The physical location of Prp43 is consistent with the model in which Prp43 grabs and pulls the middle portion of the intron lariat. The disordered intron sequences are represented by the dotted lines.

(C) Prp43 may disassemble the ILS complex by grabbing and pulling the 3′ end sequences of U6 snRNA. The close proximity of the 3′ end sequences of U6 snRNA and the RNA-binding tunnel of Prp43 lends support to this alternative model.

(D) Prp43 is located far away from Snu114. Notably, the N-terminal and C-terminal ends of Ntr1 contact the ATPase Prp43 and the GTPase Snu114, respectively. Brr2 remains unidentified in our structure of the ILS complex.

(E) A working model of the spliceosome disassembly. In this model, Prp43 serves as the primary engine to disassemble the ILS complex but the helicase activity of Prp43 must be activated by the G-patch motif of Ntr1. The recruitment of Ntr2 and Ntr1 into the spliceosome serves two purposes: to ensure the correct spliceosomal complex is chosen for disassembly and to activate the helicase Prp43. The presence of Ntr2 and Ntr1 also facilitates the recruitment of Prp43. Prp43 may act by pulling either the intron lariat or the 3′ sequences of U6 snRNA.

the ILS complexes (Figure 5) may faithfully reflect those between the P and the ILS complexes.

Prp43 and/or the Ntr complex were reported to be recruited into the spliceosome before the Prp16 and Prp22 stages of splicing (Chen et al., 2013a; Fabrizio et al., 2009; Koodathingal et al., 2010; Mayas et al., 2010; Pandit et al., 2006). In the Prp16 stage (the C complex), the Ntr complex can bind to the spliceosome because the binding sites for all three components of the Ntr complex remain unoccupied (Figure 4C). In the Prp22 stage (the C* complex), however, the binding site for the helix-loop-helix motif of Ntr2 is partially occupied by Prp22. In this case, the recruitment of the Ntr complex may be weakened if Prp22 is first recruited into the spliceosome, and vice versa. Notably, the recruitment of the Ntr components does not necessarily mean its activation. Because the helicase Prp43 remains inactive until it is activated by the G-patch motif of Ntr1, we speculate that the recruitment of Ntr1 into the spliceosome and its productive interaction with Prp43 are the key events for spliceosome disassembly. In other words, components of the Ntr complex can be recruited into earlier spliceosomes, but the productive interactions among these components are yet to be formed.

The cryo-EM structure of the ILS complex from *S. cerevisiae* follows those for the B^{act} complex (Rauhut et al., 2016; Yan et al., 2016), the C complex (Galej et al., 2016; Wan et al., 2016), the C* complex (Fica et al., 2017; Yan et al., 2017), and, more recently, the B complex (Plaschka et al., 2017). Although detailed structural information on the spliceosomal B* and P complexes is yet to be elucidated, their overall structures should closely resemble those of the C and C* complexes, respectively. In this regard, we have obtained a comprehensive picture on the catalytic cycle of pre-mRNA splicing. For mechanistic understanding of a complete splicing cycle, however, we have only reached the end of the beginning. The branching reaction is reasonably well understood; but, how the exon ligation occurs remains to be structurally characterized. Although we can speculate how the various ATPase/helicases may remodel the spliceosome, the details remain to be elucidated through structural and biophysical studies. Nonetheless, the detailed structural information on the various conformational states of the spliceosome provides an invaluable framework for mechanistic understanding of the spliceosome function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARDING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
- METHOD DETAILS
  - TAP tagging of Yju2 in *S. cerevisiae*
  - Purification of the spliceosomal complex
  - DSS cross-linking
  - Mass spectrometry analysis
  - EM data acquisition and processing
  - Image processing
  - Model building and refinement
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.029.

AUTHOR CONTRIBUTIONS

R.W. and R.B. purified the yeast ILS complex and prepared the cryo-EM sample. R.W., R.B., J.L., and C.Y. collected the EM micrographs and processed the data. C.Y. calculated the cryo-EM map and built the atomic model. All authors contributed to project discussion and structure analysis. R.W., R.B., and C.Y. contributed to manuscript preparation. Y.S. designed and guided the project and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Min Huang and Catherine Wong for help with mass spectrometric analysis and the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for the cryo-EM facility and the computational facility support on the cluster of Bio-Computing Platform. This work was supported by funds from the National Natural Science Foundation of China (31621092 and 31430020) and the Ministry of Science and Technology (2016YFA0501100 to J.L.).

Received: July 14, 2017
Revised: August 7, 2017
Accepted: August 15, 2017
Published: September 14, 2017

REFERENCES

Amunts, A., Brown, A., Bai, X.C., Llácer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. Science *343*, 1485–1489.

Aravind, L., and Koonin, E.V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. Trends Biochem. Sci. *24*, 342–344.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc. Natl. Acad. Sci. USA *94*, 11798–11802.

Bertram, K., Agafonov, D.E., Liu, W.T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318–323.

Bohnsack, M.T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E., and Tollervey, D. (2009). Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. Mol. Cell *36*, 583–592.

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuvéglise, C., Munsterkotter, M., Guldener, U., Mewes, H.W., Van Helden, J., Dujon, B., and Gaillardin, C. (2003). Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. Nucleic Acids Res. *31*, 1121–1135.

Boon, K.L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J.D., Droop, A.P., Dez, C., and Beggs, J.D. (2006). Yeast ntr1/spp382 mediates prp43 function in postspliceosomes. Mol. Cell. Biol. *26*, 6016–6023.

Chen, H.C., Tseng, C.K., Tsai, R.T., Chung, C.S., and Cheng, S.C. (2013a). Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol. Cell. Biol. *33*, 514–525.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H., and Henderson, R. (2013b). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy *135*, 24–35.

Chou, F.C., Sripakdeevong, P., Dibrov, S.M., Hermann, T., and Das, R. (2013). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. Nat. Methods *10*, 74–76.

Christian, H., Hofele, R.V., Urlaub, H., and Ficner, R. (2014). Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. Nucleic Acids Res. *42*, 1162–1179.

Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487–493.

Cordin, O., Hahn, D., and Beggs, J.D. (2012). Structure, function and regulation of spliceosomal RNA helicases. Curr. Opin. Cell Biol. *24*, 431–438.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375–383.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.org.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol. Cell *36*, 593–608.

Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature *503*, 229–234.

Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. Nature *542*, 377–380.

Fourmann, J.B., Schmitzová, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev. *27*, 413–428.

Fourmann, J.B., Dybkov, O., Agafonov, D.E., Tauchert, M.J., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2016). The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction. eLife *5*, e15564.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197–201.

Garrey, S.M., Katolik, A., Prekeris, M., Li, X., York, K., Bernards, S., Fields, S., Zhao, R., Damha, M.J., and Hesselberth, J.R. (2014). A homolog of lariat-debranching enzyme modulates turnover of branched RNA. RNA *20*, 1337–1348.

Gietz, R.D., and Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 35–37.

Gordon, P.M., Sontheimer, E.J., and Piccirilli, J.A. (2000). Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. RNA *6*, 199–205.

Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell *37*, 415–427.

Grant, T., and Grigorieff, N. (2015). Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. J. Struct. Biol. *192*, 204–208.

Hahn, D., and Beggs, J.D. (2010). Brr2p RNA helicase with a split personality: insights into structure and function. Biochem. Soc. Trans. *38*, 1105–1109.

Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. Science *349*, 1191–1198.

He, Y., Staley, J.P., Andersen, G.R., and Nielsen, K.H. (2017). Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. RNA *23*, 1110–1124.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem. Sci. *36*, 19–29.

Kimanius, D., Forsberg, B.O., Scheres, S.H., and Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. eLife *5*, e18722.

Koodathingal, P., Novak, T., Piccirilli, J.A., and Staley, J.P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. Mol. Cell *39*, 385–395.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Zhu, H., Dyer, D.W., Roe, B.A., and Murphy, J.W. (2004). Introns and splicing elements of five diverse fungi. Eukaryot. Cell *3*, 1088–1100.

Li, X., Zheng, S., Agard, D.A., and Cheng, Y. (2015). Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. J. Struct. Biol. *192*, 174–178.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. J. Biol. Chem. *277*, 17743–17750.

Mayas, R.M., Maita, H., Semlow, D.R., and Staley, J.P. (2010). Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proc. Natl. Acad. Sci. USA *107*, 10020–10025.

McPheeters, D.S., Schwer, B., and Muhlenkamp, P. (2000). Interaction of the yeast DExH-box RNA helicase prp22p with the 3′ splice site during the second step of nuclear pre-mRNA splicing. Nucleic Acids Res. *28*, 1313–1321.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nicholls, R.A., Fischer, M., McNicholas, S., and Murshudov, G.N. (2014). Conformation-independent structural comparison of macromolecules with ProSMART. Acta Crystallogr. D Biol. Crystallogr. *70*, 2487–2499.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science *225*, 898–903.

Pandit, S., Lynn, B., and Rymond, B.C. (2006). Inhibition of a spliceosome turnover pathway suppresses splicing defects. Proc. Natl. Acad. Sci. USA *103*, 13700–13705.

Pandit, S., Paul, S., Zhang, L., Chen, M., Durbin, N., Harrison, S.M., and Rymond, B.C. (2009). Spp382p interacts with multiple yeast splicing factors, including possible regulators of Prp43 DExD/H-Box protein function. Genetics *183*, 195–206.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. Nature *546*, 617–621.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218–229.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B., et al. (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science *353*, 1399–1405.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Sahi, C., Lee, T., Inada, M., Pleiss, J.A., and Craig, E.A. (2010). Cwc23, an essential J protein critical for pre-mRNA splicing with a dispensable J domain. Mol. Cell. Biol. *30*, 33–42.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol. Cell *30*, 743–754.

Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J. *17*, 2086–2094.

Semlow, D.R., Blanco, M.R., Walter, N.G., and Staley, J.P. (2016). Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell *164*, 985–998.

Shi, Y. (2017). The Spliceosome: A Protein-Directed Metalloribozyme. J. Mol. Biol. *429*, 2640–2653.

Silverman, E.J., Maeda, A., Wei, J., Smith, P., Beggs, J.D., and Lin, R.J. (2004). Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. Mol. Cell. Biol. *24*, 10101–10110.

Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol. Cell *23*, 389–399.

Sontheimer, E.J., Sun, S., and Piccirilli, J.A. (1997). Metal ion catalysis during splicing of premessenger RNA. Nature *388*, 801–805.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221–234.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. USA *90*, 6498–6502.

Tanaka, N., Aronova, A., and Schwer, B. (2007). Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. Genes Dev. *21*, 2312–2325.

Tauchert, M.J., Fourmann, J.B., Lührmann, R., and Ficner, R. (2017). Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. eLife *6*, e21510.

Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., and Cheng, S.C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. *19*, 2991–3003.

Tsai, R.T., Tseng, C.K., Lee, P.J., Chen, H.C., Fu, R.H., Chang, K.J., Yeh, F.L., and Cheng, S.C. (2007). Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. Mol. Cell. Biol. *27*, 8027–8037.

Wagner, J.D., Jankowsky, E., Company, M., Pyle, A.M., and Abelson, J.N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J. *17*, 2926–2937.

Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y., and Leulliot, N. (2010). Prp43p contains a processive

helicase structural architecture with a specific regulatory domain. EMBO J. *29*, 2194–2204.

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. Science *353*, 895–904.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, a003707.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. Science *349*, 1182–1191.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 Å resolution. Science *353*, 904–911.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. Science *355*, 149–155.

Yang, B., Wu, Y.J., Zhu, M., Fan, S.B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.X., Chen, H.F., et al. (2012). Identification of cross-linked peptides from complex samples. Nat. Methods *9*, 904–906.

Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature *408*, 881–884.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J. Struct. Biol. *193*, 1–12.

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. Cell *169*, 918–929.e14.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

Article

Cell

A Spatial Interactome Reveals the Protein Organization of the Algal CO_2-Concentrating Mechanism

Luke C.M. Mackinder,[1,6] Chris Chen,[1,2] Ryan D. Leib,[3] Weronika Patena,[1,4] Sean R. Blum,[1,7] Matthew Rodman,[2] Silvia Ramundo,[5] Christopher M. Adams,[3] and Martin C. Jonikas[1,2,4,8,*]

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA
[2]Department of Biology, Stanford University, Stanford, CA 94305, USA
[3]Stanford University Mass Spectrometry, Stanford University, Stanford, CA 94305, USA
[4]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[5]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA
[6]Present address: Department of Biology, University of York, York YO10 5DD, UK
[7]Present address: Department of Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA 95064, USA
[8]Lead Contact
*Correspondence: mjonikas@princeton.edu
http://dx.doi.org/10.1016/j.cell.2017.08.044

SUMMARY

Approximately one-third of global CO_2 fixation is performed by eukaryotic algae. Nearly all algae enhance their carbon assimilation by operating a CO_2-concentrating mechanism (CCM) built around an organelle called the pyrenoid, whose protein composition is largely unknown. Here, we developed tools in the model alga *Chlamydomonas reinhardtii* to determine the localizations of 135 candidate CCM proteins and physical interactors of 38 of these proteins. Our data reveal the identity of 89 pyrenoid proteins, including Rubisco-interacting proteins, photosystem I assembly factor candidates, and inorganic carbon flux components. We identify three previously undescribed protein layers of the pyrenoid: a plate-like layer, a mesh layer, and a punctate layer. We find that the carbonic anhydrase CAH6 is in the flagella, not in the stroma that surrounds the pyrenoid as in current models. These results provide an overview of proteins operating in the eukaryotic algal CCM, a key process that drives global carbon fixation.

INTRODUCTION

Over the past three billion years, the carbon-fixing enzyme Rubisco drew down atmospheric concentrations of CO_2 to trace levels (Dismukes et al., 2001), in effect starving itself of its substrate. In parallel, the oxygenic reactions of photosynthesis have caused the appearance of abundant O_2, which competes with CO_2 for the active site of Rubisco and results in a loss of fixed CO_2 via photorespiration (Bauwe et al., 2010). To overcome these challenges of CO_2 assimilation in today's atmosphere, many photosynthetic organisms increase CO_2 levels in the vicinity of Rubisco by operating CO_2-concentrating mechanisms (CCMs). Such mechanisms increase the CO_2:O_2 ratio at the active site of Rubisco, enhancing CO_2 fixation and decreasing photorespiration. CCMs are found in nearly all marine photoautotrophs, including cyanobacteria and eukaryotic algae (Reinfelder, 2011), which together account for ~50% of global carbon fixation (Field et al., 1998).

In cyanobacterial CCMs, inorganic carbon in the form of bicarbonate (HCO_3^-) is pumped into the cytosol to a high concentration. This HCO_3^- is then converted into CO_2 in specialized icosahedral compartments called carboxysomes, which are packed with Rubisco (Price and Badger, 1989). The components of the cyanobacterial CCMs have largely been identified, facilitated in part by the organization of the genes encoding them into operons (Price et al., 2008). Knowledge of these components has enabled the detailed characterization of the structure and assembly pathway of the beta carboxysome (Cameron et al., 2013).

Analogous to the cyanobacterial CCM, the eukaryotic green algal CCM concentrates HCO_3^- in a microcompartment containing tightly-packed Rubisco, called the pyrenoid. The pyrenoid is located in the chloroplast, surrounded by a starch sheath and traversed by membrane tubules that are continuous with the surrounding photosynthetic thylakoid membranes (Engel et al., 2015). Associated with the pyrenoid tubules is a carbonic anhydrase that converts HCO_3^- to CO_2 for fixation by Rubisco (Karlsson et al., 1998). The mechanism of delivery of HCO_3^- to the pyrenoid thylakoids remains unknown. In contrast to the prokaryotic CCM, the protein composition of the eukaryotic algal CCM and the structural organization of the pyrenoid remain largely uncharacterized.

In this study, we developed a high-throughput fluorescence protein tagging and affinity purification mass spectrometry (AP-MS) pipeline for the model green alga *Chlamydomonas reinhardtii* (Figure 1A). With this pipeline, we determined the localizations of 135 candidate CCM proteins and the physical interactions of 38 core CCM components. Our microscopy data reveals an unexpected localization for the carbonic anhydrase CAH6, identifies three previously undescribed pyrenoid protein

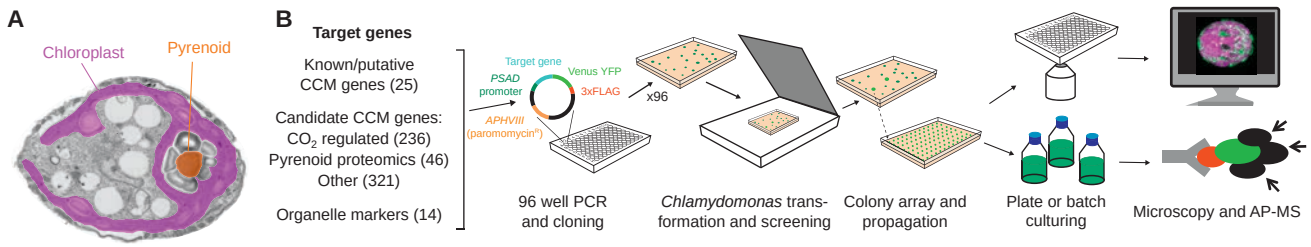


Figure 1. We Developed a High-Throughput Pipeline to Determine the Localization and Physical Interactions of Algal Proteins

(A) A false-color transmission electron micrograph of a *Chlamydomonas reinhardtii* cell. The chloroplast is highlighted in magenta and the pyrenoid matrix in orange.

(B) Tagging and mass spectrometry pipeline. Target genes were amplified by PCR and Gibson assembled in frame with Venus-3xFLAG, under the constitutive *PSAD* promoter. Transformants were screened for fluorescence using a scanner and arrayed to allow robotic propagation. Lines were either imaged using confocal microscopy to determine their spatial distribution or batch cultured for affinity purification-mass spectrometry (AP-MS).

See also Figure S1 and Table S1.

layers and suggests that the pyrenoid shows size selectivity for stromal proteins. The AP-MS data produce a spatially resolved protein-protein interaction map of the CCM and pyrenoid, identifying novel protein complexes including a complex between inorganic carbon transporters LCI1 and HLA3 and suggesting CCM functions for multiple proteins. These results transform our basic knowledge of the eukaryotic CCM and advance the prospects of transferring this system into higher plants to improve crop production (Atkinson et al., 2016; Long et al., 2015).

RESULTS AND DISCUSSION

We Developed a High-Throughput Pipeline for Systematic Localization of Proteins in *Chlamydomonas*

To allow the parallel cloning of hundreds of genes, we designed an expression cassette that enabled high-throughput seamless cloning via Gibson assembly (Gibson et al., 2009). Open reading frames (ORFs) were amplified by PCR from genomic DNA and cloned in frame with a C-terminal Venus YFP and a 3xFLAG epitope, driven by the strong PsaD promoter. These constructs were transformed into wild-type *Chlamydomonas*, where they inserted into random locations in the genome (Figure 1B). To allow dual tagging of different proteins in the same cell, we developed a second expression vector with an mCherry fluorophore and a hygromycin selection marker (Figure S1A). Potential caveats of our system include loss of the endogenous transcriptional regulation of the protein, including information encoded in the promoter, terminator, and genomic locus. Additionally, the C-terminal protein tag could obscure subcellular targeting signals or disrupt functional domains.

Our Data Reveal Guidelines for Protein Localization in *Chlamydomonas*

Given the notorious difficulties with expressing tagged genes in *Chlamydomonas* (Fuhrmann et al., 1999; Neupert et al., 2009), we started with the understanding that we would only succeed in a fraction of cases and sought to maximize the total number of proteins localized. We selected target genes from three sources: (1) genes currently thought to be involved in the CCM (see Wang et al., 2015 for a review), (2) candidate CCM genes, including those identified from both transcriptomic (Brueggeman et al., 2012; Fang et al., 2012) and proteomic (Mackinder et al., 2016) studies, and (3) organelle markers (Figure 1B; Table S1). We were able to determine the localizations of 146 out of the 624 target genes (23%).

We sought to leverage the large scale of this study to uncover factors that may contribute to cloning and tagging success in *Chlamydomonas*. We successfully cloned 298 of the 624 target genes (48%). Our cloning success rate decreased with gene size (Figure S1B). Intriguingly, cloning success was higher for genes with high expression levels (Figures S1C and S1D) ($p = 4 \times 10^{-13}$, Mann-Whitney *U* test), suggesting that intrinsic properties of a gene that influence endogenous expression may also affect PCR efficiency.

We successfully transformed and acquired protein localization data for 146 of the 298 cloned genes (49%). The two main factors correlated with our ability to obtain localization data were: (1) high endogenous gene expression level (Figures S1E and S1F) ($p = 6 \times 10^{-14}$, Mann-Whitney *U* test), and (2) absence of upstream in-frame ATGs (Figure S1G) (Cross, 2016). The failure to obtain localization data for genes with in-frame uATGs is likely due to the absence of the correct translational start site in the cloned construct, resulting in a truncated protein that can be functionally impaired, structurally unstable or lacking essential organelle targeting sequence(s). These data suggest that transcript abundance is predictive for localization success and that future protein expression studies will benefit substantially from improved annotation of *Chlamydomonas* translation start sites.

146 Tagged Proteins Show 29 Distinct Localization Patterns

To aid in the classification of unknown proteins to subcellular regions, we tagged a series of conserved, well-characterized organelle and cellular structure proteins (Table S1). We then employed a decision tree (Figure 2A) to classify visually the localization of 135 additional proteins into 29 distinct subcellular regions, representing nearly all of the known organelles and cellular structures of *Chlamydomonas* (Figure 2B). The protein localizations from our study are available at https://sites.google.com/site/chlamyspatialinteractome/.

Interestingly, 12 proteins were not confined to one organelle but were seen in multiple compartments (Figure 2C; Table S2). If these multiple localizations are not artifacts of our expression system, they may represent proteins that function in multiple compartments or are involved in inter-organelle signaling. Additionally, we observed diverse cytosolic localizations, with subtle differences between localization patterns (Figure S2A).

Localization Assignments Agree with Previous Studies for 39/41 Proteins

To evaluate the accuracy of our method, we compared our results with published localizations of individual proteins. Our data shared 25 proteins with the validated "training" set of chloroplast, mitochondria, and secretory pathway proteins from Tardif et al. (2012). Nearly all (24/25) matched our localization data, with the only exception being ACP2 (Cre13.g577100). Whereas we saw ACP2 in the chloroplast (Figure 2D), Tardif et al. (2012) saw ACP2 in isolated mitochondria. However, previous studies have either failed to detect ACP2 in mitochondria (Atteia et al., 2009) or saw it in approximately equal abundances in isolated chloroplasts and mitochondria (Terashima et al., 2010). The ambiguity in the previously published data leaves open the possibility that our ACP2 localization data may in fact be correct. We further compared our data with previously published localizations of CCM components and found that 15 of 16 localizations matched. The strong overlap with previously known localizations indicates that our dataset is of high quality (>95% accurate) and that C-terminal tagging of *Chlamydomonas* proteins results in minimal localization artifacts.

CAH6 Localizes to the Flagella

Carbonic anhydrases, which catalyze the reversible reaction of HCO_3^- to CO_2, play a critical role in CCMs (Badger, 2003). Our successful localization of nine *Chlamydomonas* carbonic anhydrases shows that they are found in a diverse range of cellular locations (Figure S2B). In all current models of the CCM (Moroney et al., 2011; Wang et al., 2015), the carbonic anhydrase CAH6 is in the chloroplast stroma, where it has been proposed to convert CO_2 to HCO_3^-.

Surprisingly, in our study, CAH6 localized to the flagella in two independent transformation lines (Figures 2D and S2B) and produced no detectable signal in the chloroplast. To exclude the possibility that our observation is due to an artifact (e.g., due to the C-terminal Venus tag), we analyzed the localization of CAH6 in existing proteomic datasets. CAH6 is present in the flagellar proteome (Pazour et al., 2005) and has been shown to be an abundant intraflagellar transport (IFT) cargo (Engel et al., 2012), providing independent validation of CAH6 localization to the flagella. Additionally, CAH6 is absent from both the chloroplast proteome (Terashima et al., 2010) and the mitochondrial proteome (Atteia et al., 2009), further suggesting that levels in the chloroplast are low or non-existent.

Previous evidence for CAH6 in the stroma came from immunogold labeling experiments, in which Mitra et al. (2004) found a 4.7-fold enrichment of gold particles associated with chloroplast starch relative to control pre-immune serum. This could be an artifact due to cross-reactivity of the immunized serum with another epitope. Alternatively, CAH6 may be an abundant flagellar protein, but present at very low levels in the chloroplast.

The apparent absence of carbonic anhydrase in the stroma may be a requirement of the *Chlamydomonas* CCM. A stromal carbonic anhydrase could risk short-circuiting the CCM by promoting the release of CO_2 from HCO_3^- in areas that are not in close proximity to Rubisco. In fact, it has been shown that the expression of carbonic anhydrase in the cyanobacterial cytosol, the likely functional equivalent of the chloroplast stroma, results in the disruption of the cyanobacterial CCM (Price and Badger, 1989).

Instead of directly participating in the CCM, CAH6 could be involved in inorganic carbon sensing. *Chlamydomonas* was recently shown to chemotax toward HCO_3^- (Choi et al., 2016), and carbonic anhydrases have been previously implicated in inorganic carbon sensing (Hu et al., 2010). Localization of sensing machinery to the flagella, which are found at the leading edge of swimming cells, could facilitate chemotaxis.

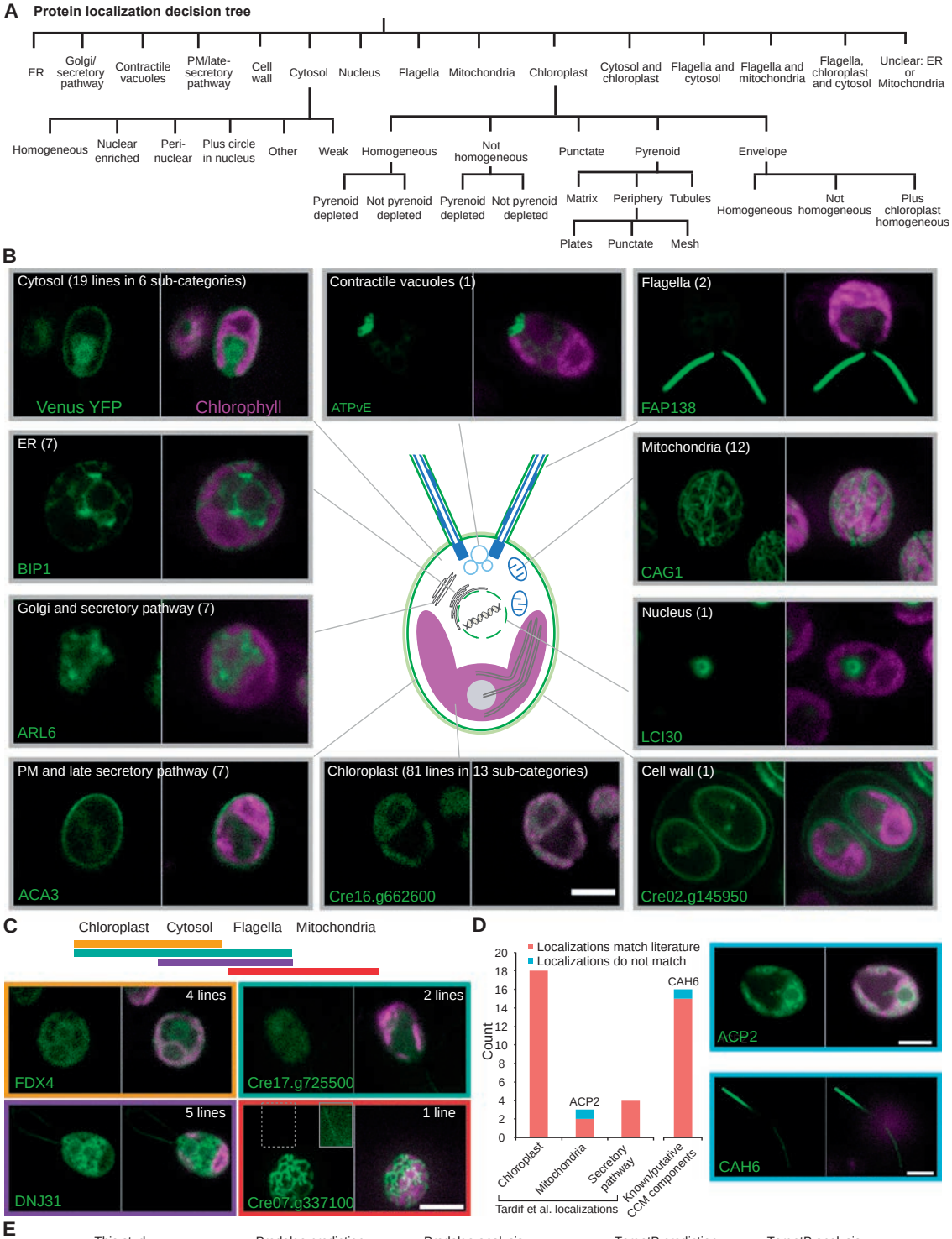
PredAlgo Is the Best Protein Localization Predictor for *Chlamydomonas*

The excellent agreement of our localization data with previous studies provided an opportunity to test the accuracy of the two main localization prediction algorithms used for *Chlamydomonas* proteins, PredAlgo (Tardif et al., 2012) and TargetP (Emanuelsson et al., 2000). For proteins that we observed in the chloroplast, PredAlgo predicted a chloroplast localization for 90% of them, whereas TargetP only predicted a chloroplast localization for 31% (Figure 2E). For mitochondrial proteins, the accuracy dropped to 31% for PredAlgo and 15% for TargetP. For secretory pathway proteins, the accuracy was 38% for PredAlgo and 24% for TargetP. These results highlight that PredAlgo is the best localization predictor for *Chlamydomonas* proteins, but its accuracy drops off significantly when proteins localize to compartments other than the chloroplast.

We Assigned 82 Proteins to 13 Sub-chloroplast Locations

Approximately 56% (82/146) of our proteins localized to the chloroplast. We assigned these 82 proteins to 13 sub-chloroplast locations (Figures 2A and 3A; Table S1). Chloroplast envelope proteins showed three subcategories of localization: (1) envelope homogeneous (signal observed evenly throughout the chloroplast envelope), (2) envelope non-homogeneous, and (3) envelope plus chloroplast homogeneous (signal observed throughout the chloroplast in addition to the envelope). LCIA (low CO_2 Inducible A) and LCI20 both showed some homogeneous chloroplast signal in addition to a clear envelope signal, suggesting the possibility that these proteins are functional in both the chloroplast envelope and thylakoid membranes.

Three proteins produced similar patterns of punctate dots throughout the chloroplast (Figure S3A): a protein with predicted 50S ribosome-binding GTPase activity (Cre12.g524950), histone-like protein 1 (HLP1; Cre06.g285400) (Karcher et al., 2009), and the fatty acid biosynthesis enzyme acetyl-CoA biotin carboxyl carrier (BCC2; Cre01.g037850). The similarity of the localization patterns of these proteins suggests that chloroplast translation, chloroplast DNA, and fatty acid synthesis may be co-localized in the chloroplast.



E

This study		Predalgo prediction				Predalgo analysis		TargetP prediction				TargetP analysis	
		C	M	SP	O	Match	False negative rate	C	M	SP	O	Match	False negative rate
Chloroplast	67	60	3	2	2	90%	10%	21	41	0	5	31%	69%
Mitochondria	13	3	4	0	6	31%	69%	4	2	0	7	15%	85%
Secretory pathway	21	2	0	8	11	38%	62%	0	3	5	13	24%	76%
Other	41	15	0	2	24	59%	41%	7	6	1	27	66%	34%

(legend on next page)

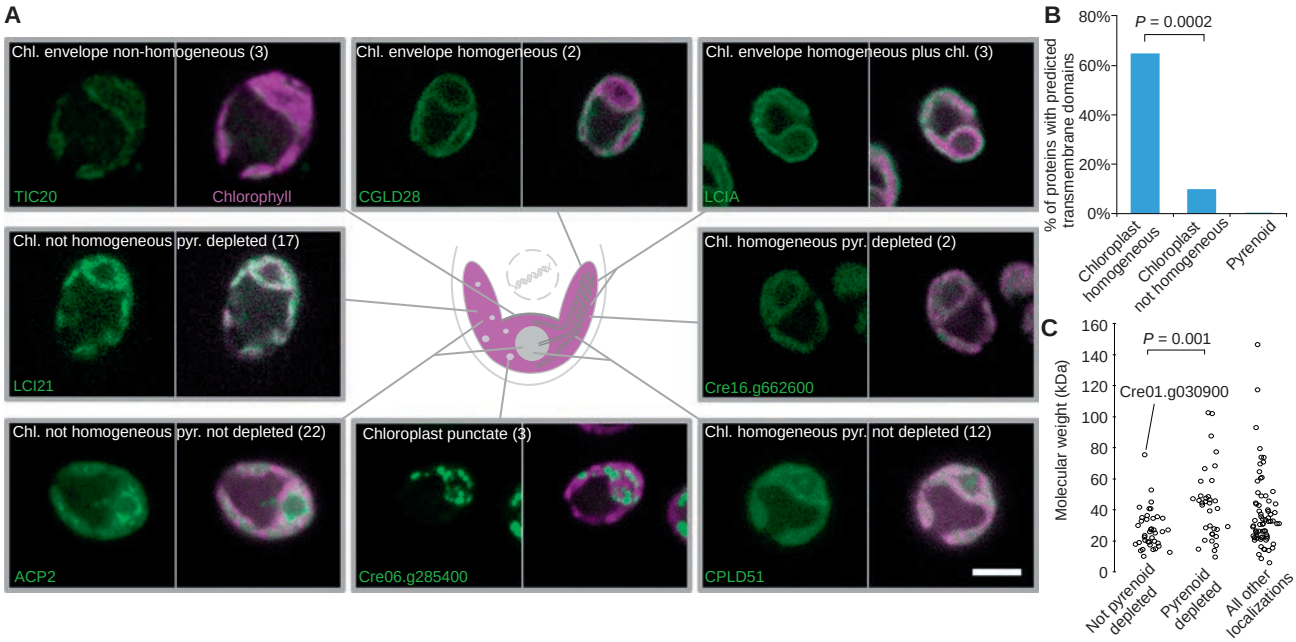


Figure 3. Chloroplast Proteins Show 13 Different Localization Patterns

(A) Representative images of proteins localized to different chloroplast regions. The number of proteins showing each pattern is in parentheses. Scale bar, 5 μm.

(B) The percentage of proteins with predicted transmembrane domains is shown for different localization patterns. Bracket shows a significant difference using Fisher's exact test.

(C) Predicted molecular weight of proteins is shown as a function of pyrenoid signal intensity. Cre01.g030900 that has a pyrenoid signal and is above the 50 kDa cut-off is labeled. Bracket shows significant difference using a Mann-Whitney *U* test.

See also Figure S3.

We found that proteins with specific patterns of localization were often enriched in certain physical properties. As expected, all eight chloroplast envelope proteins contained one or more transmembrane domains (Table S1). Interestingly, proteins showing homogeneous chloroplast localization (Figure 3B) were enriched in transmembrane domains, found in 9/14 homogeneous proteins versus 4/39 for chloroplast non-homogeneous proteins ($p = 0.0002$, Fisher's exact test). This observation suggests that proteins with homogeneous localization are most likely thylakoid membrane-associated.

The Pyrenoid Appears to Show Selectivity to Stromal Contents

Because the pyrenoid is a non-membrane-bound organelle, its protein composition cannot be regulated by a membrane translocation step. We therefore sought to understand whether pyrenoid proteins are enriched for any specific physicochemical properties. We classified chloroplast localized proteins into two groups: (1) pyrenoid-depleted, where the signal from the pyrenoid was weaker than the surrounding chloroplast, and (2) not pyrenoid-depleted, where the signal from the pyrenoid was comparable to or brighter than the surrounding chloroplast. Interestingly, the two groups showed different protein molecular weight distributions ($p = 0.001$, Mann-Whitney *U* test). The 39 proteins that are not pyrenoid-depleted are almost all smaller than ~50 kDa (Figure 3C; the value of ~50 kDa excludes the Venus YFP region, therefore the effective molecular weight is ~78 kDa), suggesting that the pyrenoid may exclude larger proteins.

We Identified Multiple New Pyrenoid Components

Electron microscopy-based techniques have shown that the *Chlamydomonas* pyrenoid contains a dense matrix of Rubisco surrounded by a starch sheath and traversed by membrane tubules formed from merged thylakoids (Figure 4A) (Engel et al., 2015). Currently, seven proteins have been unambiguously localized to three different regions of the pyrenoid: the pyrenoid matrix, periphery, and tubules. The pyrenoid matrix contains the

Figure 2. Tagged Proteins Localized to a Diverse Range of Cellular Locations and Revealed that CAH6 Localizes to Flagella

(A) A decision tree was used to assign proteins to specific subcellular locations.

(B) Representative images of proteins localized to different cellular locations. The number of different lines showing each localization pattern is in parentheses.

(C) Representative images of proteins that localized to more than one compartment. The solid outer line inset in the Cre07.g337100 image is an overexposure of the region surrounded by a dashed line to highlight flagellar fluorescence.

(D) Comparison of our observations with published localizations. Images show the two proteins that did not match their published locations. All scale bars, 5 μm.

(E) Comparison of our observations with localizations predicted by PredAlgo and TargetP. Table excludes proteins found in the PredAlgo training set.

See also Figure S2 and Table S2.

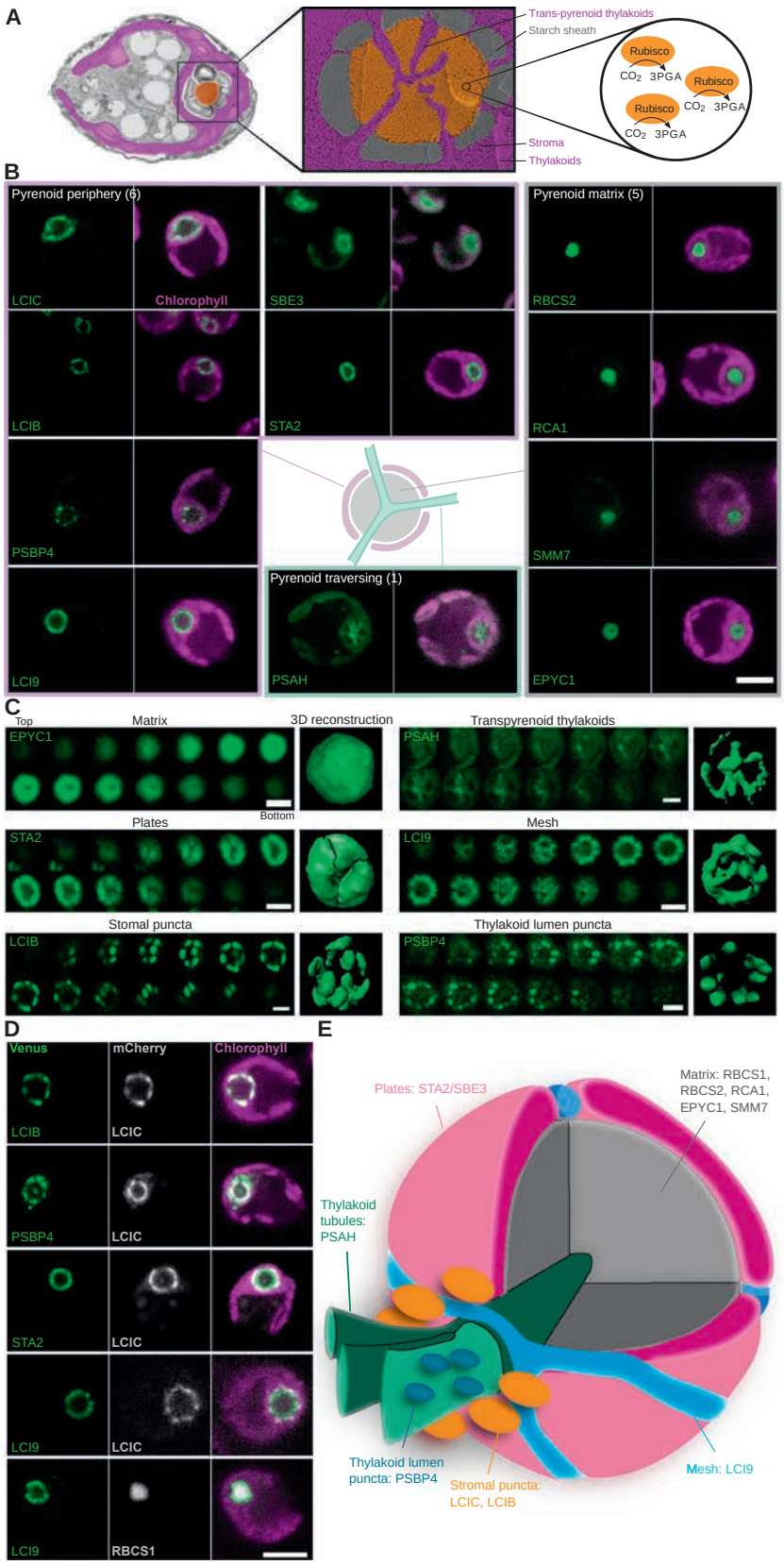


Figure 4. Pyrenoid Proteins Show at Least Six Distinct Localization Patterns and Reveal Three New Protein Layers

(A) A false-color transmission electron micrograph and deep-etched freeze-fractured image of the pyrenoid highlight the pyrenoid tubules, starch sheath, and pyrenoid matrix where the principal carbon fixing enzyme, Rubisco, is located. Images courtesy of Moritz Meyer, Ursula Goodenough, and Robyn Roth.

(B) Proteins showing various localization patterns within the pyrenoid are illustrated. Scale bar, 5 μm.

(C) Confocal sections distinguish different localization patterns within the pyrenoid. Each end panel is a space-filling reconstruction. Scale bars, 2 μm.

(D) Dual tagging refined the spatial distribution of proteins in the pyrenoid. Scale bar, 5 μm.

(E) A proposed pyrenoid model highlighting the distinct spatial protein-containing regions.

Rubisco holoenzyme (RBCS/RbcL), its chaperone Rubisco activase (RCA1), essential pyrenoid component 1 (EPYC1), a Rubisco linker protein important for Rubisco packaging in the pyrenoid (Mackinder et al., 2016), and a protein of unknown function (Cre06.g259100) (Kobayashi et al., 2015). Under very low-CO_2 conditions, the LCIB/LCIC complex, whose role is still uncertain (Jin et al., 2016), is known to form puncta around the pyrenoid periphery (Yamano et al., 2010). Recently, a Ca^{2+}-binding protein, CAS, has been shown to specifically localize to the pyrenoid tubules at low CO_2 (Wang et al., 2016). Here, we identify seven additional pyrenoid-localized components and three previously undescribed sub-pyrenoid localization patterns (Figures 4B–4D).

The Pyrenoid Has at Least Four Distinct Outer Protein Layers

Our data suggest that the pyrenoid is surrounded by at least four distinct outer protein layers: (1) LCIB and LCIC localize to puncta around the periphery, (2) PSBP4 (photosystem II subunit P4) localizes to a different set of puncta, (3) STA2 (starch synthase 2) and SBE3 (starch branching enzyme 3) localize to plate-like structures, and (4) LCI9 localizes to a mesh-like structure (Figures 4C–4E).

LCIB, LCIC, and PSBP4 showed punctate outer pyrenoid patterns, whereas SBE3, STA2, and LCI9 showed a more homogeneous distribution around the pyrenoid periphery (Figure 4B). LCIB and LCIC were co-localized (Figure 4D), supporting the previous finding that they are part of the same complex in the stroma (Yamano et al., 2010).

PSBP4-Venus did not co-localize with LCIC-mCherry (Figure 4D), indicating that PSBP4 is in a different structure or complex. PPD1, the *Arabidopsis* homolog of PSBP4, has been shown to be in the thylakoid lumen (Liu et al., 2012). Therefore, the PSBP4 puncta likely represent proteins located in the thylakoid lumen. Consistent with this possibility, we also see a small amount of PSBP4-Venus signal within the pyrenoid, and this signal forms a network-like pattern reminiscent of pyrenoid tubules.

Our data suggest that both STA2 and SBE3 localize to the starch sheath. Co-localization indicated that STA2 was localized within the perimeter described by LCIC (Figure 4D). STA2 formed a clearly defined plate-like pattern around the pyrenoid core (Figure 4C). SBE3 also displayed this plate pattern but was generally more diffuse than STA2 (Figure 4B).

LCI9 was tightly apposed to the pyrenoid matrix and, like STA2, also localized within the perimeter described by LCIC (Figure 4D). However, analysis of Z sections showed that unlike STA2 and SBE3, LCI9 formed a mesh structure around the pyrenoid (Figure 4C). Intriguingly, the complementary localizations of STA2 and LCI9 suggest that LCI9 may be part of a protein layer that fills the gaps between the starch plates.

A Putative Methyltransferase Localizes to the Pyrenoid Matrix

We discovered that SMM7 (Cre03.g151650), a putative methyltransferase, localized to the pyrenoid matrix. This is intriguing because another putative methyltransferase, CIA6 (Cre10.g437829), was found to be required for pyrenoid assembly (Ma et al., 2011), although its localization was not determined. Unlike *CIA6*, *SMM7* is strongly transcriptionally upregulated under low-CO_2 conditions (Brueggeman et al., 2012; Fang et al., 2012). Identification of the protein targets of CIA6 and SMM7 will likely provide critical insights into pyrenoid biogenesis and regulation.

Pyrenoid Tubules Are Enriched in PSAH, a Component of Photosystem I

Traversing the pyrenoid are pyrenoid tubules, which are thought to deliver CO_2 at a high concentration to the matrix (Wang et al., 2015). Previous work using Immunogold labeling and photosystem (PS) I and PSII activity assays suggested that the pyrenoid tubules from several different algal lineages contain active PSI components and are depleted in PSII components (McKay and Gibbs, 1991). In contrast to these findings, we found that PSII components (PSBP3, PSBQ, PSBR) showed similar pyrenoid localization patterns to those of PSI (PSAG, PSAK and FDX1), cytochrome b_6f (CYC6), and ATP synthase (ATPC) components (Figure S3B).

Strikingly, we found that unlike other PSI components, the PSI protein PSAH was enriched within the pyrenoid tubules (Figure 4B). PSAH is a 130 amino-acid protein with a single transmembrane helix that in land plants binds to the core PSI at the site where light harvesting complex II (LHCII) docks in state transitions (Ben-Shem et al., 2003; Lunde et al., 2000). The enrichment of PSAH in the pyrenoid tubules could indicate an additional, pyrenoid-related, role for this protein in algae. Together, our localization data for pyrenoid components allow us to propose a model for the spatial organization of the pyrenoid (Figure 4E).

We Generated a Spatially Defined Protein-Protein Interaction Network of the CCM

To understand the interconnectivity of the protein components of the CCM, we developed a large-scale affinity purification mass spectrometry (AP-MS) approach. We chose 38 candidates for AP-MS, focusing on proteins previously implicated in the CCM and on those we found in the pyrenoid (Table S3). We affinity purified fusion proteins using their 3xFLAG tag.

To aid in filtering out nonspecific bait-prey interactions from true interactions, we used ^{15}N labeling. We affinity purified baits and associated proteins from lines grown in ^{14}N media, and, before mass spectrometry, we mixed each sample with affinity-purified Venus-3xFLAG and associated proteins from lines grown in ^{15}N media. We quantified our confidence in each protein-prey interaction with a modified weighted D-score (WD-score) (Behrends et al., 2010), which incorporates the reproducibility, specificity and abundance of each interaction (Figure 5A; see the STAR Methods).

To identify high confidence interactions, we assumed that interactions between baits and preys localized to different organelles in our study are nonspecific, and thus the distribution of their WD-scores approximates the distribution of WD-scores for false-positive interactions. We took the highest WD-score value of 47.5 in this subset and used it as a cut-off. Approximately 3.8% of the interactions had WD-scores above this value, giving 513 interactions involving 398 proteins (Figures 5B and 5C). These proteins were considered high-confidence

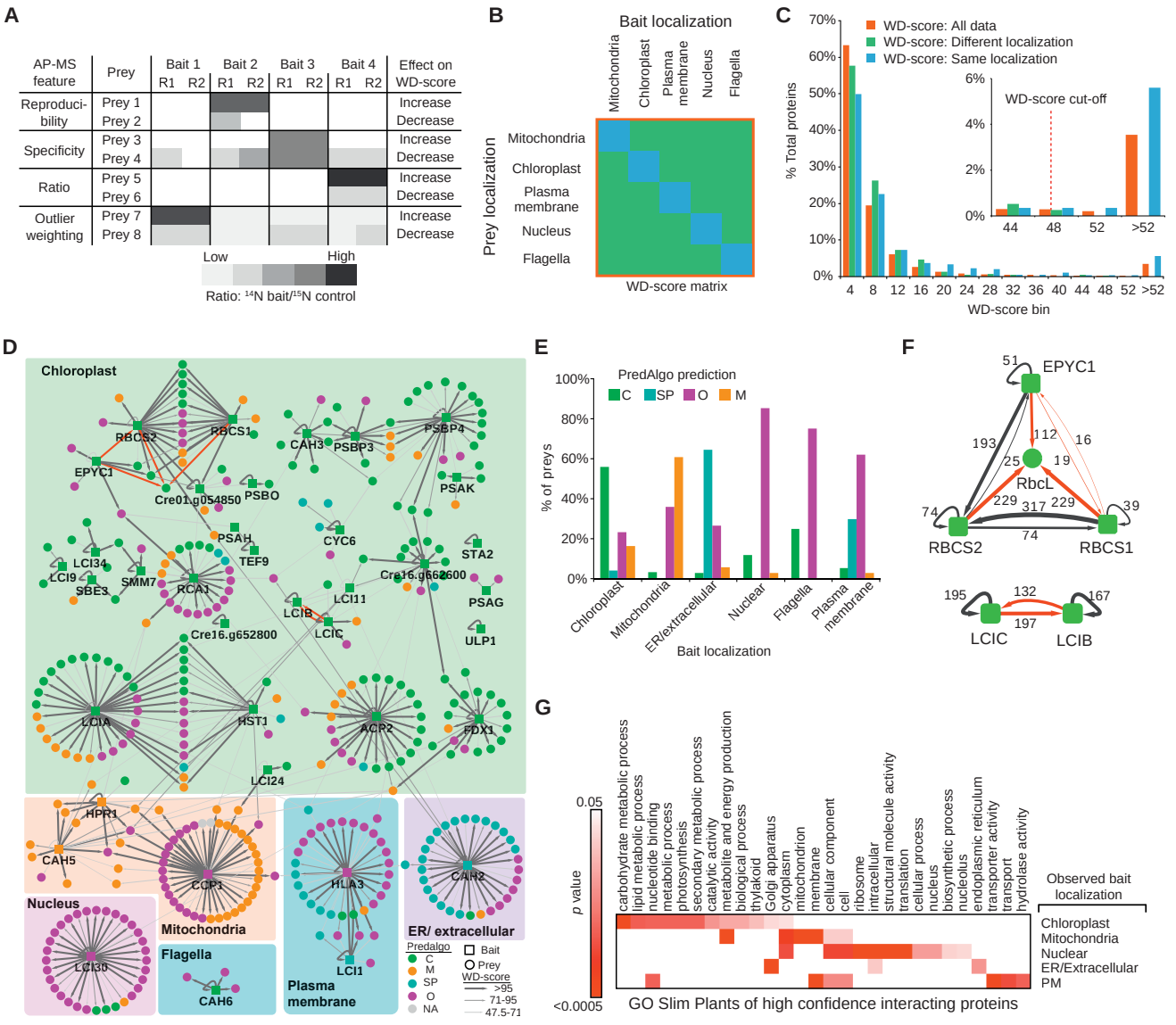


Figure 5. The AP-MS Data Are of High Quality

(A) Illustration of the influence of different AP-MS features (reproducibility, specificity, ratio, and outlier weighting) on the WD-score. R1 and R2 represent replica 1 and 2.

(B) To determine a WD-score cut-off value, a bait-prey matrix of WD-scores was formed containing only baits and preys whose localizations were determined in this study. The WD-scores from this matrix were then used to generate (C).

(C) A histogram of WD-scores for "All data," "Different localization," "Same localization." A conservative WD-score cut-off was chosen as the point where all data fell above the highest "Different localization" WD-score. Proteins with a WD-score greater than the cut-off are classified as high confidence interacting proteins (HCIPs).

(D) Protein-protein interaction network of baits and HCIPs. Bait proteins are grouped according to their localization pattern as determined by confocal microscopy. Baits and preys are colored based on their predicted localization by PredAlgo. Previously known interactions are indicated by red arrows.

(E) Comparison of prey PredAlgo predictions with bait localization. C, chloroplast; SP, secretory pathway; O, Other; M, mitochondria.

(F) Confirmation of known interactions from the literature (red arrows). Values are WD-scores.

(G) Significantly enriched gene ontology (GO) terms for interactors of baits localized to different cellular structures.

See also Figure S4 and Table S3.

interacting proteins (HCIPs). This method is more stringent than previous methods in which a simulated dataset was used to determine a cut-off, resulting in ~5% of data being considered as HCIPs (Behrends et al., 2010; Sowa et al., 2009). One inherent limitation of AP-MS is that it cannot distinguish between direct and indirect interactions, for example this can result in large protein complexes being affinity purified even though a bait protein only directly interacts with one member of the complex.

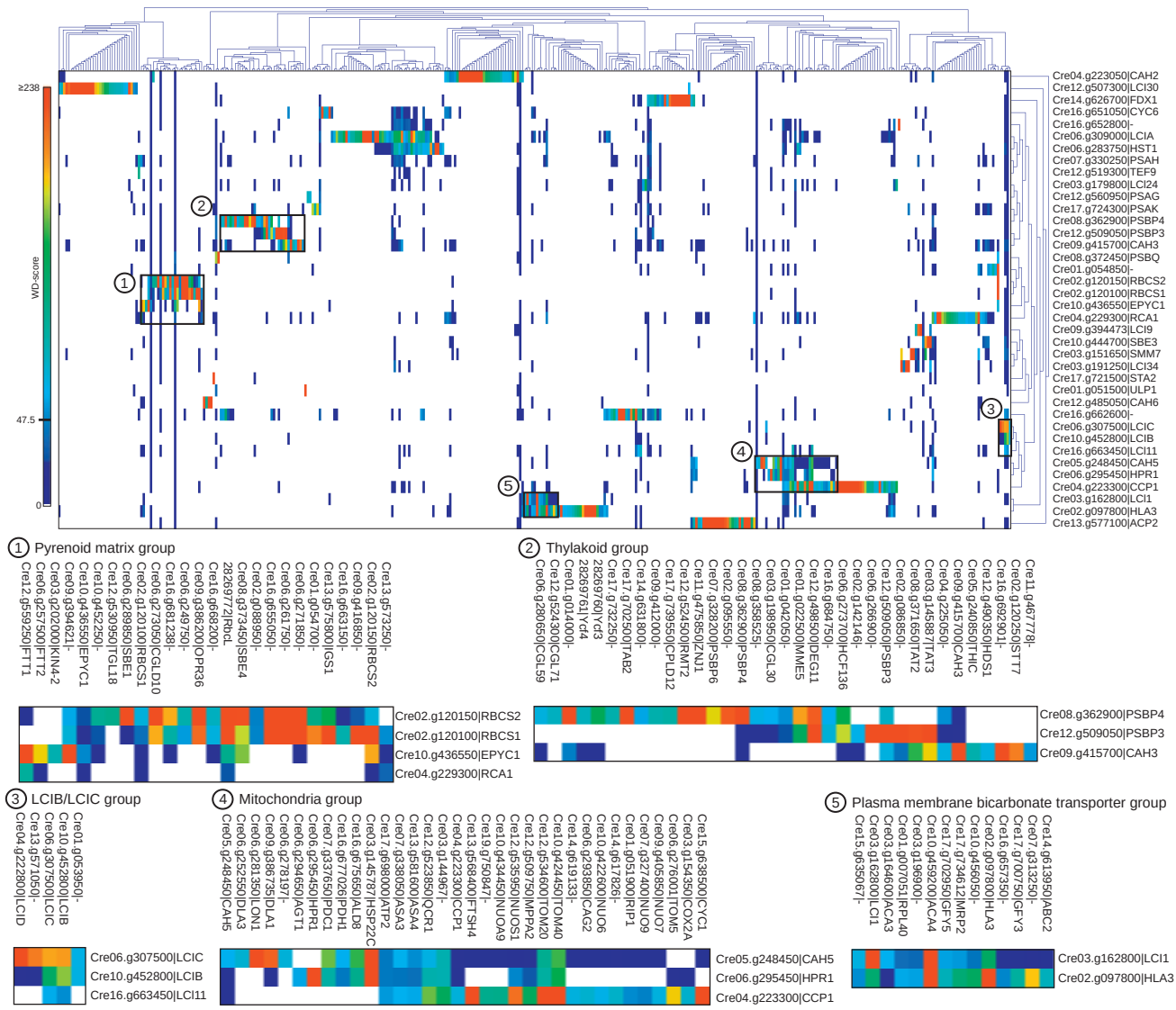


Figure 6. The AP-MS Data Reveal Previously Undescribed Physical Interactions including that Inorganic Carbon Transporters LCI1 and HLA3 Form a Physical Complex

Hierarchical clustering of all 38 baits with 398 HCIP preys. Specific groups of interest are boxed and highlighted below. Clustering of all baits and preys with interaction WD-scores ≥ 1 is provided in Figure S5.

See also Figures S6 and S7 and Tables S4 and S5.

We Used Multiple Approaches to Validate the Network

HCIPs of baits were enriched for proteins with the same PredAlgo predicted localizations (Figures 5D and 5E). HCIPs recapitulated previously known physical interactions of Rubisco subunits, EPYC1, LCIB, and LCIC (Figure 5F). HCIPs of baits from a specific compartment (i.e., chloroplast) are significantly enriched in Gene Ontology function and localization terms related to that compartment (Figure 5G). Finally, as expected from tight transcriptional control of subunit stoichiometry in most complexes (Jansen et al., 2002), most HCIPs were transcriptionally co-regulated with their baits in response to high CO_2 (Figure S4).

We Identified Many Novel Rubisco-Interacting Proteins

To identify novel protein complexes and new members of known complexes, we performed hierarchical clustering on HCIPs (Figure 6; see Figure S5 for all bait-prey interactions with a WD-score ≥ 1). The baits RBCS1 and RBCS2 clustered together and shared 15 HCIPs, 4 of which were also HCIPs of EPYC1. RBCS1- and RBCS2-associated proteins were enriched in uncharacterized proteins. Several of these interactors have homologs in other green algae but lack any conserved domains (Cre01.g054700, Cre01.g054850, Cre02.g088950, Cre16.g655050). We found that Cre16.g655050 contains a predicted N-terminal RbcX fold, which is found in a class of Rubisco

chaperones, and the rest of the protein is predicted to be disordered (Figure S6). A BLAST analysis using Cre16.g655050 as the query showed that its full sequence is conserved in the closely related species *Volvox carteri* and *Gonium pectorale.* The N-terminal RbcX-like region is conserved in several more evolutionarily distant Chlorophytes such as *Micromonas pusilla* (Table S4). Whether Cre16.g655050 is a chaperone for Rubisco or performs an alternative function is unknown.

Carbohydrate binding domains were found in three Rubisco interactors, including the two starch branching enzymes, SBE1 and SBE4, the latter of which also interacts with EPYC1. Given the concave shape of the pyrenoid-surrounding starch sheaths, there may be variation in starch synthesis and/or breakdown occurring between the two faces. One way to target a subset of starch metabolic enzymes to the inner concave face would be through a binding interaction with pyrenoid matrix proteins. The functional roles of the different SBE isoforms in *Chlamydomonas* have yet to be determined.

Interestingly, RBCS1 and RBCS2 interact with an ATP binding cassette (ABC) family transporter (Cre06.g271850). The specific role of this protein may help elucidate transmembrane transport processes occurring across pyrenoid tubules.

EPYC1 Interacts with a Kinase and Two 14-3-3 Proteins

The putative Rubisco linker protein EPYC1 is phosphorylated at low CO_2 (Turkina et al., 2006). Interestingly, we see that EPYC1 associates with a predicted serine/threonine protein kinase (KIN4-2; Cre03.g202000). Understanding the role of this kinase may shed light on post-translational modifications associated with pyrenoid biogenesis and/or function.

EPYC1 interacts with two 14-3-3 proteins FTT1 and FTT2. 14-3-3 proteins are known to bind phosphorylated proteins; hence the interaction of 14-3-3 proteins with EPYC1 could potentially be regulated by the phosphorylation state of EPYC1. 14-3-3 proteins can influence the stability, function, interactions, and localization of their targets (Chevalier et al., 2009). It is therefore possible that these 14-3-3 proteins are regulating an interaction between EPYC1 and Rubisco, possibly by changing the availability of protein-binding domains.

CAH3 Interacts with TAT proteins and STT7

The carbonic anhydrase CAH3 is essential for the CCM (Karlsson et al., 1998) and is thought to convert HCO_3^- to CO_2 in the thylakoid membranes that traverse the pyrenoid, supplying the pyrenoid with a high concentration of CO_2. In our study, CAH3 associated with the TAT2 and TAT3 proteins of the twin-arginine translocation (Tat) pathway (Figures 6 and 7; Table S5), which delivers substrate proteins to the thylakoid lumen. This observation is consistent with work showing that CAH3 contains a predicted Tat signal peptide (Benlloch et al., 2015) and with previous biochemical studies suggesting that CAH3 localizes to the thylakoid lumen (Karlsson et al., 1998).

At low CO_2, CAH3 is phosphorylated, and this phosphorylation correlates with increased CA activity and localization to the pyrenoid (Blanco-Rivero et al., 2012). Here, we find that CAH3 has a strong interaction (WD-score = 209) with the kinase STT7 (Figure 6). The role of STT7 in LHCII phosphorylation and state transitions is well documented (Depège et al., 2003). However, it is unlikely that STT7 directly phosphorylates CAH3, because the kinase domain of STT7 has been shown to be on the stromal side (Lemeille et al., 2009), and CAH3 is thought to be localized in the lumen (Karlsson et al., 1998). A direct interaction between STT7 and CAH3 may be occurring via the N terminus of STT7, which is thought to be luminal via a single membrane traversing domain (Lemeille et al., 2009).

PSBP4 Is in a Complex with PSI Assembly Factors

PSBP4 is a PsbP domain (PPD)-containing protein whose *Arabidopsis* homolog is essential for photosystem I assembly and function (Liu et al., 2012). In our data, PSBP4 interacted with four proteins associated with PSI assembly: Ycf3, Ycf4, CGL71, and TAB2 (Heinnickel et al., 2016; Rochaix et al., 2004), suggesting that PSBP4 and these factors form a PSI assembly complex. PSBP4 also interacts with three uncharacterized conserved green lineage proteins (CGL30, CGL59, and CPLD12) and nine other proteins of unknown function (Figure 7), indicating that these proteins may have roles in PSI assembly and function. Notably, PSBP4's localization suggests that PSI assembly occurs at the pyrenoid periphery.

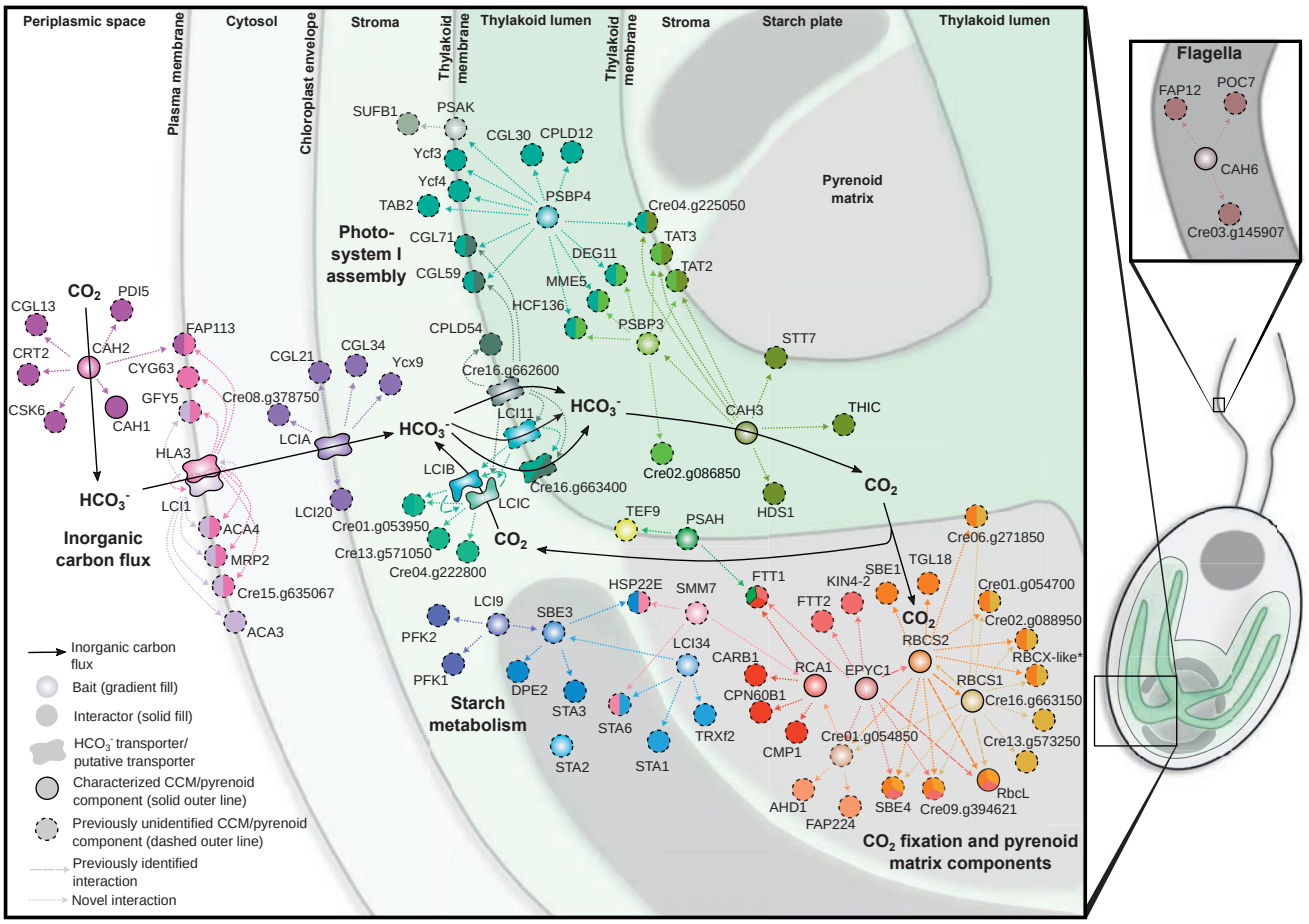
The LCIB/LCIC Complex Interacts with Two Bestrophin-like Proteins

Our data confirm that LCIB and LCIC, known stromal soluble proteins, are in a complex (Yamano et al., 2010). The *lcib* mutant has an "air-dier" phenotype: it exhibits wild-type (WT) growth in either very low CO_2 (0.01% CO_2 v/v) or high CO_2 (3% v/v) but dies in air levels of CO_2 (0.04%) (Wang and Spalding, 2006). The functional role of the LCIB/C complex is still unknown. This complex is hypothesized to either form a CO_2 leakage barrier at the pyrenoid periphery or to act as a vectorial CO_2 to HCO_3^- conversion module to recapture CO_2 that escapes from the pyrenoid (Wang et al., 2015). A role in the conversion of CO_2 to HCO_3^- is likely, as several homologs of LCIB were recently shown to be functional β-carbonic anhydrases. However, recombinant LCIB/C had no carbonic anhydrase function (Jin et al., 2016), suggesting that the complex may be tightly regulated or may require additional factors for proper function.

Both LCIB and LCIC interact with LCI11 (Cre16.g663450), and LCIC also interacts with Cre16.g662600 (Figures 6 and 7). Both LCI11 and Cre16.g662600 are putative bestrophins, which typically transport chloride but have been shown to be permeable to HCO_3^- (Qu and Hartzell, 2008). Furthermore, both proteins are upregulated at low CO_2 levels (Figure S4; Table S1). LCI11 and Cre16.g662600 directly interact and both also interact with another bestrophin-like protein, Cre16.g663400.

LCI9 Interacts with PFK1, PFK2, and SBE3 to Form a Carbohydrate Metabolism Module

As described above, LCI9 forms a mesh structure, likely in the gaps between starch plates. LCI9 contains two CBM20 (carbohydrate binding module 20) domains and is predicted to function as a glucan 1,4-α-glucosidase. Glucan 1,4-α-glucosidases hydrolyze glucosidic bonds, releasing glucose monomers from glucan chains. Therefore, LCI9 most likely plays a role in starch breakdown at the pyrenoidal starch plate junctions. AP-MS analysis shows that the strongest HCIPs of LCI9 are PFK1 and PFK2



Pyrenoid core

*Cre16.g655050|RBCX-like, Cre06.g273050|CGLD10, Cre09.g386200|OPR36, + 5 more HCIPs

Cre10.g429100, Cre03.g213313

Cre10.g452250

Cre09.g416850

Cre03.g156750|CCT5, Cre03.g168450|CCT8, Cre05.g232750|CNK4, + 16 more HCIPs

Thylakoids

Cre11.g467778, Cre16.g692901

Cre06.g266900, Cre02.g142146

Cre16.g684750

Cre01.g014000

Cre12.g524500|RMT2, Cre07.g328200|PSBP6, Cre11.g475850|ZNJ1, + 6 more HCIPs

Cre17.g729800|ALB3.2, Cre01.g056696, Cre03.g176833

Cre07.g339750|FeC, Cre19.g750547|NDA2, Cre02.g111450|TEF4, + 8 more HCIPs

Chloroplast envelope

Cre05.g230850, 28269781|Ycf78, Cre13.g587100|PCYA1, + 34 more HCIPs

Plasma membrane

Cre08.g360600|ERM4, Cre14.g613950|ABC2, Cre16.g657350, Cre17.g700750|GFY3, + 25 more HCIPs

Extracellular

Cre03.g144564|MMP13, Cre02.g093750|NRX2, Cre09.g393700|MMP3, + 21 more HCIPs

Figure 7. Combining Localization, Protein-Protein Interaction, and Protein Function Data Reveal a Spatially Defined Interactome of the *Chlamydomonas* CCM

A spatially defined protein-protein interaction model of the CCM. Baits have a gradient fill, prey have a solid fill. Each bait has a unique color. Prey are colored according to their bait, with proteins that interact with multiple baits depicted as pies with each slice colored according to one of their interacting baits. Interactors are connected to their bait by a dashed line representing the direction of interaction. Baits are arranged based on their localization observed in this study. Interactors with predicted transmembrane domains are placed on membranes. Prey of membrane localized baits lacking transmembrane domains are arranged according to their PredAlgo localization prediction. Solid black arrows indicate inorganic carbon flux through the cell. For clarity, a selection of interactors are not included in the map but are highlighted below. All interaction data with corresponding WD-scores can be found in Table S5.

(phosphofructokinases 1 and 2). PFK is a key regulator of glycolysis and is important for maintaining cellular ATP levels (Johnson and Alric, 2013). The exact metabolic role of an LCI9, PFK1, and PFK2 assemblage is still unclear. LCI9 also associates with SBE3, which in turn associates with STA3 and DPE2 (disproportionating enzyme 2), a putative α-1,4-glucanotransferase.

Because SBE3 and its HCIPs are involved in starch synthesis and modification, enzymes catalyzing starch breakdown and starch synthesis are potentially in close proximity, allowing tight regulation of starch structure. It should be noted that a caveat of performing AP-MS on proteins containing CBMs is that proteins could co-precipitate due to binding a common carbohydrate substrate, not due to direct protein-protein interactions.

Bicarbonate Transporters LCI1 and HLA3 Form a Complex with a P-type ATPase

HLA3 (high light activated 3) and LCI1 have both been implicated in HCO_3^- uptake at the plasma membrane (Ohnishi et al., 2010; Yamano et al., 2015). HLA3 is an ABC transporter, and its absence under low-CO_2 conditions results in a reduced uptake of inorganic carbon by *Chlamydomonas* cells (Yamano et al., 2015). HLA3 expressed in *Xenopus* oocytes showed moderate uptake of HCO_3^- (Atkinson et al., 2016). LCI1 lacks any conserved functional or structural domains and contains four predicted transmembrane regions. Knock-down of LCI1 protein resulted in a small reduction in inorganic carbon uptake (Ohnishi et al., 2010); however, the function of LCI1 has not been demonstrated in a heterologous system.

Unexpectedly, we found that HLA3 and LCI1 are found together in a complex. The two proteins showed a reciprocal, strong interaction, each having WD-scores >125. In addition, they appear to be in a complex with ACA4 (autoinhibited Ca^{2+}-ATPase 4; Cre10.g459200), a P-type ATPase/cation transporter. Alignment of ACA4 with functionally characterized P-type ATPases shows that it is a member of the group IIIA family of P-type ATPases (Figure S7). Group IIIA members are known H^+-exporting ATPases (Thever and Saier, 2009). ACA4 may be aiding HCO_3^- uptake either by maintaining a H^+ gradient that HLA3 and/or LCI1 is using to drive HCO_3^- uptake, or by generating localized cytosolic alkaline regions similar to those that form near anion exchanger I during HCO_3^- uptake (Johnson and Casey, 2011). A localized alkaline region could decrease HCO_3^- to CO_2 conversion and hence diffusion out of the cell.

The regulation of inorganic carbon transport is critical for the efficiency of the CCM. Recent work has shown that Ca^{2+} signaling is key for proper regulation of the CCM, with the Ca^{2+}-binding protein CAS1 transcriptionally regulating HLA3 and other components (Wang et al., 2016). One HCIP of HLA3 is an EF-hand-containing Ca^{2+}/calmodulin-dependent protein kinase (Cre13.g571700), which could potentially regulate HLA3 post-translationally. Additionally, HLA3 physically interacts with an adenylate/guanylate cyclase (CYG63: Cre05.g236650). Adenylate and guanylate cyclases are known to play a role in sensing inorganic carbon across a broad range of taxa (Tresguerres et al., 2010). Thus, Cre13.g571700 and Cre05.g236650 may represent another mode of CCM regulation, possibly by sensing inorganic carbon availability at the plasma membrane.

Perspective

By developing an efficient fluorescent protein-tagging and AP-MS pipeline in *Chlamydomonas*, we have generated a spatially defined network of the *Chlamydomonas* CCM. This large-scale approach gives a comprehensive view of the CCM by revealing missing components, by redefining the localization of others, and by identifying specific protein-protein interactions. Our work also provides insight into the function and regulation of these known and newly discovered CCM proteins and represents a valuable resource for their further characterization.

Our observation that the pyrenoid matrix appears to exclude proteins larger than ~78 kDa may be related to the liquid-like nature of the matrix (Freeman Rosenzweig et al., 2017). Interestingly, another liquid-like non-membrane organelle, the *C. elegans* P granule, shows size exclusion of fluorescently labeled dextrans 70 kDa and larger (Updike et al., 2011). This behavior may result from surface tension generated by the proteins that produce the liquid phase (Bergeron-Sandoval et al., 2016).

Our results suggest changes to the existing model of inorganic carbon flux to the pyrenoid (Figure 7). The apparent absence of carbonic anhydrase in the chloroplast stroma aligns the *Chlamydomonas* CCM model more with the cyanobacterial model, in which the absence of carbonic anhydrase in the cytosol is critical for inorganic carbon accumulation in the form of HCO_3^- (Price and Badger, 1989; Price et al., 2008). The localization of the carbonic anhydrase CAH6 in flagella suggests potential roles in inorganic carbon sensing. Furthermore, the discovery that HLA3 and LCI1 form a complex and the identification of potential regulatory factors of this complex will aid in the characterization and ultimately the reconstitution of this key plasma membrane bicarbonate transport pathway.

Due to a rapidly rising global population and a finite agricultural land area, novel approaches are essential to maintain food security. One potential approach for improving yields is the transfer of a CCM into higher plants to increase CO_2-fixation rates (Long et al., 2015). Recent work has found that nearly all algal CCM proteins localize correctly in higher plants with no changes to their protein sequence, suggesting that the transfer of algal components could be relatively straightforward (Atkinson et al., 2016). However, engineering efforts were constrained by our limited knowledge of the components of the algal CCM. The work we present here provides a detailed blueprint of the algal CCM, revealing dozens of new targets for transfer into crop plants to improve carbon fixation and enhancing our basic molecular understanding of a fundamental cellular process that drives global biogeochemical cycles.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Strains and Culturing
- METHOD DETAILS
  - Plasmid Construct and Cloning
  - Microscopy
  - Affinity Purification
  - Mass Spectrometry

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Mass Spectrometry Data Analysis
  - Comparison of Localization Data with PredAlgo and TargetP
  - Gene Expression Values and Presence of Upstream ATGs
  - P-Type ATPase Tree Assembly
  - GO Term Analysis
  - Transmembrane and Protein Disorder Prediction
  - Pyrenoid Enrichment Analysis
  - Statistical tests
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.044.

AUTHOR CONTRIBUTIONS

L.C.M.M. and M.C.J. designed and supervised the study. L.C.M.M., C.C., and M.R. performed the cloning. L.C.M.M. did the microscopy. L.C.M.M. and C.C. carried out the AP-MS. S.R. and L.C.M.M. developed the affinity purification protocol. W.P. and S.R.B. provided bioinformatics support. R.D.L. and C.M.A. oversaw the mass spectrometry and peptide mapping. L.C.M.M., C.C., and M.C.J. analyzed and interpreted the data. L.C.M.M created the figures. C.C. created the online viewing platform. L.C.M.M. and M.C.J. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We thank Jonikas laboratory members for helpful discussions, Z. Friedberg and R. Vasquez for help with gene cloning, A. Okumu for mass spectrometry sample preparation, H. Cartwright for microscopy support, and U. Goodenough, R. Milo, A. Smith, N. Wingreen, T. Silhavy, M. Meyer, and A. McCormick for comments on the manuscript. Stanford University Mass Spectrometry is thankful to the NIH NCRR (S10RR027425) for assistance in purchasing the mass spectrometer. The project was funded by NSF (EF-1105617 and IOS-1359682), NIH (7DP2GM119137-02), the Simons Foundation and HHMI (55108535), Princeton University (M.C.J.), the University of York (L.C.M.M), the EMBO LTF (S.R.), the SNSF APM (S.R.), and the Carnegie Institution for Science (L.C.M.M. and M.C.J.). The authors wish to note that the Carnegie Institution for Science has submitted a provisional patent application on aspects of the findings.

Received: April 13, 2017
Revised: June 30, 2017
Accepted: August 22, 2017
Published: September 21, 2017

REFERENCES

Atkinson, N., Feike, D., Mackinder, L.C.M., Meyer, M.T., Griffiths, H., Jonikas, M.C., Smith, A.M., and McCormick, A.J. (2016). Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. Plant Biotechnol. J. *14*, 1302–1315.

Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., Martin, W., et al. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. Mol. Biol. Evol. *26*, 1533–1548.

Badger, M. (2003). The roles of carbonic anhydrases in photosynthetic CO_2 concentrating mechanisms. Photosynth. Res. *77*, 83–94.

Bauwe, H., Hagemann, M., and Fernie, A.R. (2010). Photorespiration: players, partners and origin. Trends Plant Sci. *15*, 330–336.

Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. Nature *466*, 68–76.

Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. Nature *426*, 630–635.

Benlloch, R., Shevela, D., Hainzl, T., Grundström, C., Shutova, T., Messinger, J., Samuelsson, G., and Sauer-Eriksson, A.E. (2015). Crystal structure and functional characterization of photosystem II-associated carbonic anhydrase CAH3 in *Chlamydomonas reinhardtii*. Plant Physiol. *167*, 950–962.

Bergeron-Sandoval, L.-P., Safaee, N., and Michnick, S.W. (2016). Mechanisms and consequences of macromolecular phase separation. Cell *165*, 1067–1079.

Blanco-Rivero, A., Shutova, T., Román, M.J., Villarejo, A., and Martinez, F. (2012). Phosphorylation controls the localization and activation of the lumenal carbonic anhydrase in *Chlamydomonas reinhardtii*. PLoS ONE *7*, e49063.

Brueggeman, A.J., Gangadharaiah, D.S., Cserhati, M.F., Casero, D., Weeks, D.P., and Ladunga, I. (2012). Activation of the carbon concentrating mechanism by CO_2 deprivation coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. Plant Cell *24*, 1860–1875.

Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell *155*, 1131–1140.

Chevalier, D., Morris, E.R., and Walker, J.C. (2009). 14-3-3 and FHA domains mediate phosphoprotein interactions. Annu. Rev. Plant Biol. *60*, 67–91.

Choi, H.I., Kim, J.Y.H., Kwak, H.S., Sung, Y.J., and Sim, S.J. (2016). Quantitative analysis of the chemotaxis of a green alga, *Chlamydomonas reinhardtii*, to bicarbonate using diffusion-based microfluidic device. Biomicrofluidics *10*, 014121.

Cross, F.R. (2016). Tying down loose ends in the Chlamydomonas genome: functional significance of abundant upstream open reading frames. G3 (Bethesda) *6*, 435–446.

Depège, N., Bellafiore, S., and Rochaix, J.-D. (2003). Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in *Chlamydomonas*. Science *299*, 1572–1575.

Dismukes, G.C., Klimov, V.V., Baranov, S.V., Kozlov, Y.N., DasGupta, J., and Tyryshkin, A. (2001). The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. Proc. Natl. Acad. Sci. USA *98*, 2170–2175.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods *4*, 207–214.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. *300*, 1005–1016.

Engel, B.D., Ishikawa, H., Wemmer, K.A., Geimer, S., Wakabayashi, K., Hirono, M., Craige, B., Pazour, G.J., Witman, G.B., Kamiya, R., and Marshall, W.F. (2012). The role of retrograde intraflagellar transport in flagellar assembly, maintenance, and function. J. Cell Biol. *199*, 151–167.

Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., and Baumeister, W. (2015). Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. eLife *4*, e04889.

Fang, W., Si, Y., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., Ladunga, I., Liu, P., and Spalding, M.H. (2012). Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1. Plant Cell *24*, 1876–1893.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. Science *281*, 237–240.

Freeman Rosenzweig, E.S., Xu, B., Cuellar, L.K., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H.N., Ronceray, P., Plitzko, J.M., Förster, F., et al. (2017). The eukaryotic CO_2-concentrating organelle is liquid like and exhibits dynamic reorganization. Cell *171*, this issue, 148–162.

Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. Plant J. *19*, 353–361.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Heinnickel, M., Kim, R.G., Wittkopp, T.M., Yang, W., Walters, K.A., Herbert, S.K., and Grossman, A.R. (2016). Tetratricopeptide repeat protein protects photosystem I from oxidative disruption during assembly. Proc. Natl. Acad. Sci. USA *113*, 2774–2779.

Hu, H., Boisson-Dernier, A., Israelsson-Nordström, M., Böhmer, M., Xue, S., Ries, A., Godoski, J., Kuhn, J.M., and Schroeder, J.I. (2010). Carbonic anhydrases are upstream regulators of CO_2-controlled stomatal movements in guard cells. Nat. Cell Biol. *12*, 87–93, 1–18.

Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. Genome Res. *12*, 37–46.

Jin, S., Sun, J., Wunder, T., Tang, D., Cousins, A.B., Sze, S.K., Mueller-Cajar, O., and Gao, Y.-G. (2016). Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. Proc. Natl. Acad. Sci. USA *113*, 14716–14721.

Johnson, X., and Alric, J. (2013). Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch. Eukaryot. Cell *12*, 776–793.

Johnson, D.E., and Casey, J.R. (2011). Cytosolic H^+ microdomain developed around AE1 during AE1-mediated Cl^-/HCO_3^- exchange. J. Physiol. *589*, 1551–1569.

Karcher, D., Köster, D., Schadach, A., Klevesath, A., and Bock, R. (2009). The *Chlamydomonas* chloroplast HLP protein is required for nucleoid organization and genome maintenance. Mol. Plant *2*, 1223–1232.

Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., Moroney, J.V., and Samuelsson, G. (1998). A novel α-type carbonic anhydrase associated with the thylakoid membrane in *Chlamydomonas reinhardtii* is required for growth at ambient CO_2. EMBO J. *17*, 1208–1216.

Kobayashi, Y., Takusagawa, M., Harada, N., Fukao, Y., Yamaoka, S., Kohchi, T., Hori, K., Ohta, H., Shikanai, T., and Nishimura, Y. (2015). Eukaryotic components remodeled chloroplast nucleoid organization during the green plant evolution. Genome Biol. Evol. *8*, 1–16.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S.S., and Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. Plant J. *66*, 770–780.

Lemeille, S., Willig, A., Depège-Fargeix, N., Delessert, C., Bassi, R., and Rochaix, J.D. (2009). Analysis of the chloroplast protein kinase Stt7 during state transitions. PLoS Biol. *7*, e45.

Liu, J., Yang, H., Lu, Q., Wen, X., Chen, F., Peng, L., Zhang, L., and Lu, C. (2012). PsbP-domain protein1, a nuclear-encoded thylakoid lumenal protein, is essential for photosystem I assembly in Arabidopsis. Plant Cell *24*, 4992–5006.

Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell *161*, 56–66.

Lunde, C., Jensen, P.E., Haldrup, A., Knoetzel, J., and Scheller, H.V. (2000). The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. Nature *408*, 613–615.

Ma, Y., Pollock, S.V., Xiao, Y., Cunnusamy, K., and Moroney, J.V. (2011). Identification of a novel gene, CIA6, required for normal pyrenoid formation in *Chlamydomonas reinhardtii*. Plant Physiol. *156*, 884–896.

Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O., Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., et al. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc. Natl. Acad. Sci. USA *113*, 5958–5963.

McKay, R.M.L., and Gibbs, S.P. (1991). Composition and function of pyrenoids: cytochemical and immunocytochemical approaches. Can. J. Bot. *69*, 1040–1052.

Mitra, M., Lato, S.M., Ynalvez, R.A., Xiao, Y., and Moroney, J.V. (2004). Identification of a new chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiol. *135*, 173–182.

Moroney, J.V., Ma, Y., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., DiMario, R.J., Yang, J., and Mukherjee, B. (2011). The carbonic anhydrase isoforms of *Chlamydomonas reinhardtii*: intracellular location, expression, and physiological roles. Photosynth. Res. *109*, 133–149.

Morris, J.H., Knudsen, G.M., Verschueren, E., Johnson, J.R., Cimermancic, P., Greninger, A.L., and Pico, A.R. (2014). Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. Nat. Protoc. *9*, 2539–2554.

Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant J. *57*, 1140–1150.

Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and Fukuzawa, H. (2010). Expression of a low CO_2-inducible protein, LCI1, increases inorganic carbon uptake in the green alga *Chlamydomonas reinhardtii*. Plant Cell *22*, 3105–3117.

Pazour, G.J., Agrin, N., Leszyk, J., and Witman, G.B. (2005). Proteomic analysis of a eukaryotic cilium. J. Cell Biol. *170*, 103–113.

Price, G.D., and Badger, M.R. (1989). Expression of human carbonic anhydrase in the Cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype: evidence for a central role for carboxysomes in the CO_2 concentrating mechanism. Plant Physiol. *91*, 505–513.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J. Exp. Bot. *59*, 1441–1461.

Qu, Z., and Hartzell, H.C. (2008). Bestrophin Cl^- channels are highly permeable to HCO_3^-. Am. J. Physiol. Cell Physiol. *294*, C1371–C1377.

Reinfelder, J.R. (2011). Carbon concentrating mechanisms in eukaryotic marine phytoplankton. Annu. Rev. Mar. Sci. *3*, 291–315.

Rochaix, J.-D., Perron, K., Dauvillée, D., Laroche, F., Takahashi, Y., and Goldschmidt-Clermont, M. (2004). Post-transcriptional steps involved in the assembly of photosystem I in *Chlamydomonas*. Biochem. Soc. Trans. *32*, 567–570.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403.

Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., et al. (2012). PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol. Biol. Evol. *29*, 3625–3639.

Terashima, M., Specht, M., Naumann, B., and Hippler, M. (2010). Characterizing the anaerobic response of *Chlamydomonas reinhardtii* by quantitative proteomics. Mol. Cell. Proteomics *9*, 1514–1532.

Thever, M.D., and Saier, M.H., Jr. (2009). Bioinformatic characterization of p-type ATPases encoded within the fully sequenced genomes of 26 eukaryotes. J. Membr. Biol. *229*, 115–130.

Tresguerres, M., Buck, J., and Levin, L.R. (2010). Physiological carbon dioxide, bicarbonate, and pH sensing. Pfluegers Arch. *460*, 953–964.

Turkina, M.V., Blanco-Rivero, A., Vainonen, J.P., Vener, A.V., and Villarejo, A. (2006). CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. Proteomics *6*, 2693–2704.

Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in the *C. elegans* germ line. J. Cell Biol. *192*, 939–948.

Wang, Y., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for acclimation specific to air levels of CO_2 in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA *103*, 10110–10115.

Wang, Y., Stessman, D.J., and Spalding, M.H. (2015). The CO_2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against the gradient. Plant J. *82*, 429–448.

Wang, L., Yamano, T., Takane, S., Niikawa, Y., Toyokawa, C., Ozawa, S.I., Tokutsu, R., Takahashi, Y., Minagawa, J., Kanesaki, Y., et al. (2016). Chloroplast-mediated regulation of CO_2-concentrating mechanism by Ca^{2+}-binding protein CAS in the green alga *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA *113*, 12586–12591.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010). Light and low-CO_2-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Cell Physiol. *51*, 1453–1468.

Yamano, T., Sato, E., Iguchi, H., Fukuda, Y., and Fukuzawa, H. (2015). Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA *112*, 7315–7320.

Cell

Article

The Eukaryotic CO_2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization

Elizabeth S. Freeman Rosenzweig,[1,2] Bin Xu,[3,11] Luis Kuhn Cuellar,[4,11] Antonio Martinez-Sanchez,[4] Miroslava Schaffer,[4] Mike Strauss,[5] Heather N. Cartwright,[2] Pierre Ronceray,[6] Jürgen M. Plitzko,[4] Friedrich Förster,[4,9] Ned S. Wingreen,[7,8,*] Benjamin D. Engel,[4,*] Luke C.M. Mackinder,[2,10,12] and Martin C. Jonikas[1,2,8,11,12,13,*]

[1]Department of Biology, Stanford University, Stanford, CA 94305, USA
[2]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA
[3]Department of Physics, Princeton University, Princeton, NJ 08544, USA
[4]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[5]Cryo-EM Facility, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[6]Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA
[7]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[8]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[9]Present address: Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 Utrecht, the Netherlands
[10]Present address: Department of Biology, University of York, York YO10 5DD, UK
[11]These authors contributed equally
[12]These authors contributed equally
[13]Lead Contact
*Correspondence: wingreen@princeton.edu (N.S.W.), engelben@biochem.mpg.de (B.D.E.), mjonikas@princeton.edu (M.C.J.)
http://dx.doi.org/10.1016/j.cell.2017.08.008

SUMMARY

Approximately 30%–40% of global CO_2 fixation occurs inside a non-membrane-bound organelle called the pyrenoid, which is found within the chloroplasts of most eukaryotic algae. The pyrenoid matrix is densely packed with the CO_2-fixing enzyme Rubisco and is thought to be a crystalline or amorphous solid. Here, we show that the pyrenoid matrix of the unicellular alga *Chlamydomonas reinhardtii* is not crystalline but behaves as a liquid that dissolves and condenses during cell division. Furthermore, we show that new pyrenoids are formed both by fission and *de novo* assembly. Our modeling predicts the existence of a "magic number" effect associated with special, highly stable heterocomplexes that influences phase separation in liquid-like organelles. This view of the pyrenoid matrix as a phase-separated compartment provides a paradigm for understanding its structure, biogenesis, and regulation. More broadly, our findings expand our understanding of the principles that govern the architecture and inheritance of liquid-like organelles.

INTRODUCTION

Although pyrenoids mediate approximately one-third of global carbon fixation (Mackinder et al., 2016), the molecular structure and biogenesis of these biogeochemically fundamental organelles remain largely unknown. Pyrenoids are non-membrane-bound, proteinaceous structures that contain a matrix packed with the CO_2-fixing enzyme Rubisco. In many species, the matrix is traversed by membrane tubules that are continuous with the photosynthetic thylakoid membranes (Gibbs, 1962; Ohad et al., 1967; Engel et al., 2015). Pyrenoids are a central feature of the algal CO_2-concentrating mechanism, which supplies Rubisco with a high concentration of its substrate CO_2, enabling more efficient carbon capture than that of most land plants (Badger et al., 1998; Meyer et al., 2012). Pyrenoids are found within the chloroplasts of a diverse array of photosynthetic eukaryotes, including nearly all freshwater and marine algae, as well as a group of non-vascular plants (Brown et al., 1967; Wang et al., 2015). Although pyrenoids were among the first organelles to be scientifically documented (Vaucher, 1803), they have remained largely uncharacterized at a molecular level due to the limited availability of genetic tools for algae.

Based on results in the model unicellular green alga *Chlamydomonas reinhardtii* (*Chlamydomonas* hereafter), we recently proposed that a protein called Essential Pyrenoid Component 1 (EPYC1; also known as LCI5) links Rubisco holoenzymes together to form the pyrenoid matrix (Mackinder et al., 2016). EPYC1 localizes to the pyrenoid matrix, is of similar abundance to Rubisco, and is required for Rubisco's localization in the pyrenoid matrix, supporting a possible structural role. EPYC1 binds Rubisco, and the EPYC1 protein sequence consists of four nearly identical ~60 amino acid repeats, suggesting a model in which each of EPYC1's repeats contains a Rubisco binding site, allowing EPYC1 to link multiple Rubisco holoenzymes together. The molecular details of the EPYC1-Rubisco binding interaction, and the packing organization of EPYC1 and Rubisco in the matrix, remain unknown.

Electron micrographs of several species suggest that the densely packed pyrenoid matrix is crystalline (Holdsworth, 1968; Kowallik, 1969; Bertagnolli and Nadakavukaren, 1970), while micrographs of other species suggest the matrix is amorphous (Griffiths, 1970; Meyer et al., 2012). Our recent cryo-electron

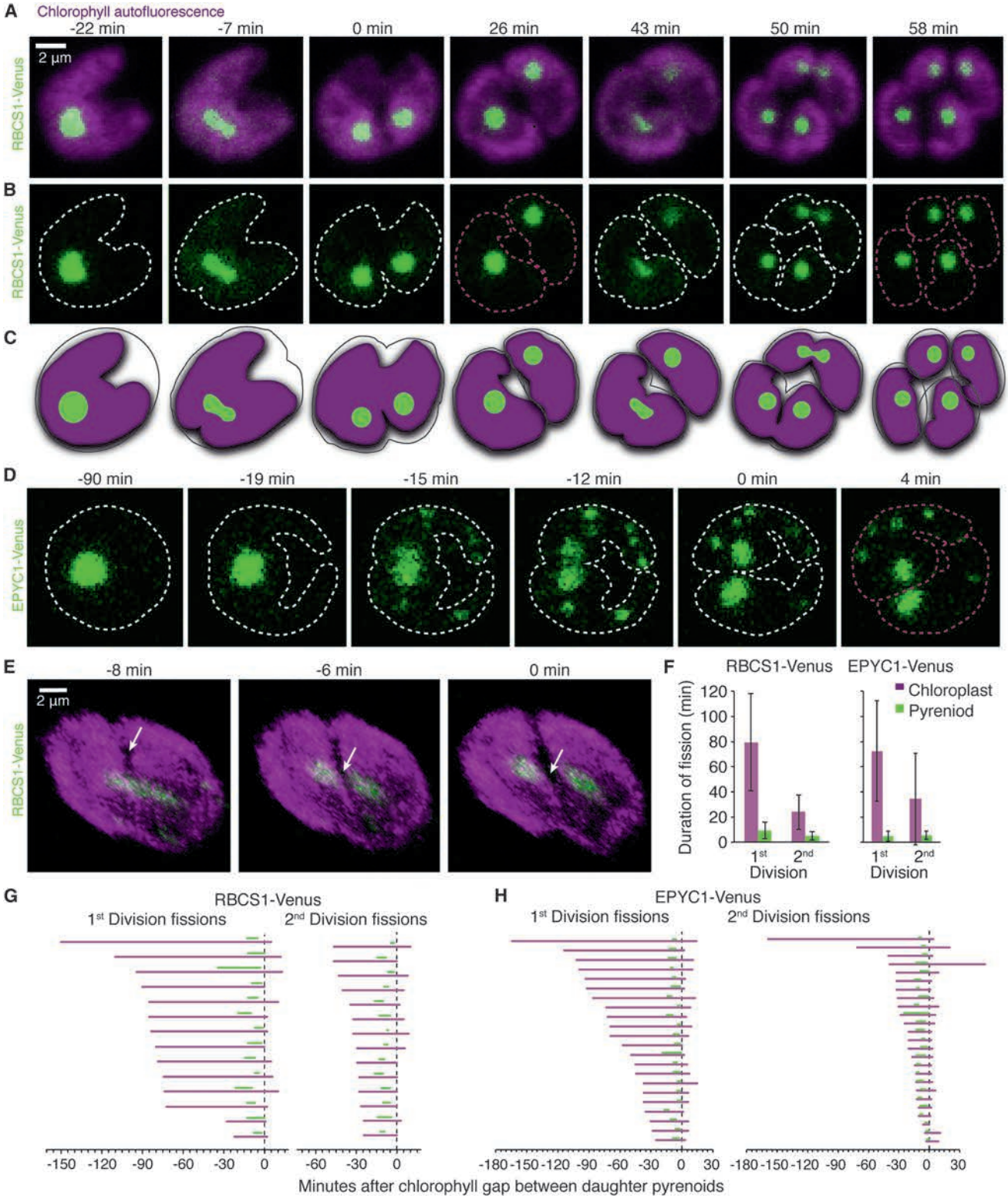


Figure 1. The Pyrenoid Is Inherited Primarily by Fission

(A and B) Confocal *Z* sum images of pyrenoid divisions by fission, with chlorophyll autofluorescence shown in magenta (A) and RBCS1-Venus in green (A and B); $t = 0$ is the first observation of a gap in chlorophyll between the daughter pyrenoids in the first division shown. Dashed curves (B) represent approximate chloroplast outlines in the mother (white) and daughter (pink) cells.

(legend continued on next page)

tomography (cryo-ET) study of *Chlamydomonas* revealed that the average packing of Rubisco molecules within the pyrenoid matrix resembles a hexagonal close-packed lattice (Engel et al., 2015). We have proposed that EPYC1 could link Rubisco holoenzymes together into this arrangement (Mackinder et al., 2016). However, if Rubisco is immobilized within a rigid lattice, it is not clear how Rubisco Activase, which regenerates Rubisco's active sites (Pollock et al., 2003) but is significantly less abundant than Rubisco in pyrenoids (McKay et al.,1991; Wienkoop et al., 2010; Mackinder et al., 2016), could reach enough Rubisco molecules to maintain efficient carbon fixation.

Moreover, pyrenoid biogenesis has not been definitively observed in living cells. Classic electron microscopy studies on fixed cells have suggested that pyrenoids may arise *de novo* (Brown et al., 1967; Retallack and Butler, 1970) in some species and by fission (Brown et al., 1967; Goodenough, 1970) in others.

Here, we describe our study of pyrenoid inheritance in living cells and how our findings inspired us to investigate the *Chlamydomonas* pyrenoid as a liquid. Examination of pyrenoid ultrastructure with improved *in situ* cryo-ET revealed that the pyrenoid matrix is not crystalline, but exhibits liquid-like local order. Further live-cell fluorescence experiments showed that the matrix mixes internally and disperses into the surrounding stroma during cell division, demonstrating that the matrix exhibits liquid-like properties. Finally, using a simple model inspired by the pyrenoid, we predict that phase transitions in some multicomponent, liquid-like biological systems are governed by changes in the ratio of available multivalent binding sites on the constituent macromolecules. This work provides insights into both algal CO_2-concentrating mechanisms and liquid-like biological systems.

RESULTS

The Pyrenoid Exhibits Both Fission and *De Novo* Assembly

To enable the first observations of pyrenoid matrix dynamics in living cells, we expressed pyrenoid matrix proteins tagged with the fluorescent protein Venus and imaged them in 3D with fluorescence time-lapse microscopy during photoautotrophic growth. We tracked inheritance of the pyrenoid matrix by monitoring Venus-tagged Rubisco small subunit 1 (RBCS1) or Venus-tagged EPYC1 and recorded chlorophyll autofluorescence to follow cellular orientation and chloroplast division. The fusion proteins localized to the pyrenoid (Figures 1A and 1D) and did not measurably perturb pyrenoid function (Figures S1A and S1B). During our observation, cells often divided twice in rapid succession, which is typical for *Chlamydomonas* (Cross and Umen, 2015).

We observed that approximately two-thirds of daughter cells inherited their pyrenoid by elongation and then fission of the mother cell's pyrenoid (Figures 1, 2E, 2F, and S2; Movie S1). The duration of pyrenoid fission, defined as the time between visible elongation of the mother pyrenoid and separation of the daughter pyrenoids, was ~7 min, much faster than chloroplast division (~30–80 min, Figure 1F). All pyrenoid fissions occurred during the final minutes of chloroplast division (Figures 1G and 1H).

In all 88 cell divisions in which pyrenoid fission was observed, the completion of pyrenoid division was quickly followed by a gap in the chlorophyll signal between the daughter pyrenoids, generated by the chloroplast cleavage furrow (Figures 1E, 1G, and 1H). Conversely, in the 44 cell divisions in which the furrow did not bisect the pyrenoid, pyrenoid fission did not occur. These findings support the hypothesis that pyrenoid fission is driven by the constrictive force of the chloroplast division furrow (Goodenough, 1970).

In cases where pyrenoid fission did not occur, one of the daughter cells typically inherited the mother cell's whole pyrenoid (~16%–21% of daughter cells; Figures 2A–2F; Movie S2), leaving the other daughter without a pyrenoid. Some of the cells that failed to inherit a pyrenoid remained without a visible pyrenoid for the duration of observation (~8%–9% of all daughter cells; Figures 2A, 2C, 2E, and 2F; Movie S2). In other cases, one or more fluorescent puncta appeared *de novo* and grew or coalesced into an apparent pyrenoid (~6%–7% of all daughter cells; Figures 2B, 2D, and 2E–2I; Movie S3). The proportions of each inheritance pattern in RBCS1-Venus and EPYC1-Venus were not significantly different (Figures 2E and 2F), indicating these patterns are not artifacts related to a particular construct.

Several characteristics of pyrenoid behavior during division are reminiscent of liquids. Toward the end of fission, a "bridge" of matrix material connecting the two lobes of a dividing pyrenoid is briefly visible (Figure 3). After the bridge ruptures, daughter pyrenoids quickly revert to spherical shapes, similar to the behavior of liquid droplets (Stone, 1994; Yanashima et al., 2012). Furthermore, during apparent *de novo* pyrenoid formation, we typically observed that smaller puncta shrank while larger ones grew until the cell contained a single pyrenoid (Figures 2B, 2D, and 2G–2I)—much like Ostwald ripening in systems containing multiple liquid droplets (Hyman et al., 2014)—indicating that components likely exchange between the puncta. Both fission and *de novo* formation have been observed in established liquid-like organelles such as *C. elegans* P granules (Brangwynne et al., 2009; Saha et al., 2016) and *Xenopus* oocyte nucleoli (Brangwynne et al., 2011). The similarity of the pyrenoid's behavior to that of such phase-separated biological liquid droplets suggested the possibility that pyrenoids may also be liquid, rather than static

(C) A cartoon of the approximate locations of the pyrenoid (green), chloroplast (magenta), and cell membrane (black outline).
(D) Example of pyrenoid fission in EPYC1-Venus, annotated as in (B).
(E) Example of the progressing chloroplast cleavage furrow (arrows) appearing to separate daughter pyrenoids. Images are 2D snapshots of 3D *Z* stack reconstructions.
(F) Average and SD of the durations of chloroplast (magenta) and pyrenoid (green) fissions in RBCS1-Venus (left; n = 28 1st and 2nd divisions) and EPYC1-Venus (right; n = 22 1st and 25 2nd divisions).
(G and H) Duration and relative timing of chloroplast (magenta) and pyrenoid (green) division for the pyrenoid fissions plotted in (F). Each bar represents a different division.
See also Figures S1 and S2 and Movie S1.

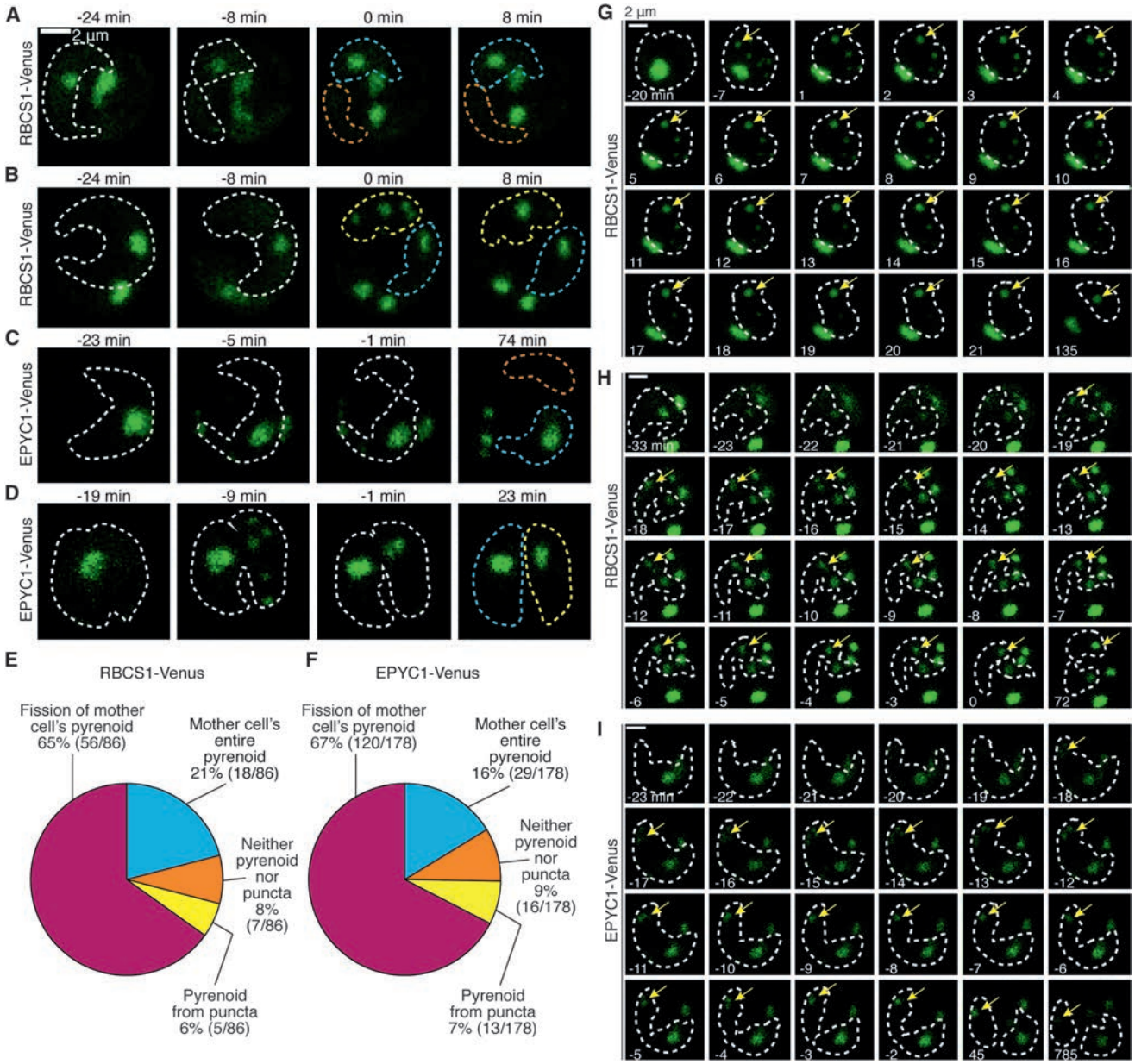


Figure 2. Pyrenoids Can Also Be Inherited by Other Means

(A–D) Examples of other types of pyrenoid inheritance patterns observed in RBCS1-Venus (A and B) and EPYC1-Venus (C and D) cells. (A and C) One daughter (blue) inherits an entire pyrenoid from the mother cell (white) and another daughter (orange) inherits neither a pyrenoid nor puncta. (B and D) One daughter (blue) inherits the entire pyrenoid, and puncta appear in the other daughter (yellow) and coalesce into a new pyrenoid.

(E and F) Proportion of observed RBCS1-Venus (E) and EPYC1-Venus (F) daughter cells that exhibited each observed inheritance pattern; the distribution of inheritance patterns in EPYC1-Venus cells was not significantly different from that of RBCS1-Venus cells (Chi-square test, $p = 0.8$).

(G–I) Stills from time course image captures in which pyrenoids were observed to grow or coalesce from puncta that appeared in the chloroplast stroma during division. (G and H) RBCS1-Venus. (I) EPYC1-Venus. Images are 2D projections of the sum of pixel values in each channel in a *Z* stack through the whole cell at each time point. The chloroplast of the dividing cell of interest in each series is outlined in white. Arrows point to growing or coalescing pyrenoids. $t = 0$ is defined as the first minute at which the daughter chloroplasts are observed to be distinct in 3D.

See also Figures S1 and S2 and Movies S2 and S3.

solids or crystals, as has been proposed (Holdsworth 1968; Kowallik 1969; Bertagnolli and Nadakavukaren 1970). We therefore decided to investigate the fine-scale structure and potential liquid-like dynamics of the pyrenoid matrix.

The Matrix Is Not Crystalline and Has Liquid-like Organization

The different reports of crystalline and amorphous pyrenoid matrices emerged from classical electron micrographs

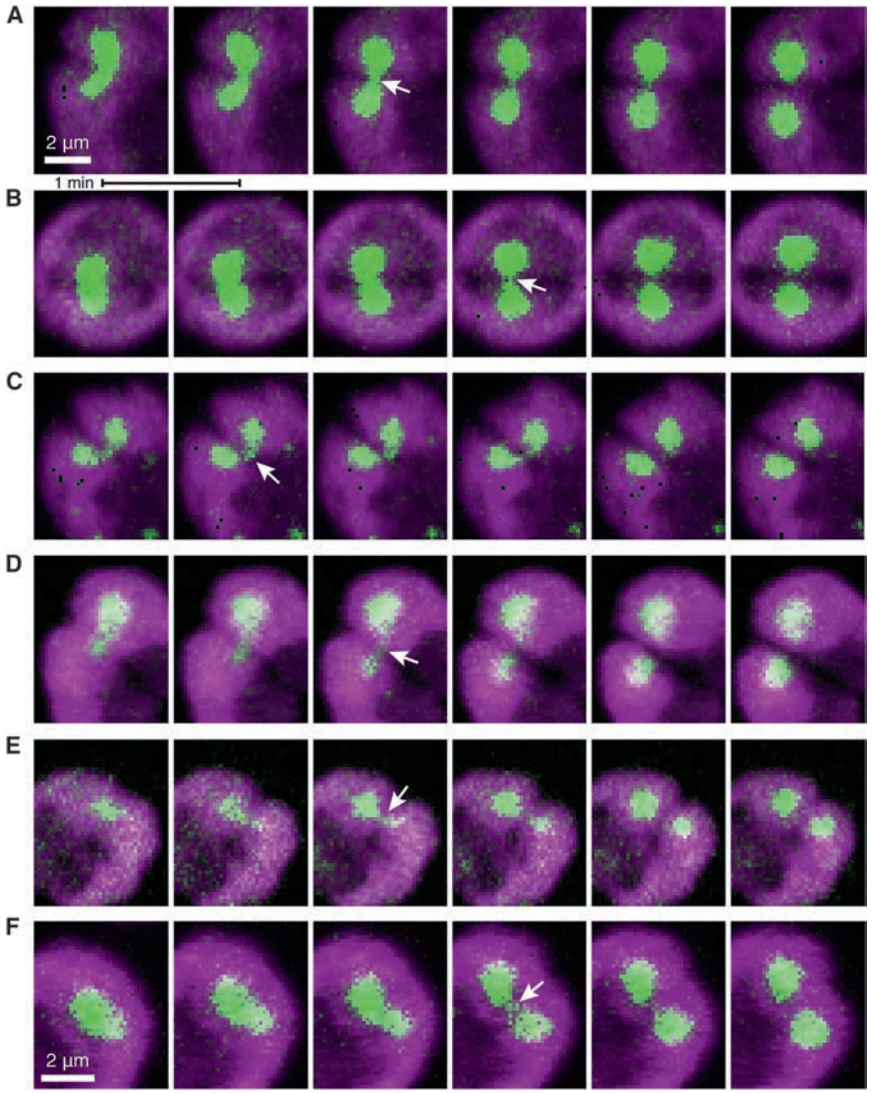


Figure 3. A "Bridge" of Matrix Material Connects Nascent Daughter Pyrenoids during Fission

(A–E) Examples of pyrenoid fissions in RBCS1-Venus cells, with each row corresponding to a separate cell division. Magenta is chlorophyll autofluorescence; green is RBCS1-Venus. Frames are 1 min apart. Scale bar in (A) represents scale in (A)–(E).

(F) Example of pyrenoid fission in an EPYC1-Venus cell. Magenta is chlorophyll autofluorescence and green is EPYC1-Venus.

Images represent the sum of pixel values in each channel from a *Z* stack through the whole cell. See also Movie S1.

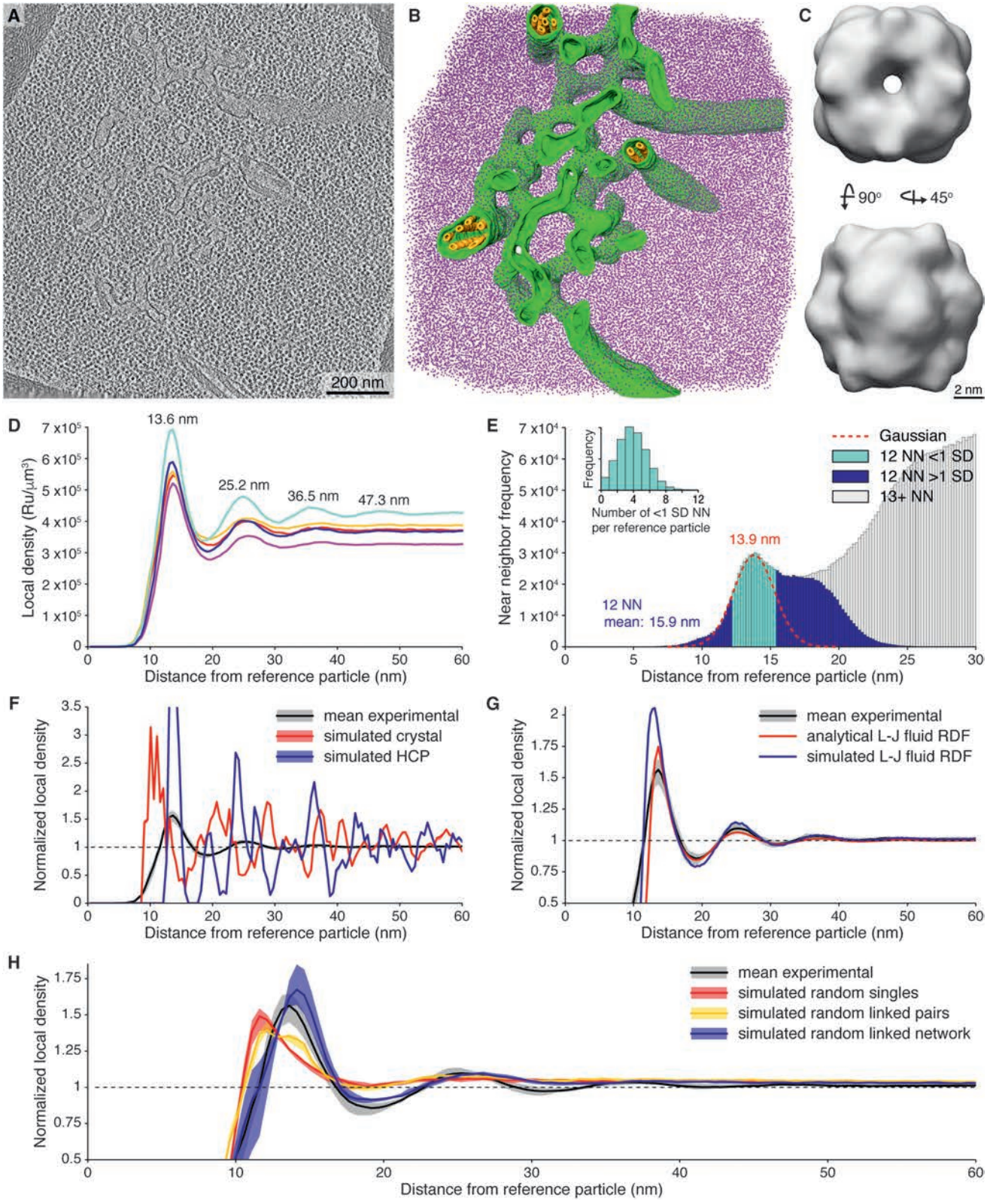
(Holdsworth, 1968; Kowallik, 1969; Bertagnolli and Nadakavukaren, 1970; Griffiths, 1970). However, these micrographs have limited resolution, and the native arrangement of Rubisco may be compromised by artifacts from the sample preparation procedure, which involves chemical fixation, dehydration, plastic embedding, and heavy metal staining (Crang and Klomparens, 1988).

To avoid such artifacts, we analyzed the positions of individual Rubisco holoenzymes within the *Chlamydomonas* pyrenoid matrix by *in situ* cryo-ET (Figure 4A; Movie S4), which provides 3D views of the native cellular environment at molecular resolution (Asano et al., 2016; Schaffer et al., 2017). In order to achieve the complete freezing vitrification required for high-quality cryo-ET, we examined *mat3-4* cells (Umen and Goodenough, 2001), which are smaller due to premature cell division but have normal pyrenoid morphology and a functional CO_2-concentrating mechanism. The high fidelity of direct electron detector imaging enabled us to localize ~97.5% of the Rubisco holoenzymes within five pyrenoid tomograms, totaling 192,100 particles (Figures 4B, 4C, S3, and S4A–S4C).

As a quantitative metric for the packing of Rubisco within the matrix, we calculated the local concentration of holoenzymes in an expanding shell around every Rubisco. Our analysis sampled only the pyrenoid matrix, avoiding potential effects from the traversing membrane tubules and pyrenoid borders. Quasiperiodic near-neighbor peaks of Rubisco density eroded and then vanished as distance increased from the reference particle (Figure 4D), suggesting that the pyrenoid lacks the long-range order characteristic of a crystal.

We previously reported that the average local neighborhood in the pyrenoid matrix resembles a lattice of hexagonal close packing (HCP) with ~15 nm between particle centers (Engel et al., 2015). However, due to limitations in imaging, this conclusion was not reached by determining the positions of individual Rubisco holoenzymes, but rather by averaging together larger volumes that contained neighborhoods of multiple Rubisco particles. The improved quality of our new tomograms enabled us to pinpoint the positions of each Rubisco, revealing that the HCP we previously observed is actually an average of a much more heterogeneous environment. In HCP, the 12 nearest neighbors to any given particle are equidistant. However, when we plotted the center-to-center distances between holoenzymes and their 12 nearest neighbors, we observed a specific peak corresponding to 4.4 ± 1.8 neighbors that were 13.9 ± 1.5 nm away, with most of the remaining neighbors at a range of longer distances, for an average distance of 15.9 ± 2.9 nm (Figure 4E). The diameter of Rubisco is ~10–13 nm, depending on orientation (Taylor et al., 2001); thus, there are ~1–4 nm between a Rubisco and its nearest neighbors in the specific peak.

To directly compare the observed Rubisco packing to that of a crystal, we performed the same local density analysis on



(legend on next page)

simulated data in which the exact volume of the pyrenoid matrix from the tomograms was replaced with one of two crystalline matrix models: one based on HCP with 13.9 nm between particle centers, and the other based on the packing observed in crystallized Rubisco (Taylor et al., 2001) (Figures 4F, S4E, and S4F). Compared to the experimental data, both crystalline simulated datasets produced much taller short-range peaks along with pronounced long-range peaks. Furthermore, the Rubisco concentration observed in the pyrenoid matrix (377,856 ± 37,823 Rubiscos/μm^3; 628 ± 63 μM) was 28% lower than in the HCP simulated data (526,039 ± 709 Rubiscos/μm^3; 874 ± 1 μM) (Figure S3). Therefore, we conclude that the arrangement of Rubisco in the pyrenoid is neither HCP nor that observed in crystallized Rubisco.

To explore whether the pyrenoid matrix has liquid-like organization, we next compared the observed Rubisco packing to the distribution of molecules within known fluids. The local density of Rubisco within the pyrenoid matrix fit well to the radial distribution function of a simple model for liquid molecular interactions, the Lennard-Jones fluid (Johnson et al., 1993) (Figure 4G), determined both analytically and by molecular dynamics simulations (see the STAR Methods). Additionally, similar radial distribution functions have been measured for colloidal liquids (Gu et al., 2010; Thorneywork et al., 2014). Therefore, we conclude that the radial distribution of Rubisco within the matrix resembles that of a liquid.

We sought to determine whether effects at the macromolecular scale could produce the liquid-like organization that we observed in the pyrenoid matrix. We first examined whether random packing of Rubisco particles could yield the local density seen in our experimental data. Using Monte Carlo simulations, we randomly placed the same number of Rubisco particles as the experimental data throughout identical pyrenoid matrix volumes (Figures 4H and S4G). The local density of this simulated data fit our experimental data poorly, showing that the organization of pyrenoid matrix is not random.

We next randomly mapped Rubisco particles into the matrix volumes while imposing specific distance constraints between either linked pairs of particles or interconnected linked networks (Figures 4H and S4H–S4J). As we increased the number of constraints on particle spacing from the single constraint between paired particles to the numerous constraints between particles in a network, the local density of Rubisco much more closely resembled the observed distribution in the pyrenoid matrix. This result suggests that holoenzymes could be linked in a network with a favored Rubisco-to-Rubisco distance. EPYC1 may be the molecular structure that enforces this spacing.

The Pyrenoid Mixes Internally

The Rubisco packing distribution we observed is consistent with either a liquid or an amorphous solid state. In order to discriminate between these states, a dynamic assay is necessary. The most direct test of liquid-like behavior is whether the constituent components exhibit internal mixing (Brangwynne et al., 2009), which is traditionally assayed *in vivo* by fluorescence recovery after photobleaching (FRAP) (van Royen et al., 2009; Hyman et al., 2014).

Strikingly, our FRAP experiments revealed that the pyrenoid matrix mixes internally (Figure 5). After bleaching approximately half the volume of the matrix (Figures S5A–S5C), we observed signal re-homogenization on a timescale of ~20 s (Figures 5A, 5C, 5G, and 5I–5L; Movie S5), similar to that of established liquid-like compartments (Brangwynne et al., 2009; Li et al., 2012; Kroschwald et al., 2015; Saha et al., 2016). During recovery, the signal from the unbleached region decreased as the signal in the bleached region increased, and a "wave" of fluorescence could be seen moving from the unbleached to the bleached region (Figures 5C and 5E), suggesting that recovery is primarily due to internal rearrangements, rather than import of stromal RBCS1-Venus. In contrast, pyrenoids in chemically crosslinked cells showed no recovery (Figures 5B, 5D, 5F, and 5G; Movie S5).

Further FRAP experiments demonstrated that the other major components of the pyrenoid matrix, EPYC1 and Rubisco Activase (McKay et al., 1991) (RCA1), are also mobile (Figures 5G–5L and S5D–S5F). Differences in the initial recovery rates of the three matrix components (Figure 5H) suggest that they may have distinct mobilities within the matrix, apparently inversely

Figure 4. The Pyrenoid Matrix Is Not Crystalline but Exhibits Short-Range Liquid-like Order

(A) Slice through a tomographic volume of the native *Chlamydomonas* pyrenoid.

(B) Segmentation of the tomogram shown in (A) with localized positions of 46,567 Rubisco holoenzymes (magenta) mapped into the volume. Green and yellow: pyrenoid tubule membranes.

(C) *In situ* subtomogram average of Rubisco (16.5 Å resolution; Figure S4A) generated from 30,000 particles extracted from the tomogram shown in (A).

(D) The local density of neighbor Rubisco particles as a function of the distance from the reference particle. Each line represents a separate tomogram, showing the sum of the local densities around every Rubisco. The distances to peaks of high local Rubisco concentration are indicated.

(E) Histogram of distances from reference particles to their nearest neighbors (NN), summed from all five tomograms. Red dashed line, Gaussian distribution fit to the 13.9 nm NN peak; light blue bars, distance to the 12 NN within 1 SD (<1 SD) of the 13.9 nm peak; dark blue bars, distance to the 12 NN beyond 1 SD (>1 SD) from the 13.9 nm peak; gray bars, distance to further (13+ NN) neighbors. Mean distance to the 12 NN = 15.9 nm. Inset: distribution of the number of neighbors per reference particle (mean = 4.4 neighbors) that are <1 SD from the 13.9 nm peak.

(F) The normalized local density of neighbor particles (local density divided by the global density), showing the mean value ± 99% confidence interval (CI) of the experimental data (black) compared to crystalline simulated data generated within the same tomogram volumes (Figures S4E and S4F). Red, crystal structure packing (Taylor et al., 2001); blue, 13.9 nm-spaced HCP.

(G) The mean value ± 99% CI of the experimental data's normalized local density (black) fit with the radial distribution function of a Lennard-Jones fluid generated by an analytical model (Morsali et al., 2005) (red) and by molecular dynamics simulations (Plimpton, 1995) (blue).

(H) The normalized local density of neighbor particles, showing the mean value ± 99% CI of the experimental data (black) compared to random simulated data generated within the same tomogram volumes (Figures S4G–S4J). Red, single particles; yellow, pairs linked by 13.9 ± 1.5 nm; blue, linked networks.

See also Figures S3 and S4 and Movie S4.

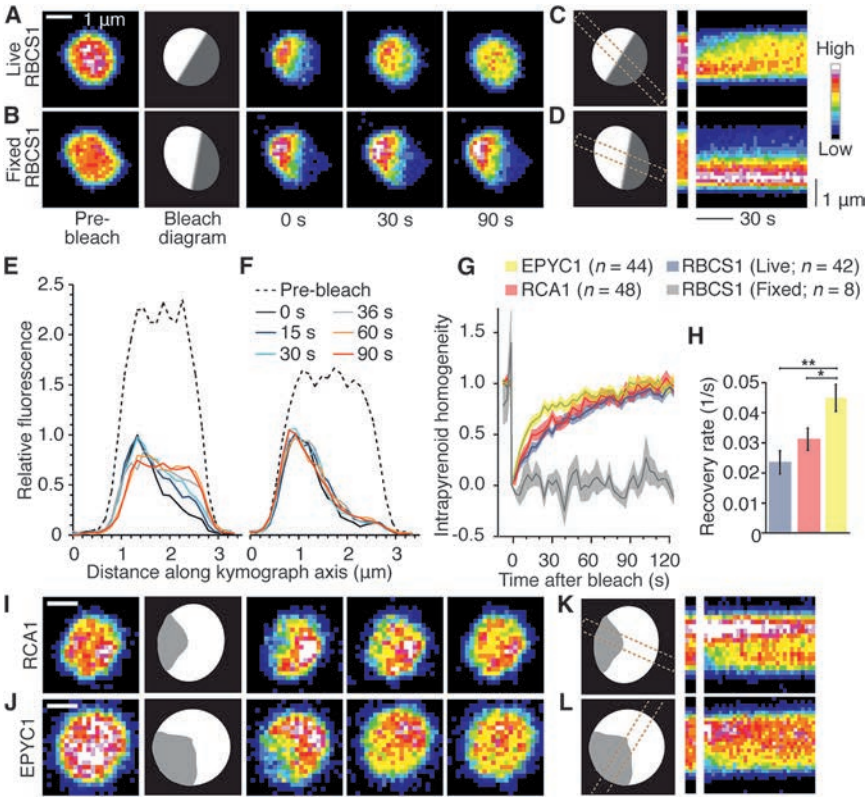


Figure 5. Pyrenoid Matrix Components Mix Internally

(A and B) FRAP in live (A) and fixed (B) RBCS1-Venus pyrenoids. Cartoons depict the approximate bleached region (dark gray). Different intensity display scales are used in the pre- and post-bleach image sets.

(C and D) Kymographs of the live (C) and fixed (D) pyrenoids shown in (A) and (B), respectively. From left to right: pyrenoid cartoon showing the region used to create the kymographs (dashed rectangle), the pre-bleach section of the kymographs, and the post-bleach kymographs.

(E and F) Fluorescence recovery occurs from within the pyrenoid in live pyrenoids (E), but does not occur in fixed pyrenoids (F). The x axis is µm along the dashed regions in (C) and (D).

(G) Average fluorescence recovery profiles ± SEM for pyrenoids in live RBCS1-Venus (blue), RCA1-Venus (red), or EPYC1-Venus (yellow) cells and in fixed RBCS1-Venus cells (gray).

(H) Average recovery rates ± SEM over the first 12 s in (G). $^{**}p < 0.005$; $^{*}p < 0.05$ (one-way ANOVA and post hoc Bonferroni means comparison).

(I–L) Examples of half-pyrenoid FRAP in live RCA-Venus (I and K) and EPYC1-Venus (J and L) cells, with images from the recovery time courses (I and J) and corresponding kymographs (K and L) as shown in (A)–(D). Physical and temporal scales are the same as in (A)–(D).

See also Figure S5 and Movie S5.

proportional to their molecular weights (Rubisco holoenzymes are ~540 kDa, RCA1 hexamers are ~270 kDa, EPYC1 is ~35 kDa). The different mobilities suggest that although these matrix components can bind to each other, the major mobile unit is not a complex between EPYC1 and Rubisco; thus, EPYC1-Rubisco binding interactions (Mackinder et al., 2016) are transient. Alongside our observations of liquid-like fission (Figures 1 and 3) and local order (Figure 4), the internal mixing of matrix components demonstrates that the *Chlamydomonas* pyrenoid matrix behaves like a liquid.

The Pyrenoid Disperses during Cell Division

A remarkable property of liquid-like protein compartments is their ability to transition between an aggregated liquid phase and a dispersed soluble phase (Brangwynne et al., 2009; Li et al., 2012). Surprisingly, our experiments revealed that the *Chlamydomonas* pyrenoid matrix also appears to undergo such a phase transition during division. A portion of the RBCS1-Venus and EPYC1-Venus signals rapidly dispersed from the pyrenoid matrix into the stroma for ~20 min near the end of chloroplast and pyrenoid division (Figures 6 and S6). During this time, the Venus signal in the chloroplast stroma rose dramatically, while the signal in the pyrenoid dropped by ~35%–50%. The Venus intensity in the pyrenoid subsequently recovered, coincident with a reduction in the stromal signal (Figures 6B and 6C). The increased stromal signal during chloroplast division is unlikely to be due to new protein synthesis because these transitions occurred on the timescale of 1–5 min and were not associated with an increase in the total fluorescence signal (Figures 6B and S6). Pyrenoid fission nearly always occurred during the time of increased stromal signal (Figures 6B, 6D, and S6), raising the intriguing possibility that the partial dispersal may be associated with a decrease in matrix surface tension or viscosity, which could facilitate the progression of the cleavage furrow through the pyrenoid.

We observed pyrenoid component dispersal regardless of the mode of pyrenoid inheritance (Figure S6). During dispersal, puncta of matrix material often appeared transiently throughout the stroma (Figures 1D, S1C, and S1D; Movie S3). These *de novo* matrix protein puncta may correspond to the "dense regions" that were observed in electron micrographs of dividing *Chlamydomonas* cells (Goodenough, 1970). In some daughter cells that failed to inherit part or all of the maternal pyrenoid during chloroplast division, we observed that these puncta grew into apparent new pyrenoids (Figures 2B, 2D, and 2G–2I). Our observations suggest a model in which the dispersal of the building blocks of the matrix serves as a redundant mechanism to fission, facilitating equal partitioning of the pyrenoid matrix to daughter chloroplasts. The inheritance of dispersed pyrenoid components enables rapid *de novo* pyrenoid assembly in daughter chloroplasts that fail to inherit a pyrenoid by other means.

A "Magic Number" Effect in Multicomponent Phase Separation

The liquid-like nature of the pyrenoid fits well with the principle that biological liquid compartments are formed by weak

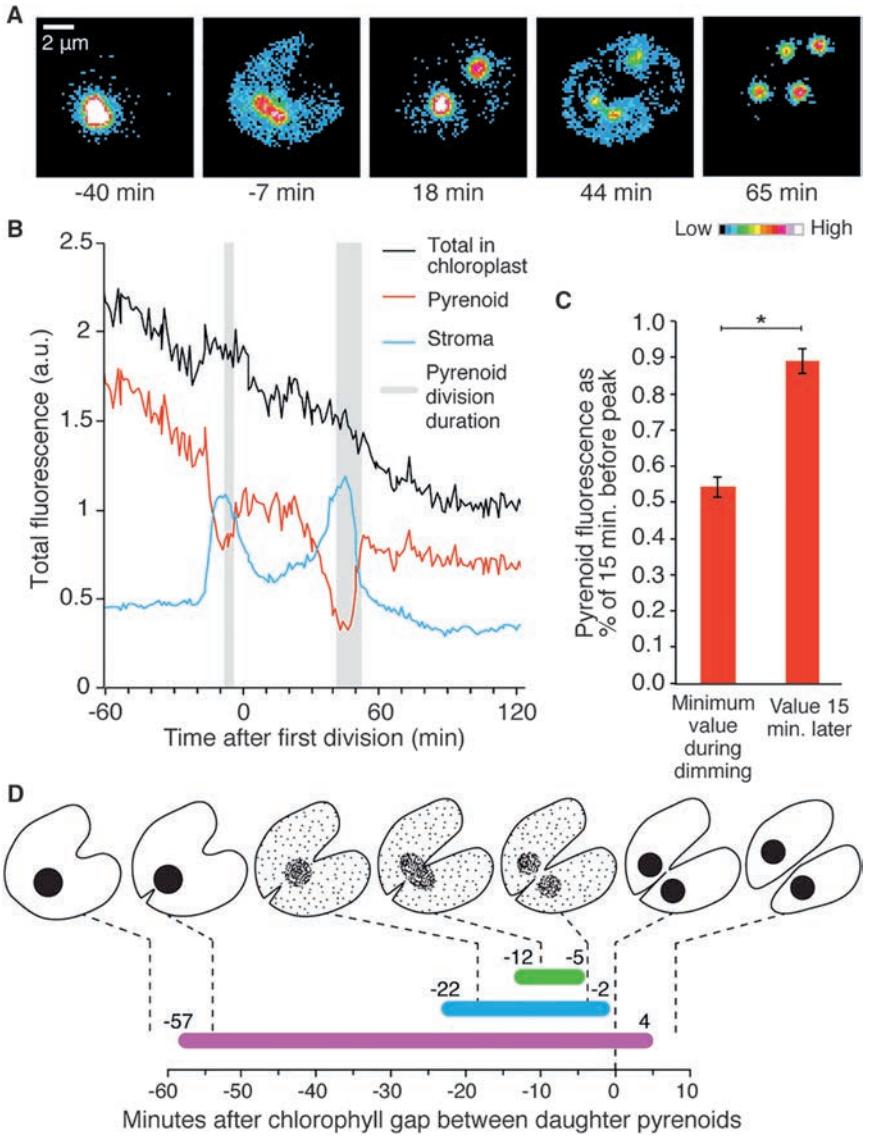


Figure 6. The Pyrenoid Matrix Disperses and Re-aggregates during Division

(A) Heatmaps of the RBCS1-Venus signal during the divisions in Figures 1A–1C. Times are defined as in Figure 1.

(B) Raw signal from (A) plotted over time by regions of interest, representing the sum through the whole z stack in each masked region over time. Times of pyrenoid divisions are highlighted in gray.

(C) The average signal in the pyrenoid during division is significantly lower than that 15 min later (Wilcoxon matched-pairs signed-ranks test; $^{*}p \leq 2 \times 10^{-6}$; n = 31), shown ± SEM.

(D) Timeline of an average cell division with pyrenoid fission. Chloroplast division (magenta), pyrenoid dissolution (cyan), and pyrenoid fission (green) are displayed relative to the moment the chloroplast division furrow passes between the daughter pyrenoids (t = 0). Cartoons depicting the aggregation state of the pyrenoid matrix are shown above each stage, with the chloroplast outlined in black, aggregated matrix components shown as filled black circles, and partially dispersed matrix components as speckles.

See also Figure S6 and Movies S1, S2, and S3.

multivalent binding between constituent proteins (Li et al., 2012). The two major constituents of the pyrenoid, Rubisco and EPYC1, likely have multiple binding sites for each other. The octameric symmetry of the Rubisco holoenzyme makes it plausible that Rubisco has eight binding sites for EPYC1. Additionally, the four repeats of the EPYC1 protein suggest that EPYC1 has four binding sites for Rubisco (Mackinder et al., 2016). We have previously proposed that dynamic regulation of the availability of EPYC1's binding sites could cause a change in the aggregation state of the pyrenoid (Mackinder et al., 2016).

In the model liquid-like system involving binding between the SRC homology 3 (SH3) domain and a proline-rich motif ligand (PRM), it has been established that decreasing the number of binding sites on flexible linker proteins promotes a phase transition from an aggregated to a dispersed state (Li et al., 2012). Surprisingly, modeling of the EPYC1-Rubisco interaction predicts the existence of an additional effect that governs phase transitions in multicomponent liquid-like biological systems.

Because the binding mechanism between Rubisco and EPYC1 is uncharacterized, we developed a simplified computational model to investigate how interactions between these two components may drive pyrenoid aggregation and dissolution. In our model, "Rubisco holoenzymes" and "EPYC1 molecules" occupy a 2D square grid. Each Rubisco holoenzyme is modeled as a 2 × 4 rectangle of 8 binding sites for EPYC1, and each EPYC1 is modeled as a flexible chain of 4 binding sites for Rubisco. Each binding site on Rubisco can bind to a single EPYC1 site and vice versa. Binding occurs when a Rubisco site and an EPYC1 site occupy the same grid position. While our model is simplified and abstract, it allows the observation of certain fundamental behaviors.

Our simulation produced the expected increased aggregation when we increased the number of binding sites on EPYC1 from 4 to 5 (Figures 7B and 7C). However, contrary to the established paradigm for liquid-like systems, we also observed increased aggregation when we decreased the number of binding sites on EPYC1 from 4 to 3 (Figures 7A and 7B).

Further investigation into this behavior revealed a "magic number" effect in the simulated EPYC1-Rubisco system. Magic numbers occur in various contexts in chemistry and physics (Sakurai et al., 1999; Steppenbeck et al., 2013), in which a certain number of particles form an unusually stable state, such as filled electron shells in atoms. However, to our knowledge, magic numbers have not previously been observed in a biological

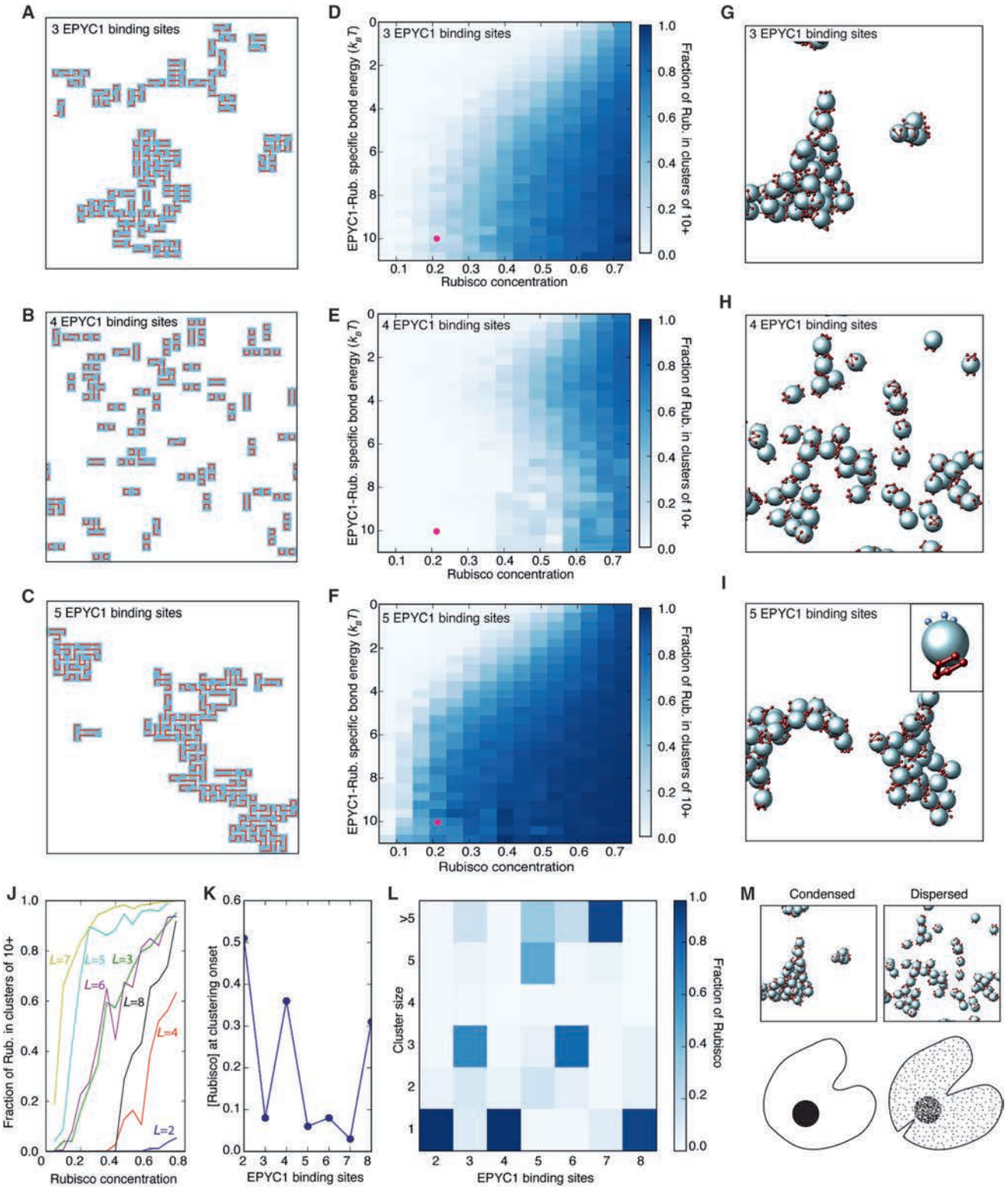


Figure 7. Simulations of an EPYC1-Rubisco System Reveal an Effect of Binding Site Stoichiometry on the Aggregation State

(A–C), Snapshots of simulations with 3 (A), 4 (B), and 5 (C) Rubisco binding sites on EPYC1. "Rubiscos" (blue rectangles) and "EPYC1s" (red polymers) bind when they occupy the same sites in a 2D grid. Snapshots are from simulations with 10 k_BT specific bonds and 0.1 k_BT nearest-neighbor non-specific bonds.

(legend continued on next page)

system. In our model, the magic number effect arises because all 8 binding sites of one Rubisco can be exactly saturated by two EPYC1s with 4 binding sites each to form a stable trimer, resulting in minimal aggregation. When the magic number stoichiometry is broken, either with more or fewer binding sites on EPYC1, larger aggregates form.

To ensure that the magic number effect is not an artifact of the lattice geometry or of two dimensions, we also simulated the system using a more computationally intensive but more realistic three-dimensional, off-lattice representation. Similar to the lattice simulation results, we observed more aggregation when we increased the number of binding sites on EPYC1 from 4 to 5 (Figures 7H and 7I) and when we decreased it from 4 to 3 (Figures 7G and 7H). Additionally, small trimer complexes containing one Rubisco and two EPYC1s were commonly observed when the number of binding sites on EPYC1 was 4, but these complexes were rarely seen when the number of binding sites on EPYC1 was 3 or 5. Thus, the magic number effect persists in three dimensions and off-lattice. Furthermore, the effect is robust for a wide range of protein concentrations and model parameters (Figures 7D–7F and S7; Table S1; Movie S6).

To explore the generality of this effect, we used the 2D lattice model to consider an even wider range of the number of Rubisco binding sites on EPYC1 (Figures 7J–7L) and found additional magic numbers at 2 and 8 binding sites per EPYC1. When the number of binding sites on EPYC1 is a magic number, small heterocomplexes are favored; consequently, higher protein concentrations are required in order to form large aggregates (Figures 7J–7L). These results are robust to constraints such as restricting EPYC1 to only bind one end of each Rubisco (Figure S7). Thus, in addition to the previously established general trend that increasing the number of binding sites in a flexible linker protein promotes phase separation, there are strong magic number effects that impact the phase diagram.

DISCUSSION

Phase Separation Helps Explain Known Pyrenoid Physiology

Phase separation provides a framework for understanding the structure, biogenesis, and inheritance of the pyrenoid. Our observations suggest that the re-localization of Rubisco from the pyrenoid to the stroma in response to high CO_2 (Borkhsenious et al., 1998) or darkness (Mitchell et al., 2014) is a phase transition of the pyrenoid matrix. The use of phase transitions could allow rapid reorganization of Rubisco to enhance CO_2 fixation in fluctuating environmental conditions. Furthermore, the fluidity of the pyrenoid matrix resolves the paradox of how Rubisco Activase chaperones can efficiently access the more abundant Rubisco active sites throughout the matrix.

EPYC1-Rubisco Interactions May Drive Pyrenoid Phase Separation

The pyrenoid's liquid-like nature may be mediated by binding between Rubisco and EPYC1's sequence repeats. Each EPYC1 repeat displays low complexity, with 73% of the repeat regions consisting of just alanine, serine, proline, and arginine (Mackinder et al., 2016). Low complexity domains and repeat regions have been widely implicated in mediating the liquid-like properties and phase separation of nonmembrane-bound organelles (Li et al., 2012; Mitrea and Kriwacki, 2016). The exact nature of the binding between Rubisco and EPYC1 is unknown, but it is likely that surface-exposed α-helices on the Rubisco small subunit play a role in the binding interaction because a specific sequence in these helices is required for Rubisco aggregation in *Chlamydomonas* (Meyer et al., 2012). Weak multivalent interactions have been implicated in mediating biological phase separation in other liquid-like organelles (Li et al., 2012; Hyman et al., 2014). Thus, if EPYC1-Rubisco interactions are weak, then these two binding partners could explain both the structure and the fluidity of the matrix.

A Magic Number Effect Could Facilitate Phase Transitions

We speculate that a magic number effect could help explain the rapid phase transitions of the pyrenoid matrix that we observed experimentally (Figure 6). The rapid partial dissolution and condensation of the pyrenoid suggests regulation at the level of EPYC1-Rubisco interactions. EPYC1 phosphorylation changes in response to CO_2 availability (Turkina et al., 2006; Wang et al., 2014), and phosphorylation could change the affinity of EPYC1 binding sites for Rubisco. As a result of the magic

(D–F) Heatmaps of the fraction of Rubiscos that are in clusters of >10 Rubiscos connected by EPYC1s with 3 (D), 4 (E), or 5 (F) Rubisco binding sites. The fraction of grid sites occupied by Rubiscos (x axis) is varied, with an equal fraction of grid sites occupied by EPYC1s. The specific bond energy (y axis) is varied, while the nonspecific bond energy is fixed at 0.1 k_BT. Red dots indicate the parameters used for the snapshots in (A)–(C).

(G–I), Snapshots of off-lattice 3D simulations with 3 (G), 4 (H), and 5 (I) Rubisco binding sites on EPYC1 for Rubisco. "Rubiscos" (blue spheres with 4 binding sites on each end) and "EPYC1s" (red polymers) bind when their binding sites overlap. Inset in (I): zoom-in with one Rubisco and one EPYC1 with 5 binding sites; 4 of the 5 binding sites of the EPYC1 are in specific bonds with the Rubisco. Snapshots are from simulations with 10 k_BT specific bonds and a Lennard-Jones nonspecific interaction with $\epsilon = 0.1\ k_BT$. The molecules occupy ~2% of the total space in these simulations, with equal total numbers of EPYC1 and Rubisco binding sites.

(J) Fraction of Rubiscos in clusters of >10 Rubiscos for EPYC1s with different numbers of binding sites in the 2D model. The specific bond energy is 10 k_BT and the nonspecific bond energy is 0.1 k_BT.

(K) The concentration of Rubisco at which clustering begins for systems with different numbers of EPYC1 binding sites in the 2D model. The onset is determined from the curves in (J) (see Figure S7).

(L) Heatmap of the distribution of Rubisco cluster sizes for different numbers of EPYC1 binding sites in the 2D model. Each column depicts the normalized cluster-size distribution at [Rubisco] = 0.15, with 10 k_BT specific bond energy and 0.1 k_BT nonspecific bond energy.

(M) Schematic of a possible mechanism by which magic numbers could regulate the formation and dissolution of pyrenoids: EPYC1s with 3 binding sites favor Rubisco clustering in the pyrenoid (left), while EPYC1s with 4 binding sites form stable 2:1 complexes of EPYC1:Rubisco that dissolve into the chloroplast stroma (right).

See also Figure S7, Table S1, and Movie S6.

number effect, only a single binding site on each EPYC1 may need to be modified to trigger a pronounced phase change (Figure 7).

Both lattice and off-lattice models indicate that clustering can be strongly suppressed if the number of binding sites on one species (e.g., Rubisco) is an integer multiple of the number of binding sites on the other (e.g., EPYC1), as this favors the assembly of small oligomers in which all binding sites are saturated. In general, we expect this magic number effect to pertain to multicomponent, multivalent binding systems where bonds are one-to-one and saturable. Moreover, the effect requires these specific bonds to have an energy of several times the product of the Boltzmann constant and temperature (k_BT), strong enough for most small oligomers to be fully bonded without defects (Figures 5D–5F). Because such binding energies are quite common, magic number effects can be predicted broadly for interacting protein pairs such as SH3-PRM (Li et al., 2012) and SIM-SUMO (Banani et al., 2016) as well as for RNA-protein droplets (Lin et al., 2015). Additional insights could be gained in the future by investigating the influence of polymer flexibility, stoichiometry, and mixed valency populations on the magic number effect described in this work.

Other Species May Also Have Liquid-like Pyrenoids

Our results suggest that the pyrenoid "regression" (reduction in size and disappearance) observed in some algae during chloroplast division in zoosporogenesis (Brown and Arnott, 1970) may be a phase transition from an aggregated to a soluble phase. Previous reports suggested that different species of algae undergo either pyrenoid fission or pyrenoid regression during cell division (Brown et al., 1967; Griffiths, 1970). However, our observations demonstrate that both phenomena can occur simultaneously in the same cell.

The observation of pyrenoid fission and regression in other species of algae, combined with the amorphous appearance of the matrix of many species (Griffiths, 1970), suggests that the liquid-like nature may be a general property of all pyrenoids. This hypothesis can be tested as new genetic tools become available in a broad range of algae.

Pyrenoids are not only broadly distributed in the Chloroplastida, but are also found in five out of the seven supergroups of the eukaryotic tree of life (Excavates, Stramenopiles, Alveolates, Rhizaria, and Archaeplastids) (Burki, 2014), and it is thought that pyrenoids have evolved multiple times (Villarreal and Renner, 2012; Meyer and Griffiths, 2013). Intriguingly, across sequenced algae, the presence of a pyrenoid is correlated with the presence of a protein with EPYC1-like physiochemical properties, such as a low-complexity repeat sequence, a high pI, and no transmembrane domains (Mackinder et al., 2016). If the only requirement for aggregating diffuse Rubisco into a liquid-like pyrenoid matrix is the evolution of a linker protein with multiple weak binding sites for Rubisco, pyrenoids could be relatively simple to evolve.

Pyrenoid Behavior Is Distinct from that of Carboxysomes

Cyanobacteria have a functionally similar structure to the pyrenoid called the carboxysome. Like the pyrenoid, carboxysomes contain aggregated Rubisco and are not surrounded by a membrane. There are two classes of carboxysomes, α and β, which are thought to have arisen through convergent evolution (Rae et al., 2013).

Rubisco in β-carboxysomes is thought to be linked together by CcmM, a protein containing multiple Rubisco small subunit-like domains, each of which is thought to be incorporated into a different Rubisco holoenzyme (Long et al., 2010). Such incorporation would be expected to preclude internal mixing. Indeed, fluorescently tagged Rubisco in β-carboxysomes does not recover after photobleaching (Chen et al., 2013).

Less is known about the arrangement of Rubisco in α-carboxysomes. The proposed α-carboxysome Rubisco linker protein CsoS2 (Cai et al., 2015) is a disordered repeat protein, raising the possibility that the α-carboxysome may have liquid-like characteristics. Whereas we have observed pyrenoid biogenesis by both fission and de novo assembly, biogenesis of new α- and β-carboxysomes has only been observed to occur *de novo* (Iancu et al., 2010; Chen et al., 2013). It will be interesting to see whether α-carboxysomes do exhibit liquid-like behavior or whether the underlying structural and biogenesis principles are a general distinguishing feature between carboxysomes and pyrenoids.

The Pyrenoid Provides Insights into Liquid-like Organelle Biology

To our knowledge, the pyrenoid is the only example so far of a liquid-like organelle specific to photosynthetic organisms. Like other previously characterized liquid-like organelles (Brangwynne et al., 2011; Saha et al., 2016), new pyrenoids can be generated by either *de novo* aggregation or fission into two daughters. The first characterized liquid-like organelle, the *C. elegans* P granule, leverages the properties of phase transitions to promote its own asymmetric inheritance by preferentially condensing at one end of the cell before cell division (Brangwynne et al., 2009; Saha et al., 2016). In contrast, the pyrenoid appears to leverage the liquid-like properties of partial dissolution and fission to promote symmetric inheritance, allowing both daughter cells to inherit a pyrenoid.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - RBCS1-, RCA1-, *epyc1* EPYC1-, and EPYC1-Venus *Chlamydomonas reinhardtii* strain generation and culture conditions
  - Cells used for cryo-electron tomography
- METHOD DETAILS
  - Division microscopy and analysis
  - Vitrification, cryo-focused ion beam milling, cryo-electron tomography, and tomogram reconstruction
  - Localization of Rubisco within tomograms and subtomogram averaging
  - Local density analysis and generation of simulated data

- Fitting Lennard-Jones fluid models to the experimental data
- Spot tests to assay growth of Venus-expressing strains
- Fluorescence Recovery After Photobleaching microscopy and analysis
- "Magic number" modeling

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Division image analysis
  - Tomogram analysis
  - FRAP image analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, one table, and six movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.008.

AUTHOR CONTRIBUTIONS

E.S.F.R. performed the FRAP and division experiments and data analysis. L.C.M.M. generated the fluorescent constructs and made the initial FRAP observation. E.S.F.R., L.C.M.M., and M.C.J. designed and interpreted the FRAP and division work. H.N.C. assisted confocal data acquisition and analysis. M. Schaffer and B.D.E. performed cryo-ET data acquisition. L.K.C. and M. Strauss performed the cryo-ET image analysis. A.M.-S. performed the cryo-ET statistical analysis and simulated data generation. A.M.-S. and B.D.E. performed the Lennard-Jones analytical fit. B.X. performed the Lennard-Jones simulation fit. J.M.P., F.F., and B.D.E. supervised the cryo-ET work. B.X. and N.S.W. designed and interpreted the on-lattice EPYC1-Rubisco interaction modeling work. B.X. performed the on-lattice EPYC1-Rubisco interaction modeling work. B.X. and P.R. performed the 3D off-lattice simulation. E.S.F.R., B.X., N.S.W., B.D.E., and M.C.J. wrote the text with input from all authors.

ACKNOWLEDGMENTS

We thank W. Baumeister, J. Berry, D. Ehrhardt, J. Feldman, W. Frommer, A. Grossman, Y. Meir, M.B. Mudgett, M. Scott, H. Stone, V. Walbot, and members of the Jonikas laboratory for helpful discussion and suggestions, and we thank U. Goodenough, A. Grossman, A. McCormick, and M.T. Meyer for critical reading of the manuscript. This work was supported by the National Science Foundation (EF-1105617 and IOS-1359682 to M.C.J. and PHY-1305525 to N.S.W.), the Carnegie Institution for Science (to L.C.M.M. and M.C.J.); NIH (T32GM007276 to E.S.F.R. and 7DP2GM119137-02 to M.C.J.), the Simons Foundation and HHMI (55108535), and Princeton University (to M.C.J.), a CONACyT-DAAD Graduate Scholarship (to L.K.C), a Fundación Séneca Postdoctoral Fellowship (to A.M.-S.), an Alexander von Humboldt Foundation Postdoctoral Fellowship (to B.D.E), and the Deutsche Forschungsgemeinschaft (FO 716/4-1 to F.F.).

Received: March 9, 2017
Revised: June 12, 2017
Accepted: August 4, 2017
Published: September 21, 2017

REFERENCES

Asano, S., Engel, B.D., and Baumeister, W. (2016). In situ cryo-electron tomography: a post-reductionist approach to structural biology. J. Mol. Biol. *428* (2 Pt A), 332–343.

Badger, M.R., Andrews, T.J., Whitney, S.M., Ludwig, M., Yellowlees, D.C., Leggat, W., and Price, G.D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. Can. J. Bot. *76*, 1052–1071.

Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional control of phase-separated cellular bodies. Cell *166*, 651–663.

Bertagnolli, B.L., and Nadakavukaren, M.J. (1970). An ultrastructural study of pyrenoids from *Chlorella pyrenoidosa*. J. Cell Sci. *7*, 623–630.

Bharat, T.A.M., and Scheres, S.H.W. (2016). Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. Nat. Protoc. *11*, 2054–2065.

Borkhsenious, O.N., Mason, C.B., and Moroney, J.V. (1998). The intracellular localization of ribulose-1,5-bisphosphate Carboxylase/Oxygenase in *chlamydomonas reinhardtii*. Plant Physiol. *116*, 1585–1591.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. Proc. Natl. Acad. Sci. USA *108*, 4334–4339.

Brown, R.M., Jr., and Arnott, H.J. (1970). Structure and function of the algal pyrenoid. I. Ultrastructure and cytochemistry during zoosporogenesis of *Tetracystis excentrica*. J. Phycol. *6*, 14–22.

Brown, R.M., Jr., Arnott, H.J., Bisalputra, T., and Hoffman, L.R. (1967). The pyrenoid: Its structure, distribution, and function. J. Phycol. *3* (Supp), 5–6.

Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb. Perspect. Biol. *6*, a016147.

Cai, F., Dou, Z., Bernstein, S.L., Leverenz, R., Williams, E.B., Heinhorst, S., Shively, J., Cannon, G.C., and Kerfeld, C.A. (2015). Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. Life (Basel) *5*, 1141–1171.

Chen, A.H., Robinson-Mosher, A., Savage, D.F., Silver, P.A., and Polka, J.K. (2013). The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. PLoS ONE *8*, e76127.

Cheng, S., Cetinkaya, M., and Gräter, F. (2010). How sequence determines elasticity of disordered proteins. Biophys. J. *99*, 3863–3869.

Crang, R.F.E., and Klomparens, K.L. (1988). Artifacts in Biological Electron Microscopy (Plenum Press).

Cross, F.R., and Umen, J.G. (2015). The *Chlamydomonas* cell cycle. Plant J. *82*, 370–392.

Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., and Baumeister, W. (2015). Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. eLife *4*, e04889.

Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., and Brangwynne, C.P. (2016). Coexisting liquid phases underlie nucleolar subcompartments. Cell *165*, 1686–1697.

Förster, F., Pruggnaller, S., Seybert, A., and Frangakis, A.S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. J. Struct. Biol. *161*, 276–286.

Förster, F., Han, B.G., and Beck, M. (2010). Visual proteomics. Methods Enzymol. *483*, 215–243.

Gibbs, S.P. (1962). The ultrastructure of the pyrenoids of green algae. J. Ultrastruct. Res. *7*, 262–272.

Goodenough, U.W. (1970). Chloroplast division and pyrenoid formation in *Chlamydomonas reinhardtii*. J. Phycol. *6*, 1–6.

Griffiths, D.J. (1970). The pyrenoid. Bot. Rev. *36*, 29–58.

Gu, L., Xu, S., Sun, Z., and Wang, J.T. (2010). Brownian dynamics simulation of the crystallization dynamics of charged colloidal particles. J. Colloid Interface Sci. *350*, 409–416.

Hofmann, H., Soranno, A., Borgia, A., Gast, K., Nettels, D., and Schuler, B. (2012). Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc. Natl. Acad. Sci. USA *109*, 16155–16160.

Holdsworth, R.H. (1968). The presence of a crystalline matrix in pyrenoids of the diatom, *Achnanthes brevipes*. J. Cell Biol. *37*, 831–837.

Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A.V., and Förster, F. (2012). PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. J. Struct. Biol. *178*, 177–188.

Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Iancu, C.V., Morris, D.M., Dou, Z., Heinhorst, S., Cannon, G.C., and Jensen, G.J. (2010). Organization, structure, and assembly of α-carboxysomes determined by electron cryotomography of intact cells. J. Mol. Biol. *396*, 105–117.

Johnson, J.K., Zollweg, J.A., and Gubbins, K.E. (1993). The Lennard-Jones equation of state revisited. Mol. Physiol. *78*, 591–618.

Keown, J.R., Griffin, M.D., Mertens, H.D., and Pearce, F.G. (2013). Small oligomers of ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase are required for biological activity. J. Biol. Chem. *288*, 20607–20615.

Klammt, C., Novotná, L., Li, D.T., Wolf, M., Blount, A., Zhang, K., Fitchett, J.R., and Lillemeier, B.F. (2015). T cell receptor dwell times control the kinase activity of Zap70. Nat. Immunol. *16*, 961–969.

Kowallik, K. (1969). The crystal lattice of the pyrenoid matrix of *Prorocentrum micans*. J. Cell Sci. *5*, 251–269.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. *116*, 71–76.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S.S., and Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. Plant J. *66*, 770–780.

Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D., and Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. eLife *4*, e06807.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. Mol. Cell *60*, 208–219.

Long, B.M., Tucker, L., Badger, M.R., and Price, G.D. (2010). Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. Plant Physiol. *153*, 285–293.

Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O., Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., et al. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc. Natl. Acad. Sci. USA *113*, 5958–5963.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51.

McKay, R.M.L., Gibbs, S.P., and Vaughn, K.C. (1991). RuBisCo activase is present in the pyrenoid of green algae. Protoplasma *162*, 38–45.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. J. Chem. Phys. *21*, 1087–1092.

Meyer, M., and Griffiths, H. (2013). Origins and diversity of eukaryotic CO_2-concentrating mechanisms: lessons for the future. J. Exp. Bot. *64*, 769–786.

Meyer, M.T., Genkov, T., Skepper, J.N., Jouhet, J., Mitchell, M.C., Spreitzer, R.J., and Griffiths, H. (2012). Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. Proc. Natl. Acad. Sci. USA *109*, 19474–19479.

Mitchell, M.C., Meyer, M.T., and Griffiths, H. (2014). Dynamics of carbon-concentrating mechanism induction and protein relocalization during the dark-to-light transition in synchronized Chlamydomonas reinhardtii. Plant Physiol. *166*, 1073–1082.

Mitrea, D.M., and Kriwacki, R.W. (2016). Phase separation in biology; functional organization of a higher order. Cell Commun. Signal. *14*, 1.

Morsali, A., Goharshadi, E.K., Mansoori, G.A., and Abbaspour, M. (2005). An accurate expression for radial distribution function of the Lennard-Jones fluid. Chem. Phys. *310*, 11–15.

Ohad, I., Siekevitz, P., and Palade, G.E. (1967). Biogenesis of chloroplast membranes. I. Plastid dedifferentiation in a dark-grown algal mutant (*Chlamydomonas reinhardi*). J. Cell Biol. *35*, 521–552.

Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. J. Comp. Physiol. *117*, 1–19.

Pollock, S.V., Colombo, S.L., Prout, D.L.J., Jr., Godfrey, A.C., and Moroney, J.V. (2003). Rubisco activase is required for optimal photosynthesis in the green alga *Chlamydomonas reinhardtii* in a low-CO_2 atmosphere. Plant Physiol. *133*, 1854–1861.

Rae, B.D., Long, B.M., Whitehead, L.F., Förster, B., Badger, M.R., and Price, G.D. (2013). Cyanobacterial carboxysomes: microcompartments that facilitate CO_2 fixation. J. Mol. Microbiol. Biotechnol. *23*, 300–307.

Retallack, B., and Butler, R.D. (1970). The development and structure of pyrenoids in *Bulbochaete hiloensis*. J. Cell Sci. *6*, 229–241.

Saha, S., Weber, C.A., Nousch, M., Adame-Arana, O., Hoege, C., Hein, M.Y., Osborne-Nishimura, E., Mahamid, J., Jahnel, M., Jawerth, L., et al. (2016). Polar positioning of phase-separated liquid compartments in cells regulated by an mRNA competition mechanism. Cell *166*, 1572–1584.e16.

Sakurai, M., Watanabe, K., Sumiyama, K., and Suzuki, K. (1999). Magic numbers in transition metal (Fe, Ti, Zr, Nb, and Ta) clusters observed by time-of-flight mass spectrometry. J. Chem. Phys. *111*, 235–238.

Schaffer, M., Engel, B.D., Laugks, T., Mahamid, J., Plitzko, J.M., and Baumeister, W. (2015). Cryo-focused ion beam sample preparation for imaging vitreous cells by cryo-electron tomography. Bio Protoc. *5*, e1575.

Schaffer, M., Mahamid, J., Engel, B.D., Laugks, T., Baumeister, W., and Plitzko, J.M. (2017). Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. J. Struct. Biol. *197*, 73–82.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Steppenbeck, D., Takeuchi, S., Aoi, N., Doornenbal, P., Matsushita, M., Wang, H., Baba, H., Fukuda, N., Go, S., Honma, M., et al. (2013). Evidence for a new nuclear "magic number" from the level structure of 54Ca. Nature *502*, 207–210.

Stone, H.A. (1994). Dynamics of drop deformation and breakup in viscous fluids. Annu. Rev. Fluid Mech. *26*, 65–102.

Taylor, T.C., Backlund, A., Bjorhall, K., Spreitzer, R.J., and Andersson, I. (2001). First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. J. Biol. Chem. *276*, 48159–48164.

Thévenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. *7*, 27–41.

Thorneywork, A.L., Roth, R., Aarts, D.G.A.L., and Dullens, R.P.A. (2014). Communication: radial distribution functions in a two-dimensional binary colloidal hard sphere system. J. Chem. Phys. *140*, 161106.

Turkina, M.V., Blanco-Rivero, A., Vainonen, J.P., Vener, A.V., and Villarejo, A. (2006). CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. Proteomics *6*, 2693–2704.

Umen, J.G., and Goodenough, U.W. (2001). Control of cell division by a retinoblastoma protein homolog in *Chlamydomonas*. Genes Dev. *15*, 1652–1661.

van Royen, M.E., Farla, P., Mattern, K.A., Geverts, B., Trapman, J., and Houtsmuller, A.B. (2009). Fluorescence recovery after photobleaching (FRAP) to study nuclear protein dynamics in living cells. Methods Mol. Biol. *464*, 363–385.

Vaucher, J.-P. (1803). Histoire des conferves d'eau douce, contenant leurs différens modes de reproduction, et la description de leurs principales espèces, suivie de l'histoire des trémelles et des ulves d'eau douce (J.J. Paschoud).

Villarreal, J.C., and Renner, S.S. (2012). Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. Proc. Natl. Acad. Sci. USA *109*, 18873–18878.

Wales, D.J., and Doye, J.P.K. (1997). Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. *101*, 5111–5116.

Wang, H., Gau, B., Slade, W.O., Juergens, M., Li, P., and Hicks, L.M. (2014). The global phosphoproteome of *Chlamydomonas reinhardtii* reveals complex organellar phosphorylation in the flagella and thylakoid membrane. Mol. Cell. Proteomics *13*, 2337–2353.

Wang, Y., Stessman, D.J., and Spalding, M.H. (2015). The CO_2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against the gradient. Plant J. *82*, 429–448.

Wiegand, T., and Moloney, K.A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. Oikos *104*, 209–229.

Wienkoop, S., Weiss, J., May, P., Kempa, S., Irgang, S., Recuenco-Munoz, L., Pietzke, M., Schwemmer, T., Rupprecht, J., Egelhofer, V., and Weckwerth, W. (2010). Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. Mol. Biosyst. *6*, 1018–1031.

Yanashima, R., García, A.A., Aldridge, J., Weiss, N., Hayes, M.A., and Andrews, J.H. (2012). Cutting a drop of water pinned by wire loops using a superhydrophobic surface and knife. PLoS ONE *7*, e45893.

Article

Cell

Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain

Gaylor Boulay,[1,3,5,6] Gabriel J. Sandoval,[2,3,6] Nicolo Riggi,[4] Sowmya Iyer,[1] Rémi Buisson,[5] Beverly Naigles,[1,5] Mary E. Awad,[1,5] Shruthi Rengarajan,[1,5] Angela Volorio,[1,4,5] Matthew J. McBride,[2,3] Liliane C. Broye,[4] Lee Zou,[5] Ivan Stamenkovic,[4] Cigall Kadoch,[2,3,7,*] and Miguel N. Rivera[1,3,5,7,8,*]

[1]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[2]Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA

[3]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[4]Institute of Pathology, Centre Hospitalier Universitaire Vaudois, Faculty of Biology and Medicine, University of Lausanne, 1011 Lausanne, Switzerland

[5]Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129, USA

[6]These authors contributed equally

[7]Senior author

[8]Lead Contact

*Correspondence: cigall_kadoch@dfci.harvard.edu (C.K.), mnrivera@mgh.harvard.edu (M.N.R.)

http://dx.doi.org/10.1016/j.cell.2017.07.036

SUMMARY

Alterations in transcriptional regulators can orchestrate oncogenic gene expression programs in cancer. Here, we show that the BRG1/BRM-associated factor (BAF) chromatin remodeling complex, which is mutated in over 20% of human tumors, interacts with EWSR1, a member of a family of proteins with prion-like domains (PrLD) that are frequent partners in oncogenic fusions with transcription factors. In Ewing sarcoma, we find that the BAF complex is recruited by the EWS-FLI1 fusion protein to tumor-specific enhancers and contributes to target gene activation. This process is a neomorphic property of EWS-FLI1 compared to wild-type FLI1 and depends on tyrosine residues that are necessary for phase transitions of the EWSR1 prion-like domain. Furthermore, fusion of short fragments of EWSR1 to FLI1 is sufficient to recapitulate BAF complex retargeting and EWS-FLI1 activities. Our studies thus demonstrate that the physical properties of prion-like domains can retarget critical chromatin regulatory complexes to establish and maintain oncogenic gene expression programs.

INTRODUCTION

Temporal and spatial regulation of gene expression plays a fundamental role in directing cell identity and proliferation in both normal tissues and in human disease. The striking number of genetic alterations in genes encoding transcription factors, chromatin modifiers, and histones that have been uncovered in recent whole-exome sequencing efforts have further highlighted the importance of gene regulation in cancer (Lander, 2011). Whereas these alterations can have profound consequences on cancer-specific gene expression, their precise mechanisms of action, in most cases, remain poorly understood.

In contrast to most adult tumor types, pediatric cancers are often driven by a limited number of genetic alterations (Lawrence et al., 2013). Pathognomonic chromosomal translocations represent an important class of these abnormalities and often lead to the formation of oncogenic fusion proteins that involve transcription factors or transcriptional regulators. One of the most well-characterized translocations results in the fusion of the EWSR1 gene and the FLI1 E-Twenty Six (ETS) transcription factor in Ewing sarcoma, the second most common pediatric bone cancer (Delattre et al., 1992). The EWS-FLI1 oncogenic fusion protein is often the only genetic alteration in these tumors (Brohl et al., 2014; Crompton et al., 2014; Tirode et al., 2014) and operates as an aberrant transcription factor containing the ETS DNA-binding domain of FLI1. EWSR1 has been linked to transcriptional activation and RNA binding (Kovar, 2011), yet its contribution to the function of EWS-FLI1 remains poorly defined.

Several studies have shown that EWS-FLI1 is necessary for Ewing sarcoma tumorigenicity (Herrero-Martin et al., 2011) and is sufficient for transformation of mesenchymal stem cells (MSCs) (Riggi et al., 2005, 2008). More recently, EWS-FLI1 has been shown to be a major determinant of genome-wide chromatin states in Ewing sarcoma (Riggi et al., 2014; Tomazou et al., 2015). Strikingly, EWS-FLI1 is able to activate a large set of target genes by operating as a pioneer factor at GGAA microsatellite repeats and inducing active enhancers de novo starting from a closed chromatin conformation (Gangwal et al., 2008; Guillon et al., 2009; Patel et al., 2012; Riggi et al., 2014). This process of enhancer activation requires major restructuring of the chromatin environment, suggesting the participation of chromatin remodeling proteins that have yet to be defined.

The mammalian switch/sucrose non-fermentable (SWI/SNF) (or BAF, for BRG1/BRM-associated factor) complex is an ATP-dependent chromatin remodeler composed of 12–15 subunits

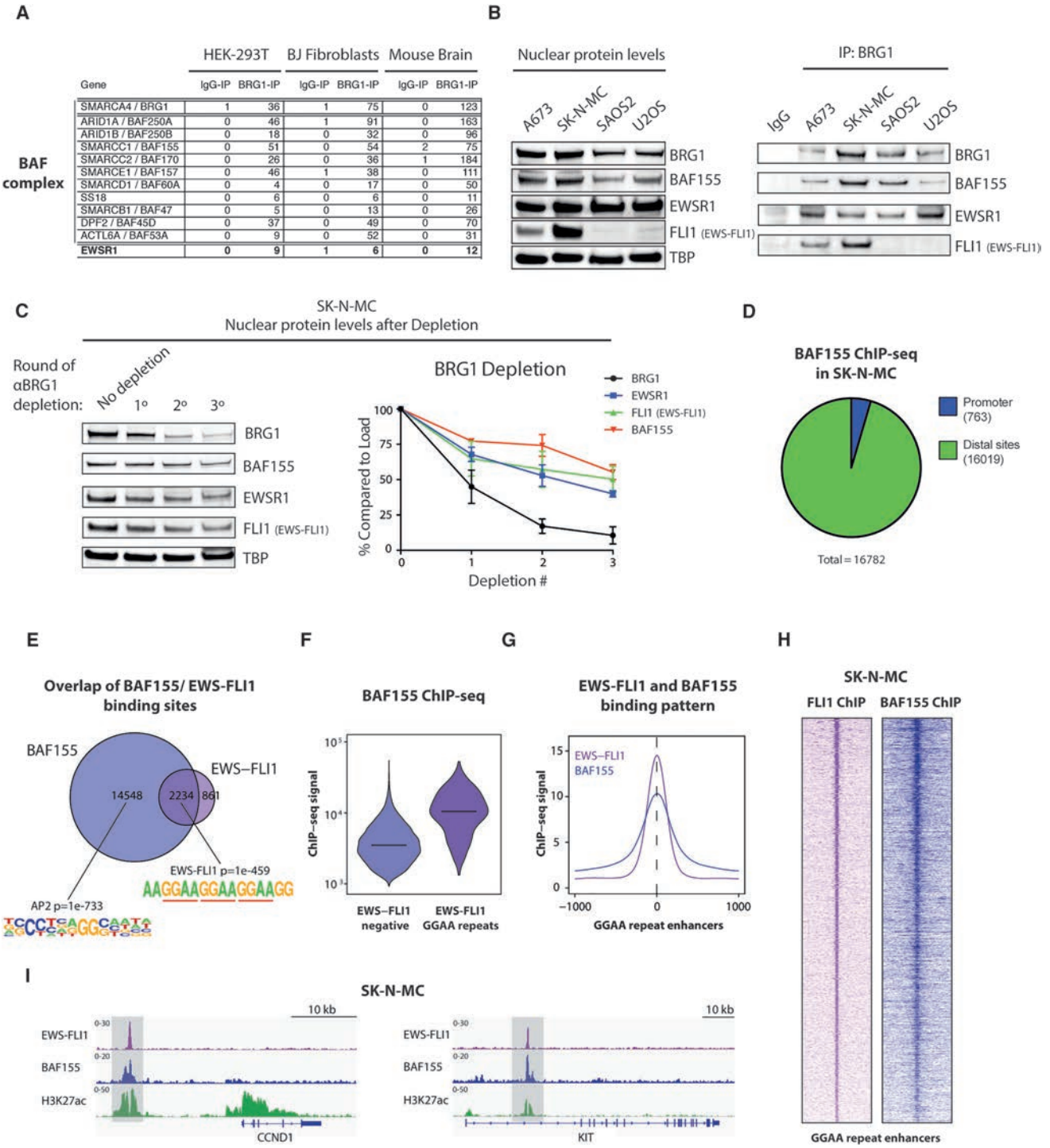


Figure 1. EWS-FLI1 Binds mSWI/SNF (BAF) Complexes and Co-localizes at Tumor-Specific GGAA Repeat Enhancer Elements in Ewing Sarcoma

(A) Table highlighting enrichment of EWSR1 peptides in anti-BRG1 immunoprecipitation/mass spectrometry studies in human cells (HEK293-T and BJ fibroblasts) and in mouse brain tissue. Highlighted is the number of peptides of EWSR1 and BAF complex members.

(B) (Left) Immunoblotting for BAF subunits, EWSR1 and EWS-FLI1, performed on nuclear extracts used for immunoprecipitation. (Right) Immunoprecipitation studies using an anti-BRG1 antibody in Ewing sarcoma (A673 and SK-N-MC) and osteosarcoma (SAOS2 and U2OS) cell lines demonstrate binding of EWSR1 (wild-type) and EWS-FLI1 to BAF complexes.

(legend continued on next page)

that regulates genomic architecture and DNA accessibility (Kadoch and Crabtree, 2015). Recent exome-sequencing studies have revealed that the genes encoding BAF complex subunits are recurrently mutated in over 20% of human cancers (Kadoch et al., 2013). Interestingly, specific subunits appear to be mutated in different cancer subtypes, suggesting tissue-specific functions (Kadoch and Crabtree, 2015; Kadoch et al., 2013; Roberts et al., 2002; Versteege et al., 1998). The high frequency of alterations in BAF complex subunits across a range of tumor types points to their critical role in controlling chromatin architecture and gene expression in cancer (Kadoch and Crabtree, 2015).

Using an unbiased mass spectrometry approach, we now show that BAF complexes interact with the wild-type protein EWSR1 in several cell types and with the EWS-FLI1 fusion protein in Ewing sarcoma. The BAF complex is specifically recruited by EWS-FLI1 to tumor-specific GGAA repeat microsatellites and is necessary for the activation of target genes. Remarkably, the ability to recruit BAF complexes and activate enhancers de novo at these repeat sites is a neomorphic property of EWS-FLI1 that depends on tyrosine residues in the EWSR1 prion-like domain, which are necessary for its interaction with wild-type EWSR1 and for its phase transition properties in vitro. These observations expand the set of human cancers in which BAF complex mistargeting contributes to oncogenesis beyond settings in which BAF complex genes themselves are mutated and show that recruitment via a prion-like domain is a powerful means of retargeting critical chromatin regulatory complexes to tumor-specific loci.

RESULTS

BAF Complexes Interact with Wild-Type EWSR1 and the Fusion Protein EWS-FLI1

BAF complexes are combinatorially assembled from a set of ubiquitously expressed core subunits as well as many cell-type- and context-specific subunits that give rise to an extensive diversity of complex configurations. In order to identify the constellation of BAF complex subunits and associated proteins, we performed endogenous capture of BAF complexes via anti-BRG1 immunoprecipitation followed by proteomic mass spectrometry in several cell types. Notably, these experiments revealed substantial enrichment of peptides corresponding to the EWSR1 protein, among several other previously unidentified proteins (Figure 1A). EWSR1 has been linked to several cellular processes, but most notably it has been shown to be directly involved in gene regulation as a frequent partner in oncogenic fusion proteins with transcription factors, such as the EWS-FLI1 protein in Ewing sarcoma (Mertens et al., 2016). Given the strong connection between BAF and gene regulation in cancer, we proceeded to confirm whether EWSR1 as well as the EWS-FLI1 fusion protein can interact with BAF complexes. Immunoprecipitation experiments using antibodies specific for BRG1 performed on nuclear extracts isolated from EWSR1 wild-type cells (SAOS2 and U2OS) or EWS-FLI1-positive Ewing sarcoma cell lines revealed that both EWSR1 and the EWS-FLI1 fusion protein interact with BAF complexes (Figure 1B). Reciprocal immunoprecipitation experiments using an antibody specific for EWSR1 confirmed these interactions (Figure S1A). Similar experiments using antibodies specific to additional BAF complex subunits, BAF155 and SS18, also confirmed these interactions with EWS-FLI1 in Ewing sarcoma (Figure S1B).

To further characterize the interaction of EWS-FLI1 with BAF complexes, we performed immunodepletion experiments using an anti-BRG1 antibody in SK-N-MC Ewing sarcoma cell nuclear extracts to determine the relative fraction of EWS-FLI1 bound to BAF complexes. We found that EWS-FLI1 was significantly depleted from nuclear lysates over three rounds of anti-BRG1 immunoprecipitation, suggesting that a high percentage of total nuclear EWS-FLI1 associates with BAF complexes (Figure 1C). Reciprocal immunodepletion experiments using an EWSR1 antibody were only able to slightly deplete BAF complex components, indicating that only a small percentage of total BAF complexes are bound to EWS-FLI1 and wild-type EWSR1 (Figure S1C), consistent with genome-wide activities for this chromatin remodeling complex that are independent of these proteins. This was further substantiated by density sedimentation experiments, which showed that neither EWSR1 nor EWS-FLI1 were core members of the BAF complex (Figure S1D). Similarly, urea denaturation studies showed that EWSR1 interactions with BAF complex subunits were decreased at ~0.5 M urea (Figure S1E) and were thus weaker than those observed between core BAF complex members ($\geq$2.5 M urea; Kadoch and Crabtree, 2013). Taken together, these data demonstrate that both wild-type EWSR1 and the oncogenic fusion protein EWS-FLI1 interact with BAF complexes in a transient manner.

(C) (Left) Immunodepletion studies performed on SK-N-MC Ewing sarcoma nuclear extracts using an anti-BRG1 antibody. (Right) Quantification of depletion experiments using quantitative densitometry is shown. Error bars represent SEM of n = 2 independent experiments.
(D) Distribution of MACS-called BAF155 ChIP-seq peaks in SK-N-MC Ewing sarcoma cells. BAF complexes (as marked by BAF155) are primarily localized at putative enhancer sites. Promoters are annotated using the Refseq promoter database.
(E) Venn diagram depicting the overlap of BAF155 and EWS-FLI1 (FLI1) MACS-called peaks in Ewing sarcoma SK-N-MC cells. The top motifs for distal sites with BAF155-only or BAF155/EWS-FLI1 overlap are shown.
(F) Total BAF155 ChIP-seq signals at BAF-155-only sites (n = 14,548) and sites co-bound with EWS-FLI1 at GGAA repeats (n = 660) as represented by violin plots.
(G) Composite plot shows EWS-FLI1 and BAF155 ChIP-seq signals at overlapping GGAA repeat binding sites. The x axis represents a 2-kb window centered on EWS-FLI1 binding sites.
(H) Heatmaps showing EWS-FLI1 and BAF155 ChIP-seq signal density in Ewing sarcoma SK-N-MC cells. 10-kb windows in each panel are centered on EWS-FLI1-bound GGAA repeat enhancer sites (n = 812).
(I) Representative examples of EWS-FLI1 and BAF155 co-occupancy shown at GGAA repeat enhancers associated with the *CCND1* and *KIT* genes. Enhancer regions are highlighted in light gray.
See also Figure S1.

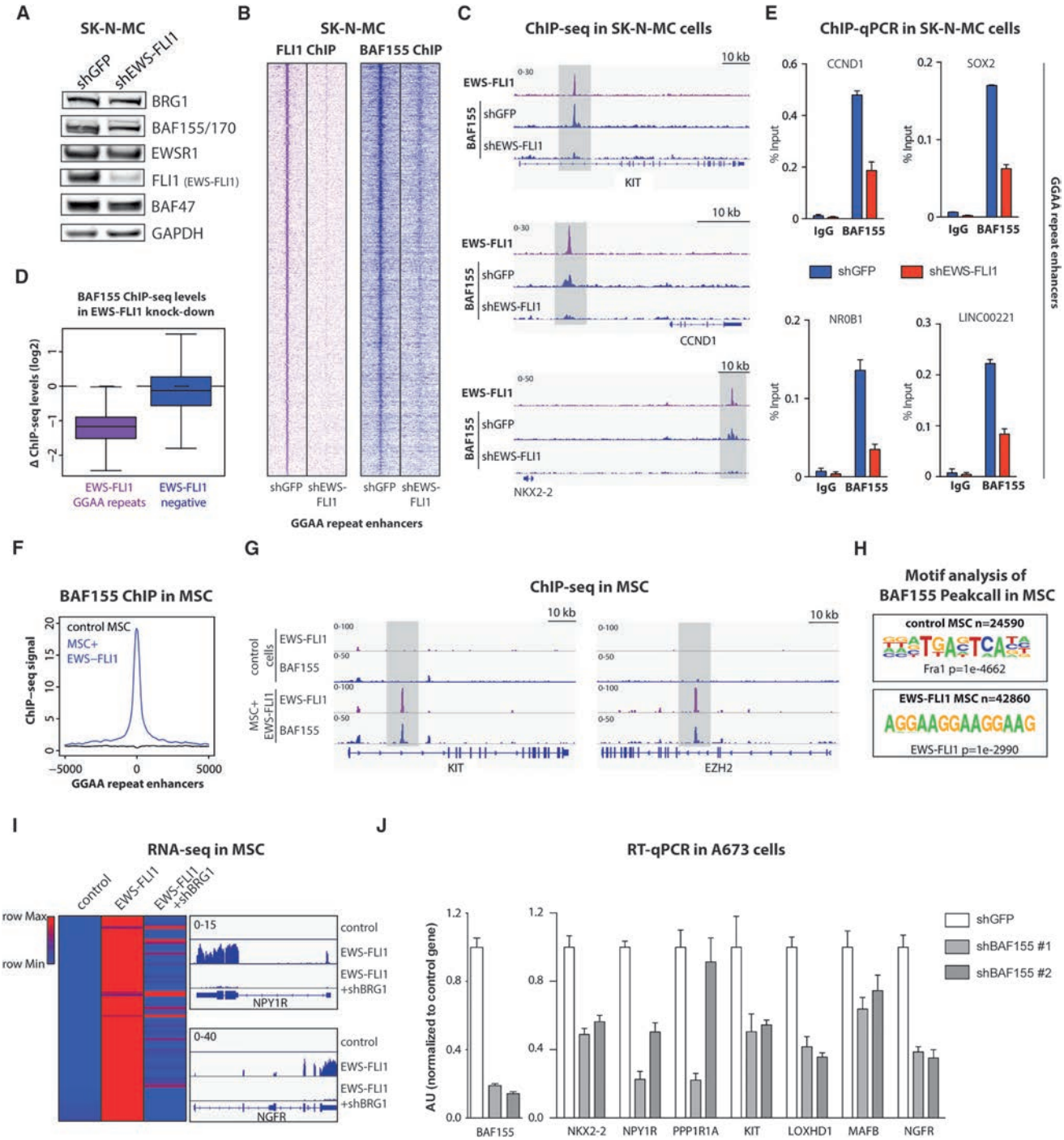


Figure 2. Interdependency of EWS-FLI1 and BAF Complexes in Driving Oncogenic Gene Expression Programs in Ewing Sarcoma

(A) shRNA-mediated suppression of EWS-FLI1 (versus shGFP as control) in SK-N-MC Ewing sarcoma cells; immunoblot for FLI1 (EWS-FLI1), EWSR1, and BAF complex subunits performed on nuclear extracts.

(B) Heatmaps showing EWS-FLI1 and BAF155 ChIP-seq signal density in SK-N-MC cells treated with either shGFP control or shEWS-FLI1 knockdown. 10-kb windows in each panel are centered on EWS-FLI1-bound GGAA repeat enhancer sites (n = 812).

(C) Example tracks demonstrating decreased binding of BAF155 at EWS-FLI1-bound GGAA repeat enhancers associated with *KIT*, *CCND1*, and *NKX2-2* in SK-N-MC cells treated with either shGFP or shEWS-FLI1 knockdown. Enhancer regions of interest are highlighted in light gray.

(D) BAF155 occupancy is decreased specifically at GGAA repeat regions following EWS-FLI1 knockdown in SK-N-MC cells. Boxplots depict the changes in BAF155 ChIP-seq signals between SK-N-MC cells treated with either shGFP or shEWS-FLI1 knockdown. BAF155 MACS-called peaks are divided into EWS-FLI1-bound GGAA repeat enhancers (n = 660 sites, purple) and BAF155-only sites (n = 14,548 peaks, blue).

(legend continued on next page)

EWS-FLI1 Recruits BAF Complexes to Tumor-Specific GGAA Microsatellite Repeat Enhancers to Activate Target Gene Expression

Owing to the biochemical interaction between BAF complexes and EWS-FLI1, we next sought to determine whether BAF complexes cooperate with EWS-FLI1 to regulate gene expression in Ewing sarcoma. We recently demonstrated that EWS-FLI1 operates as a pioneer factor to induce tumor-specific de novo enhancers at GGAA microsatellite repeats (Riggi et al., 2014), a process that may involve significant redistribution of chromatin remodeling complexes that directly govern DNA accessibility, such as BAF (Phelan et al., 1999). To address a potential collaborative role in this process, we performed chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) studies to localize BAF complexes and EWS-FLI1 occupancy in the Ewing sarcoma cell line SK-N-MC, using BAF155 and FLI1 antibodies, respectively. The majority (95.4%) of BAF155 peaks were detected at putative enhancer regions (Figures 1D and S1F). Moreover, we detected a striking degree of overlap between BAF155 and EWS-FLI1 sites genome-wide, for which the GGAA repeat was the top-ranked DNA motif (Figure 1E).

In addition to substantial co-localization, median BAF155 peak occupancy was 2-fold higher at EWS-FLI1-bound GGAA repeats compared to all other genomic locations (Figure 1F) and was centered on EWS-FLI1 peaks (Figures 1G–1I and S1G), suggesting that the localization of BAF complexes may be dependent on EWS-FLI1 binding. To test this hypothesis, we depleted EWS-FLI1 in SK-N-MC Ewing sarcoma cells using short hairpin RNAs (shRNAs) and found that suppression of EWS-FLI1 led to an almost complete disappearance of BAF complex occupancy at GGAA repeats (Figures 2A–2C and S2A), whereas other BAF155 peaks outside these GGAA enhancer regions remained unaffected (Figure 2D). These results were further validated by ChIP-qPCR for BAF155 occupancy over selected loci containing GGAA repeats, including *CCND1*, *SOX2*, *NR0B1*, and *LINC00221* (Figures 2E and S2B), as well as by using an antibody that recognizes the alternative BAF ATPase core subunits BRG1/BRM (Figures S2C–S2E). To account for possible effects of the cell cycle arrest observed after EWS-FLI1 knockdown (Tanaka et al., 1997), we arrested A673 cells in the G1 phase prior to performing ChIP-qPCR (Figure S2F) and co-immunoprecipitation assays (Figure S2G). Interactions and recruitment at chromatin were stable in these conditions, further confirming the specificity of decreased BAF complex occupancy at GGAA repeats observed in the absence of EWS-FLI1. These observations thus indicate that the presence of BAF complexes at GGAA repeat microsatellites is dependent on the EWS-FLI1 fusion protein.

To evaluate the potentially instructive nature of EWS-FLI1 in targeting BAF complexes, we assessed whether EWS-FLI1 is capable of site-specific recruitment of BAF complexes in MSCs, a likely cell of origin of Ewing sarcoma (Riggi et al., 2005, 2008). In this system, EWS-FLI1 has been shown to operate as a pioneer factor at GGAA repeats (Riggi et al., 2014). Whereas in naive MSCs, BAF complex occupancy over GGAA repeats was undetectable, we observed a substantial increase in BAF complex localization upon introduction of EWS-FLI1 (Figures 2F and 2G). Notably, GGAA repeats were the highest ranking DNA motif at newly created BAF complex sites, consistent with a major EWS-FLI1-mediated redistribution of BAF complexes (Figure 2H).

We next assessed whether BAF complexes are required for the induction of EWS-FLI1-mediated gene expression. shRNA-mediated depletion of the BRG1 ATPase prior to expression of EWS-FLI1 in MSCs caused a striking reduction in target gene activation (Figure 2I). Furthermore, knockdown of the subunit BAF155 in SK-N-MC Ewing sarcoma cells, using two independent BAF155-specific shRNAs, resulted in decreased expression of EWS-FLI1 target genes (Figures 2J and S2H) and significantly impaired viability of Ewing sarcoma cells in culture (Figure S2I). A similar decrease in viability was observed using shRNAs targeting the BRG1 ATPase subunit (Figure S2J). These studies performed in both MSCs and Ewing sarcoma cell lines indicate that BAF complexes play a major role in EWS-FLI1-mediated oncogenic gene regulation.

Recruitment of BAF Complexes to GGAA Repeats Is a Neomorphic Property of the EWS-FLI1 Fusion Compared to Wild-Type FLI1

The respective contributions of EWS and FLI1 components of the EWS-FLI1 fusion protein to chromatin reorganization have not to date been fully delineated. As such, we aimed to explore whether expression of the FLI1 transcription factor alone could recruit BAF complexes to GGAA repeats. In contrast to expression of EWS-FLI1, expression of the wild-type FLI1, which provides the ETS DNA binding domain to the fusion protein, did not induce de novo activation of enhancers at GGAA repeats in MSCs and was not associated with recruitment of

(E) ChIP-qPCR validation of decreased BAF155 occupancy at selected EWS-FLI1 GGAA repeat enhancers associated with *CCND1*, *SOX2*, *NR0B1*, and *LINC00221*. Error bars indicate SD of three technical replicates and represent at least two independent biological experiments.

(F) Introduction of EWS-FLI1 in MSCs results in recruitment of BAF complexes to GGAA microsatellite repeats. Composite plot shows BAF155 ChIP-seq signals in control MSCs (black) and EWS-FLI1-expressing MSCs (blue). The x axis represents a 10-kb window centered on EWS-FLI1 binding sites.

(G) Examples of recruitment of BAF155 by EWS-FLI1. ChIP-seq tracks illustrate EWS-FLI1 and BAF155 binding at GGAA repeat microsatellites upon introduction of EWS-FLI1 into MSCs. Enhancer regions of interest are highlighted in light gray.

(H) Motif analysis on BAF155 MACS-called peaks in control conditions and on newly created peaks after EWS-FLI1 expression in MSCs.

(I) BAF complex activity is required for activation of EWS-FLI1 target genes. (Left) Heatmap shows relative RNA-seq gene expression levels of genes associated with GGAA repeats and activated upon introduction of EWS-FLI1 in MSCs (rows, n = 79 genes). The columns show MSCs treated with either empty vector control, EWS-FLI1 + shGFP, or EWS-FLI1+ shBRG1. Expression values were normalized by row. (Right) Example RNA-seq tracks over selected genes are shown.

(J) RT-qPCR experiments show decreased mRNA expression of EWS-FLI1 target genes 48 hr post-infection with BAF155 shRNA in A673 Ewing sarcoma cells. Error bars indicate SD of three technical replicates and represent at least two independent biological experiments.

See also Figure S2.

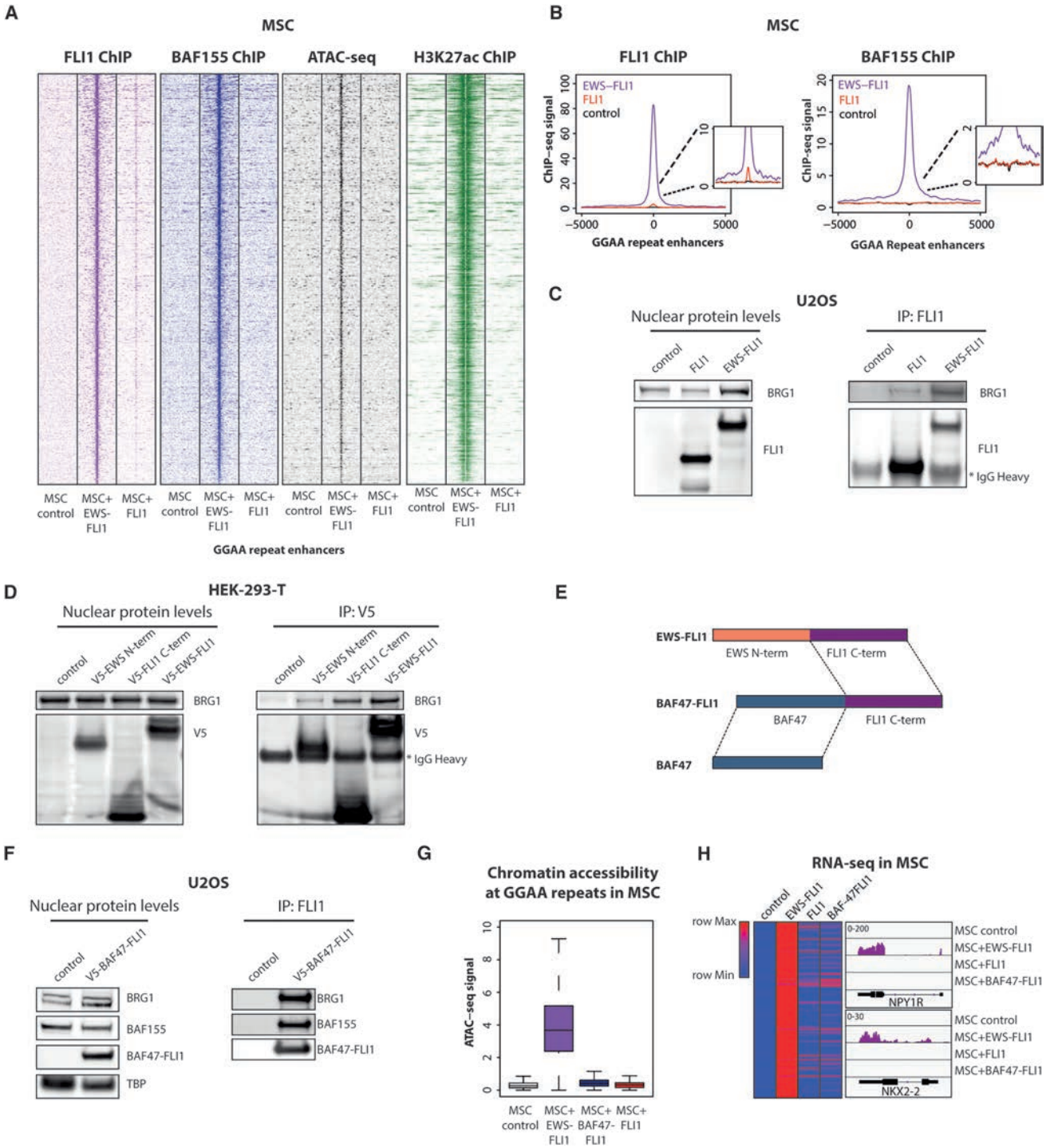


Figure 3. Recruitment of BAF Complexes to GGAA Microsatellite Repeats Is a Neomorphic Property of EWS-FLI1

(A) Heatmaps of FLI1, BAF155, H3K27ac ChIP-seq, and ATAC-seq signal densities in MSCs infected with either control vector, EWS-FLI1, or wild-type FLI1. 10-kb windows in each panel are centered on EWS-FLI1-bound GGAA repeat enhancer sites (n = 812).

(B) Composite plots show FLI1 (left) and BAF155 (right) ChIP-seq occupancy over GGAA repeat enhancers in control MSCs and MSCs expressing EWS-FLI1 or FLI1. The x axis represents a 10-kb window centered on EWS-FLI1 binding sites. Inset: 10-fold magnification shows minimal wild-type FLI1 binding over repeat enhancers but no BAF155 recruitment by FLI1.

(C) Both wild-type FLI1 and EWS-FLI1 interact with BAF complexes. (Left) Immunoblots from nuclear extracts show lentiviral expression of wild-type FLI1 or EWS-FLI1 and the levels of endogenous BRG1 in U2OS cells. (Right) Co-immunoprecipitation experiments using anti-FLI1 antibodies show interactions with BAF. * indicates immunoglobulin G (IgG) heavy chains used for immunoprecipitation.

(legend continued on next page)

BAF complexes (Figures 3A, 3B, and S3A). Furthermore, whereas FLI1 binding was readily detected at non-repeat canonical ETS binding motifs (Figures S3B and S3C), we did not observe significant binding for FLI1 at GGAA repeat sites (Figures 3A, 3B, and S3D). Accordingly, FLI1 failed to induce the expression of known EWS-FLI1 target genes (Figure S3E). Tethering to GGAA repeats and the recruitment of BAF complexes to these sites is thus a neomorphic property of the EWS-FLI1 fusion protein compared to wild-type FLI1.

In order to test whether these striking differences between EWS-FLI1 and wild-type FLI1 were linked to their differential ability to bind BAF complexes, we performed immunoprecipitation experiments comparing FLI1 and EWS-FLI1. Immunoprecipitation of lentivirally expressed FLI1 or EWS-FLI1 using an anti-FLI1 antibody showed that both proteins were capable of pulling down the BAF complex subunit BRG1 (Figure 3C). Consistent with these observations, further characterization showed that both the EWS N-terminal and FLI1 C-terminal fragments of EWS-FLI1 were able to co-immunoprecipitate BRG1 (Figure 3D). To test whether an additional interaction domain could account for the neomorphic properties of EWS-FLI1, we generated a fusion protein between the BAF complex subunit BAF47 and the FLI1 C-terminal domain (Figure 3E). This fusion protein was able to interact with the BAF complex, as demonstrated by immunoprecipitation (Figure 3F), and had the ability to bind and increase chromatin opening when expressed in MSCs (Figures S4A–S4D). However, in contrast to EWS-FLI1, the BAF47-FLI1 fusion protein failed to display significant binding to GGAA repeats (Figure S4E), to induce de novo enhancer formation (Figures 3G and S4F), or to activate target gene expression (Figures 3H and S4G). Thus, fusion of the FLI1 C terminus directly to the BAF chromatin remodeling complex is insufficient to replicate the pioneer function of the EWS-FLI1-bound BAF complexes at GGAA repeats, suggesting additional critical properties of the EWSR1 fragment of the fusion protein.

Fusion of EWSR1 to FLI1 Confers Prion-like Phase Transition Properties

Phase transition is defined as the ability of a biological system to undergo a change of phase or state, including transitions from protein solutions to liquid-like phase-separated compartments that constitute membrane-less organelles (Aguzzi and Altmeyer, 2016). EWSR1 belongs to the FUS, EWSR1, TAF15 (FET) family of proteins and is characterized by intrinsically disordered low-complexity prion-like domains that have been shown to mediate multimerization, physiological liquid-liquid phase separation, and pathological protein aggregation (Couthouis et al., 2011, 2012; Kato et al., 2012; Kwon et al., 2013; Schwartz et al., 2013). In keeping with these findings, we confirmed the multimerization potential of EWSR1 by introducing GFP-tagged EWSR1 into U2OS cells and performing anti-GFP immunoprecipitation, which captured both the exogenously introduced variant (GFP-EWSR1) and endogenous EWSR1 (Figure S5A).

Given that the FET prion-like domains can undergo multimerization and concentration-dependent phase transition (Kato et al., 2012; Patel et al., 2015) and are retained in EWS-FLI1 as well as in other oncogenic fusions between FET proteins and transcription factors (Mertens et al., 2016), we first tested whether EWS-FLI1 can multimerize in the setting of Ewing sarcoma cells. In keeping with previous studies (Embree et al., 2009; Spahn et al., 2003), immunoprecipitation of tagged EWS-FLI1 showed a strong interaction with endogenous EWSR1 (Figure 4A). In contrast, tagged wild-type FLI1 only produced weak signals in these assays (Figure 4A). This suggests that a major difference between FLI1 and the fusion protein EWS-FLI1 is the ability to interact with wild-type EWSR1. In further support of this finding, immunoprecipitation followed by mass spectrometry of endogenous EWSR1 in SK-N-MC Ewing sarcoma cells, using an antibody targeting the C terminus of EWSR1 not present in EWS-FLI1, showed strong reciprocal interactions with EWS-FLI1 and also with other FET family proteins (TAF15 and FUS) as well as several members of the BAF complex (Figure 4B).

In order to assess the phase transition potential of EWS-FLI1, we first tested its ability to precipitate in presence of biotinylated isoxazole (b-isox), a recently identified compound with the ability to precipitate proteins with low complexity domains, including FET family proteins (Han et al., 2012; Kato et al., 2012). Our experiments show that EWS-FLI1 exhibits robust, concentration-dependent precipitation, comparable to and even higher than wild-type EWSR1 in SK-N-MC Ewing sarcoma cells (Figures 4C and 4D). This observation was recapitulated in U2OS cells stably expressing EWS-FLI1, but not in U2OS cells expressing wild-type FLI1 using the same conditions (Figure 4E).

Given that the presence of RNA has previously been shown to affect some EWSR1 interactions (Spahn et al., 2003), we tested the effects of RNase treatment in our experiments. We observed that RNase reduced b-isox-induced precipitation of EWSR1 but had small effects on EWS-FLI1 (Figure S5B). Similarly,

(D) EWS-FLI1 interacts with BAF complexes through both EWS N-terminal and FLI1 C-terminal fragments. (Left) Immunoblots from nuclear extracts show the expression of transiently transfected V5-EWSR1 N-terminal, V5-FLI1 C-terminal, or V5-EWS-FLI1 and the levels of endogenous BRG1 in HEK293-T cells. (Right) Co-immunoprecipitation experiments using anti-V5 antibodies show interactions with BAF. * indicates IgG heavy chains used for immunoprecipitation.
(E) Schematic of the BAF47-FLI1 fusion protein used in experiments in relation to EWS-FLI1 and BAF47.
(F) (Left) Immunoblots from nuclear extracts show lentiviral expression of BAF47-FLI1 fusion protein and the levels of endogenous BAF proteins in U2OS cells. (Right) Anti-FLI1 immunoprecipitation confirms an interaction between BAF47-FLI1 and BAF complex subunits.
(G) ATAC-seq signal intensity indicative of chromatin accessibility at GGAA repeat microsatellites in MSCs infected with either control vector, EWS-FLI1, BAF47-FLI1, or FLI1 wild-type.
(H) The fusion of the FLI1 C-terminal region to BAF47 is not sufficient for the activation of EWS-FLI1 target genes. Heatmap shows relative RNA-seq gene expression levels of genes associated with GGAA repeats and activated upon introduction of EWS-FLI1 in MSCs (rows, n = 207 genes). The columns show MSCs treated with either control vector, EWS-FLI1, wild-type FLI1, or BAF47-FLI1. Expression values were normalized by row. (Right) Example RNA-seq tracks over selected genes are shown.
See also Figures S3 and S4.

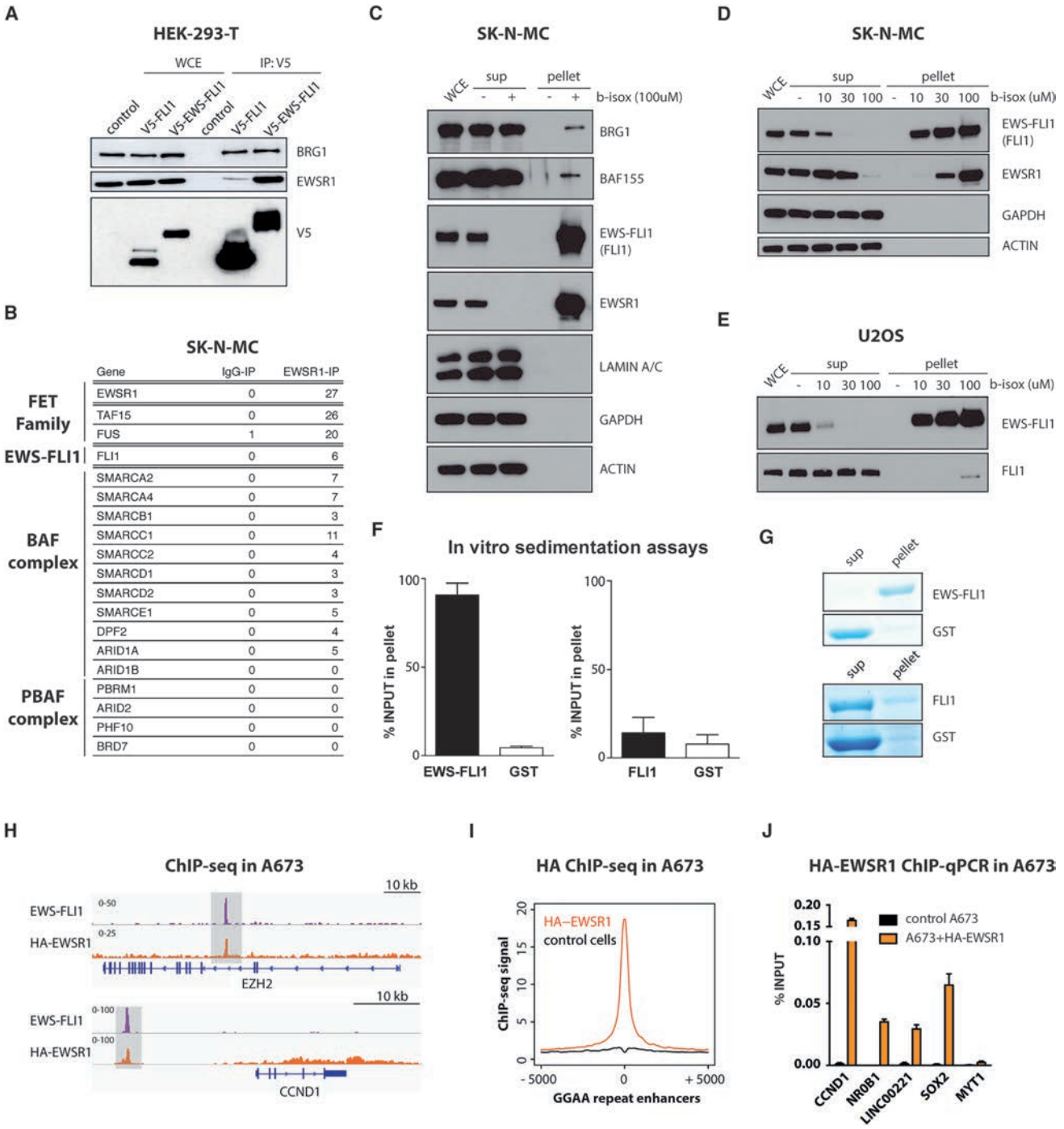


Figure 4. Fusion of EWSR1 to FLI1 Confers Multimerization and Phase Transition Properties

(A) Endogenous wild-type EWSR1 strongly interacts with EWS-FLI1 compared to wild-type FLI1. Immunoblots of whole-cell extract and anti-V5 immunoprecipitates from 293T cells transfected with either control vector, V5-FLI1, or V5-EWS-FLI1 are shown.

(B) Proteomic mass spectrometry of IgG and anti-EWSR1 immunoprecipitations performed in SK-N-MC cells. Table shows the number of unique peptides in each condition.

(C–E) EWS-FLI1 has a strong ability to precipitate in presence of b-isox compared to wild-type FLI1. (C) EWS-FLI1 precipitates in presence of 100 μM b-isox in Ewing sarcoma cell lysates. (D) EWS-FLI1 precipitates upon treatment with b-isox in a dose-dependent manner in Ewing sarcoma cell lysates. (E) Lentivirally expressed EWS-FLI1, but not wild-type FLI1, precipitates upon treatment with b-isox in U2OS osteosarcoma cell lysates.

(F and G) In vitro sedimentation assays from bacterially expressed and purified EWS-FLI1 or wild-type FLI1. (F) Quantification of two independent experiments is shown. (G) Representative examples of in vitro sedimentation assays are shown. The GST tag is cleaved as part of the assay and is used as a soluble internal control.

(legend continued on next page)

interactions between EWS-FLI1, EWSR1, and BAF complexes (BRG1) were not significantly affected by RNase treatment (Figures S5C and S5D). These results suggest distinct properties for EWS-FLI1-containing complexes and match previous in vitro results showing decreased EWSR1 homotypic interactions and stable heterotypic interactions with EWS-FLI1 upon RNase treatment (Spahn et al., 2003).

Wild-type EWSR1 has been previously shown to spontaneously precipitate in sedimentation experiments used to measure the intrinsic phase transition potential of purified proteins in vitro (Couthouis et al., 2012). We thus purified bacterially expressed glutathione S-transferase (GST)-tagged EWS-FLI1 and wild-type FLI1 to test their intrinsic phase transition potential. Sedimentation assays revealed that most EWS-FLI1 spontaneously precipitated, whereas wild-type FLI1 remained soluble in these conditions (Figures 4F, 4G, and S5E). These results demonstrate that the EWSR1 prion-like domain confers neomorphic intrinsic phase transition properties to the EWS-FLI1 oncogenic fusion protein. Supporting these observations, confocal imaging showed that EWS-FLI1 was detectable as nuclear dot-like structures after lentiviral expression in MSCs whereas wild-type FLI1 exhibited a more diffuse pattern (Figure S5F).

Having demonstrated interactions between EWS-FLI1 and EWSR1 and similar phase transition properties conferred by their prion-like domain in vitro, we expected to find these proteins in the same complexes at GGAA repeats in Ewing sarcoma cells. To assess this, we introduced hemagglutinin (HA)-tagged EWSR1 into A673 Ewing sarcoma cells, confirmed its nuclear localization and its ability to interact with endogenous EWS-FLI1 and the BAF complex ATPase subunit BRG1 (Figures S5G and S5H), and assessed its presence at EWS-FLI1-bound GGAA microsatellite repeats. Importantly, we observed substantial co-enrichment of HA-EWSR1 and EWS-FLI1 at GGAA microsatellite repeats, both by ChIP-seq and by validation ChIP-qPCR over selected target sites (Figures 4H–4J, S5I, and S5J). Furthermore, occupancy of HA-EWSR1 was decreased at these loci upon shRNA-mediated suppression of EWS-FLI1 (Figure S5K). Taken together, these observations suggest that EWS-FLI1 and wild-type EWSR1 are both present in the same macromolecular complexes at GGAA repeats in Ewing sarcoma.

Tyrosine Residues in the EWS-FLI1 Prion-like Domain Are Necessary for DNA Binding at GGAA Microsatellites and De Novo Enhancer Activation

We next determined whether phase transition mediated by the prion-like domain of EWS-FLI1 is necessary for DNA binding, BAF complex recruitment, and de novo enhancer activation at GGAA microsatellite repeats by generating a series of V5-tagged EWS-FLI1 mutant proteins lacking the ability to precipitate in vitro. The EWSR1 prion-like domain is rich in [G/S]Y[G/S] motifs (Figure S6A), and the substitution of these tyrosine residues with serine has been shown to abrogate phase transitions to hydrogels observed for the FET protein FUS (Kato et al., 2012; Kwon et al., 2013). Thus, we generated two EWS-FLI1 mutant proteins with point mutations altering either 12 or all 37 tyrosines in the prion-like domain (namely EWS(YS12)-FLI1 and EWS(YS37)-FLI1; Figures 5A and S6A). Both mutants were expressed and localized to the nucleus of lentivirally transduced MSCs as assessed by immunofluorescence in a manner comparable to that of EWS-FLI1 (Figures S6B and S6C). The EWS(YS12)-FLI1 mutant protein maintained significant interactions with wild-type EWSR1 and BRG1, and although diminished, b-isox-induced precipitation remained significantly higher than wild-type FLI1 (Figures 4E, 5B, and 5C). In contrast, EWS(YS37)-FLI1 displayed reduction in binding to wild-type EWSR1 and BRG1 as well as a profound loss of b-isox-induced precipitation to levels comparable to those of wild-type FLI1 (Figures 4E, 5B, and 5C). EWS(YS37)-FLI1, however, maintained the ability to homodimerize, a hallmark of ETS family transcription factors, as assessed by reciprocal immunoprecipitation experiments using overexpressed HA- and V5-tagged variants in HEK293-T cells (Figure S6D). In keeping with our results obtained with b-isox, in vitro sedimentation assays performed with purified GST-tagged EWS(YS37)-FLI1 showed a complete loss of its ability to spontaneously precipitate (Figures 5D and S6E).

Based on these findings, we further tested the ability of the EWS(YS37)-FLI1 mutant protein to bind GGAA microsatellite repeats and to create active enhancers once expressed in MSCs (Figure S6F). ChIP-seq experiments clearly demonstrated a dramatic reduction in binding of the EWS(YS37)-FLI1 mutant at these sites, as well as impaired BAF complex recruitment (Figures 5E and S6G). In line with this observation, DNA accessibility and marks of enhancer activity assessed by, respectively, ATAC-seq and ChIP-seq for H3K27ac were undetectable for the EWS(YS37)-FLI1 mutant (Figures 5E–5G, S6H, and S6I). Finally, consistent with the impairment of its biochemical properties, EWS(YS37)-FLI1 was not able to induce expression of GGAA microsatellite target genes after introduction in MSCs, whereas EWS(YS12)-FLI1 retained nearly full activity (Figure 5H). Taken together, our results demonstrate that the tyrosine residues in the EWS-FLI1 prion-like domain are necessary to mediate phase transitions and are required for its pioneer activity by allowing stable DNA binding, BAF complex recruitment at GGAA repeat microsatellites, and target gene activation.

Fusion of EWSR1 Prion-like Domain Fragments to the FLI1 C Terminus Is Sufficient to Recapitulate EWS-FLI1 Activity

Given that prion-like domains are unstructured, low-complexity protein sequences, we next tested whether any specific region within this domain is critical for EWS-FLI1 function. To this end, we generated a series of EWS-FLI1 internal deletion mutants (Figure S7A). Whereas the tyrosine residues that we

(H–J) EWSR1 is recruited to EWS-FLI1-bound GGAA repeat enhancers in Ewing Sarcoma. (H) Example ChIP-seq tracks show co-occupancy of EWS-FLI1 and HA-EWSR1 at GGAA repeat enhancers in A673 cells. Regions of co-occupancy are highlighted in light gray. (I) Composite plot shows HA-EWSR1 binding at EWS-FLI1 GGAA repeat enhancers in A673 cells. A 10-kb window centered on EWS-FLI1-bound repeat enhancer is shown. (J) ChIP-qPCR experiments validate HA-EWSR1 binding at EWS-FLI1 GGAA repeat enhancers associated with *CCND1*, *SOX2*, *NR0B1*, and *LINC00221*, but not a control region near *MYT1*. See also Figure S5.

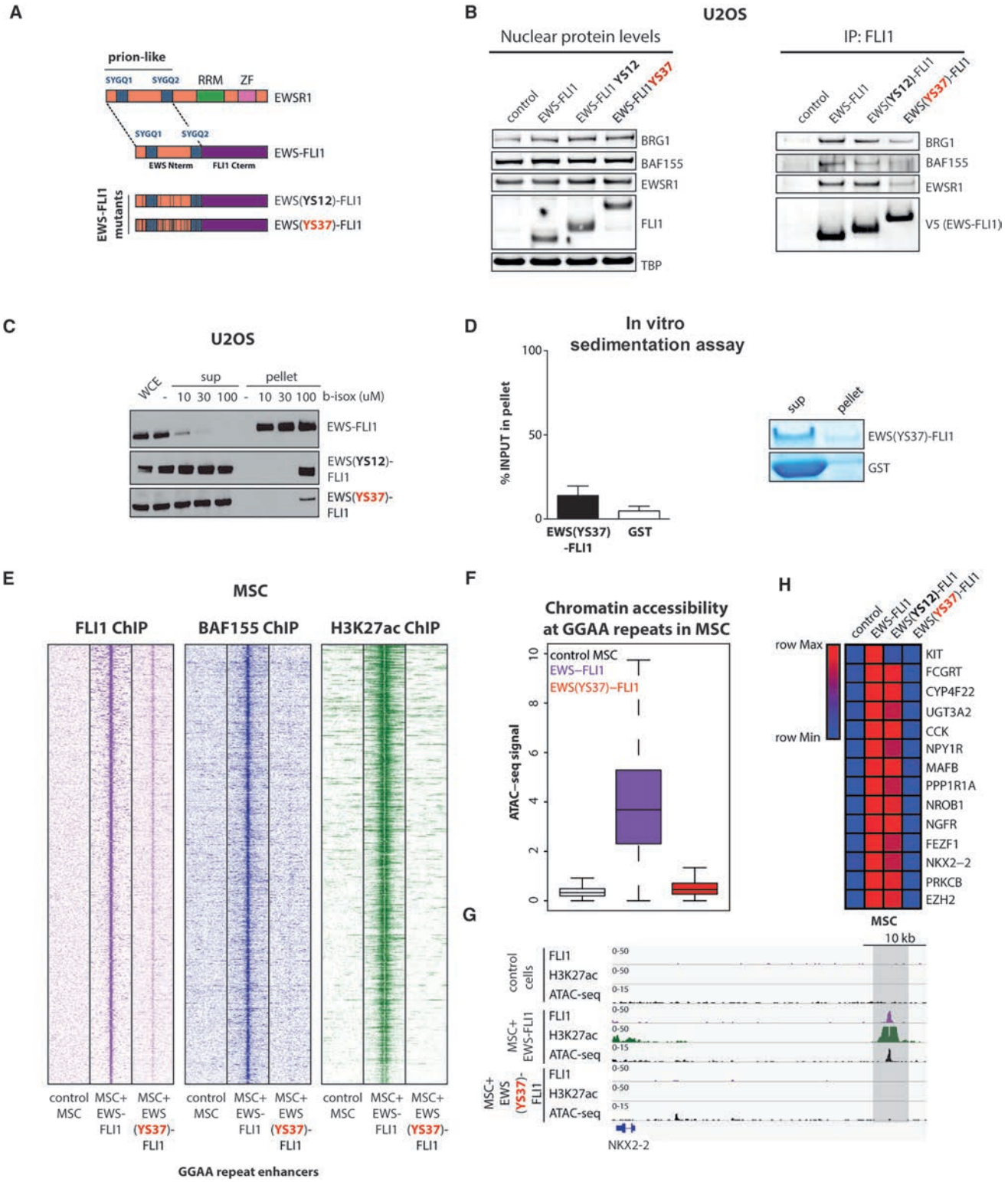


Figure 5. Tyrosine Residues in the EWS-FLI1 Prion-like Domain Are Necessary for DNA Binding at GGAA Microsatellites and Enhancer Induction

(A) Schematics of EWSR1, EWS-FLI1, and EWS-FLI1 tyrosine mutant variants used in experiments. Tyrosines (Y) mutated into serines (S) are shown as black bars within the EWS N-terminal prion-like domain. Mutants contained either 12 (YS12) or 37 (YS37) Y to S mutations. See also Figure S6A.

(legend continued on next page)

showed to be critical are mostly evenly distributed over the EWSR1 prion-like domain, there are two regions that contain exact [G/S]Y[G/S] motifs followed by a glutamine (SYGQ), which we have designated as SYGQ1 (also called FETBM1; Thomsen et al., 2013) and SYGQ2. We thus deleted these regions either independently or in combination (Figure S7A). All EWS-FLI1 deletion mutant proteins accumulated in the nucleus of MSCs and exhibited comparable binding to wild-type EWSR1 and BRG1 (Figures S7B–S7D). Similarly, EWS-FLI1 deletion mutants maintained b-isox-induced precipitation properties that were diminished when compared to EWS-FLI1 but significantly higher than those of wild-type FLI1 (Figure S7E). We next assessed changes in target gene expression after introduction of the different EWS-FLI1 mutants in MSCs and observed that a large set of target genes activated by GGAA microsatellites was still strongly induced by all constructs (Figure S7F). Thus, these experiments did not identify any singular EWSR1 subdomain necessary for EWS-FLI1 activity, suggesting that there is significant functional redundancy between different parts of the EWSR1 prion-like domain.

Given these results, we tested whether small EWSR1 fragments fused to the FLI1 C-terminal region were sufficient to recapitulate EWS-FLI1 function (Figure 6A). Strikingly, the fusion of the SYGQ1 fragment (37 amino acids) was sufficient to confer binding to wild-type EWSR1 and BRG1 (Figure S7G), b-isox-induced precipitation (Figure 6B), and induction of expression of EWS-FLI1 target genes in MSCs (Figure 6C). In line with these observations, the addition of a short EWSR1 fragment was sufficient to induce precipitation of purified GST-tagged SYGQ1-FLI1 in our in vitro sedimentation assays (Figures 6D and S7H). Fusion of the SYGQ2 fragment (64 amino acids) to the FLI1 C-terminal region was also sufficient to recapitulate EWS-FLI1 function (Figures 6A–6C and S7G). In addition, a more detailed analysis of the SYGQ2 fusion showed that this mutant protein was able to bind GGAA repeat microsatellites and to recruit BAF complexes, leading to DNA accessibility and enhancer activation, as assessed by ATAC-seq and ChIP-seq for H3K27ac, respectively (Figures 6E–6G).

Finally, the distinct abilities of mutant EWS-FLI1 proteins to recapitulate EWS-FLI1-mediated gene expression programs were evident by comparing RNA-seq expression profiles. MSCs expressing the short fragment fusion SYGQ2-FLI1 demonstrated clustering with EWS-FLI1-expressing cells (Figure 6H) whereas MSCs expressing the tyrosine mutant EWS(YS37)-FLI1 clustered with control cells infected with an empty vector (Figure 6H). In agreement with these results, cell growth arrest and phenotypic changes induced by EWS-FLI1 knockdown were rescued by the SYGQ2-FLI1 mutant protein, but not by EWS(YS37)-FLI1 (Figure S7I). These results thus demonstrate that even small isolated fragments of the EWSR1 prion-like domain are sufficient to recapitulate the function of the full EWS-FLI1 fusion on chromatin and to induce gene expression programs associated with GGAA microsatellites in Ewing sarcoma tumors.

DISCUSSION

Taken together, our studies elucidate critical mechanisms whereby EWSR1 contributes to the oncogenic activity of EWS-FLI1 (Figure 7). Low-complexity domains have been proposed to play a variety of roles in normal cellular functions and in disease states (Aguzzi and Altmeyer, 2016; March et al., 2016). In normal cells, proteins containing intrinsically disordered domains are believed to have the ability to form liquid-like compartments that, in the case of the FUS protein, have been observed in vivo in the cytoplasm upon stress and at sites of DNA damage in the nucleus (Patel et al., 2015). This process can be altered by mutations in the prion-like domain or local protein accumulation, as observed in amyotrophic lateral sclerosis (ALS), where the pathological aggregation of low-complexity proteins may be the result of the conversion from liquid to solid states (Couthouis et al., 2011, 2012; Patel et al., 2015; Sun et al., 2011). Our results suggest that a similar phase transition mechanism could allow protein accumulation and stabilization at tumor-specific DNA binding sites in Ewing sarcoma as well as in other tumor types involving fusions of FET family proteins with transcription factors. Further characterization will be necessary to pinpoint the exact nature of these transitions at chromatin. Indeed, it will be of great interest to establish whether EWS-FLI1 forms fibrils or amorphous aggregates or undergoes liquid-liquid phase separation in vivo and to what extent the disruption of these neomorphic properties may be exploited from a therapeutic standpoint.

Mutations in the genes encoding BAF complex subunits are observed in many tumor types, suggesting that tumor-specific changes in BAF complex composition and function play

(B) (Left) Immunoblots show nuclear input levels of EWSR1 and BAF proteins and the lentiviral expression of EWS-FLI1, EWS(YS12)-FLI1, or EWS(YS37)-FLI1 mutants in U2OS cells. (Right) Co-immunoprecipitation experiments using anti-FLI1 antibodies reveal that the EWS(YS37)-FLI1 mutant exhibits decreased interactions with wild-type EWSR1 and BRG1.

(C) Dose-dependent b-isox precipitation assay after lentiviral expression of either EWS-FLI1 or mutants EWS(YS12)-FLI1 or EWS(YS37)-FLI1 in U2OS cells.

(D) In vitro sedimentation assay from bacterially expressed and purified EWS(YS37)-FLI1. (Left) Quantification of two independent experiments is shown. (Right) Representative examples of in vitro sedimentation assays are shown. The GST tag is cleaved as part of the assay and is used as a soluble internal control.

(E) Heatmaps of FLI1, BAF155, and H3K27ac ChIP-seq signal densities in MSCs treated with either control vector, EWS-FLI1, or EWS(Y37)-FLI1 mutant. 10-kb windows in each panel are centered on EWS-FLI1-bound GGAA repeat enhancer sites (n = 812).

(F) ATAC-seq signal intensity indicative of chromatin accessibility at GGAA repeat microsatellites in MSCs infected with either control, EWS-FLI1, or EWS(YS37)-FLI1 mutant.

(G) Representative example ChIP-seq tracks of FLI1 (EWS-FLI1), H3K27Ac, and ATAC-seq signals over the NKX2-2 locus in MSCs expressing either control, EWS-FLI1, or EWS(YS37)-FLI1 mutant.

(H) Heatmap shows changes in expression detected by RT-qPCR for selected EWS-FLI1 target genes associated with GGAA repeats after infection of MSCs with either control vector, EWS-FLI1, EWS(YS12)-FLI1, or EWS(YS37)-FLI1 mutants.

See also Figure S6.

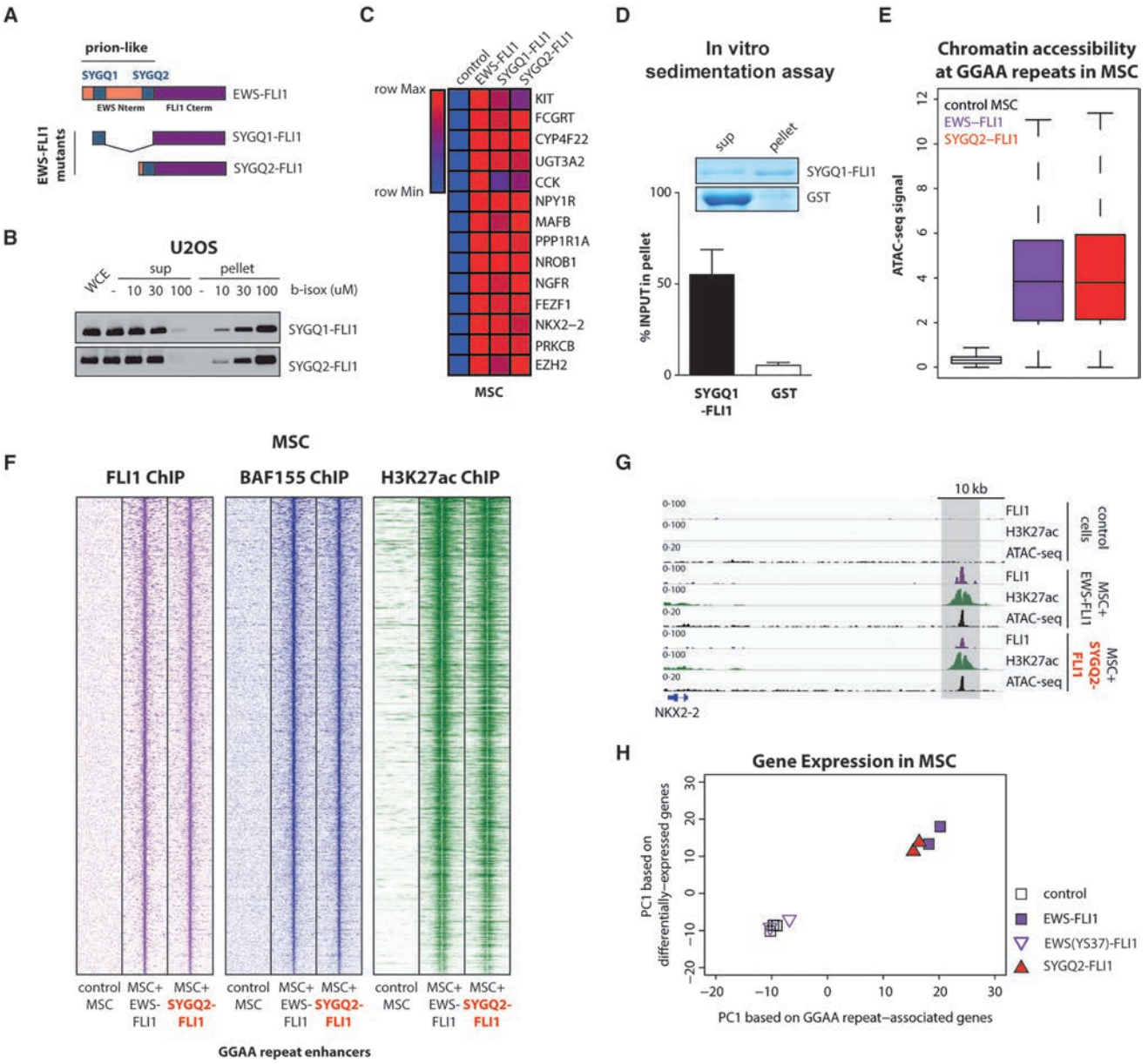


Figure 6. Fusion of Fragments of the EWSR1 Prion-like Domain to the FLI1 C Terminus Is Sufficient to Recapitulate EWS-FLI1 Activity

(A) Schematic representation of EWS-FLI1 prion-like domain mutants used in experiments. SYGQ1 or SYGQ2 fragments are fused to the FLI1 C terminus.

(B) Dose-dependent precipitation assay in presence of b-isox after lentiviral expression of fusion proteins in U2OS cells.

(C) Heatmap shows changes in expression detected by RT-qPCR for selected EWS-FLI1 target genes associated with GGAA repeats after infection of MSCs with either control vector, EWS-FLI1, SYGQ1-FLI1, or SYGQ2-FLI1 fusion proteins.

(D) In vitro sedimentation assay from bacterially expressed and purified SYGQ1-FLI1. (Top) Representative examples of in vitro sedimentation assays are shown. The GST tag is cleaved as part of the assay and is used as a soluble internal control. (Bottom) Quantification of two independent experiments is shown.

(E) ATAC-seq signal intensity indicative of chromatin accessibility at GGAA repeat microsatellites in MSCs infected with either control, EWS-FLI1, or the SYGQ2-FLI1 fusion protein.

(F) Heatmaps of FLI1, BAF155, and H3K27ac ChIP-seq signal densities in MSCs treated with either control vector, EWS-FLI1, or the SYGQ2-FLI1 fusion protein. 10-kb windows in each panel are centered on EWS-FLI1-bound GGAA repeat enhancer sites (n = 812).

(G) Example ChIP-seq tracks of FLI1 (EWS-FLI1), H3K27Ac, and ATAC-seq signals over the NKX2-2 locus in MSCs expressing either control, EWS-FLI1, or the SYGQ2-FLI1 fusion protein.

(H) Principal-component analysis (PCA) plot showing PC1 for the 207 target genes associated with EWS-FLI1 GGAA repeats sites (x axis) and PC1 for the 158 remaining differentially expressed genes in EWS-FLI1-expressing cells (y axis). RNA-seq datasets are from MSCs infected with either control vector, EWS-FLI1, EWS(YS37)-FLI1, or the SYGQ2-FLI1 fusion protein.

See also Figure S7.

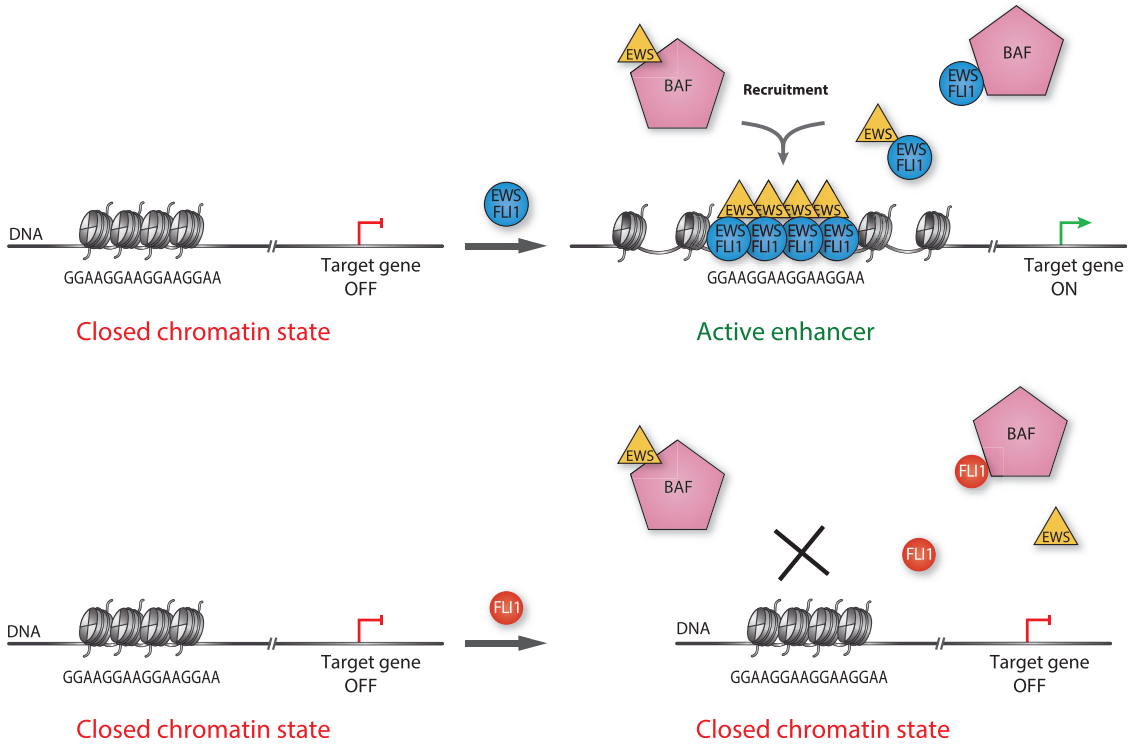


Figure 7. Model for EWS-FLI1 Binding at GGAA Repeat Microsatellites and Enhancer Activation in Ewing Sarcoma
(Top) In presence of EWS-FLI1, multimerization is required for stable binding at GGAA repeats and recruitment of BAF complexes. (Bottom) Wild-type FLI1 cannot stably bind at GGAA repeats.

important roles in human cancer (Kadoch et al., 2013). We now show that, in the absence of genetic alterations in BAF complex components, the EWS-FLI1 fusion protein utilizes an alternate mechanism to retarget BAF complexes to distal regulatory elements. This, in turn, leads to the activation of target genes and the establishment of a tumor-specific regulatory network and transcriptional program. Tumor-specific retargeting of BAF complexes may thus play an important role in many tumor types in which BAF subunits are not genetically altered, thereby expanding the already wide-spanning role of these complexes in human cancer. Given our prior results showing that EWS-FLI1 also mediates the recruitment of the acetyl-transferase P300 and mixed-lineage leukemia (MLL) complexes to GGAA repeats (Riggi et al., 2014), it will be important to determine how the interplay between these chromatin-modifying complexes leads to de novo enhancer activation.

Alterations in BAF subunits may also be associated with changes in both subunit composition and configuration of the BAF complex. This has been demonstrated in synovial sarcoma, where the SS18-SSX translocation leads to the formation of a modified version of the BAF complex that incorporates the translocation protein (Kadoch and Crabtree, 2013). Our results show that, despite strong interactions with BAF, EWS-FLI1 is not incorporated into the core BAF complex and therefore does not affect BAF composition directly. Nevertheless, the recruitment of BAF complexes to GGAA repeats in association with the potential phase transitions induced by the EWSR1 prion-like motif may involve an unusual configuration that could be exploited for the development of new therapeutic approaches.

Our experiments comparing EWS-FLI1, FLI1, and various mutant proteins in MSCs establish a direct link between the neomorphic properties conferred by the EWS prion-like domain and a tumor-specific gene regulation program mediated by de novo enhancer activation at GGAA microsatellites. Indeed, FLI1 exhibited substantial occupancy at canonical ETS binding sites but, in contrast to EWS-FLI1, was not able to bind GGAA repeats. This is in agreement with the observation that Ewing sarcoma cell lines were the only cell types with open chromatin at EWS-FLI1-bound GGAA microsatellite repeats in a survey of 112 cell types analyzed by DNase hypersensitivity (Riggi et al., 2014). These data included profiles for cells that express high levels of endogenous FLI1 and yet did not show signals at GGAA repeat sites. This is also consistent with recent studies suggesting that various types of stem cells may contain destabilized nucleosomes detected by FAIRE (formaldehyde-assisted isolation of regulatory elements) at repetitive elements, but not open chromatin (Gomez et al., 2016). Importantly, prior electrophoretic mobility shift (EMSA) assays showed binding of both EWS-FLI1 and FLI1 to GGAA repeats in vitro (Gangwal et al., 2008), suggesting that the difference between these two proteins may only be evident in the appropriate in vivo chromatin context. The application of methods for direct high-throughput genome-wide analysis of chromatin thus provides opportunities

for new insights into critical mechanisms of oncogenic gene regulation.

In conclusion, our study demonstrates that a prion-like domain can confer neomorphic properties to fusion proteins that lead to retargeting of key chromatin regulators and the establishment of an oncogenic gene regulatory program. Similar events mediated by FET family proteins or other intrinsically disordered proteins are likely to play important roles in generating tumor-specific regulatory elements in a variety of tumor types and may constitute attractive targets for therapeutic development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mesenchymal stem cells
  - Cell lines
  - Bacterial Cultures
- METHOD DETAILS
  - Lentiviral Generation
  - Transient Transfections
  - Real-Time Quantitative RT-PCR
  - Western Blot Analysis
  - Nuclear Extract Preparation
  - Immunoprecipitations
  - Depletion Studies
  - Urea Denaturation Studies
  - Density Sedimentation Analyses
  - Mass Spectrometry
  - ChIP-seq
  - ChIP-seq Bioinformatic Analysis
  - ChIP-qPCR
  - RNA-Seq
  - RNA-seq bioinformatic Processing
  - Definition of target genes associated with EWS-FLI1 bound GGAA-repeats
  - Principal Component Analysis
  - ATAC-seq Genome-Wide DNA Accessibility Assay
  - ATAC-seq bioinformatic Processing
  - Immunofluorescence Stainings
  - Cell Viability Assays
  - G1 Arrest Experiments
  - Biotinylated isoxazole-Mediated Precipitation
  - Protein purification and sedimentation assays
- QUANTIFICATION AND STATISTICAL ANALYSES
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.07.036.

AUTHOR CONTRIBUTIONS

G.B., G.J.S., C.K., and M.N.R. designed the study and wrote the manuscript. G.B., G.J.S., N.R., R.B., B.N., M.E.A., S.R., A.V., and M.J.M. performed the experiments. S.I., G.B., and M.N.R. conducted bioinformatics analyses. N.R., L.C.B., L.Z., and I.S. provided necessary reagents and conceptual advice.

ACKNOWLEDGMENTS

We thank S. Gillepsie, M. Bobbin, M. Aryee, and N. Mashtalir as well as members of the Rivera, Kadoch, Bernstein, and Suva labs for valuable guidance and advice. We thank L. Nieman of the MGH Cancer Center/Molecular Pathology Confocal Core (MGH) and members of the Taplin Mass Spec Facility (HMS) for assistance with experiments and data analysis. We thank G.R. Crabtree and A. Kuo for kindly sharing the homemade BAF155 (SMARCC1) antibody used in ChIP-seq experiments. G.J.S. is supported by the Department of Defense Prostate Cancer Research Program Postdoctoral Training Award W81XWH-15-1-0659. N.R. is supported by the Swiss National Science Foundation Professorship grant (PP00P3-157468/1) and the MEDIC Foundation. R.B. is supported by a NIH Pathway to Independence Award (1K99CA212154). C.K. is supported by the NIH DP2 New Innovator Award 1DP2CA195762-01, the American Cancer Society Research Scholar Award RSG-14-051-01-DMC, and the Pew-Stewart Scholars in Cancer Research Grant. This work was also supported by the Alex's Lemonade Stand Foundation Young Investigator Award (C.K.). M.N.R. is supported by the V Foundation for Cancer Research. M.N.R. is a consultant for Loxo and receives research support from Affymetrix. C.K. is a scientific founder, shareholder, and consultant for Foghorn Therapeutics.

Received: February 23, 2017
Revised: June 14, 2017
Accepted: July 21, 2017
Published: August 24, 2017

REFERENCES

Aguzzi, A., and Altmeyer, M. (2016). Phase separation: linking cellular compartmentalization to disease. Trends Cell Biol. *26*, 547–558.

Boulay, G., Awad, M.E., Riggi, N., Archer, T.C., Iyer, S., Boonseng, W.E., Rossetti, N.E., Naigles, B., Rengarajan, S., Volorio, A., et al. (2017). OTX2 activity at distal regulatory elements shapes the chromatin landscape of group 3 medulloblastoma. Cancer Discov. *7*, 288–301.

Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic landscape of the Ewing sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet. *10*, e1004475.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Couthouis, J., Hart, M.P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., Liu, A.X., Ramos, D., Jethava, N., Hosangadi, D., et al. (2011). A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. USA *108*, 20881–20890.

Couthouis, J., Hart, M.P., Erion, R., King, O.D., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H.J., Mojsilovic-Petrovic, J., Panossian, S., et al. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum. Mol. Genet. *21*, 2899–2911.

Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer Discov. *4*, 1326–1341.

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., et al. (1992). Gene fusion with an

ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature *359*, 162–165.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Embree, L.J., Azuma, M., and Hickstein, D.D. (2009). Ewing sarcoma fusion protein EWSR1/FLI1 interacts with EWSR1 leading to mitotic defects in zebrafish embryos and human cell lines. Cancer Res. *69*, 4363–4371.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Gangwal, K., Sankar, S., Hollenhorst, P.C., Kinsey, M., Haroldsen, S.C., Shah, A.A., Boucher, K.M., Watkins, W.S., Jorde, L.B., Graves, B.J., and Lessnick, S.L. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proc. Natl. Acad. Sci. USA *105*, 10149–10154.

Gomez, N.C., Hepperla, A.J., Dumitru, R., Simon, J.M., Fang, F., and Davis, I.J. (2016). Widespread chromatin accessibility at repetitive elements links stem cells with human cancer. Cell Rep. *17*, 1607–1620.

Guillon, N., Tirode, F., Boeva, V., Zynovyev, A., Barillot, E., and Delattre, O. (2009). The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. PLoS ONE *4*, e4932.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768–779.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Herrero-Martin, D., Fourtouna, A., Niedan, S., Riedmann, L.T., Schwentner, R., and Aryee, D.N. (2011). Factors affecting EWS-FLI1 activity in Ewing's sarcoma. Sarcoma *2011*, 352580.

Kadoch, C., and Crabtree, G.R. (2013). Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. Cell *153*, 71–85.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci. Adv. *1*, e1500447.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat. Genet. *45*, 592–601.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Kovar, H. (2011). Dr. Jekyll and Mr. Hyde: the two faces of the FUS/EWS/TAF15 protein family. Sarcoma *2011*, 837474.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049–1060.

Lander, E.S. (2011). Initial impact of the sequencing of the human genome. Nature *470*, 187–197.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

March, Z.M., King, O.D., and Shorter, J. (2016). Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. Brain Res. *1647*, 9–18.

Mertens, F., Antonescu, C.R., and Mitelman, F. (2016). Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. Genes Chromosomes Cancer *55*, 291–310.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Patel, M., Simon, J.M., Iglesia, M.D., Wu, S.B., McFadden, A.W., Lieb, J.D., and Davis, I.J. (2012). Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. Genome Res. *22*, 259–270.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. Cell *162*, 1066–1077.

Phelan, M.L., Sif, S., Narlikar, G.J., and Kingston, R.E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol. Cell *3*, 247–253.

Pohl, A., and Beato, M. (2014). bwtool: a tool for bigWig files. Bioinformatics *30*, 1618–1619.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Riggi, N., Cironi, L., Provero, P., Suvà, M.L., Kaloulis, K., Garcia-Echeverria, C., Hoffmann, F., Trumpp, A., and Stamenkovic, I. (2005). Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. Cancer Res. *65*, 11459–11468.

Riggi, N., Suvà, M.L., Suvà, D., Cironi, L., Provero, P., Tercier, S., Joseph, J.M., Stehle, J.C., Baumer, K., Kindler, V., and Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res. *68*, 2176–2185.

Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., et al. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell *26*, 668–681.

Roberts, C.W., Leroux, M.M., Fleming, M.D., and Orkin, S.H. (2002). Highly penetrant, rapid tumorigenesis through conditional inversion of the tumor suppressor gene Snf5. Cancer Cell *2*, 415–425.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell Rep. *5*, 918–925.

Spahn, L., Siligan, C., Bachmaier, R., Schmid, J.A., Aryee, D.N., and Kovar, H. (2003). Homotypic and heterotypic interactions of EWS, FLI1 and their oncogenic fusion protein. Oncogene *22*, 6819–6829.

Sun, Z., Diaz, Z., Fang, X., Hart, M.P., Chesi, A., Shorter, J., and Gitler, A.D. (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS Biol. *9*, e1000614.

Tanaka, K., Iwakuma, T., Harimaya, K., Sato, H., and Iwamoto, Y. (1997). EWS-Fli1 antisense oligodeoxynucleotide inhibits proliferation of human Ewing's sarcoma and primitive neuroectodermal tumor cells. J. Clin. Invest. *99*, 239–247.

Thomsen, C., Grundevik, P., Elias, P., Ståhlberg, A., and Aman, P. (2013). A conserved N-terminal motif is required for complex formation between FUS, EWSR1, TAF15 and their oncogenic fusion proteins. FASEB J. *27*, 4965–4974.

Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium (2014). Genomic landscape

of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. *4*, 1342–1353.

Tomazou, E.M., Sheffield, N.C., Schmidl, C., Schuster, M., Schönegger, A., Datlinger, P., Kubicek, S., Bock, C., and Kovar, H. (2015). Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. Cell Rep. *10*, 1082–1095.

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck, M.F., Ambros, P., Handgretinger, R., Aurias, A., and Delattre, O. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature *394*, 203–206.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

Article

Cell

In Situ Architecture and Cellular Interactions of PolyQ Inclusions

Felix J.B. Bäuerlein,[1] Itika Saha,[2,6] Archana Mishra,[3,6] Maria Kalemanov,[1,4] Antonio Martínez-Sánchez,[1] Rüdiger Klein,[3,5] Irina Dudanova,[3] Mark S. Hipp,[2,5] F. Ulrich Hartl,[2,5,*] Wolfgang Baumeister,[1,7,*] and Rubén Fernández-Busnadiego[1]

[1]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[2]Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[3]Department of Molecules, Signaling, and Development, Max Planck Institute of Neurobiology, 82152 Martinsried, Germany
[4]Graduate School of Quantitative Biosciences Munich, 81337 Munich, Germany
[5]Munich Cluster for Systems Neurology (SyNergy), 80336 Munich, Germany
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: uhartl@biochem.mpg.de (F.U.H.), baumeist@biochem.mpg.de (W.B.)
http://dx.doi.org/10.1016/j.cell.2017.08.009

SUMMARY

Expression of many disease-related aggregation-prone proteins results in cytotoxicity and the formation of large intracellular inclusion bodies. To gain insight into the role of inclusions in pathology and the in situ structure of protein aggregates inside cells, we employ advanced cryo-electron tomography methods to analyze the structure of inclusions formed by polyglutamine (polyQ)-expanded huntingtin exon 1 within their intact cellular context. In primary mouse neurons and immortalized human cells, polyQ inclusions consist of amyloid-like fibrils that interact with cellular endomembranes, particularly of the endoplasmic reticulum (ER). Interactions with these fibrils lead to membrane deformation, the local impairment of ER organization, and profound alterations in ER membrane dynamics at the inclusion periphery. These results suggest that aberrant interactions between fibrils and endomembranes contribute to the deleterious cellular effects of protein aggregation.

INTRODUCTION

Protein aggregation is a hallmark of many neurodegenerative disorders, including Huntington's disease (HD), amyotrophic lateral sclerosis, and Parkinson's disease (Hipp et al., 2014; Ross and Poirier, 2004). Soluble oligomeric aggregates have been linked to toxicity in many of these proteinopathies, whereas the large insoluble inclusions found in the brains of patients and in animal and cellular models are often regarded as oligomer-sequestering protective entities (Arrasate and Finkbeiner, 2012; Bucciantini et al., 2002; Haass and Selkoe, 2007). However, the exact contribution of inclusions to pathology remains poorly characterized, and detailed structural information on inclusion body (IB) organization in unperturbed cellular environments is missing. The traditional methods of sample preparation including chemical fixation, dehydration, and heavy metal staining are not compatible with high-resolution structural studies.

Recent technical advances make cryo-electron tomography (cryo-ET) ideally suited to address the challenge of studying IB structures in situ, enabling the analysis of their interactions with their cellular environment. Cryo-focused ion beam milling (cryo-FIB) opens windows into the interior of frozen-hydrated vitrified cells by producing thin lamellas that can be studied in three dimensions (3D) by cryo-ET (Mahamid et al., 2016; Rigort et al., 2012a). Furthermore, cryo-ET imaging capabilities have been dramatically expanded by the introduction of direct electron detectors (Li et al., 2013) and the Volta phase plate, which enables imaging close to focus with maximum contrast (Asano et al., 2015; Danev et al., 2014; Mahamid et al., 2016). Thus, cryo-ET allows the 3D visualization of pristinely preserved cells at molecular resolution.

Here, we capitalize on these developments to dissect IB architecture within unstained, frozen-hydrated, vitrified mammalian cells, using polyglutamine (polyQ)-expanded huntingtin (Htt) exon 1 as a model aggregating protein. The expansion of a polyQ stretch in the N-terminal region of Htt renders the mutant protein highly aggregation-prone and leads to HD in humans (Finkbeiner, 2011; Macdonald, 1993; Scherzinger et al., 1997). PolyQ-expanded Htt oligomers have been implicated in the misregulation of, among others, transcription, vesicular traffic, autophagy, and the function of the endoplasmic reticulum (ER) and mitochondria (Arrasate and Finkbeiner, 2012), but it is not known whether inclusions play any role in these phenomena. Furthermore, several other neurodegenerative disorders including spinal and bulbar muscular atrophy and spinocerebellar ataxias are also caused by polyQ expansions in otherwise non-related proteins (Orr and Zoghbi, 2007). Our results show that inclusions of polyQ-expanded Htt exon 1 are formed by fibrils that impinge on cellular membranes, especially those of the ER. These interactions result in membrane deformation and possibly rupture, as well as impaired ER organization and dynamics around inclusion bodies.

RESULTS

3D Architecture of Htt97Q Inclusion Bodies

Inclusion bodies in HD patient neurons are formed by N-terminal fragments of Htt harboring the polyQ-expanded exon 1,

expression of which suffices to recapitulate HD-like symptoms in mice (Davies et al., 1997; DiFiglia et al., 1997; Mangiarini et al., 1996; Yamamoto et al., 2000). We transfected mouse primary neurons with GFP-tagged Htt exon 1 containing 97 Q (Htt97Q-GFP). Non-apoptotic cells containing inclusions were identified by live imaging and subsequently vitrified by plunge freezing following a short incubation in medium containing 10% glycerol to ensure full vitrification. Correlative microscopy allowed us to target inclusions for cryo-FIB milling to yield 150- to 250-nm-thick cellular lamellas that were imaged in 3D by Volta phase plate cryo-ET (Figure S1). Htt97Q-GFP inclusions were roughly spherical, ~3 μm in diameter and mostly cytosolic, although nuclear inclusions were also found. Both cytosolic and nuclear inclusions consisted of a network of amyloid-like fibrils with a diameter of 7–8 nm and length of 125 ± 81 nm (mean ± SD) (Figures 1A, 1B, 1E, 2A, and 2B). Thus, despite the similarity in fibril length, the cellular organization of Htt97Q-GFP fibrils was very different from the bundles of parallel fibrils formed by Sup35 yeast prions (Kawai-Noma et al., 2010; Saibil et al., 2012). The analysis of fibril curvature provided insights into fibril flexibility. We calculated an average fibril persistence length of 2.6 ± 0.1 μm (mean ± SD) and a Young's modulus of 52 ± 2 MPa (mean ± SD) (Figure S2A), in the range of other amyloid fibrils or actin (Mahamid et al., 2016; Wegmann et al., 2010). Despite the dense appearance of the network, the fibrils occupied less than 3% of the IB volume (Figure 1F).

In situ fibrils were morphologically similar to those formed in vitro, but did not associate laterally to form wider ribbons (Scherzinger et al., 1997) (Figure 1A, inset). Interestingly, Htt97Q-GFP fibrils were decorated by regularly spaced globular densities of ~6 nm in diameter consistent in size with GFP dimers. To further investigate the nature of these densities, cells were co-transfected with untagged 97Q Htt exon 1 (Htt97Q) and mCherry-ubiquitin. The latter can be conjugated to target proteins in a manner similar to wild-type ubiquitin (Qian et al., 2002) and its recruitment to inclusion bodies (Hipp et al., 2012) served as a surrogate fluorescent marker for Htt97Q inclusions. Given the relatively low transfection rates obtained in neurons, these and further experiments were carried out in HeLa cells to increase the number of cells amenable to cryo-ET analysis. Htt97Q and Htt97Q-GFP inclusions in HeLa cells were almost identical to those in neurons in terms of overall architecture and fibril morphology (Figures 1E–1H and 3A–3E). Both in neurons and HeLa cells the fibrils were radially arranged in most, but not all inclusions (Figures 1A, 1B, 2, and 3A–3C). Although Htt97Q-GFP and untagged Htt97Q fibrils were similar in diameter and length (Figure 1E), untagged Htt97Q fibrils were not decorated by additional densities (Figure 1H). This demonstrates that the fibrils consisted of Htt97Q, and suggests a molecular organization in which the polyQ regions form the fibril core and the more flexible C-terminal sequence protrudes outward (Bugg et al., 2012; Isas et al., 2015; Lin et al., 2017). The presence of GFP resulted in a 50% reduction in fibril density within the IB (Figure 1F) and a 25% increase in fibril stiffness (Figures S2A–S2C).

Htt97Q and Htt97Q-GFP fibrils were always observed as part of inclusions, suggesting that inclusions are the main sites of fibril growth in the cell (Ossato et al., 2010). Although only a small fraction of IB volume was occupied by fibrils (Figure 1F), the core of the aggregates was mostly devoid of large macromolecules such as ribosomes, which were abundant at the IB periphery (Figures 1A and 1B).

Htt97Q Fibrils Impinge on Cellular Endomembranes

For all cytosolic inclusions analyzed (N = 5 inclusions, neurons, Htt97Q-GFP; N = 10, HeLa, Htt97Q-GFP; N = 8, HeLa, Htt97Q; Table S1), fibrils in extended areas of the IB periphery visibly contacted the membranes of various organelles. These included mitochondria, lysosomes and, most prominently, the ER (Figures 1A, 1B, 3, S3B, and S3D; Movie S1), but no observable association with autophagic structures. The electron densities of membranes and fibrils often appeared continuous at points of contact (Figures 1D, 3D, and S3D), indicating that the fibril-membrane distance was shorter than the pixel size (1.7 or 2.8 nm). Ribosome-free ER tubes often protruded into the IB, apparently interacting extensively with the fibrillar network (Figures 1B–1D, 3B, 3C, and 3G). Both the ends and sides of fibrils directly touched the membranes, and these regions displayed extremely high membrane curvature (Figures 1C, 1D, and 3D–3F). A systematic analysis showed that ER membrane curvature was higher at the IB interface than in more distal areas, and also compared to ER membranes in cells expressing diffuse Htt97Q-GFP without visible inclusions or non-pathogenic Htt25Q-GFP (Figures S2D and S4).

A heterogeneous population of vesicles was embedded in most cytosolic inclusions (100% of inclusions, neurons, and Htt97Q-GFP; 94%, HeLa and Htt97Q-GFP; 82%, HeLa and Htt97Q) at sites of interaction with organelles (Figures 1A, 1B, 1D, and 3A–3F). Many of these vesicles were highly irregular in shape and were often in contact with fibrils at regions of high membrane curvature, suggesting that they resulted from the disruption of organellar membranes following interaction with the fibrils (Figures 1D and 3D–3F). Ribosomes were bound to the membranes of some of these vesicles, suggesting that they originated from ruptured ER membranes (Figures 3E and 3F). Interestingly, similar membrane deformations at contact regions with amyloid fibrils leading to membrane disruption were previously observed in vitro with liposomes (Milanesi et al., 2012).

Comparable IB architecture and membrane interactions were found in cells expressing Htt64Q-GFP (Figures 2C and 2D), arguing against a strong influence of the exact length of the expanded polyQ tract on fibril arrangement and cellular interactions. Htt97Q and Htt97Q-GFP nuclear inclusions were similar in overall architecture to cytosolic ones but did not contain vesicles and did not contact the inner nuclear membrane (Figures 2A and 2B), indicating differential mechanisms of cellular interaction (Benn et al., 2005; Gu et al., 2015; Liu et al., 2015). In summary, the fibrils forming cytosolic polyQ inclusions have the potential to deform and perhaps disrupt cellular membranes in their proximity, particularly those of the ER, both in mouse neurons and in human cells.

Htt97Q Inclusions Alter ER Organization and Dynamics

Light microscopy in HeLa cells confirmed the association of Htt97Q-GFP inclusions with the ER (Figure 4A). Interestingly, some inclusions contained puncta positive for ER markers but disconnected from the ER network (Figure 4B), in line with the assumption that some of the vesicles found within inclusions

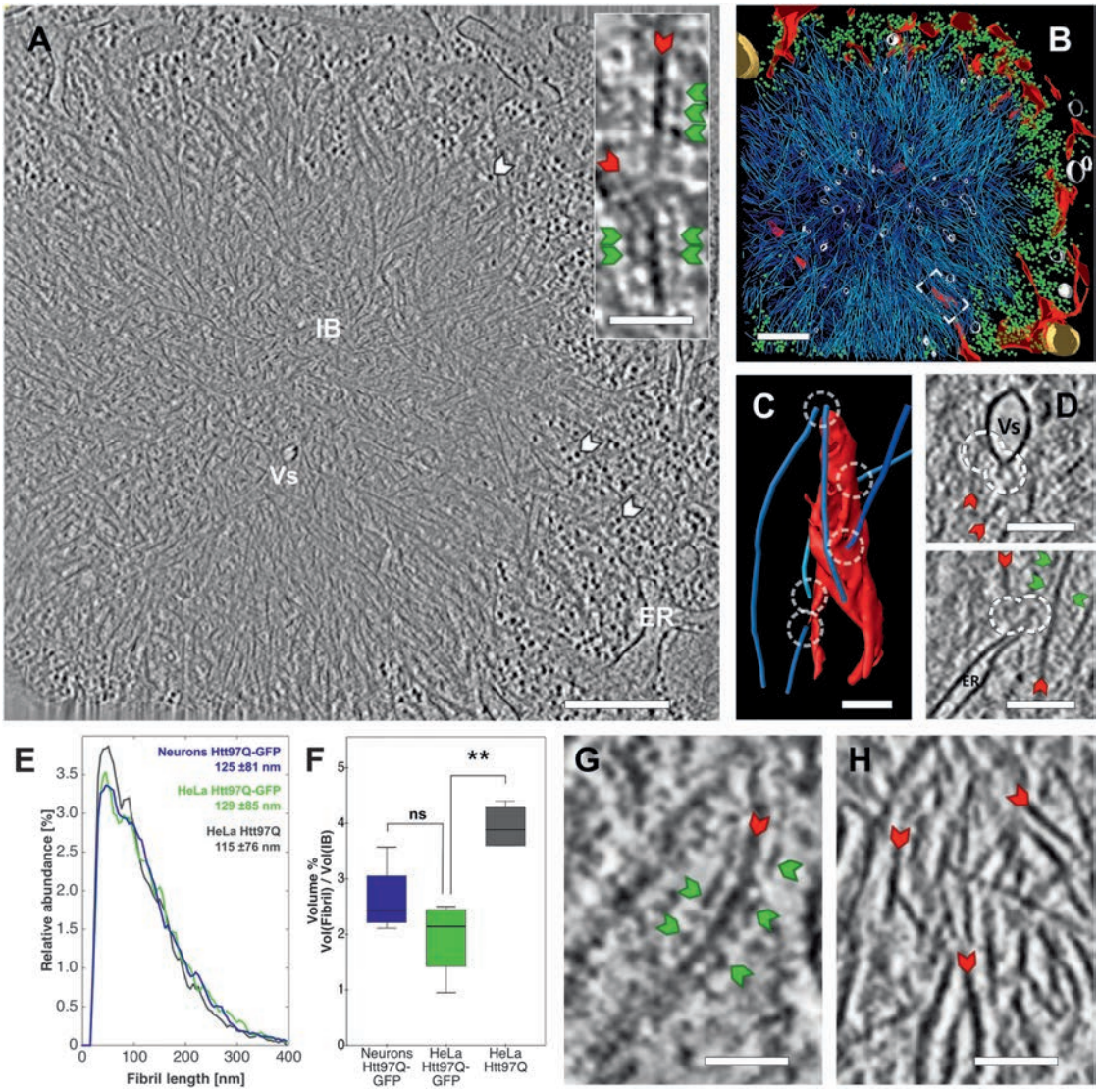


Figure 1. Cytosolic Inclusions of Htt97Q-GFP Contain Amyloid-like Fibrils

(A) Tomographic slice of an IB in an Htt97Q-GFP-transfected mouse primary neuron. The cytoplasmic electron dense particles represent ribosomes (white arrowheads). ER, endoplasmic reticulum; IB, Htt97Q-GFP inclusion body; Vs, vesicle. Inset: high-magnification image of Htt97Q-GFP fibrils (red arrowheads) decorated by globular densities (green arrowheads).

(B) 3D rendering of the tomogram shown in (A). ER membranes (red), Htt97Q-GFP fibrils (cyan), ribosomes (green), vesicles (white), and mitochondria (gold). Note that the core of the IB is largely devoid of ribosomes, which are abundant at the IB periphery. Scale bars, 400 nm in (A) and (B) and 30 nm in (A) (inset).

(C) Magnified rendering of the region marked in (B) showing interaction sites (white circles) between Htt97Q-GFP fibrils and the ER membrane. Scale bar, 50 nm.

(D) Magnified tomographic slices showing Htt97Q-GFP fibrils (red arrowheads) decorated by globular densities (green arrowheads) interacting with cellular membranes in Htt97Q-GFP-transfected neurons. Scale bars, 100 nm.

(E) Histograms of fibril length in mouse neurons expressing Htt97Q-GFP (blue), HeLa cells expressing Htt97Q-GFP (green), and HeLa cells expressing Htt97Q (gray) (number of fibrils: n = 11,481 neurons and Htt97Q-GFP; n = 7,648 HeLa and Htt97Q-GFP; n = 12,465 HeLa and Htt97Q; 4 tomograms for all conditions).

(F) Percentage of IB volume occupied by fibrils. The boxes and whiskers, respectively, indicate confidence intervals of 50% and 95% around the mean (solid line inside each box). ** indicates $p < 0.01$; ns, not significant in an ANOVA analysis with Bonferroni post hoc test.

(G) Magnified tomographic slice of an Htt97Q-GFP-transfected HeLa cell showing an Htt97Q-GFP fibril (red arrowhead) decorated by globular densities (green arrowheads).

(H) Magnified tomographic slice of an Htt97Q-transfected HeLa cell showing Htt97Q fibrils.

Scale bars, 50 nm in (G) and (H). Tomographic slices are 2.8-nm (A and D) or 1.7-nm (G and H) thick. The number of experiments and cells analyzed per condition is shown in Table S1.

See also Figures S1, S2, and S4 and Table S1.

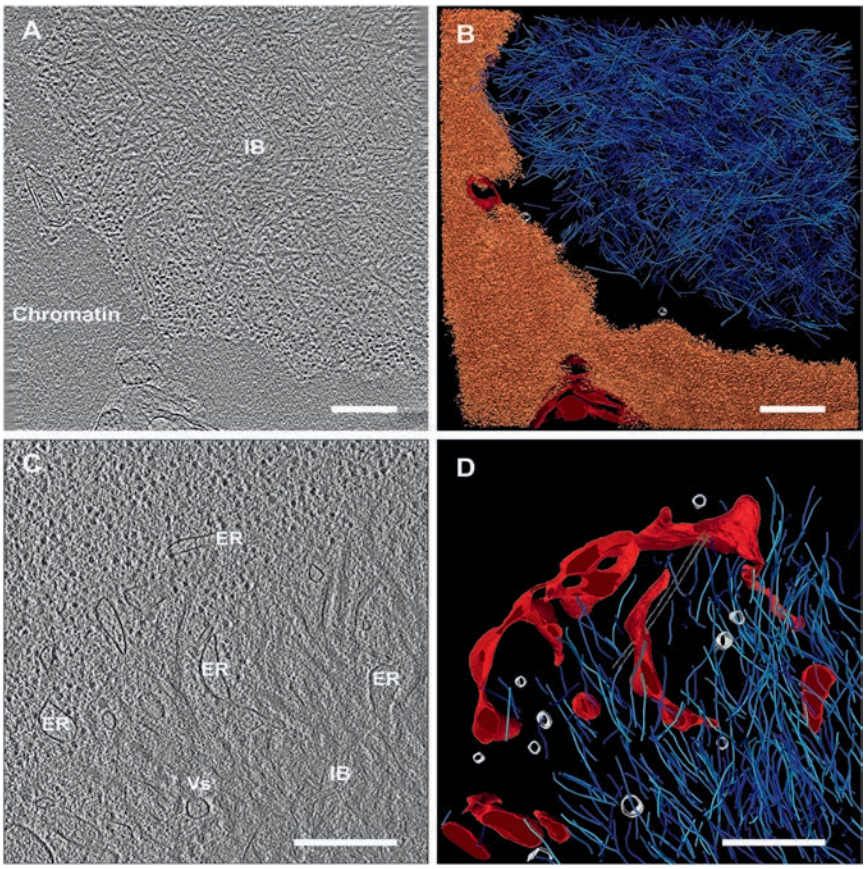


Figure 2. Nuclear Htt97Q-GFP Inclusions in Neurons and Cytosolic Htt64Q-GFP Inclusions in HeLa Cells

(A) 1.7-nm thick tomographic slice of a nuclear IB in an Htt97Q-GFP-transfected neuron. IB, Htt97Q-GFP inclusion body.

(B) 3D rendering of the tomogram shown in (A). Chromatin (orange), nuclear envelope membranes (red), vesicles in the periphery of the IB (white), and Htt97Q-GFP fibrils (cyan). Scale bars, 400 nm in (A) and (B).

(C) 2.5-nm-thick tomographic slice showing an IB in an Htt64Q-GFP-transfected HeLa cell. ER, endoplasmic reticulum; IB, Htt64Q-GFP inclusion body; Vs, vesicle.

(D) 3D rendering of the tomogram shown in (C). ER membranes (red), Htt64Q-GFP fibrils (cyan), and microtubule (gray parallel lines).

Scale bars, 250 nm in (C) and (D). The number of experiments and cells analyzed per condition is shown in Table S1.

See also Figures S2 and S4 and Table S1.

by cryo-ET originated from ruptured ER membranes. Occasionally, puncta positive for mitochondrial markers were also observed inside Htt97Q-GFP inclusions (Figure S3A), but this was not the case for other organelles, such as the Golgi apparatus (Figure S3C). Thus, our light and electron microscopy data indicate that the polyQ Htt exon 1 fibrils of inclusion bodies interact most extensively with ER membranes.

We further investigated the consequences of this membrane interaction by immunostaining for a variety of ER-resident proteins. Interestingly, some ER chaperones (e.g. calnexin) and components of the ER-associated degradation (ERAD) machinery (e.g., Erlin-2, Sel1L) were enriched in the ER domain surrounding the inclusions, whereas other ER proteins and ERAD components (e.g., BiP, calreticulin, PDI, or p97) did not show this pattern (Figure S3E; data not shown). In contrast, the translocase component Sec61 was largely excluded from the IB vicinity, consistent with the observation by cryo-ET that ER membranes in contact with the IB were essentially devoid of ribosomes (Figures 1C, 3B, 3C, and 3G). Thus, some, but not all, of the components of the so-called ''ER quality control compartment'' (Leitman et al., 2013a) accumulated around Htt97Q-GFP inclusions. Strikingly, live cell imaging revealed a complete ''freezing'' of ER dynamics in the vicinity of Htt97Q-GFP inclusions (Figures 4C–4F and S3F–S3H; Movie S2). This effect was not merely steric, as the ER was highly dynamic around other large cellular structures or the nucleus (Figures 4C and 4D).

Altogether, these data suggest that the interaction of Htt97Q-GFP fibrils with ER membranes alters cell physiology by locally impairing ER organization and dynamics.

DISCUSSION

While polyQ-expanded Htt exon 1 forms fibrils in vitro (Scherzinger et al., 1997), it has remained controversial whether Htt in cellular inclusions is present in granular or fibrillar form (Finkbeiner, 2011; Qin et al., 2004; Waelter et al., 2001). Importantly, fibril formation is thought to be required for neurodegeneration in HD mice (Gu et al., 2009). Our cryo-ET data conclusively show that in vitrified frozen hydrated murine neurons and human cells, inclusions of polyQ-expanded Htt exon 1 do consist of fibrils. The high resolution of our images allowed us to resolve additional densities decorating the fibrils formed by a GFP-labeled Htt97Q construct, and to quantify the changes in fibril density and rigidity induced by the GFP tag. Htt97Q fibrils were substantially thinner than those found in heavy-metal stained preparations (Qin et al., 2004), and sufficiently stiff to deform membranes (Roux, 2013). The structure of in situ fibrils is consistent with that proposed by recent NMR studies, in which the polyQ stretch forms the amyloid core and the flanking regions protrude outward in a bottlebrush fashion (Isas et al., 2015; Lin et al., 2017). However, our data do not allow the molecular organization of the fibril core to be resolved, possibly due to its structural heterogeneity (Hoop et al., 2016; Lin et al., 2017). A comparative cryo-ET analysis of endogenous polyQ inclusions in brain tissue is not yet technically feasible and remains a goal for future research.

Wild-type Htt is known to interact with cellular membranes (Kegel-Gleason, 2013), and fibrils of polyQ-expanded Htt exon 1 and other amyloids can cause membrane disruption in vitro (Milanesi et al., 2012; Pieri et al., 2012). Our results suggest that as a

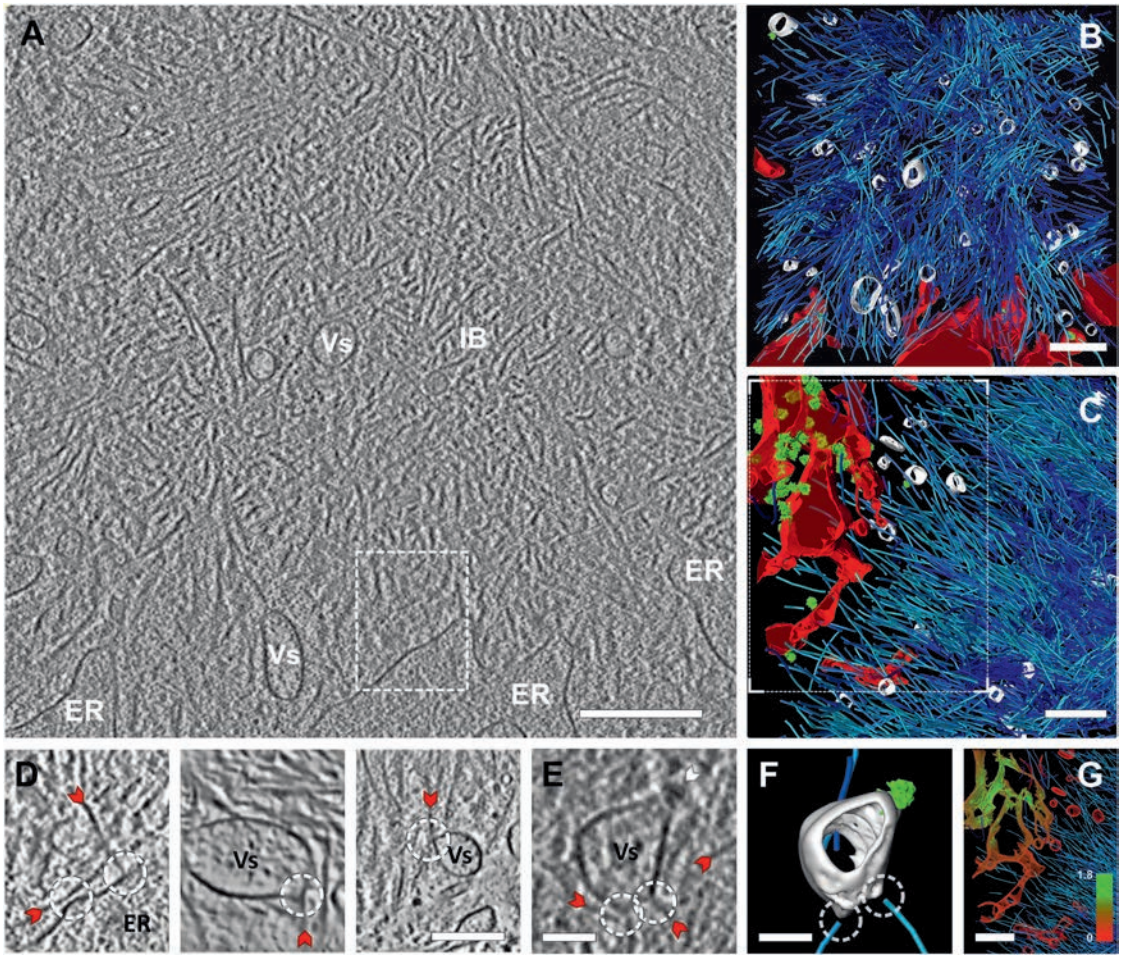


Figure 3. Htt97Q Inclusions Interact with Cellular Membranes

(A) Tomographic slice from the interaction zone between an IB and cellular membranes in an Htt97Q-transfected HeLa cell. ER, endoplasmic reticulum; IB, Htt97Q inclusion body; Vs, vesicles.

(B) 3D rendering of the tomogram shown in (A). ER membranes (red), ER-bound ribosomes (green), Htt97Q fibrils (cyan), and vesicles inside the IB (white).

(C) 3D rendering of ER membranes and ER-bound ribosomes in the vicinity of Htt97Q fibrils in a different cell. Note that ER-bound ribosomes are depleted from ER membranes directly interacting with Htt97Q fibrils but are abundant in more distal areas (see also G). Scale bars, 250 nm in (A)–(C).

(D) Magnified tomographic slices showing the sides and ends of Htt97Q (left and middle) and Htt97Q-GFP (right) fibrils (red arrowheads) interacting with cellular membranes (white circles). The left panel shows the area boxed in (A) in a different tomographic slice.

(E) Magnified tomographic slice of a putative membrane-bound ribosome (white arrowhead) on a small vesicle contacted (white circles) by the Htt97Q-GFP fibrils of an IB.

(F) 3D rendering of the vesicle shown in (E). Note the high curvature of the vesicle membrane at the sites of interaction with fibrils. Scale bars, 100 nm in (D)–(F).

(G) Visualization of the density of membrane-bound ribosomes in the area boxed in (C). Ribosome density is indicated by color and is lower (red) on ER membranes in direct contact with fibrils. Scale bar, 250 nm. Tomographic slices are 1.7-nm thick. The number of experiments and cells analyzed per condition is shown in Table S1.

See also Figures S2, S3, and S4, Table S1, and Movie S1.

cytopathological consequence of these phenomena, the fibrils of Htt exon 1 inclusions impinge on and possibly disrupt cellular endomembranes, most prominently those of the ER. As a result, inclusions drive a reorganization of the ER network in their periphery. Fibril-membrane interactions are thought to be largely lipid-mediated (Burke et al., 2013; Kegel et al., 2005; Trevino et al., 2012), and whether any specificity for ER membranes exists (Atwal et al., 2007) remains to be determined. Although it is possible that oligomers dissociating from fibril ends (Carulla et al., 2005; Martins et al., 2008) contribute to these effects, high concentrations of soluble Htt97Q-GFP did not disturb membrane morphology in the absence of inclusions. Thus, fibrils appear to be necessary for membrane deformation.

The region of the ER surrounding the inclusions showed dramatically reduced membrane dynamics, suggesting that a wide variety of cellular processes that depend on the dynamic behavior of the ER (Zhang and Hu, 2016) might be affected as a consequence. Moreover, translation is halted in these regions, as ER membranes contacting fibrils lacked ribosomes and the Sec61 translocon. The accumulation of ERAD factors and ER

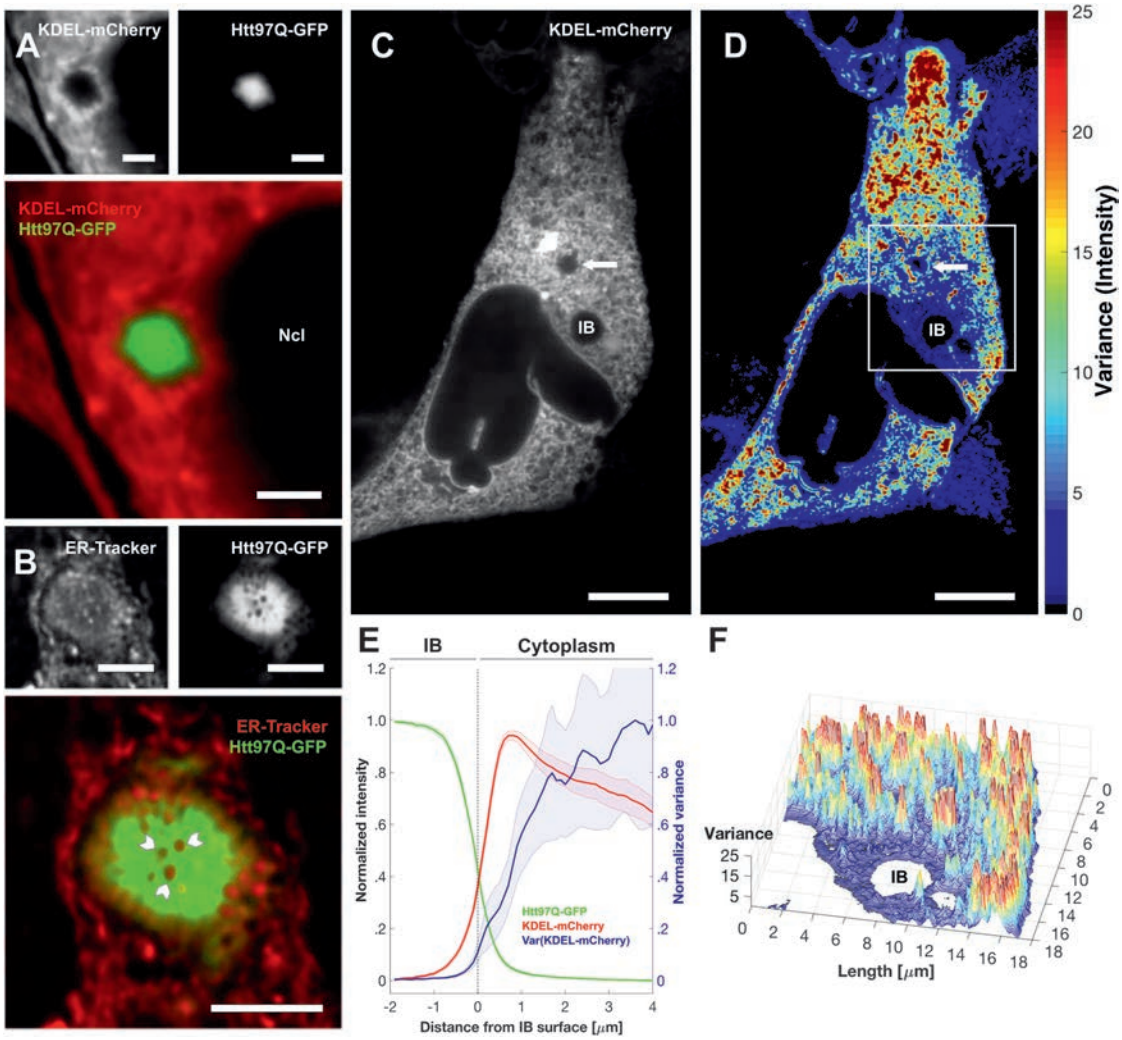


Figure 4. Htt97Q-GFP Inclusions Locally Alter ER Organization and Dynamics

(A) HeLa cell co-expressing Htt97Q-GFP (green) and the ER luminal marker KDEL-mCherry (red). Note the accumulation of ER around the IB. Ncl, nucleus.

(B) HeLa cell expressing Htt97Q-GFP (green) and labeled with ER-Tracker (red). White arrowheads mark ER-positive structures inside the IB. Scale bars, 5 μm in (A) and (B).

(C) Additional example of a HeLa cell co-expressing Htt97Q-GFP and KDEL-mCherry. A white arrow points to an Htt97Q-GFP-negative large cytoplasmic structure around which ER dynamics are normal. IB, Htt97Q-GFP inclusion body.

(D) ER dynamics of the cell shown in (C) quantified as the variance of KDEL-mCherry pixel intensity over time for 20 s. Scale bars, 10 μm in (C) and (D).

(E) Quantification of ER membrane dynamics in the vicinity of Htt97Q-GFP inclusions (n = 44 HeLa cells). Note the substantial accumulation of ER (red curve; radial average of KDEL-mCherry pixel intensity) at the IB periphery (x = 0). Membrane dynamics, assessed by the variance of KDEL-mCherry pixel intensity over time (blue curve), were markedly slower in this ER domain than in more distal regions. Radial averages (solid lines) and 95% confidence intervals (shaded areas) are shown. See Figures S3F–3H for individual traces.

(F) 3D representation of the boxed region in (D) around the IB.

See also Figure S3 and Movie S2.

chaperones suggests that cytosolic inclusions locally induce misfolding of ER luminal proteins causing ER stress. These results are consistent with previous findings that polyQ expression compromises ER function by perturbing ERAD, ER Ca^{2+} signaling and the ER redox state, leading to ER stress in cellular and animal models (Duennwald and Lindquist, 2008; Higo et al., 2010; Jiang et al., 2016; Kirstein et al., 2015; Kouroku et al., 2002; Lajoie and Snapp, 2011; Leitman et al., 2013b; Tang et al., 2003). ER stress markers are upregulated in Htt knockin mice and HD patients, and alleviating ER stress is beneficial in various HD models (Carnemolla et al., 2009; Lee et al., 2012; Leitman et al., 2014; Vidal et al., 2012). Thus, beyond the toxic role of oligomeric aggregate species, our data suggest that the formation of large fibrillar inclusions (Benn et al., 2005; Liu et al., 2015; Ramdzan et al., 2017; Woerner et al., 2016) also contributes to cytopathology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture
  - Hippocampal Neuron Culture and Transfection
- METHOD DETAILS
  - Plasmids and Antibodies
  - Immunofluorescence
  - Correlative Light Microscopy and Live Cell Imaging
  - Cell Vitrification
  - Cryo-focused Ion Beam Milling
  - Cryo-electron Tomography
  - Tomogram Reconstruction and Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Persistence Length Analysis
  - Ribosome Template Matching and Calculation of ER-bound Ribosome Density
  - Determination of ER Membrane Curvature
  - Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, one table, and two movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.009.

A video abstract is available at http://dx.doi.org/10.1016/j.cell.2017.08.009#mmc4.

AUTHOR CONTRIBUTIONS

F.J.B.B. performed all electron and light microscopy experiments, analyzed the data, and prepared the figures. I.S. and A.M. performed the cell culture experiments. M.K. and A.M.-S. contributed to the computational analysis of cryo-ET data. R.K., I.D., M.S.H., F.U.H., W.B., and R.F.-B. designed the research. I.D., M.S.H., and R.F.-B. supervised the experiments. M.S.H., F.U.H., W.B., and R.F.-B. wrote the manuscript with contributions from the other authors.

ACKNOWLEDGMENTS

This research has received funding from the European Commission (FP7 GA ERC-2012-SyG_318987–ToPAG) and the Munich Cluster for Systems Neurology (EXC 010 SyNergy). M.K. was supported by the Graduate School of Quantitative Biosciences Munich (GSC-1006). We thank Radostin Danev, Günter Pfeifer, and Jürgen Plitzko for electron microscopy support; Tom Gaitanos and Tim Laugks for assistance with light microscopy; Stefan Pfeffer for advice in data analysis; and Eri Sakata for discussions and critically reading the manuscript.

Received: March 31, 2017
Revised: June 30, 2017
Accepted: August 7, 2017
Published: September 7, 2017

REFERENCES

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Arrasate, M., and Finkbeiner, S. (2012). Protein aggregates in Huntington's disease. Exp. Neurol. *238*, 1–11.

Asano, S., Fukuda, Y., Beck, F., Aufderheide, A., Förster, F., Danev, R., and Baumeister, W. (2015). Proteasomes. A molecular census of 26S proteasomes in intact neurons. Science *347*, 439–442.

Atwal, R.S., Xia, J., Pinchev, D., Taylor, J., Epand, R.M., and Truant, R. (2007). Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. Hum. Mol. Genet. *16*, 2600–2615.

Bence, N.F., Sampat, R.M., and Kopito, R.R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. Science *292*, 1552–1555.

Benn, C.L., Landles, C., Li, H., Strand, A.D., Woodman, B., Sathasivam, K., Li, S.H., Ghazi-Noori, S., Hockly, E., Faruque, S.M., et al. (2005). Contribution of nuclear and extranuclear polyQ to neurological phenotypes in mouse models of Huntington's disease. Hum. Mol. Genet. *14*, 3065–3078.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature *416*, 507–511.

Bugg, C.W., Isas, J.M., Fischer, T., Patterson, P.H., and Langen, R. (2012). Structural features and domain organization of huntingtin fibrils. J. Biol. Chem. *287*, 31739–31746.

Burke, K.A., Hensal, K.M., Umbaugh, C.S., Chaibva, M., and Legleiter, J. (2013). Huntingtin disrupts lipid bilayers in a polyQ-length dependent manner. Biochim. Biophys. Acta *1828*, 1953–1961.

Carnemolla, A., Fossale, E., Agostoni, E., Michelazzi, S., Calligaris, R., De Maso, L., Del Sal, G., MacDonald, M.E., and Persichetti, F. (2009). Rrs1 is involved in endoplasmic reticulum stress response in Huntington disease. J. Biol. Chem. *284*, 18167–18173.

Carulla, N., Caddy, G.L., Hall, D.R., Zurdo, J., Gairí, M., Feliz, M., Giralt, E., Robinson, C.V., and Dobson, C.M. (2005). Molecular recycling within amyloid fibrils. Nature *436*, 554–558.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. Proc. Natl. Acad. Sci. USA *111*, 15635–15640.

Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L., and Bates, G.P. (1997). Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. Cell *90*, 537–548.

Deo, N. (2016). Graph Theory with Applications to Engineering and Computer Science (Dover Publications).

DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science *277*, 1990–1993.

Duennwald, M.L., and Lindquist, S. (2008). Impaired ERAD and ER stress are early and specific events in polyglutamine toxicity. Genes Dev. *22*, 3308–3319.

Fernandez, J.J., Laugks, U., Schaffer, M., Bäuerlein, F.J., Khoshouei, M., Baumeister, W., and Lucic, V. (2016). Removing contamination-induced reconstruction artifacts from cryo-electron tomograms. Biophys. J. *110*, 850–859.

Finkbeiner, S. (2011). Huntington's disease. Cold Spring Harb. Perspect. Biol. *3*, 3.

Fukuda, Y., Laugks, U., Lučić, V., Baumeister, W., and Danev, R. (2015). Electron cryotomography of vitrified cells with a Volta phase plate. J. Struct. Biol. *190*, 143–154.

Gu, X., Greiner, E.R., Mishra, R., Kodali, R., Osmand, A., Finkbeiner, S., Steffan, J.S., Thompson, L.M., Wetzel, R., and Yang, X.W. (2009). Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in HD mice. Neuron *64*, 828–840.

Gu, X., Cantle, J.P., Greiner, E.R., Lee, C.Y., Barth, A.M., Gao, F., Park, C.S., Zhang, Z., Sandoval-Miller, S., Zhang, R.L., et al. (2015). N17 Modifies mutant Huntingtin nuclear pathogenesis and severity of disease in HD BAC transgenic mice. Neuron *85*, 726–741.

Haass, C., and Selkoe, D.J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nat. Rev. Mol. Cell Biol. *8*, 101–112.

Higo, T., Hamada, K., Hisatsune, C., Nukina, N., Hashikawa, T., Hattori, M., Nakamura, T., and Mikoshiba, K. (2010). Mechanism of ER stress-induced brain damage by IP(3) receptor. Neuron *68*, 865–878.

Hipp, M.S., Patel, C.N., Bersuker, K., Riley, B.E., Kaiser, S.E., Shaler, T.A., Brandeis, M., and Kopito, R.R. (2012). Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease. J. Cell Biol. *196*, 573–587.

Hipp, M.S., Park, S.H., and Hartl, F.U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. Trends Cell Biol. *24*, 506–514.

Holmberg, C.I., Staniszewski, K.E., Mensah, K.N., Matouschek, A., and Morimoto, R.I. (2004). Inefficient degradation of truncated polyglutamine proteins by the proteasome. EMBO J. *23*, 4307–4318.

Hoop, C.L., Lin, H.K., Kar, K., Magyarfalvi, G., Lamley, J.M., Boatz, J.C., Mandal, A., Lewandowski, J.R., Wetzel, R., and van der Wel, P.C. (2016). Huntingtin exon 1 fibrils feature an interdigitated β-hairpin-based polyglutamine core. Proc. Natl. Acad. Sci. USA *113*, 1546–1551.

Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A.V., and Förster, F. (2012). PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. J. Struct. Biol. *178*, 177–188.

Isas, J.M., Langen, R., and Siemer, A.B. (2015). Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. Biochemistry *54*, 3942–3949.

Jiang, Y., Chadwick, S.R., and Lajoie, P. (2016). Endoplasmic reticulum stress: the cause and solution to Huntington's disease? Brain Res. *1648* (Pt B), 650–657.

Kawai-Noma, S., Pack, C.G., Kojidani, T., Asakawa, H., Hiraoka, Y., Kinjo, M., Haraguchi, T., Taguchi, H., and Hirata, A. (2010). In vivo evidence for the fibrillar structures of Sup35 prions in yeast cells. J. Cell Biol. *190*, 223–231.

Kegel, K.B., Sapp, E., Yoder, J., Cuiffo, B., Sobin, L., Kim, Y.J., Qin, Z.H., Hayden, M.R., Aronin, N., Scott, D.L., et al. (2005). Huntingtin associates with acidic phospholipids at the plasma membrane. J. Biol. Chem. *280*, 36464–36473.

Kegel-Gleason, K.B. (2013). Huntingtin interactions with membrane phospholipids: strategic targets for therapeutic intervention? J. Huntingtons Dis. *2*, 239–250.

Kirstein, J., Morito, D., Kakihana, T., Sugihara, M., Minnen, A., Hipp, M.S., Nussbaum-Krammer, C., Kasturi, P., Hartl, F.U., Nagata, K., and Morimoto, R.I. (2015). Proteotoxic stress and ageing triggers the loss of redox homeostasis across cellular compartments. EMBO J. *34*, 2334–2349.

Koenderink, J.J., and van Doorn, A.J. (1992). Surface shape and curvature scales. Image Vis. Comput. *10*, 557–564.

Kouroku, Y., Fujita, E., Jimbo, A., Kikuchi, T., Yamagata, T., Momoi, M.Y., Kominami, E., Kuida, K., Sakamaki, K., Yonehara, S., and Momoi, T. (2002). Polyglutamine aggregates stimulate ER stress signals and caspase-12 activation. Hum. Mol. Genet. *11*, 1505–1515.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. *116*, 71–76.

Lajoie, P., and Snapp, E.L. (2011). Changes in BiP availability reveal hypersensitivity to acute endoplasmic reticulum stress in cells expressing mutant huntingtin. J. Cell Sci. *124*, 3332–3343.

Lee, H., Noh, J.Y., Oh, Y., Kim, Y., Chang, J.W., Chung, C.W., Lee, S.T., Kim, M., Ryu, H., and Jung, Y.K. (2012). IRE1 plays an essential role in ER stress-mediated aggregation of mutant huntingtin via the inhibition of autophagy flux. Hum. Mol. Genet. *21*, 101–114.

Leitman, J., Ron, E., Ogen-Shtern, N., and Lederkremer, G.Z. (2013a). Compartmentalization of endoplasmic reticulum quality control and ER-associated degradation factors. DNA Cell Biol. *32*, 2–7.

Leitman, J., Ulrich Hartl, F., and Lederkremer, G.Z. (2013b). Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. Nat. Commun. *4*, 2753.

Leitman, J., Barak, B., Benyair, R., Shenkman, M., Ashery, U., Hartl, F.U., and Lederkremer, G.Z. (2014). ER stress-induced eIF2-alpha phosphorylation underlies sensitivity of striatal neurons to pathogenic huntingtin. PLoS ONE *9*, e90803.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Lin, H.K., Boatz, J.C., Krabbendam, I.E., Kodali, R., Hou, Z., Wetzel, R., Dolga, A.M., Poirier, M.A., and van der Wel, P.C.A. (2017). Fibril polymorphism affects immobilized non-amyloid flanking domains of huntingtin exon1 rather than its polyglutamine core. Nat. Commun. *8*, 15462.

Liu, K.Y., Shyu, Y.C., Barbaro, B.A., Lin, Y.T., Chern, Y., Thompson, L.M., James Shen, C.K., and Marsh, J.L. (2015). Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington's disease. Hum. Mol. Genet. *24*, 1602–1616.

Macdonald, M.; The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72*, 971–983.

Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Förster, F., Hyman, A.A., Plitzko, J.M., and Baumeister, W. (2016). Visualizing the molecular sociology at the HeLa cell nuclear periphery. Science *351*, 969–972.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S.W., and Bates, G.P. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell *87*, 493–506.

Martinez-Sanchez, A., Garcia, I., Asano, S., Lucic, V., and Fernandez, J.J. (2014). Robust membrane detection based on tensor voting for electron tomography. J. Struct. Biol. *186*, 49–61.

Martins, I.C., Kuperstein, I., Wilkinson, H., Maes, E., Vanbrabant, M., Jonckheere, W., Van Gelder, P., Hartmann, D., D'Hooge, R., De Strooper, B., et al. (2008). Lipids revert inert Abeta amyloid fibrils to neurotoxic protofibrils that affect learning in mice. EMBO J. *27*, 224–233.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51.

Milanesi, L., Sheynis, T., Xue, W.F., Orlova, E.V., Hellewell, A.L., Jelinek, R., Hewitt, E.W., Radford, S.E., and Saibil, H.R. (2012). Direct three-dimensional visualization of membrane disruption by amyloid fibrils. Proc. Natl. Acad. Sci. USA *109*, 20455–20460.

Nagashima, H., and Asakura, S. (1980). Dark-field light microscopic study of the flexibility of F-actin complexes. J. Mol. Biol. *136*, 169–182.

Nickell, S., Förster, F., Linaroudis, A., Net, W.D., Beck, F., Hegerl, R., Baumeister, W., and Plitzko, J.M. (2005). TOM software toolbox: acquisition and analysis for electron tomography. J. Struct. Biol. *149*, 227–234.

Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annu. Rev. Neurosci. *30*, 575–621.

Ossato, G., Digman, M.A., Aiken, C., Lukacsovich, T., Marsh, J.L., and Gratton, E. (2010). A two-step path to inclusion formation of huntingtin peptides revealed by number and brightness analysis. Biophys. J. *98*, 3078–3085.

Page, D.L., Sun, Y., Koschan, A.F., Paik, J., and Abidi, M.A. (2002). Normal vector voting: crease detection and curvature estimation on large, noisy meshes. Graph. Models *64*, 199–229.

Pfeffer, S., Brandt, F., Hrabe, T., Lang, S., Eibauer, M., Zimmermann, R., and Förster, F. (2012). Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. Structure *20*, 1508–1518.

Pieri, L., Madiona, K., Bousset, L., and Melki, R. (2012). Fibrillar α-synuclein and huntingtin exon 1 assemblies are toxic to the cells. Biophys. J. *102*, 2894–2905.

Qian, S.B., Ott, D.E., Schubert, U., Bennink, J.R., and Yewdell, J.W. (2002). Fusion proteins with COOH-terminal ubiquitin are stable and maintain dual functionality in vivo. J. Biol. Chem. *277*, 38818–38826.

Qin, Z.H., Wang, Y., Sapp, E., Cuiffo, B., Wanker, E., Hayden, M.R., Kegel, K.B., Aronin, N., and DiFiglia, M. (2004). Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction. J. Neurosci. *24*, 269–281.

Ramdzan, Y.M., Trubetskov, M.M., Ormsby, A.R., Newcombe, E.A., Sui, X., Tobin, M.J., Bongiovanni, M.N., Gras, S.L., Dewson, G., Miller, J.M.L., et al. (2017). Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis. Cell Rep. *19*, 919–927.

Rigort, A., Bäuerlein, F.J., Villa, E., Eibauer, M., Laugks, T., Baumeister, W., and Plitzko, J.M. (2012a). Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. Proc. Natl. Acad. Sci. USA *109*, 4449–4454.

Rigort, A., Günther, D., Hegerl, R., Baum, D., Weber, B., Prohaska, S., Medalia, O., Baumeister, W., and Hege, H.C. (2012b). Automated segmentation of electron tomograms for a quantitative description of actin filament networks. J. Struct. Biol. *177*, 135–144.

Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nat. Med. *10* (*Suppl*), S10–S17.

Roux, A. (2013). The physics of membrane tubes: soft templates for studying cellular membranes. Soft Matter *9*, 6726–6736.

Saibil, H.R., Seybert, A., Habermann, A., Winkler, J., Eltsov, M., Perkovic, M., Castaño-Diez, D., Scheffer, M.P., Haselmann, U., Chlanda, P., et al. (2012). Heritable yeast prions have a highly organized three-dimensional architecture with interfiber structures. Proc. Natl. Acad. Sci. USA *109*, 14906–14911.

Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M., and Hartl, F.U. (2004). Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. Mol. Cell *15*, 95–105.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H., and Wanker, E.E. (1997). Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell *90*, 549–558.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Tang, T.S., Tu, H., Chan, E.Y., Maximov, A., Wang, Z., Wellington, C.L., Hayden, M.R., and Bezprozvanny, I. (2003). Huntingtin and huntingtin-associated protein 1 influence neuronal calcium signaling mediated by inositol-(1,4,5) triphosphate receptor type 1. Neuron *39*, 227–239.

Trevino, R.S., Lauckner, J.E., Sourigues, Y., Pearce, M.M., Bousset, L., Melki, R., and Kopito, R.R. (2012). Fibrillar structure and charge determine the interaction of polyglutamine protein aggregates with the cell surface. J. Biol. Chem. *287*, 29722–29728.

Vidal, R.L., Figueroa, A., Court, F.A., Thielen, P., Molina, C., Wirth, C., Caballero, B., Kiffin, R., Segura-Aguilar, J., Cuervo, A.M., et al. (2012). Targeting the UPR transcription factor XBP1 protects against Huntington's disease through the regulation of FoxO1 and autophagy. Hum. Mol. Genet. *21*, 2245–2262.

Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H., and Wanker, E.E. (2001). Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. Mol. Biol. Cell *12*, 1393–1407.

Wegmann, S., Jung, Y.J., Chinnathambi, S., Mandelkow, E.M., Mandelkow, E., and Muller, D.J. (2010). Human Tau isoforms assemble into ribbon-like fibrils that display polymorphic structure and stability. J. Biol. Chem. *285*, 27302–27313.

Woerner, A.C., Frottin, F., Hornburg, D., Feng, L.R., Meissner, F., Patra, M., Tatzelt, J., Mann, M., Winklhofer, K.F., Hartl, F.U., and Hipp, M.S. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. Science *351*, 173–176.

Yamamoto, A., Lucas, J.J., and Hen, R. (2000). Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. Cell *101*, 57–66.

Zhang, H., and Hu, J. (2016). Shaping the endoplasmic reticulum into a social network. Trends Cell Biol. *26*, 934–943.

Cell

Article

Load Adaptation of Lamellipodial Actin Networks

Jan Mueller,[1] Gregory Szep,[1] Maria Nemethova,[1] Ingrid de Vries,[1] Arnon D. Lieber,[2] Christoph Winkler,[3,4] Karsten Kruse,[5] J. Victor Small,[6] Christian Schmeiser,[3,4] Kinneret Keren,[2,7] Robert Hauschild,[1] and Michael Sixt[1,8,*]

[1]Institute of Science and Technology Austria (IST Austria), am Campus 1, 3400 Klosterneuburg, Austria
[2]Department of Physics and Russell Berrie Nanotechnology Institute, Technion, Israel Institute of Technology, Haifa 32000, Israel
[3]RICAM, Austrian Academy of Sciences, Apostelgasse 23, 1030 Vienna, Austria
[4]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
[5]NCCR Chemical Biology, Departments of Biochemistry and Theoretical Physics, University of Geneva, 30, quai Ernest-Ansermet, 1211 Geneva, Switzerland
[6]Institute of Molecular Biotechnology GmbH (IMBA), Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[7]Network Biology Research Laboratories, Technion, Israel Institute of Technology, Haifa 32000, Israel
[8]Lead Contact
*Correspondence: sixt@ist.ac.at
http://dx.doi.org/10.1016/j.cell.2017.07.051

SUMMARY

Actin filaments polymerizing against membranes power endocytosis, vesicular traffic, and cell motility. *In vitro* reconstitution studies suggest that the structure and the dynamics of actin networks respond to mechanical forces. We demonstrate that lamellipodial actin of migrating cells responds to mechanical load when membrane tension is modulated. In a steady state, migrating cell filaments assume the canonical dendritic geometry, defined by Arp2/3-generated 70° branch points. Increased tension triggers a dense network with a broadened range of angles, whereas decreased tension causes a shift to a sparse configuration dominated by filaments growing perpendicularly to the plasma membrane. We show that these responses emerge from the geometry of branched actin: when load per filament decreases, elongation speed increases and perpendicular filaments gradually outcompete others because they polymerize the shortest distance to the membrane, where they are protected from capping. This network-intrinsic geometrical adaptation mechanism tunes protrusive force in response to mechanical load.

INTRODUCTION

From endocytic processes and vesicle trafficking to cellular locomotion and intracellular pathogen movement, the actomyosin cytoskeleton mediates most of the mechanical responses of eukaryotic cells (Pollard and Cooper, 2009). Actomyosin networks have two principal mechanisms of force generation: (1) polymerization can expand the network by elongating filaments against a load, and (2) myosin motors can contract the network or transport cargo along filaments. Every actin filament is born by a nucleation event. Nucleation is catalyzed by a number of molecular machines, the most prominent being formins and the Arp2/3 complex. While formins achieve *de novo* linear nucleation, the Arp2/3 complex binds to an existing filament and branches off a new filament at a 70° angle. Elongation, the addition of new monomers to the growing (barbed) end of the filament, is mainly mediated by formins or by proteins of the VASP family. Both nucleation and elongation are enhanced at the interface to membranes, where growing filaments are shielded from capping proteins, which otherwise terminate filament elongation by sealing the filament's barbed end (Carlsson, 2010; Pollard, 2007).

Arp2/3-dependent actin polymerization is a force-sensitive process. This has been demonstrated *in vitro*, in minimalist reconstituted systems, whereby increasing the mechanical load on a protruding reconstituted network led to higher network density (Bieling et al., 2016; Démoulin et al., 2014; Parekh et al., 2005). Thus, the network structure was able to adapt to the ambient mechanical conditions and to support force generation and mechanical resilience under varying loads. It has been shown that such adaptation is partially mediated by differential force sensitivities of nucleation, elongation, and capping, leading to enhanced branching under higher loads (Bieling et al., 2016). However, these kinetic effects were only partially able to explain the changes in network density, and it was speculated that the remaining adaptation might be mediated by spatial rearrangements of the network.

The mechanochemical response of protruding actin networks has not yet been studied in living cells. However, indirect evidence suggests that the leading edge of a motile cell is able to respond to load. When the lamellipodium, the flat protrusive front of a migrating cell, encounters an elastic obstacle (e.g., the vertical cantilever of an atomic force microscope), it was shown to substantially increase its pushing force, until collapsing when the load became excessive. For a migrating cell encountering a barrier, this means that its front pushes harder until the obstacle is pushed away or the leading edge stalls and the cell either retracts or circumnavigates the barrier (Heinemann et al., 2011; Prass et al., 2006).

In fibroblasts or epithelial cells, where lamellipodia typically undergo cycles of protrusion and retraction, it has been suggested that during each protrusive phase the polymerizing actin filaments experience a gradually growing load, because increase

in projected cell area is accompanied by increase in membrane tension (the cell stretches out the rather inextensible bag of plasma membrane) (Diz-Muñoz et al., 2013; Gauthier et al., 2011; Raucher and Sheetz, 2000; Sens and Plastino, 2015). This growing load on the pushing filaments is accompanied by accelerated incorporation of the actin nucleating Arp2/3 complex, which might, in turn, increase actin network density and boost force generation until the filaments stall and the leading edge retracts (Ji et al., 2008; Lee et al., 2015; Ryan et al., 2012). While these response patterns might result from complex mechanosensitive signaling processes, theoretical considerations put forward the alternative idea that branched actin networks respond to varying polymerization kinetics by changing their geometry (Maly and Borisy, 2001; Schaus et al., 2007; Weichsel and Schwarz, 2010). Such responses would argue that load adaptation is an emergent geometrical property of branched actin.

Here, we use the lamellipodium of fish keratocytes to quantitatively examine how protruding actin networks respond to varying forces. We use a combination of quantitative light and electron microscopy to describe geometrical changes of lamellipodial actin upon varying load regimes and employ stochastic modeling to elucidate how the network structure geometrically adapts to counter-forces.

RESULTS

Temporal Fluctuations in Actin Density Arise at the Leading Edge and Are Correlated with Projected Cell Area and Protrusion Speed

To quantitatively explore how lamellipodial actin reacts to changes in load we employed fish keratocytes migrating on planar surfaces (Mogilner and Keren, 2009). We established protocols to derive keratocytes from adult zebrafish, allowing access to stable transgenesis. On serum-coated coverslips, these cells migrate with exceptional speed, persistence, and morphological stability (Keren et al., 2008). We first chose a correlative approach and observed single migrating keratocytes expressing the actin reporter lifeact:GFP (Riedl et al., 2008) with high spatiotemporal resolution (Figure 1A; Movie S1). This allowed us to monitor fluctuations in actin network dynamics and density together with protrusion speed and morphological parameters (Figures 1A–1D, S1A, and S1B). The projected area of a keratocyte typically fluctuated ±5% around a baseline, which was consistent with earlier studies (Keren et al., 2008) (Figure S1A). These limited fluctuations supported the notion that the keratocyte's plasma membrane does not harbor major membrane reservoirs, which might be mobilized upon mechanical stretch (Lieber et al., 2013). When we measured lifeact:GFP intensity within a ~1-μm broad zone behind the leading front, intensity fluctuations were more substantial than area changes and varied ±20% (Figures 1D and S1A). Strikingly, the normalized, temporal fluctuations in projected cell area were tightly correlated with lifeact:GFP intensities as shown by cross-correlation peaks between 0.6–0.7 at time lag zero (Figures 1D, 1E, and S1B). This correlation was also observed in actin:GFP-expressing keratocytes, but not when the plasma membrane was uniformly labeled (Figure S2).

Correlation analysis of consecutive ~1-μm broad lamellipodial regions showed that the time delay between correlation peaks increased with the distance between the measured regions (Figures S1C–S1E). Dividing the respective lag times by the speed of each cell yields distances matching those between adjacent lamellipodial regions (Figure S1F). This analysis showed that, first, as previously demonstrated (Theriot and Mitchison, 1991; Wilson et al., 2010) keratocytes exhibited minimal retrograde flow of actin in relation to the substrate. Second, temporal fluctuations of actin density arise at the leading edge and propagate rearward. We next quantified protrusion speed by optical flow analysis at each pixel along the cell front. Protrusion speed was negatively correlated with both projected cell area and lifeact:GFP intensity with a negative cross-correlation peak between −0.3 and −0.4 at time lag zero (Figures 1B–1E). Lifeact:GFP intensities were equally co-fluctuating with area changes when corrected for the fluctuations in protrusion speed (Figure S1A).

Taken together, our quantitative analysis demonstrates that keratocytes migrate slower and produce denser actin networks during intervals when their projected area is larger.

Experimental Manipulations of Membrane Tension Reveal a Lamellipodial Response to Altered Load

Among several interpretations of these correlative data we considered that a fluctuation in projected cell area might correspond to changes in membrane tension (Raucher and Sheetz, 2000). To dissect the causal relationship between actin network density and membrane tension, we experimentally increased membrane tension by aspirating the trailing edge of migrating keratocytes with a micropipette (Houk et al., 2012) (Figure 2A; Movie S2). As membrane tension equilibrates almost instantaneously over the whole cell, this manipulation allowed us to tune lateral tension at the lamellipodial tip (Diz-Muñoz et al., 2013). Experiments were performed by transiently applying four different vacuum levels ranging from −10 to −40 mbar. In all cases, we found that aspiration caused an increase in lifeact:GFP signal concomitant with a moderate decrease in protrusion speed (Figures 2A–2C). Actin density and protrusion speed were dependent on the applied vacuum (Figures 2D and 2E). Following transient aspiration, the band of denser actin network traveled backward (in the cell frame of reference) with the actin flow (Figures 2A and 2C). Under the given parameters the cells kept protruding during aspiration, demonstrating that the increase in tension was below the stall force of the lamellipodium. These results suggested that lamellipodial actin responds to an increase in membrane tension by increasing network density.

To experimentally decrease membrane tension we took advantage of the fact that on adhesive substrates cells occasionally form tethered trailing edges, which leads to a stretched morphology (Figure 2F). When these tethers spontaneously detach or when we cut them with a pulsed laser, tension is released and the tether snaps forward. Here, the projected cell area shrunk within a few seconds when the cell went from a stretched to a more compact configuration. Such high recoil velocities indicate the rapid drop in tension, which had previously built up in the stretched cell. To directly test if shrinkage is accompanied by changes in membrane tension, membrane

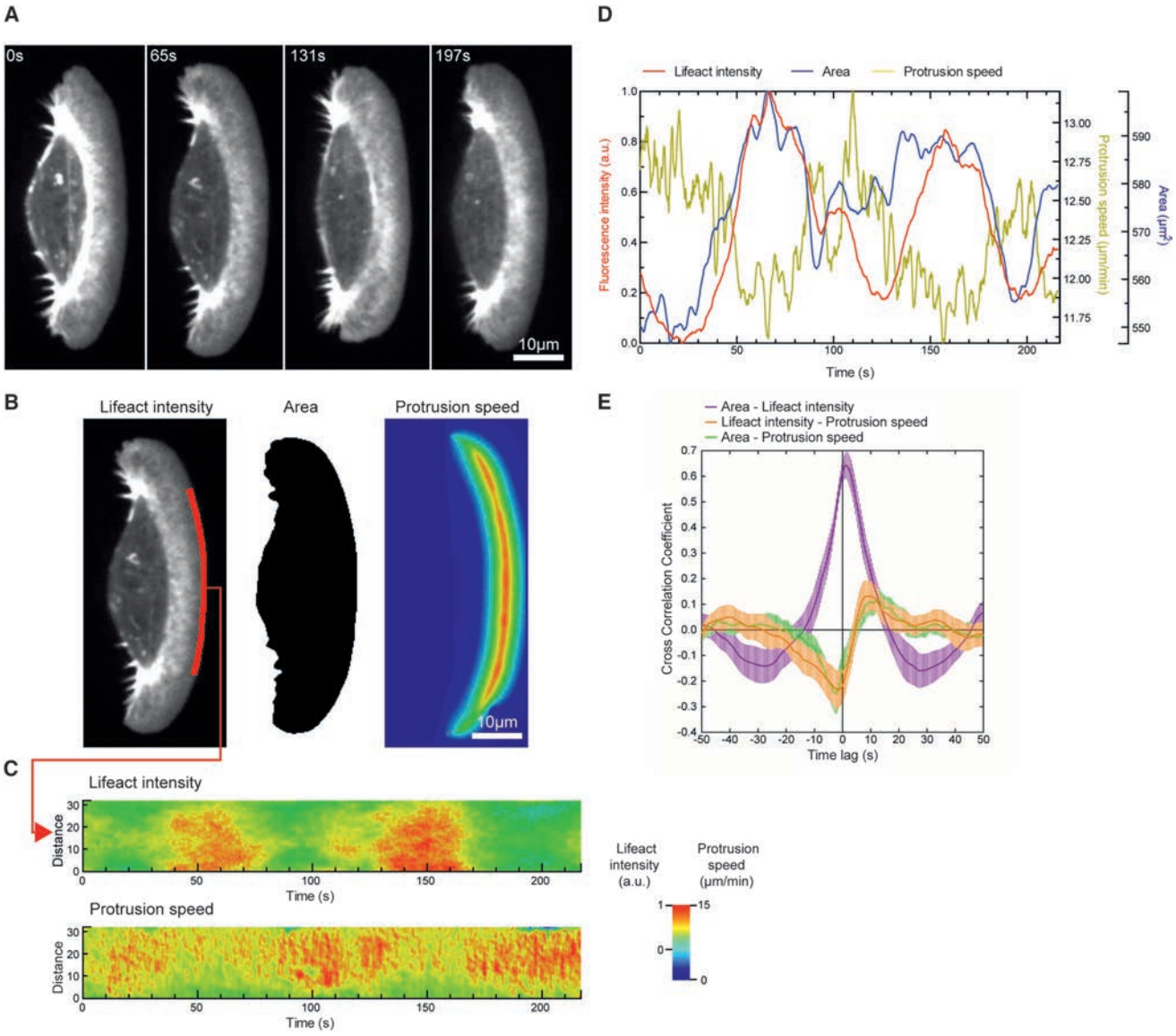


Figure 1. Correlative Analysis of Actin, Projected Cell Area, and Protrusions in Migrating Zebrafish Keratocytes

(A) Confocal imaging of lifeact:GFP-expressing keratocyte moving on a glass coverslip.

(B) Time frame from (A) showing the 1.09-μm wide area of lifeact:GFP intensity measurements (left), a binary mask used for quantifying the area (middle), and a pseudo-colored Horn-Schunck optical flow analysis image of the leading edge (right).

(C) Fluorescence intensity and leading edge velocity maps for the time lapse of the cell shown in (A).

(D) Temporal fluctuations of cell area, lifeact:GFP intensity, and protrusion speed averaged across the analyzed region shown in red in (B).

(E) Average of temporal cross-correlation functions of 21 migrating keratocytes. Mean and SEM are shown.

See also Figures S1 and S2 and Movie S1.

tethers were pulled from the plasma membrane of migrating keratocytes using an optical trap (Lieber et al., 2013) (Figure S3A). While in steady-state migrating cells there was no consistent relation between projected area and tether force, abrupt retraction events were accompanied by area shrinkage concurrent with a drop in tether force and thus membrane tension (Figure S3B). During abrupt retraction, the previously observed correlation between projected cell area and actin intensity at the leading edge manifested in a sudden drop in lamellipodial lifeact:GFP intensity at the moment of area shrinkage (Figures 2F and 2G). The resultant abrupt change in actin density was maintained while it steadily traveled backward (in the cell frame of reference) with the actin flow, further substantiating that altered polymerization originated at the leading edge of the cell (Figures 2F and 2H; Movie S2). Cell shrinkage was also accompanied by a rapid but very transient increase in protrusion speed (Figure 2G). These results suggested that lamellipodial actin responds to a decrease in membrane tension by decreasing network density.

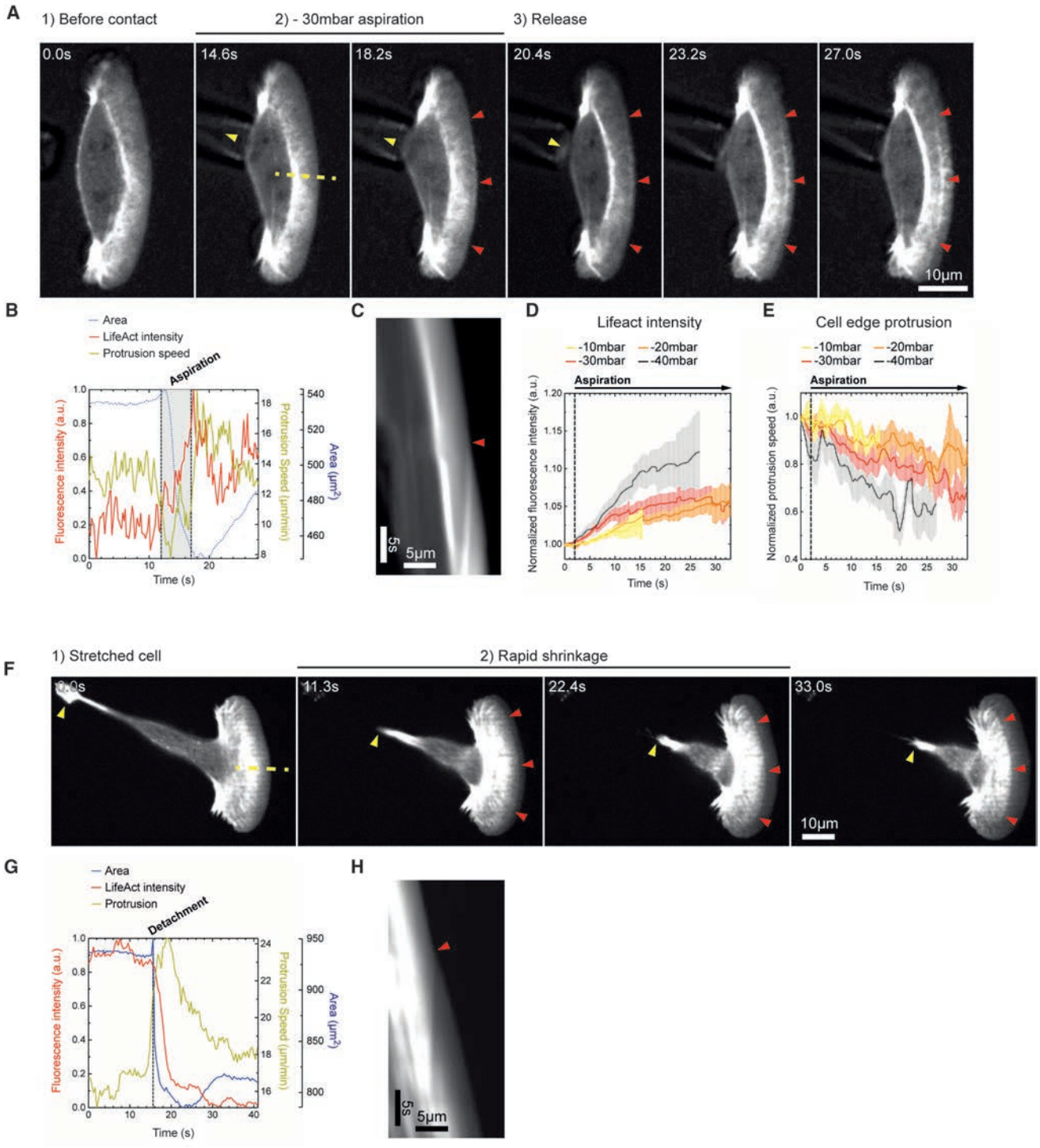


Figure 2. Cytoskeletal Response to Manipulations of Membrane Tension

(A) Lifeact:GFP imaging following micropipette aspiration. (1) Migrating keratocyte before contact with the micromanipulator, (2) aspiration of membrane, and (3) release of vacuum.

(B) Temporal fluctuations of lifeact:GFP intensity at the leading edge defined as in Figure 1B, area, and protrusion speed. The area between the dotted lines shows where the cell was aspirated.

(C) Kymograph along dashed yellow line in (A).

(D) Change in lifeact:GFP signal following aspiration with four different vacuum levels.

(E) Change in cell edge protrusion for the same cells as in (D). Mean and SEM are shown for 28 aspiration events in 13 individual cells in (D) and (E).

(legend continued on next page)

As our mechanical manipulations of membrane tension might have caused additional changes in the cells, we chose an independent approach to challenge our hypothesis. Migrating cells were exposed to cycles of altered osmolarity (Diz-Muñoz et al., 2013), by supplementing the medium with sucrose versus pure water. Monitoring lifeact:GFP at the leading edge showed that increased osmolarity (water efflux–cell shrinkage) was accompanied by a decrease in lifeact:GFP signal and this was reversed when water was added (water influx–cell swelling) (Figures S3C and S3D).

Together, our data show that lamellipodial actin responds to changes in membrane tension by altering its density: density increases following an increase in tension and decreases following a decrease in tension.

Quantitative Analysis of Lamellipodial Ultrastructure during Steady-State Migration

To gain quantitative insight into lamellipodial architecture at the single filament level we employed three-dimensional (3D) electron microscopy of fixed and negatively stained keratocytes (Vinzenz et al., 2012). This approach allows tracing single filaments in 3D space (Winkler et al., 2012). In keratocyte lamellipodia, filaments are consistently oriented with their barbed ends toward the front (Narita et al., 2012; Sivaramakrishnan and Spudich, 2009; Svitkina et al., 1997), which is in line with the polar forward growth of actin at the leading cell front (Figures S1C–S1F) (Keren, 2011; Lai et al., 2008; Theriot and Mitchison, 1991). In the digitalized tomograms we defined a barbed end as the end of a filament proximal to the leading edge and a pointed end as the one distal to the leading edge. Any barbed end could either represent an actively growing or a capped filament. A pointed end represents (1) a nucleation event generated by Arp2/3 (a branch), (2) a nucleation event generated by an alternative nucleation factor, or (3) a debranched or severed filament (Ydenberg et al., 2013, 2011). Throughout the lamellipodium of steady-state migrating keratocytes we found that the numbers of barbed and pointed ends were balanced (Figures 3A–3D and S4A). Only at the leading front, barbed ends occurred in excess, which is expected due to ongoing polymerization at this site. Next, we analyzed the angles, which the filaments assumed relative to the leading membrane and found that the distribution of their orientations peaked at ±35° (with 0° defined as perpendicular to the leading membrane) (Figure S4B). This configuration predominated in keratocytes as shown by a global order parameter defined as ((Filaments 0°–20°)−(Filaments 30°–50°))/((Filaments 0°–20°)+(Filaments 30°–50°)) as a readout for network orientation (Weichsel and Schwarz, 2010) (Figures 3E and 3F).

Together, our tomography data provide a quantitative description of lamellipodial actin as derived from filament configuration. We conclude that in a steady-state migrating keratocyte, nucleation, capping and elongation create an expanding network with a canonical branch geometry prescribed by the 70° branch structure of the Arp2/3 complex (Svitkina et al., 1997).

Network Configuration following Load Increase

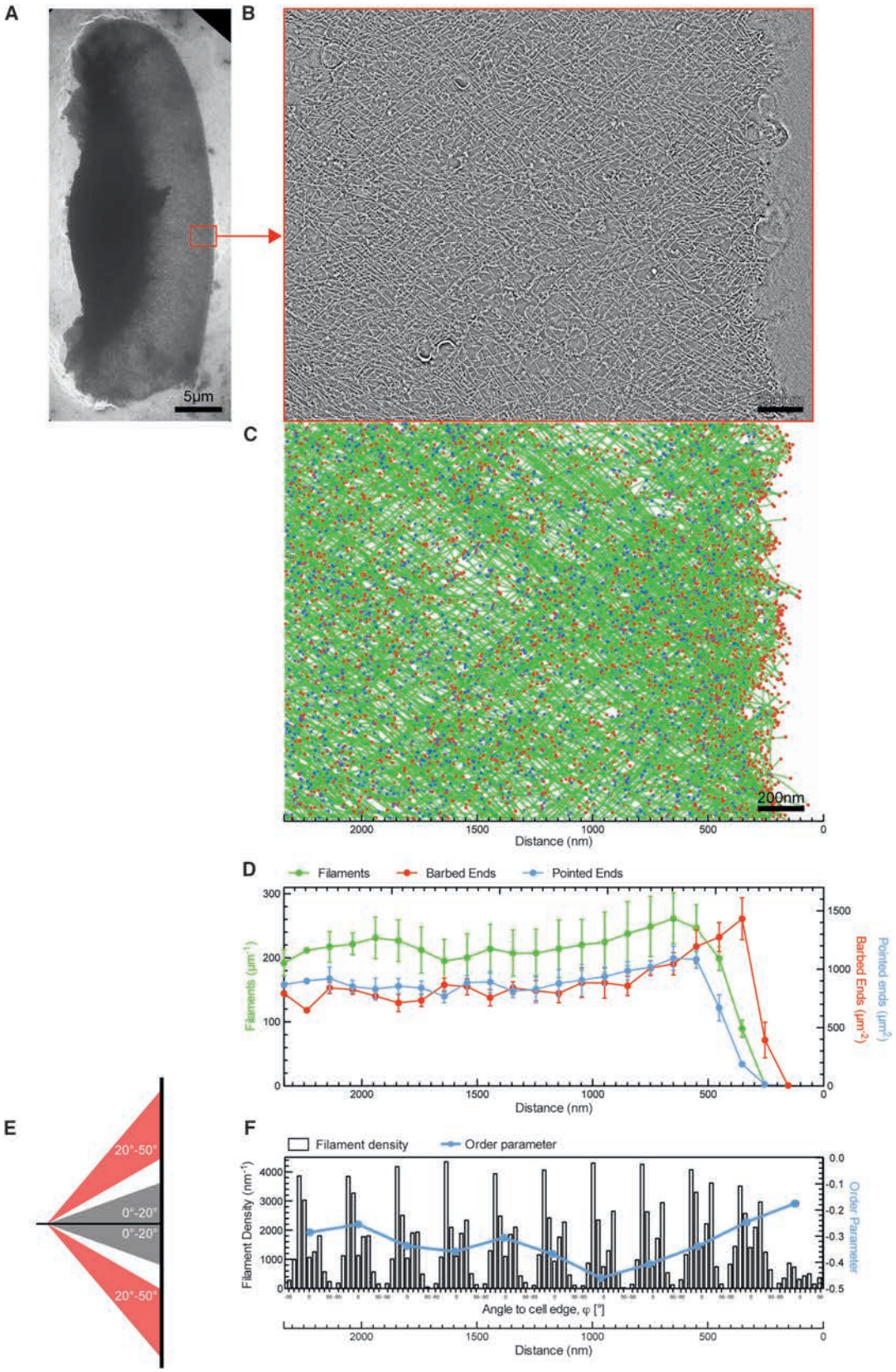
We next used trailing edge aspiration to interrogate the ultrastructural changes in lamellipodial networks polymerizing against increased load. To this end we employed correlated fluorescence live-cell imaging and electron tomography on lifeact:GFP-expressing keratocytes migrating on electron microscopy grids. After 4 s of −30mbar aspiration, cells were released, fixed and prepared for electron microscopy (Figure 4A; Movie S3). Under the given conditions where density fluctuations arise from the leading edge, the temporal growth-history of lamellipodial actin is spatially encoded in the network: the pre-aspiration zone (proximal to the nucleus) represents steady-state growth, the aspiration zone the increase in load, and the post-aspiration zone at the leading edge the phase at which the load drops again due to the release of the vacuum.

In the low-magnification electron micrograph, the transient increase in density during aspiration was clearly visible as a darker zone running parallel to the leading edge (Figure 4A). This was consistent with the lifeact:GFP live cell imaging (Figures 2A–2C). Electron tomography with automated tracking of actin filaments allowed us to trace the temporal evolution of the load response at the single filament level. The pre-aspiration zone resembled a steady-state lamellipodium with a canonical 70° branched network. Within an 800nm broad aspiration zone (corresponding to a 4-s aspiration time) filament density was markedly increased. At the beginning of the aspiration zone (distal from the leading front, hence earlier in time) a transient drop in the number of barbed ends and a gradual increase in pointed ends created a mismatch between filament birth and death, along with the increase in filament density (Figures 4C and 4D). This increase in filament density was also accompanied by an increase in the 0°–20° and 50°–70° fraction of filament angles, whereas filaments at intermediate angles of 20°–50° increased only moderately (Figures 4E and S4C). At the leading front, filament density decreased to reach values comparable to levels before aspiration (Figure 4D) and filament angles reverted to the canonical branch-pattern (Figures 4E, S4C, and S4D). These data show that an increase in membrane tension causes an increase in network density and a change in geometry, with filaments growing at steeper angles toward the plasma membrane at higher membrane tensions.

Network Configuration following Load Decrease

We followed the same correlative visualization strategy to analyze the network following a decrease in load. Similar to the laser cutting used in Figures 2F–2H rapid cell shrinkage could also be triggered by mechanically detaching one side of a cell with a micropipette (Figure 5A; Movie S4). This maneuver allowed sufficient experimental control to fix the cell seconds after

(F) Lifeact:GFP signal following fast cell shrinkage. (1) Keratocyte exhibiting stretched morphology, while its trailing edge is tethered to the substrate. (2) The trailing edge is cut with a pulsed laser (yellow arrowhead).
(G) Temporal fluctuations of leading edge lifeact:GFP intensity, area, and protrusion speed. The rapid retraction is marked with a dotted line.
(H) Kymograph along dashed yellow line in (F).
See also Figure S3 and Movie S2.



(legend on next page)

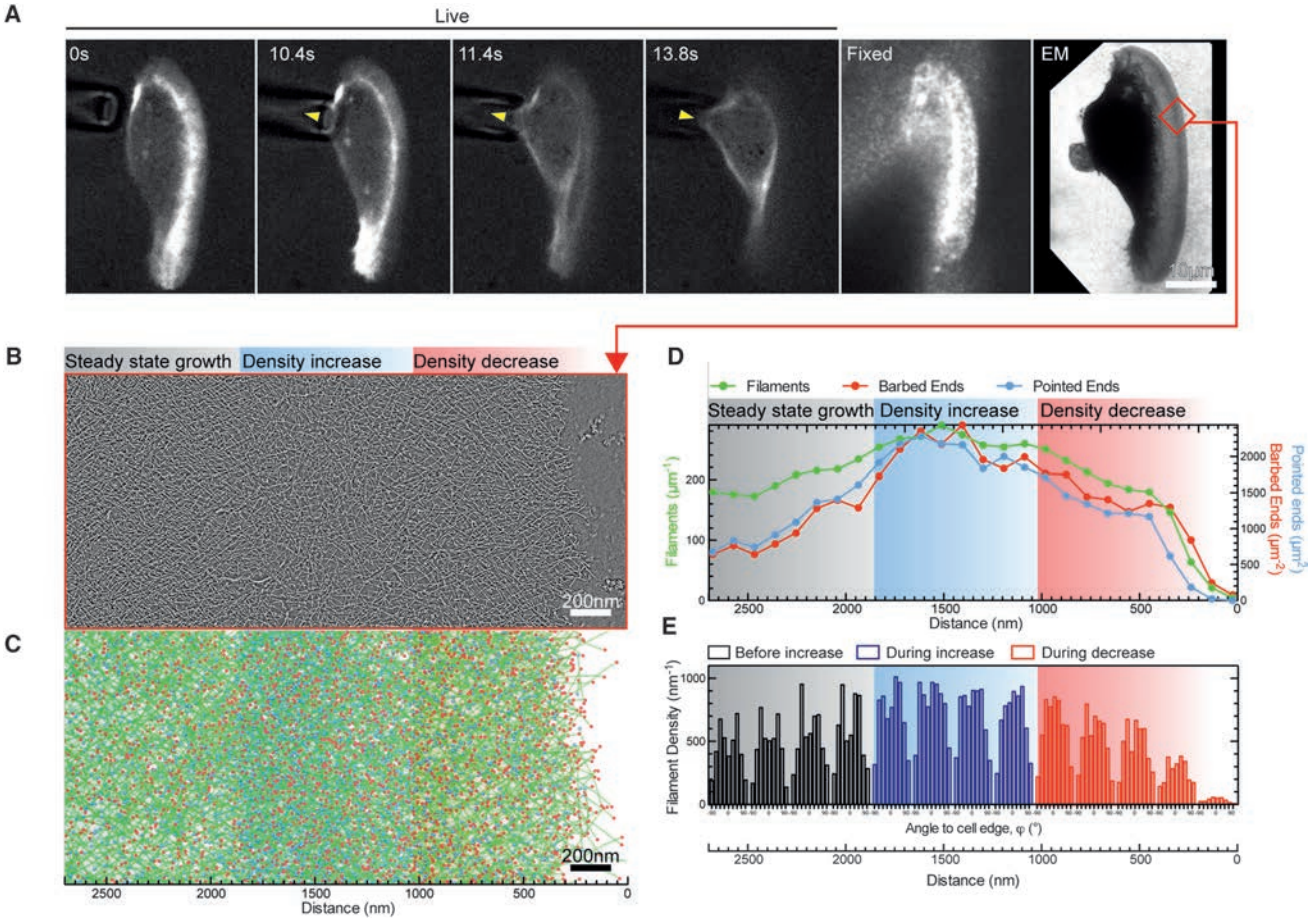


Figure 4. Correlated Live Microscopy-Electron Tomography of Ultrastructural Changes in Networks with Increasing Filament Density

(A) A migrating keratocyte expressing lifeact:GFP was aspirated at the rear with a micropipette, fixed within ≈3 s, and prepared for electron microscopy. The cell shifted out of focus because of the manipulating procedure. A transient increase in actin density can be seen in the low-magnification electron micrograph on the right.

(B) 5.5-nm tomogram slice of the region marked by a red box in (A). The rear of the cell is toward the left side of the picture, and the cell front is seen on the right. Regions of steady-state density, increased density, and decreased density are marked with black, blue, and red throughout the whole figure.

(C) Filament tracks of the lamellipodium shown in (B), with actin filaments shown in green, barbed ends in red, and pointed ends in blue.

(D) Filament numbers and densities of barbed and pointed ends in 106-nm-wide spatial bins throughout the lamellipodium shown in (B) and (C).

(E) Histogram showing filament angles to the cell edge in 212-nm distance bins.

See also Figures S4C and S4D and Movie S3.

shrinkage. Analogous to the increase in actin density following aspiration, the decrease in actin density in the time window of rapid shrinkage was clearly visible as a transition zone in lifeact:GFP signal as well as phalloidin staining (Figure 5A). Both signals showed a sharp decrease indicating the drop in filament density. This was substantiated by low-magnification views of the same cell after extraction and negative stain, which showed a sharp decrease in electron density, overlapping with the fluorescent imaging (Figures 5A and S6C). Quantification of manually as well as automatically traced filaments together with automated image segmentation showed a drop in filament numbers in the transition zone (Figures 5C, 5D, S5, and S6). As expected,

Figure 3. Electron Tomography of Migrating Wild-Type Keratocytes

(A) Overview electron micrograph of migrating keratocyte with acquired tomogram montage marked in red.

(B) 5.5-nm slice of a negatively stained tomogram of the actin network behind the leading edge.

(C) Automated tracking results of the same region with filaments shown in green, barbed ends in red, and pointed ends in blue.

(D) Normalized densities of filaments, barbed ends, and pointed ends in 106-nm-wide bins of four averaged tomogram montages. Graph shows mean and SEM.

(E) Scheme showing the filament angle bins used for calculating the global order parameter.

(F) Histogram of combined filament length growing at indicated angle toward the cell membrane (black) is shown together with a global order parameter (blue) in 212-nm distance bins defined as ((Filaments 0°–20°)−(Filaments 30°–50°))/((Filaments 0°–20°)+(Filaments 30°–50°)).

See also Figures S4A and S4B.

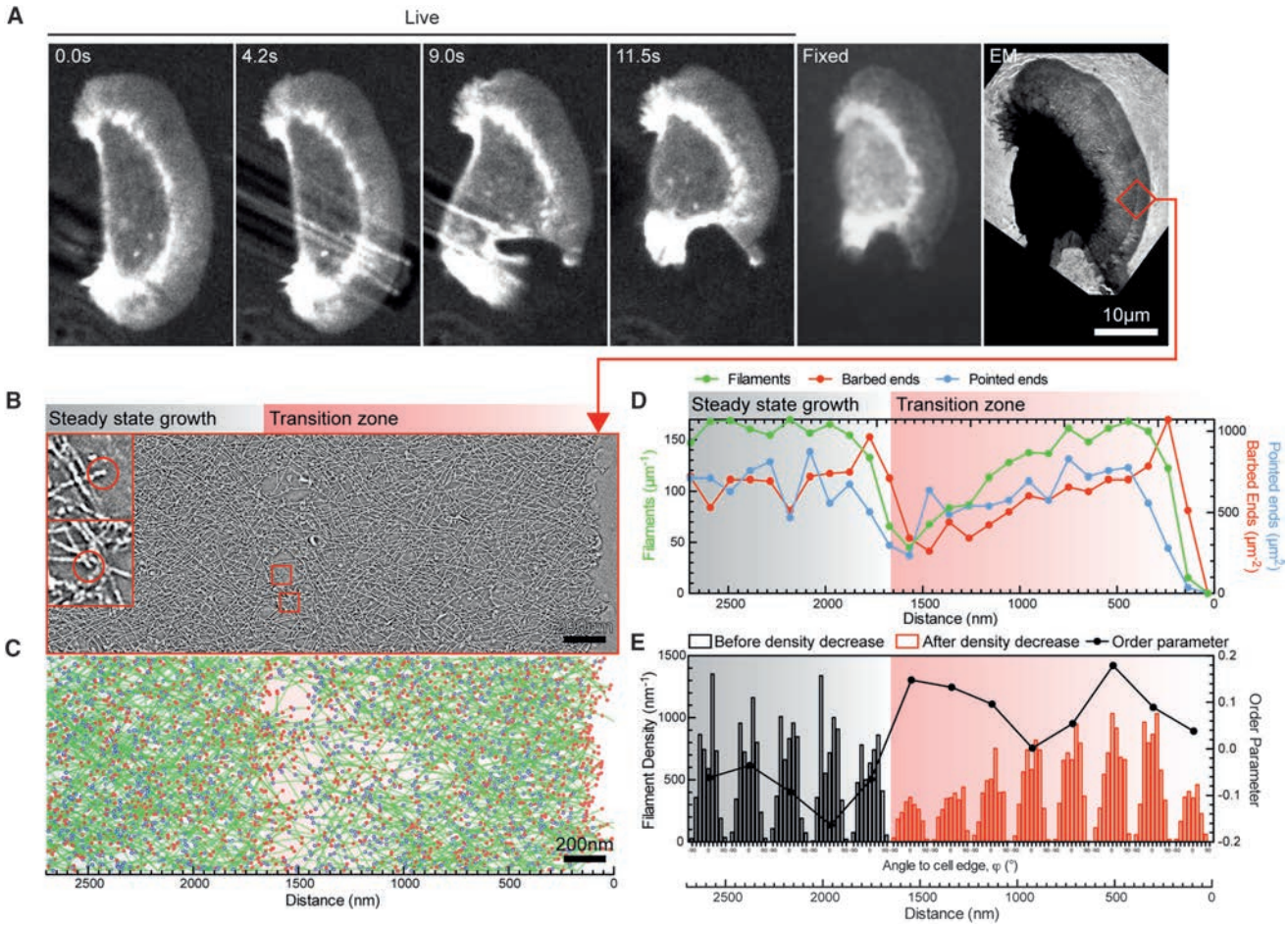


Figure 5. Correlated Live Microscopy-Electron Tomography of Ultrastructural Changes in Networks with Decreasing Filament Density

(A) Migrating keratocyte manipulated with a microneedle to induce a rapid decrease in projected cell area with an accompanying decrease in lifeact:GFP signal, fixed within ≈3 s, and prepared for electron microscopy. A rapid decrease in the lifeact:GFP signal is preserved in the fixed lifeact:GFP sample and the low-magnification electron micrograph.

(B) 5.5-nm tomogram slice showing the region marked with a red box in (A). The cell edge is seen on the right side, and the region of lower density is distinguishable toward the middle of the micrograph. Region of steady-state network density and decreased density are marked in black and red throughout the whole figure.

(C) Filament tracks of the lamellipodium shown in (B), with actin filaments shown in green, barbed ends in red, and pointed ends in blue.

(D) Filament numbers and densities of barbed and pointed ends in 106-nm-wide spatial bins throughout the lamellipodium shown in (B) and (C).

(E) Histogram showing filament densities growing at the indicated angle from the membrane in 212-nm distance bins. An order parameter (blue) in 212 nm is defined like in Figure 3.

See also Figures S5 and S6 and Movie S4.

the pre-transition zone (prior to mechanical manipulation) showed the quantitative signature of the steady-state lamellipodium. The onset of the transition zone was marked by a transient increase of barbed ends, which preceded a drop in pointed ends (Figure 5D). Here, the resultant gap between putative capping and nucleation events explained the concomitant drop in filament density. The transition zone was followed by a leveling of barbed and pointed ends and a continuous recovery of the network toward the cell front.

Notably, the transition zone was characterized by radically changed network geometry. When we quantified and parameterized filament angles depending on their position in the network we found that at the onset of cell shrinkage the canonical ±35° dominated network abruptly changed into a configuration dominated by filaments growing perpendicularly (0°) to the membrane. Toward the cell front the network again recovered gradually to canonical (Figures 5E and S6D). The change in network geometry was due to selective elimination of filaments: in the transition zone the rate of filament survival strictly depended on filament orientation: filaments at a higher angle were preferentially eliminated, whereas low angle filaments had a higher rate of survival (Figures S5E and S6D). These data show that a decrease in membrane tension causes a decrease in network density and a change in geometry, with more filaments growing perpendicularly to the plasma membrane.

A Stochastic 2D Model of Actin-Based Protrusion Recapitulates Light and Electron Microscopy Data

To gain a quantitative understanding of the experimental data we built on earlier work by Maly and Borisy (2001) and Weichsel and Schwarz (2010) and developed a stochastic model of lamellipodial network growth. We formulated a two-dimensional (2D) model taking into account filament elongation and Arp2/3-mediated (70°) nucleation at the leading front. Elongation is terminated by capping whenever filaments detach from the membrane and thereby leave the zone in which they are protected from capping (by elongation factors like VASP or formins). Filaments push the membrane forward according to a force-velocity relation based on thermal fluctuations as described in Dickinson (2009) and Mogilner and Oster (1996, 2003) (for details see STAR Methods and Tables S1 and S2). Here, network geometry dictates that increased protrusion velocity (as caused by decreased load) rapidly decreases filament density in an angle-dependent manner: A filament growing at an angle φ has to polymerize with $(1/cos\varphi)$ the speed of the advancing membrane in order to keep contact with the membrane. Hence, with increasing protrusion speed filaments growing at higher angles will lag behind the leading edge and therefore detach from the membrane where they are protected from capping (Figures 6A, 6B, S7D, and S7E). Consequently, higher angle filaments will be capped at higher rates and the family of low angle filaments (and their accompanying ±70° side branches) will outcompete the ±35° network. Slowing down the membrane has the opposite effect: More filaments at higher angles are able to catch up with the membrane, leading to a denser network consisting of filaments growing at a broad range of angles. The stochastic simulation is able to reproduce the correlative electron tomography findings of filament orientation and density by using a sharp decrease from 300 pN as force input. 300 pN were taken as a realistic force reported for steady-state migrating fish keratocytes (Lieber et al., 2013). Together, our stochastic simulation faithfully reproduced both kinetics and geometry of network parameters upon externally induced changes in load (Figures 6C, 6D, S7A, and S7B; Tables S1 and S2; Movie S5).

Next we wanted to test if the stochastic simulation can produce the co-fluctuations of protrusion speed, actin density and projected area as we observed them in our light microscopy experiments in unperturbed migrating keratocytes (Figure 1). The amount of decrease in tension necessary to obtain the correct response at the filament level in the correlated experiments was used as a scaling factor to relate membrane tension to projected area. The experimental decrease and increase of membrane tension (Figure 2) could be recapitulated by simulating representative protrusion events of 10 s with increasing or decreasing tension (Figure S7C). By using the measured temporal fluctuations in projected cell area (Figure 1D) with the obtained scaling factor as a proxy for tension we could recapitulate the positive correlation of membrane tension and actin density as well as the negative correlations of membrane tension, actin density, and protrusion speed (Figures 7A and 7B).

This showed that our filament-level model is able to recapitulate lamellipodial dynamics as we observed them at the cellular scale and using fluorescent markers. Our model predicts that keratocytes migrating on an adhesive substrate in 2D exhibit a canonically 70° branched network in their lamellipodia the vast majority of the time. This is consistent with the measurements on steady state migrating keratocytes. However, with a low frequency, periods of perpendicular dominated networks (with an order parameter above 0) during phases of high protrusion speed are predicted (Figure 7C).

DISCUSSION

We demonstrate that the lamellipodial actin network undergoes profound structural changes when filaments pushing against the membrane experience varying load. These changes happen at short time-scales and can be a sheer consequence of the geometrical and kinetic properties of Arp2/3-nucleated branched networks. The key-ingredients of the mechanism we propose are the force-velocity relation determining filament elongation rates, the geometry of protruding networks, and the protection from capping near the membrane. Filaments polymerize faster when growing against decreased loads (here, the load is due to membrane tension). With decreasing load the expanding network reaches a speed, where filaments growing perpendicularly to the membrane gradually ''outrun'' filaments pushing at steeper angles because these have to travel a longer distance. Hence, these steeper-angle filaments are eliminated, because they lose contact with the membrane, where the polymerizing factors (VASP and formins) protect them from capping proteins. If the perpendicular filaments originate from ''noise'' in the angular distributions of a purely Arp2/3-nucleated network or if they represent a family nucleated by other factors like formins remains open, as ''non-branch'' pointed ends in the tomograms might not only represent Arp2/3-independent filaments but could also be Arp2/3 nucleated, but later severed or unbranched.

Whereas this proposed mechanism will operate in any Arp2/3-nucleated process like endocytosis, phagocytosis, vesicle trafficking, intracellular pathogen transport, dendritic spine formation, etc., its role in cell motility is most intuitive. Unlike in the idealized keratocyte system, cells migrating in physiological environments will almost never experience mechanically isotropic environments. First, a cell usually pushes against an interstitium with inhomogeneous viscoelastic features, and here it has been shown that lamellipodia can increase their force when counter-resistance increases locally (Heinemann et al., 2011; Prass et al., 2006). Second, the adhesiveness of the substrate is often variable (e.g., in any 3D fibrillar environment). As the actin network slides back in areas of the cell, which are less coupled to the substrate, the load experienced by filaments polymerizing against the leading membrane will be reduced. It has been shown that such local inhomogeneities can be compensated by adaptations in polymerization speed, which ''fill up the space'' resulting from retrograde slippage of the network (Barnhart et al., 2011; Graziano and Weiner, 2014; Renkawitz et al., 2009). Thus, for a cell migrating through a complex 3D environment, the local adaptations we describe serve to keep the cell edge and its actin network coherent. This geometrical adaptation might provide another remarkable feature: wherever two adjacent regions of the network show differential flow speed, the faster network will be less crosslinked to the neighboring, slower, regions, because it is less interconnected by branches. This might

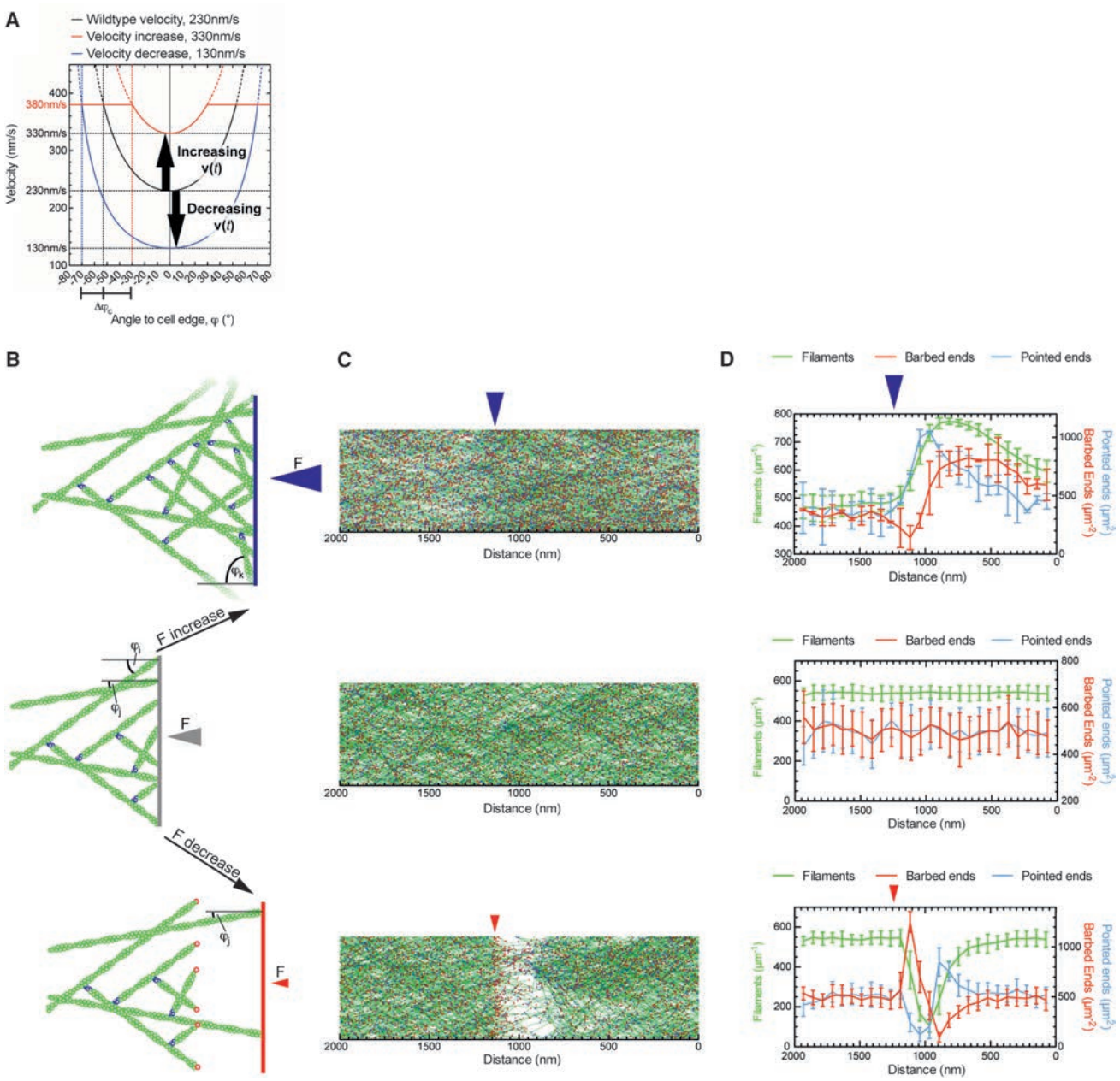


Figure 6. Stochastic Model Recapitulating Light and Electron Microscopy Results

(A) The critical angle at which filaments are not able to keep up with the advancing membrane, φ_c, depends on the maximal polymerization rate, set here as 450 nm/s, and the angle of the filament toward the membrane, φ (black line). A drop in external force leads to a transient increase in protrusion speed leading to a decrease in φ_c and a subsequent, angle-dependent decrease in filaments (red line). Force increase causes a decrease in protrusion speed, leading to an increase in φ_c and filament density (blue line).

(B) Scheme of filaments growing at steady state at intermediate load (middle, gray arrow). Following an increase in load (top, blue arrow) filament density increases and the angle distribution broadens, whereas at a decrease in load (bottom, red arrow) filaments at a higher angle φ_i are preferentially capped and density decreases.

(C) Visualization of model results with filaments shown in green, barbed ends in red, and pointed ends in blue. Visualization of model results in a regime with force increase (top) and decrease (bottom), ΔF, at around 900 nm. Subsequent panels are aligned next to the respective force regimes.

(D) Filament numbers, barbed, and pointed end density quantified for the three conditions like in Figures 3, 4, and 5. In the unperturbed situation with fixed external force, F, the network assumes a steady state with balanced filament density, branching, and capping. After a decrease in F, capping increases, nucleation decreases, and the network is thinned out, whereas force increase causes a nucleation peak and an increase in filament density.

For all graphs, mean and SD for an average of 20 runs are shown.

See also Figure S7, Tables S1 and S2, Movie S5, and the STAR Methods.

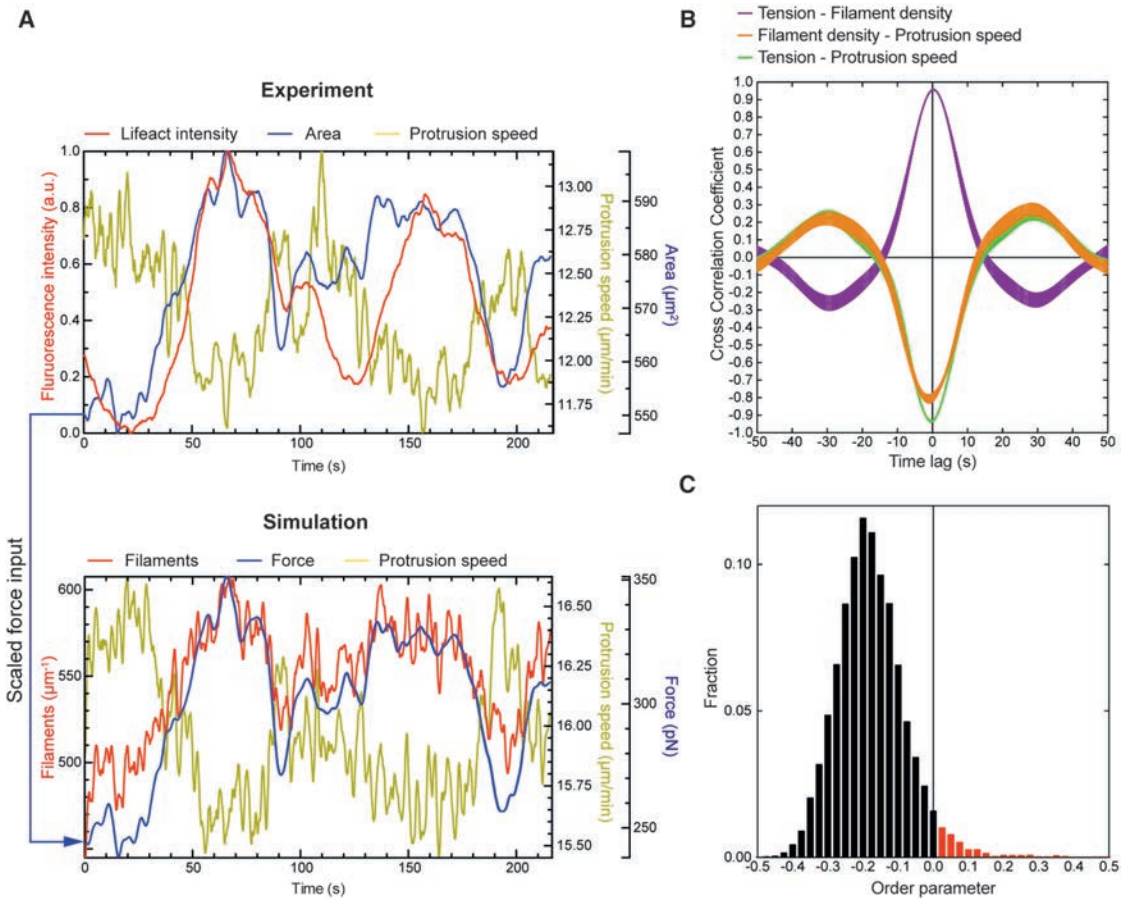


Figure 7. The Stochastic Model Recapitulates Cell Migration Parameters of Fish Keratocytes

(A) Example of temporal fluctuations in lifeact:GFP intensity, area, and protrusion speed for the cell shown in Figure 1 compared to the migration parameters produced by a simulation run with the projected cell area as scaled tension input.

(B) Average of temporal cross-correlation functions of the simulations obtained with the projected cell area of the 21 migrating keratocytes shown in Figure 1. Temporal correlation peaks at time lag zero between lifeact:GFP intensity, projected cell area, and protrusion match the ones observed *in vivo*. Mean and SEM are shown.

(C) Histogram of the order parameters obtained for the 21 steady-state migrating keratocytes. The stochastic model predicts cells migrating with a predominantly negative order parameter, with only occasional switches to positive order parameter values as cells undergo rapid shrinkage and a transient burst in protrusion speed.

See also Figure S7.

potentially allow the two networks to slide relative to each other and might explain how lamellipodia or actin cortices can harbor differentially flowing populations of actin as shown previously (Ponti et al., 2004; Vitriol et al., 2015).

In our work we used manipulations of membrane tension as a proxy for altered load. For the protruding lamellipodium membrane tension likely plays the same mechanical role as an external load. However, lateral membrane tension cannot act locally as it equilibrates instantaneously across the cell surface (Diz-Muñoz et al., 2013; Sens and Plastino, 2015). Based on this potential signaling function it has been proposed that membrane tension acts as a global long-range regulator of cell polarity. Whenever lamellipodial actin stretches the leading membrane it suppresses polymerization at the sides and the trailing edge. Here, polymerization will stall under high membrane tension as it is not boosted by the biochemical enhancers driving the lamellipodium (Houk et al., 2012). This mechanism to suppress competing protrusions has a pure mechanical component (filaments stall under high load) but is also complemented by a signaling mechanism, where membrane tension triggers mTORC2 signaling, which in turn dampens polymerization (Diz-Muñoz et al., 2010). Although this mechanism seems to counteract the geometrical adaptations we describe, it likely acts at much higher force regimes, where filaments ultimately fail to protrude the membrane.

We think that the geometrical rules we describe here represent the most basic regulatory framework of Arp2/3-mediated actin polymerization. They are a fundamental feature of the network structure, which demonstrates that mechanosensitivity is inherently built into the architecture of branched networks. Additional layers of regulation by chemoattractant-sensing and mechanosensitive signaling pathways will act on top of this basic

mechanism and might add modulations e.g., by altering the filament distribution via different nucleators and elongators.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Zebrafish
- METHOD DETAILS
  - Keratocyte preparation
  - mRNA injection
  - Transfection
  - Membrane dye
  - Live cell imaging
  - Micromanipulators
  - Osmotic treatment
  - Tether pulling experiments and force measurements
  - Image analysis
  - Electron tomography
  - Actin filament tracking and analysis
  - Correlated light and electron microscopy
  - Stochastic Simulation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and five movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.07.051.

AUTHOR CONTRIBUTIONS

Conceptualization: J.M. and M.S.; Methodology: J.M., G.S., A.D.L., K. Keren, and R.H.; Software: C.W., K. Kruse, and C.S.; Formal Analysis: J.M., G.S., and R.H.; Investigation: J.M., M.N., I.d.V., and A.D.L.; Writing – Original Draft: J.M. and M.S.; Writing – Review & Editing: all authors; Visualization: J.M. and G.S.; Project Administration: M.S.; Funding Acquisition: M.S.

ACKNOWLEDGMENTS

We thank the scientific support facilities of IST Austria and Biocenter Vienna for technical support and CP Heisenberg for fish lines. This work was supported by the European Research Council (ERC StG 281556), a grant from the Austrian Science Foundation (FWF) (to M.S.), and a grant from Vienna Science and Technology Fund (WWTF) (No. LS13-029 to C.S. and M.S.). Research in the lab of K. Keren was supported by a grant from the United States-Israel Binational Science Foundation (No. 2013275 with Alex Mogilner).

Received: January 11, 2017
Revised: May 21, 2017
Accepted: July 31, 2017
Published: August 31, 2017

REFERENCES

Barnhart, E.L., Lee, K.C., Keren, K., Mogilner, A., and Theriot, J.A. (2011). An adhesion-dependent switch between mechanisms that determine motile cell shape. PLoS Biol. *9*, e1001059.

Behrndt, M., Salbreux, G., Campinho, P., Hauschild, R., Oswald, F., Roensch, J., Grill, S.W., and Heisenberg, C.P. (2012). Forces driving epithelial spreading in zebrafish gastrulation. Science *338*, 257–260.

Bieling, P., Li, T.-D., Weichsel, J., McGorty, R., Jreij, P., Huang, B., Fletcher, D.A., and Mullins, R.D. (2016). Force feedback controls motor activity and mechanical properties of self-assembling branched actin networks. Cell *164*, 115–127.

Carlsson, A.E. (2010). Actin dynamics: from nanoscale to microscale. Annu. Rev. Biophys. *39*, 91–110.

Démoulin, D., Carlier, M.-F., Bibette, J., and Baudry, J. (2014). Power transduction of actin filaments ratcheting in vitro against a load. Proc. Natl. Acad. Sci. USA *111*, 17845–17850.

Dickinson, R.B. (2009). Models for actin polymerization motors. J. Math. Biol. *58*, 81–103.

Diz-Muñoz, A., Krieg, M., Bergert, M., Ibarlucea-Benitez, I., Muller, D.J., Paluch, E., and Heisenberg, C.-P. (2010). Control of directed cell migration in vivo by membrane-to-cortex attachment. PLoS Biol. *8*, e1000544.

Diz-Muñoz, A., Fletcher, D.A., and Weiner, O.D. (2013). Use the force: membrane tension as an organizer of cell shape and motility. Trends Cell Biol. *23*, 47–53.

Gauthier, N.C., Fardin, M.A., Roca-Cusachs, P., and Sheetz, M.P. (2011). Temporary increase in plasma membrane tension coordinates the activation of exocytosis and contraction during cell spreading. Proc. Natl. Acad. Sci. USA *108*, 14467–14472.

Graziano, B.R., and Weiner, O.D. (2014). Self-organization of protrusions and polarity during eukaryotic chemotaxis. Curr. Opin. Cell Biol. *30*, 60–67.

Heinemann, F., Doschke, H., and Radmacher, M. (2011). Keratocyte lamellipodial protrusion is characterized by a concave force-velocity relation. Biophys. J. *100*, 1420–1427.

Houk, A.R., Jilkine, A., Mejean, C.O., Boltyanskiy, R., Dufresne, E.R., Angenent, S.B., Altschuler, S.J., Wu, L.F., and Weiner, O.D. (2012). Membrane tension maintains cell polarity by confining signals to the leading edge during neutrophil migration. Cell *148*, 175–188.

Ji, L., Lim, J., and Danuser, G. (2008). Fluctuations of intracellular forces during cell protrusion. Nat. Cell Biol. *10*, 1393–1400.

Keren, K. (2011). Cell motility: the integrating role of the plasma membrane. Eur. Biophys. J. *40*, 1013–1027.

Keren, K., Pincus, Z., Allen, G.M., Barnhart, E.L., Marriott, G., Mogilner, A., and Theriot, J.A. (2008). Mechanism of shape determination in motile cells. Nature *453*, 475–480.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. *116*, 71–76.

Kreshuk, A., Straehle, C.N., Sommer, C., Koethe, U., Cantoni, M., Knott, G., and Hamprecht, F.A. (2011). Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. PLoS ONE *6*, e24899.

Lai, F.P.L., Szczodrak, M., Block, J., Faix, J., Breitsprecher, D., Mannherz, H.G., Stradal, T.E.B., Dunn, G.A., Small, J.V., and Rottner, K. (2008). Arp2/3 complex interactions and actin network turnover in lamellipodia. EMBO J. *27*, 982–992.

Lee, C.T., Hoopes, M.F., Diehl, J., Gilliland, W., Huxel, G., Leaver, E.V., McCann, K., Umbanhowar, J., and Mogilner, A. (2001). Non-local concepts and models in biology. J. Theor. Biol. *210*, 201–219.

Lee, K., Elliott, H.L., Oak, Y., Zee, C.-T., Groisman, A., Tytell, J.D., and Danuser, G. (2015). Functional hierarchy of redundant actin assembly factors revealed by fine-grained registration of intrinsic image fluctuations. Cell Syst. *1*, 37–50.

Lieber, A.D., Yehudai-Resheff, S., Barnhart, E.L., Theriot, J.A., and Keren, K. (2013). Membrane tension in rapidly moving cells is determined by cytoskeletal forces. Curr. Biol. *23*, 1409–1417.

Lou, S.S., Diz-Muñoz, A., Weiner, O.D., Fletcher, D.A., and Theriot, J.A. (2015). Myosin light chain kinase regulates cell polarization independently of membrane tension or Rho kinase. J. Cell Biol. *209*, 275–288.

Maître, J.L., Berthoumieux, H., Krens, S.F.G., Salbreux, G., Jülicher, F., Paluch, E., and Heisenberg, C.P. (2012). Adhesion functions in cell sorting by mechanically coupling the cortices of adhering cells. Science *338*, 253–256.

Maly, I.V., and Borisy, G.G. (2001). Self-organization of a propulsive actin network as an evolutionary process. Proc. Natl. Acad. Sci. USA *98*, 11324–11329.

Manhart, A., Oelz, D., Schmeiser, C., and Sfakianakis, N. (2015). An extended filament based lamellipodium model produces various moving cell shapes in the presence of chemotactic signals. J. Theor. Biol. *382*, 244–258.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51.

Mogilner, A., and Keren, K. (2009). The shape of motile cells. Curr. Biol. *19*, R762–R771.

Mogilner, A., and Oster, G. (1996). Cell motility driven by actin polymerization. Biophys. J. *71*, 3030–3045.

Mogilner, A., and Oster, G. (2003). Force generation by actin polymerization II: the elastic ratchet and tethered filaments. Biophys. J. *84*, 1591–1605.

Narita, A., Mueller, J., Urban, E., Vinzenz, M., Small, J.V., and Maéda, Y. (2012). Direct determinaftion of actin polarity in the cell. J. Mol. Biol. *419*, 359–368.

Parekh, S.H., Chaudhuri, O., Theriot, J.A., and Fletcher, D.A. (2005). Loading history determines the velocity of actin-network growth. Nat. Cell Biol. *7*, 1219–1223.

Pollard, T.D. (2007). Regulation of actin filament assembly by Arp2/3 complex and formins. Annu. Rev. Biophys. Biomol. Struct. *36*, 451–477.

Pollard, T.D., and Cooper, J.A. (2009). Actin, a central player in cell shape and movement. Science *326*, 1208–1212.

Ponti, A., Machacek, M., Gupton, S.L., Waterman-Storer, C.M., and Danuser, G. (2004). Two distinct actin networks drive the protrusion of migrating cells. Science *305*, 1782–1786.

Prass, M., Jacobson, K., Mogilner, A., and Radmacher, M. (2006). Direct measurement of the lamellipodial protrusive force in a migrating cell. J. Cell Biol. *174*, 767–772.

Raucher, D., and Sheetz, M.P. (2000). Cell spreading and lamellipodial extension rate is regulated by membrane tension. J. Cell Biol. *148*, 127–136.

Renkawitz, J., Schumann, K., Weber, M., LAmmermann, T., Pflicke, H., Piel, M., Polleux, J., Spatz, J.P., and Sixt, M. (2009). Adaptive force transmission in amoeboid cell migration. Nat. Cell Biol. *11*, 1438–1443.

Riedl, J., Crevenna, A.H., Kessenbrock, K., Yu, J.H., Neukirchen, D., Bista, M., Bradke, F., Jenne, D., Holak, T.A., Werb, Z., et al. (2008). Lifeact: a versatile marker to visualize F-actin. Nat. Methods *5*, 605–607.

Ryan, G.L., Petroccia, H.M., Watanabe, N., and Vavylonis, D. (2012). Excitable actin dynamics in lamellipodial protrusion and retraction. Biophys. J. *102*, 1493–1502.

Schaus, T.E., Taylor, E.W., and Borisy, G.G. (2007). Self-organization of actin filament orientation in the dendritic-nucleation/array-treadmilling model. Proc. Natl. Acad. Sci. USA *104*, 7086–7091.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods. *9*, 676–682.

Sens, P., and Plastino, J. (2015). Membrane tension and cytoskeleton organization in cell motility. J. Phys. Condens. Matter *27*, 273103.

Sivaramakrishnan, S., and Spudich, J.A. (2009). Coupled myosin VI motors facilitate unidirectional movement on an F-actin network. J. Cell Biol. *187*, 53–60.

Small, J.V., Herzog, M., and Anderson, K. (1995). Actin filament organization in the fish keratocyte lamellipodium. J. Cell Biol. *129*, 1275–1286.

Svitkina, T.M., Verkhovsky, A.B., McQuade, K.M., and Borisy, G.G. (1997). Analysis of the actin-myosin II system in fish epidermal keratocytes: mechanism of cell body translocation. J. Cell Biol. *139*, 397–415.

Theriot, J.A., and Mitchison, T.J. (1991). Actin microfilament dynamics in locomoting cells. Nature *352*, 126–131.

Vinzenz, M., Nemethova, M., Schur, F., Mueller, J., Narita, A., Urban, E., Winkler, C., Schmeiser, C., Koestler, S.A., Rottner, K., et al. (2012). Actin branching in the initiation and maintenance of lamellipodia. J. Cell Sci. *125*, 2775–2785.

Vitriol, E.A., McMillen, L.M., Kapustina, M., Gomez, S.M., Vavylonis, D., and Zheng, J.Q. (2015). Two functionally distinct sources of actin monomers supply the leading edge of lamellipodia. Cell Rep. *11*, 433–445.

Weichsel, J., and Schwarz, U.S. (2010). Two competing orientation patterns explain experimentally observed anomalies in growing actin networks. Proc. Natl. Acad. Sci. USA *107*, 6304–6309.

Wilson, C.A., Tsuchida, M.A., Allen, G.M., Barnhart, E.L., Applegate, K.T., Yam, P.T., Ji, L., Keren, K., Danuser, G., and Theriot, J.A. (2010). Myosin II contributes to cell-scale actin network treadmilling through network disassembly. Nature *465*, 373–377.

Winkler, C., Vinzenz, M., Small, J.V., and Schmeiser, C. (2012). Actin filament tracking in electron tomograms of negatively stained lamellipodia using the localized radon transform. J. Struct. Biol. *178*, 19–28.

Ydenberg, C.A., Smith, B.A., Breitsprecher, D., Gelles, J., and Goode, B.L. (2011). Cease-fire at the leading edge: new perspectives on actin filament branching, debranching, and cross-linking. Cytoskeleton *68*, 596–602.

Ydenberg, C.A., Padrick, S.B., Sweeney, M.O., Gandhi, M., Sokolova, O., and Goode, B.L. (2013). GMF severs actin-Arp2/3 complex branch junctions by a cofilin-like mechanism. Curr. Biol. *23*, 1037–1045.

Cell

Regulatory Innate Lymphoid Cells Control Innate Intestinal Inflammation

Shuo Wang,[1,7,*] Pengyan Xia,[1,7] Yi Chen,[3] Yuan Qu,[1] Zhen Xiong,[1] Buqing Ye,[1] Ying Du,[1] Yong Tian,[2,4] Zhinan Yin,[5] Zhiheng Xu,[6] and Zusen Fan[1,2,8,*]

[1]CAS Key Laboratory of Infection and Immunity, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 100101 Beijing, China

[2]University of Chinese Academy of Sciences, 100049 Beijing, China

[3]Department of Gastrointestinal Surgery, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, 610041 Chengdu, China

[4]Key Laboratory of RNA Biology of CAS, Institute of Biophysics, Chinese Academy of Sciences, 100101 Beijing, China

[5]The First Affiliated Hospital, Biomedical Translation Research Institute and Guangdong Province Key Laboratory of Molecular Immunology and Antibody Engineering, Jinan University, 510632 Guangzhou, China

[6]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 100101 Beijing, China

[7]These authors contributed equally

[8]Lead Contact

*Correspondence: wangshuo@moon.ibp.ac.cn (S.W.), fanz@moon.ibp.ac.cn (Z.F.)

http://dx.doi.org/10.1016/j.cell.2017.07.027

SUMMARY

An emerging family of innate lymphoid cells (termed ILCs) has an essential role in the initiation and regulation of inflammation. However, it is still unclear how ILCs are regulated in the duration of intestinal inflammation. Here, we identify a regulatory subpopulation of ILCs (called ILCregs) that exists in the gut and harbors a unique gene identity that is distinct from that of ILCs or regulatory T cells (Tregs). During inflammatory stimulation, ILCregs can be induced in the intestine and suppress the activation of ILC1s and ILC3s via secretion of IL-10, leading to protection against innate intestinal inflammation. Moreover, TGF-β1 is induced by ILCregs during the innate intestinal inflammation, and autocrine TGF-β1 sustains the maintenance and expansion of ILCregs. Therefore, ILCregs play an inhibitory role in the innate immune response, favoring the resolution of intestinal inflammation.

INTRODUCTION

The intestine represents a major gateway for potential pathogens, which also contains dietary antigens and an extensive and diverse microbial flora that need to be tolerated. Due to these requirements, the gut constitutes the largest lymphoid organ in the body, encompassing an extensive network of secondary lymphoid organs and is home to an enormous number of lymphocytes (Izcue et al., 2009). Intestinal immune responses take place in the epithelium, lamina propria, and gut-associated lymphoid tissue (GALT). The gut harbors several intestine-specific subpopulations specialized in antigen presentation, antimicrobial immunity, and maintenance of tolerance. Breakdown in the regulatory pathways leads to chronic intestinal inflammation (Kaser et al., 2010). For example, dysregulation of mucosal T cell responses may cause a loss of tolerance, leading to harmful intestinal inflammation reminiscent of human inflammatory bowel disease (IBD) (Strober et al., 2002). Moreover, the regulatory T cells (Tregs) play a critical role in the maintenance of intestinal homeostasis and self-tolerance (Asseman et al., 1999; Sakaguchi, 2005). However, how the intestinal innate immune system is regulated during the intestinal inflammation remains elusive.

Innate lymphoid cells (ILCs) are located in mucosal surfaces to potentiate immune responses, sustain mucosal integrity and promote lymphoid organogenesis (Diefenbach et al., 2014; Eberl et al., 2015). Group 1 (ILC1) cells are characterized by their capacity to secrete interferon γ (IFN-γ) responding to interleukin-12 (IL-12), IL-15, and IL-18 (Gordon et al., 2012; Klose et al., 2014). Group 2 (ILC2) cells generate type 2 T helper (Th2) cell cytokines, such as IL-5, IL-9, and IL-13, in response to IL-25 and IL-33 stimulation (Brestoff et al., 2015; Moro et al., 2010). Group 3 (ILC3) cells produce IFN-γ, IL-17, and IL-22 following stimulation with IL-1β and IL-23 (Buonocore et al., 2010; Klose et al., 2013). Given that ILCs produce substantial effector cytokines when stimulated, they play a critical role in the regulation of type 1, type 2, and type 3 (or Th17 cell) responses, controlling host protective immunity and intestinal homeostasis (Bedoui et al., 2016; Gasteiger and Rudensky, 2014; Shih et al., 2016).

Accumulating evidence shows that inflammatory responses can be induced and regulated independently of adaptive immunity (Sonnenberg and Artis, 2015). Rag (V[D]J recombination activation gene)-deficient mice develop colitis after treatment with anti-CD40 antibody or dextran sodium sulfate (DSS) (Buonocore et al., 2010; Wirtz et al., 2007). Intraepithelial ILC1s undergo expansion and produce large amounts of IFN-γ in the human Crohn's disease. Blocking of ILC1 with anti-NK1.1 antibody in mice is able to attenuate colitis (Fuchs et al., 2013). In addition,

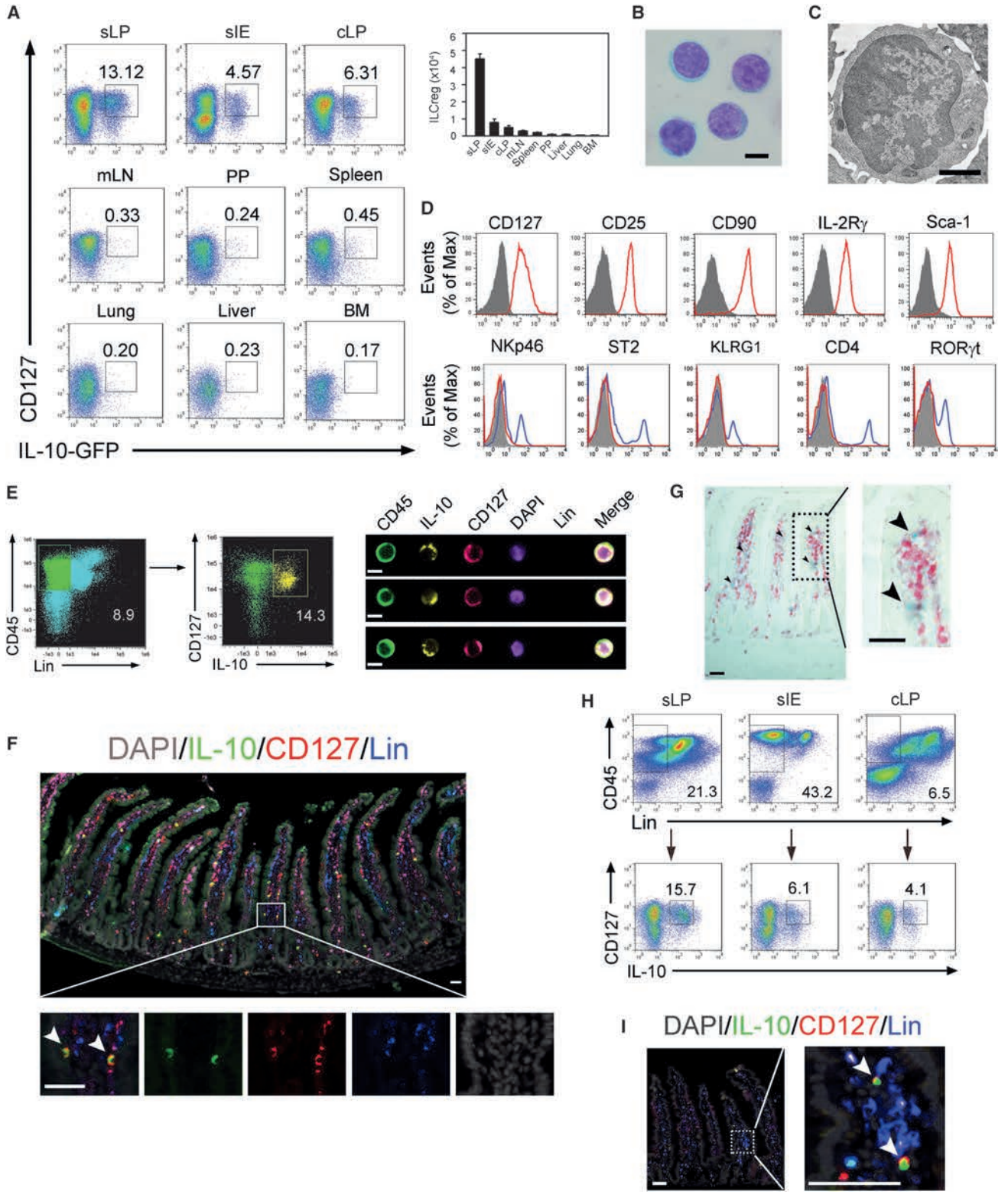


Figure 1. ILCregs Exist in Mouse and Human Intestines

(A) Flow cytometry analysis of ILCregs in different tissues of IL-10-GFP reporter mice. $Lin^{-}CD45^{+}$ lymphocytes were gated out for analysis of CD127 versus IL-10-GFP. The total number of ILCregs was calculated and is shown as the mean ± SD. sLP, small intestine lamina propria; sIE, small intestine intra-epithelial

(legend continued on next page)

ILC3s are implicated in the innate intestinal inflammation via secretion of IL-17 and IL-22. IL-17-producing ILC3s are responsible for the pathogenesis of *Tbx21*$^{-/-}$*Rag2*$^{-/-}$ ulcerative colitis mice (Garrett et al., 2007; Powell et al., 2012). Depletion of ILC3s by anti-Thy1 antibody abrogates innate colitis (Buonocore et al., 2010), suggesting a critical role of ILC3s in the pathogenesis of intestinal inflammation. However, it is unknown whether a regulatory subset of ILCs exists and how they function in the regulation of intestinal inflammation. Here, we identify a novel regulatory subpopulation of ILCs (called ILCregs) that exists in the gut and expands in the intestine following pathogenic stimulation. ILCregs exert an inhibitory role in the innate immune response against intestinal inflammation.

RESULTS

ILCregs Exist in Mouse and Human Intestines

Given that Treg cell-mediated suppression through secretion of their feature cytokines IL-10 and TGF-β (Izcue et al., 2009; Josefowicz et al., 2012), we then gated out Lin$^-$CD45$^+$CD127$^+$ ILCs from multiple tissues in IL-10-GFP reporter mice and analyzed IL-10 expression (Kamanaka et al., 2006). We noticed that a unique subset of ILCs constitutively expressed IL-10 in the intestine (Figures 1A and S1A). These Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILCs were mainly located in the lamina propria of the small intestine (sLP), and some of them in the small intestinal epithelium (sIE) and colon lamina propria (cLP) (Figure 1A). Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILCs were small in size and had a high nucleus/cytoplasm ratio and scanty cytoplasm, which are all characteristics of lymphoid morphology (Figures 1B and 1C). Of note, Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILCs were devoid of expression of CD4 and FoxP3 (Figure S1B), which are signature markers of Tregs (Sakaguchi et al., 2008), indicating that the Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILC subset was different from the Tregs subset. In addition, Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILCs also expressed ILC markers, such as CD25 (IL-2Rα) and CD90 (Thy1). This population also highly expressed IL-2Rγ and Sca-1, but lacked ILC1 markers (NK1.1 and NKp46), ILC2 markers (ST2 and KLRG1), ILC3 markers (NKp46, CD4, and RORγt), or other leukocyte lineage markers (Gury-BenAri et al., 2016; Robinette et al., 2015) (Figures 1D and S1C). These Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ ILCs represented a new IL-10-producing ILC population that we have named regulatory ILCs (ILCregs).

IL-10-producing ILCregs were further confirmed through imaging flow cytometry by using anti-IL-10 antibody (Figure 1E). Furthermore, ILCregs with co-expression of CD127 and IL-10 indeed existed in mouse small and large intestines by immunofluorescence staining (Figures 1F and S1D). These observations were further validated through immunohistochemical staining (Figure 1G). In addition, ILCregs also existed in the human intestine from biopsies (Figures 1H, 1I, and S1E). Collectively, the Lin$^-$CD45$^+$CD127$^+$IL-10$^+$ cells are a new ILC subset that exists in mouse and human intestines.

ILCregs Are Distinct from ILCs and Tregs

We isolated ILCregs from mouse sLP and performed transcriptome microarray assays. Comparing existing transcriptome datasets of ILCs with natural killer (NK) cells (GEO: GSE37448) and Tregs (GEO: GSE68009) highlights that ILCregs do not show any significant gene profile similarity to known ILC subsets and Tregs (Figure 2A). Of note, ILCregs do express high levels of *Il10* and several ILC feature markers, including *Il7r* (encoding CD127), *Il2ra* (encoding CD25), *Il2rg*, and *Ly6a* (encoding Sca-1), which is consistent with our previous observations (Figure 1D). ILCregs uniquely express transcription factors, such as *Id3* and *Sox4* (Figure 2A), and they also highly expressed *Id2*, which is required for development of ILCs (Diefenbach et al., 2014; Spits and Cupedo, 2012). However, ILCregs lack typical transcription factors of other ILCs and Tregs, such as *Rorc* (encoding RORγt), *Tbx21* (encoding T-bet), *Gata3*, and *Foxp3*, respectively (Serafini et al., 2015). Intriguingly, ILCregs constitutively express *Tgfbr1*, *Tgfbr2*, *Il2rb*, and *Il2rg* (Figure 2A), suggesting they responded to TGF-β and IL-2 signaling. Principal component analysis further verified that ILCregs harbor a unique gene profile (Figure 2B). Similarly, human ILCregs show distinct gene profiles of transcription factors and cytokines compared to other ILCs (Figure 2C). Additionally, some feature cytokines, such as IL-10 and transforming growth factor β1 (TGF-β1), are also highly expressed in human ILCregs as well.

High expression levels of TGF-βRI, TGF-βRII, IL-2Rβ, and Id2 (inhibitor of DNA binding 2) in ILCregs were validated by flow cytometry (Figures S1F and S1G). Moreover, *Il2rg*- or *Id2*-deficient

cells; cLP, colon lamina propria; mLN, mesenteric lymph nodes; PP, Peyer's patches; BM, bone marrow. (Lin = CD3e, CD4, CD8, CD19, NK1.1, CD11b, CD11c, Gr1, Ter119).

(B) Giemsa staining of sorted Lin$^-$CD45$^+$CD127$^+$IL-10-GFP$^+$ ILCreg cells from IL-10-GFP mice. Scale bar, 5 μm.

(C) Ultrastructures of ILCregs were analyzed by electron microscopy. Lin$^-$CD45$^+$CD127$^+$IL-10-GFP$^+$ ILCregs were sorted out and visualized by electron microscopy. Scale bar, 2 μm.

(D) Analysis of surface and transcription markers on ILCregs by flow cytometry. Gray histograms depict isotype control of each antibody. Red lines indicate ILCregs. Blue lines denote Lin$^-$CD45$^+$CD127$^+$ cells as positive controls.

(E) Imaging flow cytometry analysis of ILCregs. More than 90% of cells exhibited similar distribution of CD45, IL-10, and CD127 in Lin$^-$CD45$^+$IL-10$^+$CD127$^+$ gate, and three of them were shown. Scale bar, 5 μm.

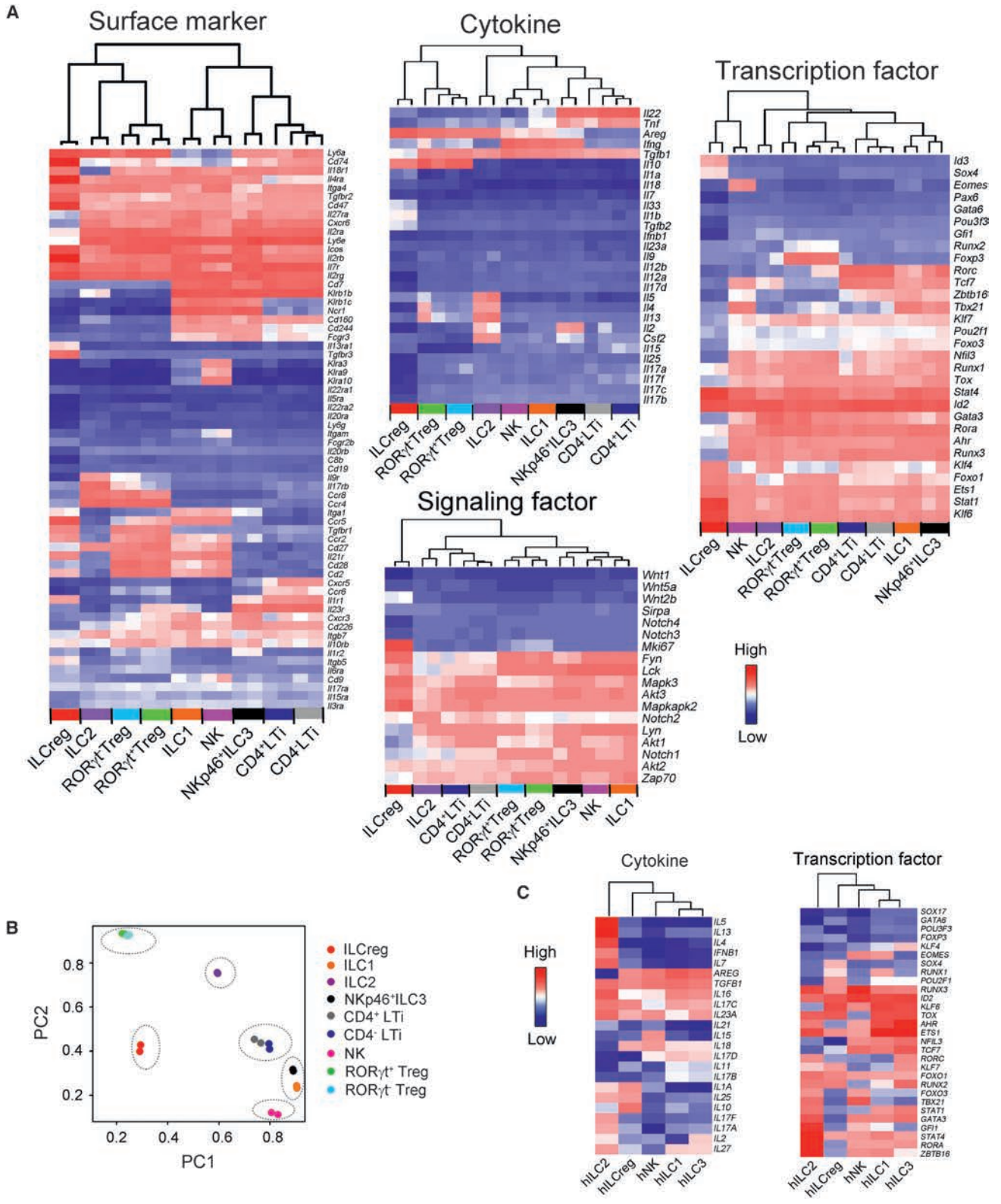
(F) Immunofluorescence staining of ILCregs in mouse small intestine. Intestinal sections were stained with primary antibodies against Lin (blue), CD127 (red), and IL-10 (green). DAPI (gray) denotes nucleus. The white arrowhead indicates ILCregs (Lin$^-$CD127$^+$IL-10$^+$). Scale bar, 50 μm.

(G) Immunohistochemical staining of ILCregs in mouse intestine by using anti-Lin (red) and anti-IL-10 (green) antibodies. The black arrowhead indicates ILCregs (Lin$^-$IL-10$^+$). Scale bar, 50 μm.

(H) Analysis of ILCregs in human intestines by flow cytometry. Lin$^-$CD45$^+$ lymphocytes from human duodenum or colon were gated out for CD127 versus IL-10 analysis.

(I) Immunofluorescence staining of ILCregs in human duodenum. The white arrowhead indicates ILCregs (Lin$^-$CD127$^+$IL-10$^+$). Scale bar, 50 μm. All data are representative of at least three independent experiments.

See also Figure S1.



(legend on next page)

mice abrogated ILCregs (Figure S1H), which are common characters similar to other ILC subsets (Serafini et al., 2015). In contrast, *Rag1*$^{-/-}$ mice show a normal number of ILCregs compared to wild-type (WT) mice (Figure S1H), and ILCregs of *Rag1*$^{-/-}$ mice do not express immature T cell markers (data not shown), suggesting they develop independent of Rag. By contrast, *Id2* deficiency failed to impact the development of Tregs (Figure S1I). To further determine whether ILCregs are derived from myeloid or T cells, we generated *Rosa26-STOP-YFP;Lyz2-Cre;IL-10-GFP* mice and *Rosa26-STOP-YFP;CD4-Cre;IL-10-GFP* mice. We found that ILCregs were indeed negative for YFP signals in the above two mouse strains, whereas their respective lineage positive cells virtually displayed YFP signals (Figures S2A and S2B), indicating that ILCregs are not derived from myeloid or T cells. Recently, integrin $\alpha4\beta7^{+}Id2^{high}$ CHILP (common helper-like innate lymphoid precursor) was identified as a common precursor for all helper-like ILCs, including LTi and ILC1-3 subsets (Klose et al., 2014), whereas PLZF^{+} ILCP (common ILC precursor) is the precursor for ILC1-3s, only giving rise to ILC1s, ILC2s, and ILC3s (Ishizuka et al., 2016). Next, we crossed *Rosa26-STOP-YFP* mice with different Cre mice to determine the precursor of ILCregs. We observed that only *Rosa26-STOP-YFP;Id2-CreERT2* mice displayed YFP-positive ILCregs in the presence or absence of DSS treatment (Figures S2C and S2D). However, ILCregs in *Rosa26-STOP-YFP;PLZF-Cre* mice or *Rosa26-STOP-YFP; RORγt-Cre* mice were YFP negative (Figures S2C and S2D). These data suggest that ILCregs are derived from Id2-driven CHILPs, but not ILCPs or ILC3s.

To further validate IL-10-producing ILCregs represented a distinct lineage as opposed to a further differentiation state of ILC1, ILC2, or ILC3 cells, we established Id2- CreERT2;Rosa^{DTR}, PLZF-Cre;Rosa^{DTR}, and RORγt-Cre;Rosa^{DTR} mice, followed by adoptive transfer assays. With diphtheria toxin (DT) treatment, Id2-CreERT2;Rosa^{DTR} mice abrogated ILCreg cells (Figure S2E), suggesting ILCregs were generated from the progenitor CHILP (Klose et al., 2014). However, deletion of PLZF- or RORγt-expressing cells does not affect the development of ILCregs (Figure S2E), indicating that ILCregs do not differentiate from the progenitor of LTi, ILC1, ILC2, or ILC3 cells (Artis and Spits, 2015). To further verify the progenitor of ILCregs, we transferred CLP, αLP, CHILP, and ILCP progenitors into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice for adoptive transfer assays. We noticed that CLP, αLP, and CHILP were able to produce ILCregs, whereas ILCPs could not generate ILCregs (Figure S2F). Depletion of CHILPs by DT treatment in adoptively reconstituted *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice abrogated generation of ILCregs (Figure S2G). Collectively, ILCregs are derived from CHILPs, but not ILCPs, which are a separate subset distinct from Tregs and other ILC lineages.

ILCregs Contribute to the Resolution of Innate Intestinal Inflammation

Since ILCregs mainly reside in the intestine and develop independent of Rag, we next used innate intestinal inflammation models to monitor ILCreg functions. We treated *Rag1*$^{-/-}$ mice with several inflammatory stimuli, including DSS, anti-CD40 antibody, *Salmonella typhimurium* (*S. typhimurium*), and *Citrobacter Rodentium* (*C. rodentium*). We noticed that ILCregs could be induced in the intestines of *Rag1*$^{-/-}$ mice with these inflammatory stimuli (Figure 3A). Notably, ILCregs gradually expanded in the intestines during the process of inflammation (Figures 3B, 3C, and S3A), reaching a peak level at 8 days post-stimulation, which came up at peak inflammation (Figure S3B). By contrast, the number of ILC1s and ILC3s immediately increased at early stages post-stimulation and decreased soon afterward (Figure S3C), whereas the number of ILC2s did not change during the process of inflammation (Figure S3C). These data indicate that ILCregs can be induced by innate intestinal inflammation.

It has been reported that IL-10-deficient mice display spontaneous colitis (Josefowicz et al., 2012; Kühn et al., 1993). Since IL-10 was a signature cytokine of ILCregs, we then generated *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice to test the inhibitory role of IL-10 in the regulation of innate colitis. We observed that *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice caused more severe colitis with DSS stimulation than *Rag1*$^{-/-}$ mice alone (Figures 3D and 3E), indicating that IL-10 was involved in the suppression of innate colitis. We then transferred WT ILCregs into *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice, and ILCregs and IL-10 were successfully rescued in *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice (Figures S3D and S3E). Moreover, engrafted ILCregs could proliferate in vivo with DSS stimulation (Figure S3F). We observed that ILCreg transfer was able to prevent pathogenic innate colitis of their rescued *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice (Figures 3D–3F). Finally, ILCreg transfer exhibited a higher survival rate with DSS treatment (Figure 3G). Similar observations were achieved by infection with *S. typhimurium* (Figures S3G and S3H). Collectively, ILCregs are able to protect mice from innate immune colitis.

ILCregs Produce IL-10 and TGF-β1 during Intestinal Inflammation

Next, we analyzed cytokine expression profiles of ILCregs in *Rag1*$^{-/-}$ mice during inflammatory stimulation. We observed that ILCregs in *Rag1*$^{-/-}$ mice generated high levels of IL-10 and TGF-β1 after treatment with the previously used stimuli

Figure 2. ILCregs Harbor a Unique Gene Profile Distinct from ILCs and Tregs

(A) Gene expression profiling of ILCregs. mRNA of ILCregs was extracted and analyzed by gene expression profiling by using Mouse Transcriptome Array (v.1.0) (Affymetrix). Genes related to mouse ILCs and mouse Tregs were selected out for heatmap analysis. Biological replicates from two independent experiments are shown.

(B) Principal component analysis of the selected genes in (A) from the indicated cell populations.

(C) Transcriptome microarray analysis of cytokines and transcription factors of human ILCregs. Human ILCregs were isolated, and mRNA of ILCregs was extracted followed by gene expression profiling by using Human Transcriptome Array (v.2.0) (Affymetrix). Genes related to human ILCs were selected for heatmap analysis.

See also Figures S1 and S2.

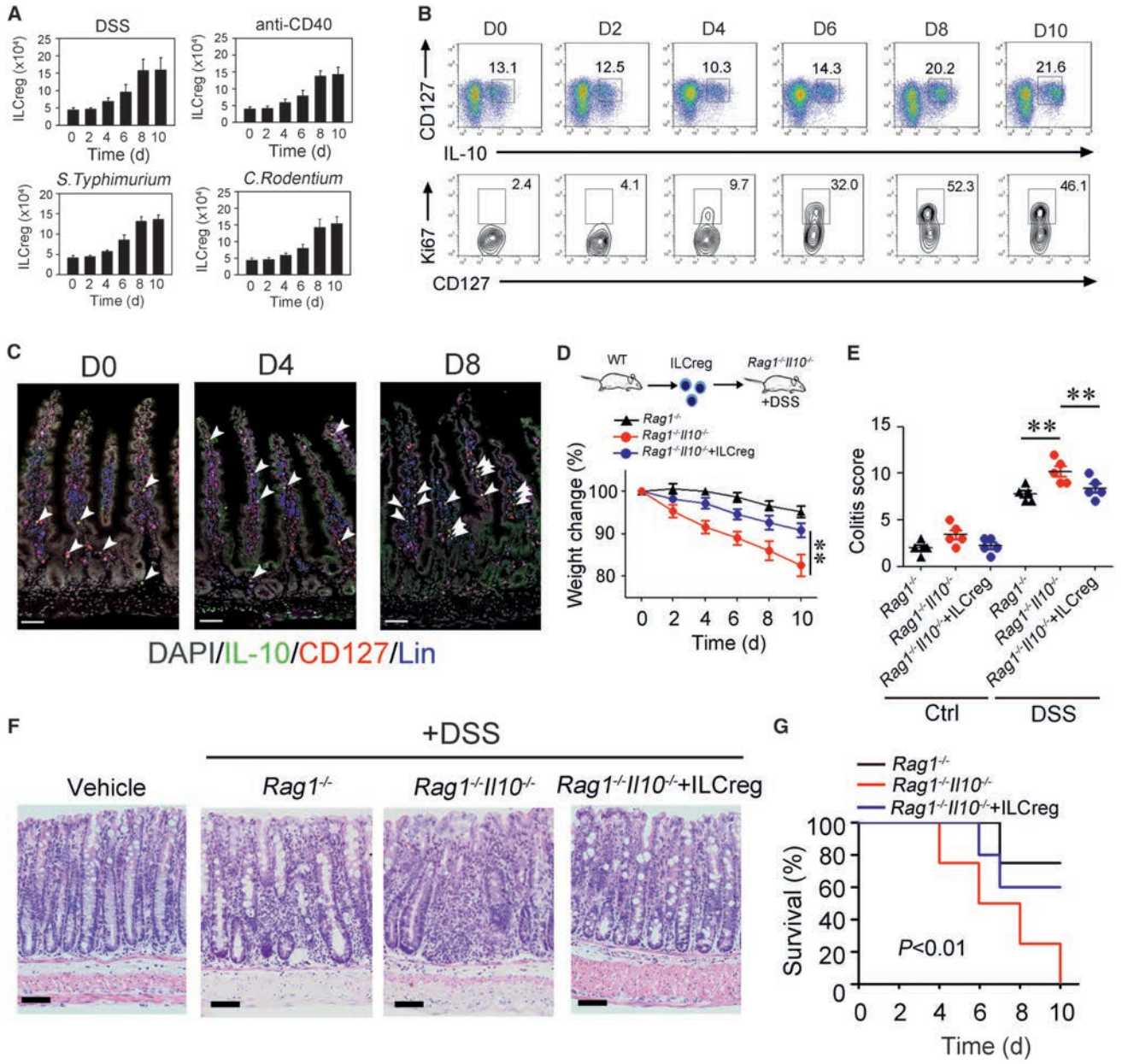


Figure 3. ILCregs Favor the Resolution of Intestinal Inflammation

(A) *Rag1*$^{-/-}$ mice were treated with 3% DSS in drinking water for the indicated days, 150 μg anti-CD40 antibody (intraperitoneally [i.p.]) on day 0, or infected by oral gavage of 5 × 10^4 colony-forming units (CFUs). *S. typhimurium* or 1 × 10^9 CFUs. *C. rodentium* on day 0. Cell numbers of ILCregs were calculated and are shown as means ± SD. n = 6 for each group.

(B) *Rag1*$^{-/-}$ mice were treated with 3% DSS for the indicated days, and ILCregs were analyzed by flow cytometry. Lin$^-$CD45$^+$ lymphocytes were gated out for IL-10 versus CD127 staining for ILCregs. ILCregs were gated out for analysis of Ki67 by flow cytometry.

(C) Immunofluorescence staining of ILCregs in mouse intestine after treatment with DSS by using antibodies against Lin (blue), CD127 (red), and IL-10 (green). The white arrowhead indicates ILCregs. Scale bar, 50 μm.

(D) ILCregs from IL-10-GFP mice were isolated and adoptively transferred into *Rag1*$^{-/-}$*Il10*$^{-/-}$ mice. After 2 weeks, mice were treated with 3% DSS, and body weight changes of mice were calculated and are shown as means ± SD. n = 10 for each group. **p < 0.01 by a non-parametric Mann-Whitney U-test.

(E) Colitis score of the indicated mice by 8 days' oral gavage of water (Ctrl) or DSS. **p < 0.01 by one-way ANOVA. n = 5 for each group.

(F) H&E staining of colon of indicated mice treated like in (D). Scale bar, 50 μm.

(G) Survival rates of adoptively transferred mice. n = 10 for each group. p values (*Rag1*$^{-/-}$*Il10*$^{-/-}$ versus *Rag1*$^{-/-}$*Il10*$^{-/-}$+ILCreg) were analyzed by the Kaplan-Meier method. All data are representative of at least three independent experiments.

See also Figure S3.

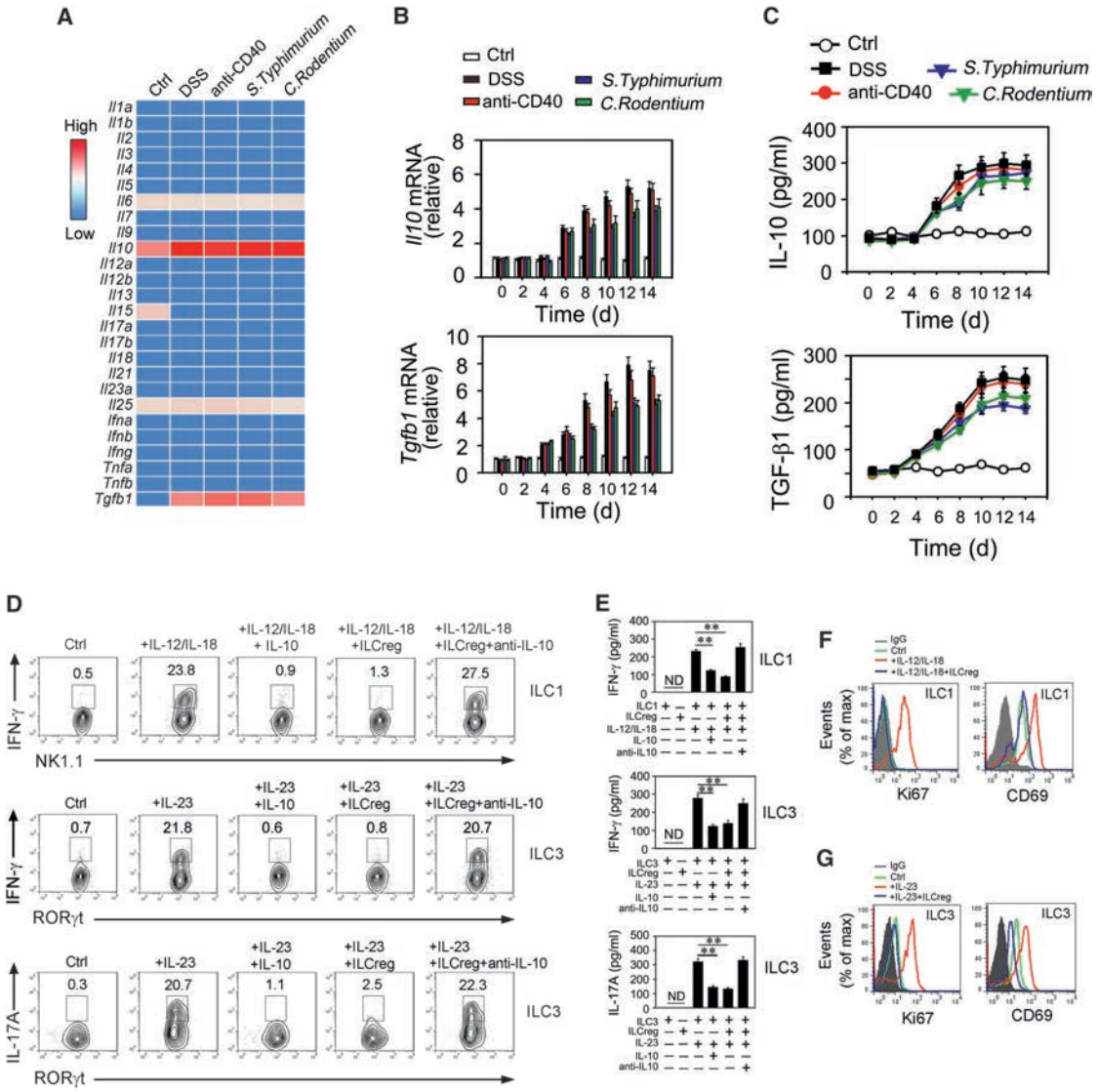


Figure 4. ILCregs Produce IL-10 and TGF-β1 during Intestinal Inflammation

(A) Cytokine expression profiles of ILCregs after inflammatory stimulation. mRNA of ILCregs from *Rag1*$^{-/-}$ mice with indicated treatments for 8 days was extracted and subjected to RT-PCR.

(B and C) *Rag1*$^{-/-}$ mice were stimulated with indicated reagents or pathogens for different days. ILCregs were isolated and subjected to mRNA extraction and RT-PCR of *Il10* and *Tgfb1* mRNA (B) or cultured in media for 24 hr followed by an ELISA of secreted IL-10 and TGF-β1 by ILCregs (C). Untreated *Rag1*$^{-/-}$ mice served as controls (Ctrl).

(D and E) ILCregs inhibit cytokine production of ILC1s and ILC3s in vitro. Isolated ILC1s (Lin^{-}CD127^{+}NK1.1^{+}NKp46^{+}) or ILC3s (Lin^{-}CD45^{+}CD127^{+}RORγt-GFP^{+} from RORγt-GFP mice) were incubated with ILCregs from DSS-treated mice in the presence or in the absence of the indicated reagents for 3 days. Intracellular cytokines were analyzed by flow cytometry (D), and secreted cytokines were analyzed by ELISA (E). Data are shown as means ± SD. ND, not detectable. **p < 0.01 by one-way repeated-measures (RM) ANOVA.

(F and G) Flow cytometry analysis of activation markers on ILC1s (F) and ILC3s (G) after treatment with cytokines in the presence or absence of ILCregs from DSS-treated mice. All data are representative of at least three independent experiments.

See also Figure S4 and Table S1.

(Figures 4A–4C; Table S1). Of note, TGF-β1 was induced earlier than IL-10 during innate immune responses (Figures 4B and 4C). Moreover, ILCregs significantly expanded in the intestine with DSS stimulation (Figure S4A). Additionally, ILCregs secreted both IL-10 and TGF-β1 in the small and large intestines of *Rag1*$^{-/-}$ mice with DSS treatment (Figures S4A and S4B). Altogether, ILCregs generate IL-10 and TGF-β during innate immune responses.

Growing evidence indicates that ILCs mediate T-cell-independent innate colitis (Buonocore et al., 2010). Our above observations demonstrated that ILCregs protected against innate colitis. We thus hypothesized that ILCregs might inhibit the functions of

ILCs to prevent innate immune responses. We isolated ILCregs from sLP of DSS-treated mice and incubated them with activated ILC1s, ILC2s, and ILC3s, respectively. We noticed that ILCregs were able to inhibit cytokine secretion by ILC1s (IFN-γ) and ILC3s (IFN-γ and IL-17A), but not by ILC2s (Figures 4D, 4E, S4C, and S4D). Of note, anti-IL-10 antibody treatment abrogated inhibitory roles of ILCregs (Figures 4D and 4E). However, ILCregs failed to suppress IL-22 secretion by ILC3s (Figures S4E and S4F). Furthermore, IL-10 alone could suppress the activation of ILC1s and ILC3s (Figures 4D and 4E). Of note, via titration assays, ILCregs effectively inhibited the activities of ILC1/3 at ratios around 1:2 to 2:1 (Figure S4G). Ki67 and CD69 are two feature molecules that denote the activation of ILCs (Munneke et al., 2014). We found that ILCregs also inhibited the expression of Ki67 and CD69 on activated ILC1s and ILC3s, respectively (Figures 4F and 4G). By contrast, addition of TGF-β1 did not impact the activation of ILC1s and ILC3s (Figures S4H and S4I). These data indicate that ILCregs suppress the activation of ILC1s and ILC3s via secretion of IL-10.

ILCregs Suppress the Activation of ILC1s and ILC3s via Secretion of IL-10

To further validate the inhibitory role of ILCregs in vivo, we generated *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice that were devoid of T, B, NK, and ILC cells. We then transferred activated ILC1s (aILC1) or activated ILC3s (aILC3), together with ILCregs from DSS-treated mice into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure 5A). We noticed that transfer of aILC1 or aILC3 alone was able to cause innate colitis (Figures 5A–5E). As expected, co-transfer of ILCregs could protect mice from innate colitis caused by activated ILCs (Figure 5A–5E). Finally, ILCreg transplantation suppressed the secretion of inflammatory cytokines by ILCs (Figures 5F and 5G). As expected, neutralization of IL-10 by anti-IL-10 antibody was able to block the inhibitory role of ILCregs (Figures 5A–5G). Next, we generated *Il10ra*$^{flox/flox}$*;Id2-CreERT2* mice, followed by adoptive transfer assays. Transfer of IL-10Rα-deleted ILC1s and ILC3s with WT ILCregs into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice caused much more severe colitis and more elevated activities of ILC1/3 (Figures S5A–S5C). In addition, we used CRISRP/Cas9 technology to delete IL-10 in ILCregs. We found that transplantation of IL-10-deleted ILCregs in adoptively reconstituted *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice also caused much more severe colitis and more elevated activities of ILC1/3 (Figures S5D–S5G). Collectively, ILCregs inhibit the activation of ILCs and protect against intestinal innate inflammation in an IL-10-dependent manner.

We then isolated *FoxP3-DTR-GFP* Treg cells from *FoxP3-DTR-GFP* mice and transferred them together with ILC1s and ILC3s into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure S6A). With DT treatment, Tregs were almost depleted in recipient mice (Figure S6A). We noticed that mice with depletion of Tregs did not affect innate intestinal inflammation, which was comparable to that of mice with replete Tregs (Figures S6B and S6C). Consistently, depletion of Tregs did not impact the activities of ILC1s and ILC3s in engrafted mice (Figure S6D). Based on the microarray analysis, ILCregs constitutively expressed Lck (Figure 2A). We then generated *Lck-Cre;Rosa*^{DTR}*;IL-10-GFP* mice for isolation of ILCregs with DTR (named ILCreg^{DTR} cells), followed by adoptive transfer assays (Figure S6E). DT treatment completely depleted ILCreg cells in the intestine (Figure S6E). We observed that mice with depletion of ILCregs caused much more severe intestinal inflammation compared to those of vehicle-treated control mice (with ILCregs) (Figures S6F–S6J). Consequently, ILCreg depletion remarkably promoted the activation of ILC1s and ILC3s in engrafted mice (Figure S6H). Collectively, Tregs have no significant inhibitory effect on ILC1s and ILC3s-mediated innate inflammation, but ILCregs really attenuate ILC1s and ILC3s-mediated innate inflammation.

In order to further define the physiological role of ILCregs in a lymphocyte replete setting, we isolated ILCreg^{DTR} cells from *Lck-Cre;Rosa*^{DTR}*;IL-10-GFP* mice and transferred them together with T, B, NK, and ILC cells of WT mice into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure S6K). ILCreg^{DTR}, T, B, NK, and ILC cells were successfully reconstituted in *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure S6K). With DT treatment, ILCregs were almost lost in engrafted *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure S6L). Following DSS challenge, mice with depletion of ILCregs caused more severe intestinal inflammation compared to those mice with replete ILCregs (Figures S6M and S6N). Consequently, depletion of ILCregs markedly augmented the activation of ILC1s and ILC3s with DSS treatment (Figure S6O). We conclude that ILCregs play an inhibitory role in the regulation of innate immunity.

Id3 Is Required for the Development of ILCregs

Based on our above microarray analysis (Figure 2A), ILCregs uniquely, highly expressed the translational regulator Id3 (inhibitor of DNA binding 3). Id3 belongs to the Id family members that contain an HLH dimerization domain but lack the basic DNA-binding region. The Id proteins (Id1-4) modulate the DNA-binding activity of E proteins that act as transcriptional activators or repressors (Rivera et al., 2000). Accumulating evidence shows that Id2 and Id3 play critical roles in regulating the developmental progression of T cells (Li et al., 2004; Miyazaki et al., 2014). We further confirmed the expression of Id3 in ILCregs at mRNA and protein levels (Figures 6A and 6B). Next, we wanted to test the role of Id3 during ILCreg development. We generated *Id3*$^{flox/flox}$*;Id2-CreERT2* mice, followed by tamoxifen (TMX) treatment. We found that Id3 deletion abrogated ILCreg generation, but did not impact the development of other ILCs and T cells (Figure 6C), whose phenotype was different from that of Id2-deficient mice. In addition, we rescued WT ILCregs into *Id3*$^{flox/flox}$*;Id2-CreERT2* mice (Figure 6D). Of note, deficiency of Id3 abrogated ILCreg population and displayed severe colitis, which could be attenuated by rescue of WT ILCregs (Figures 6E and 6F). Activation of ILC1/3s was weakened after adoptive transfer of ILCregs (Figure 6G), indicating that deficiency of Id3 impairs ILCreg-related functions. We conclude that Id3 is a unique fate determination factor to drive the development and/or maintenance of ILCregs.

Autocrine TGF-β1 Is Required for the Expansion of ILCregs during Inflammation

Next, we wanted to examine which cytokine signaling induced the development and/or maintenance of ILCregs. As ILCregs expressed high levels of IL-2R and TGF-βR (Figure S1F), we then isolated ILCregs from IL-10-GFP reporter mice and incubated them with common γ-chain-related cytokines and TGF-β1. We

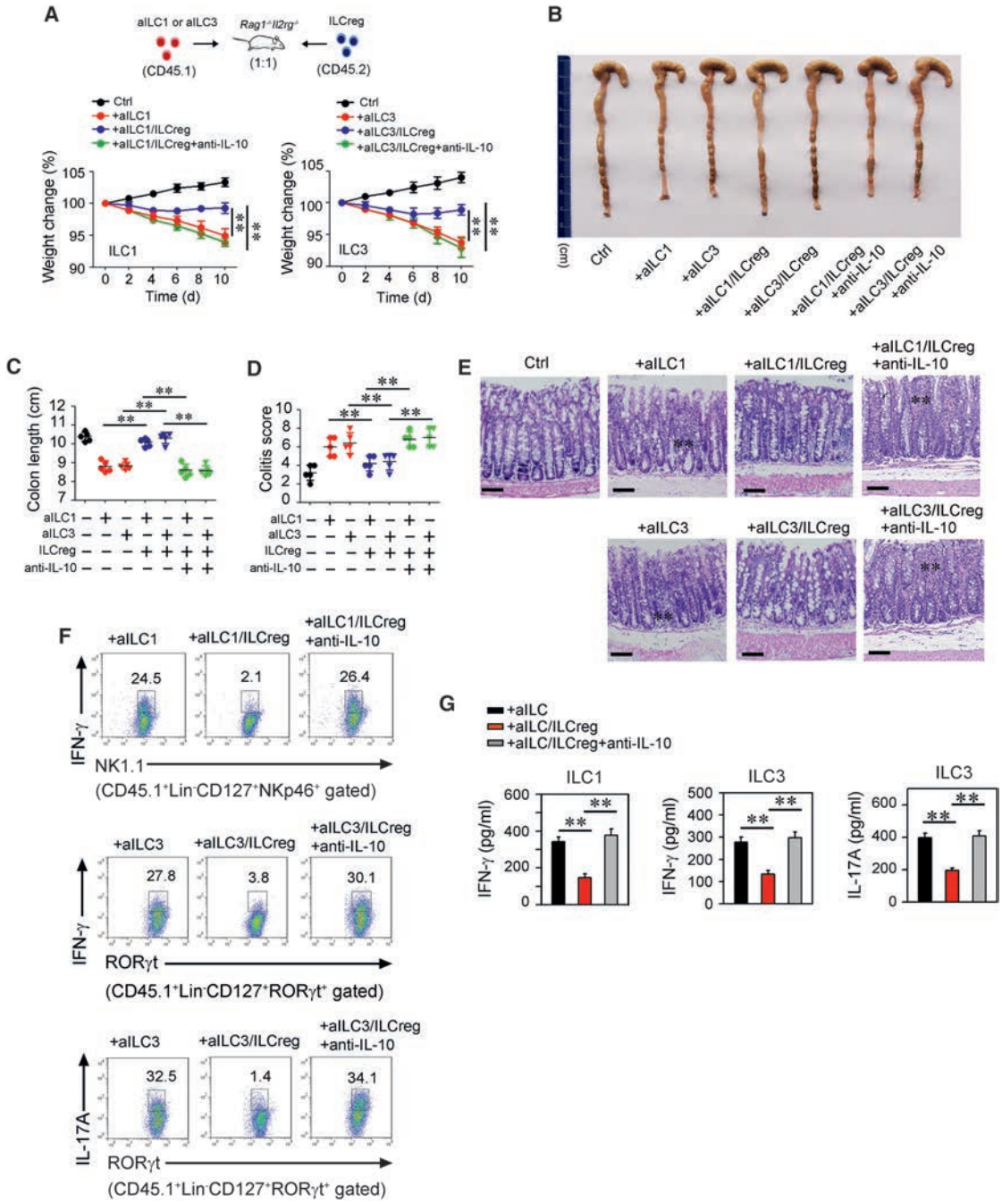


Figure 5. ILCregs Suppress the Activation of ILC1s and ILC3s via Secretion of IL-10

(A) IL-12/IL-18-activated ILC1s (aILC1) or IL-23-activated ILC3s (aILC3) isolated from CD45.1 mice were intravenously injected alone or together with ILCregs from DSS-treated CD45.2 mice (1:1 ratio) into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice. Transferred mice were injected i.p. with PBS or 200 μg of anti-IL-10 antibody daily. Body weight changes were calculated and are shown as means ± SD. n = 5 for each group. **p < 0.01 by one-way ANOVA.

(B and C) Colon images of indicated mice after adoptive transfer for 10 days. Colon length of mice was measured (C). **p < 0.01 by one-way ANOVA. n = 5 for each group.

(D) Colitis score of mice treated as in (A) were analyzed. **p < 0.01 by one-way ANOVA. n = 5 for each group.

(E) H&E staining of colon sections from the indicated mice after adoptive transfer for 10 days. Scale bar, 50 μm. The histological pictures showed inflammation (**).

(F and G) Production of IFN-γ by ILC1s or ILC3s and IL-17A by ILC3s from indicated mice were analyzed by flow cytometry (F) or ELISA (G). **p < 0.01 by a two-tailed unpaired Student's t test. All data represent at least three independent experiments.

See also Figures S5 and S6.

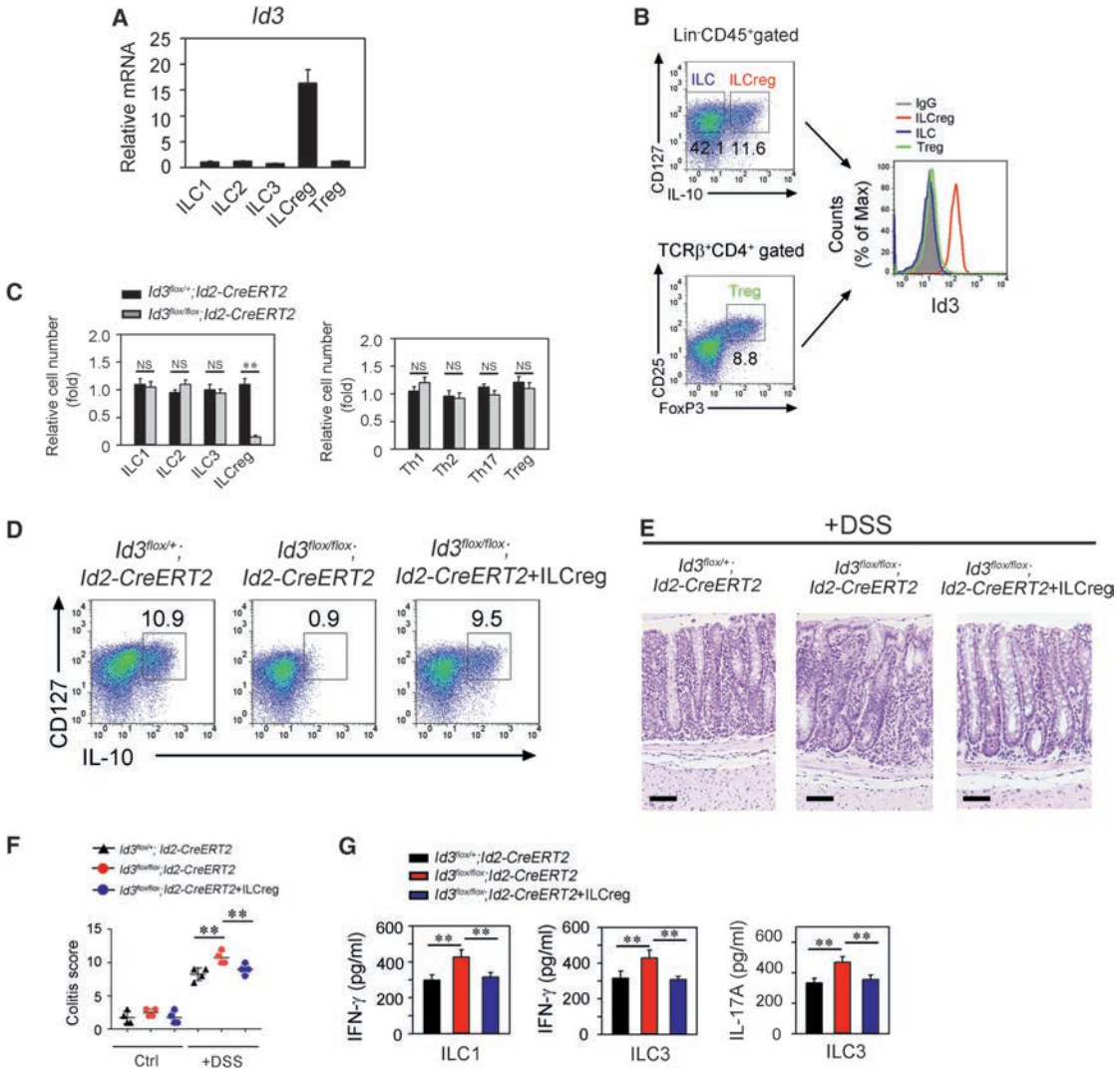


Figure 6. Id3 Is a Fate Decision Factor for ILCreg Development

(A and B) Expression levels of Id3 in different lymphocyte populations. Expression levels of Id3 in indicated populations were analyzed by RT-PCR for mRNA (A) or flow cytometry for protein levels (B).

(C) Deletion of Id3 impairs generation of ILCregs. *Id3$^{flox/+}$;Id2-CreERT2* or *Id3$^{flox/flox}$;Id2-CreERT2* mice were treated with 50 mg/kg (i.p.) tamoxifin (TMX) for 5 consecutive days followed by flow cytometry analysis. Relative cell numbers of each population compared to WT mice were calculated and are shown as means ± SD. **p < 0.01 by a two-tailed unpaired Student's t test. NS, not significant.

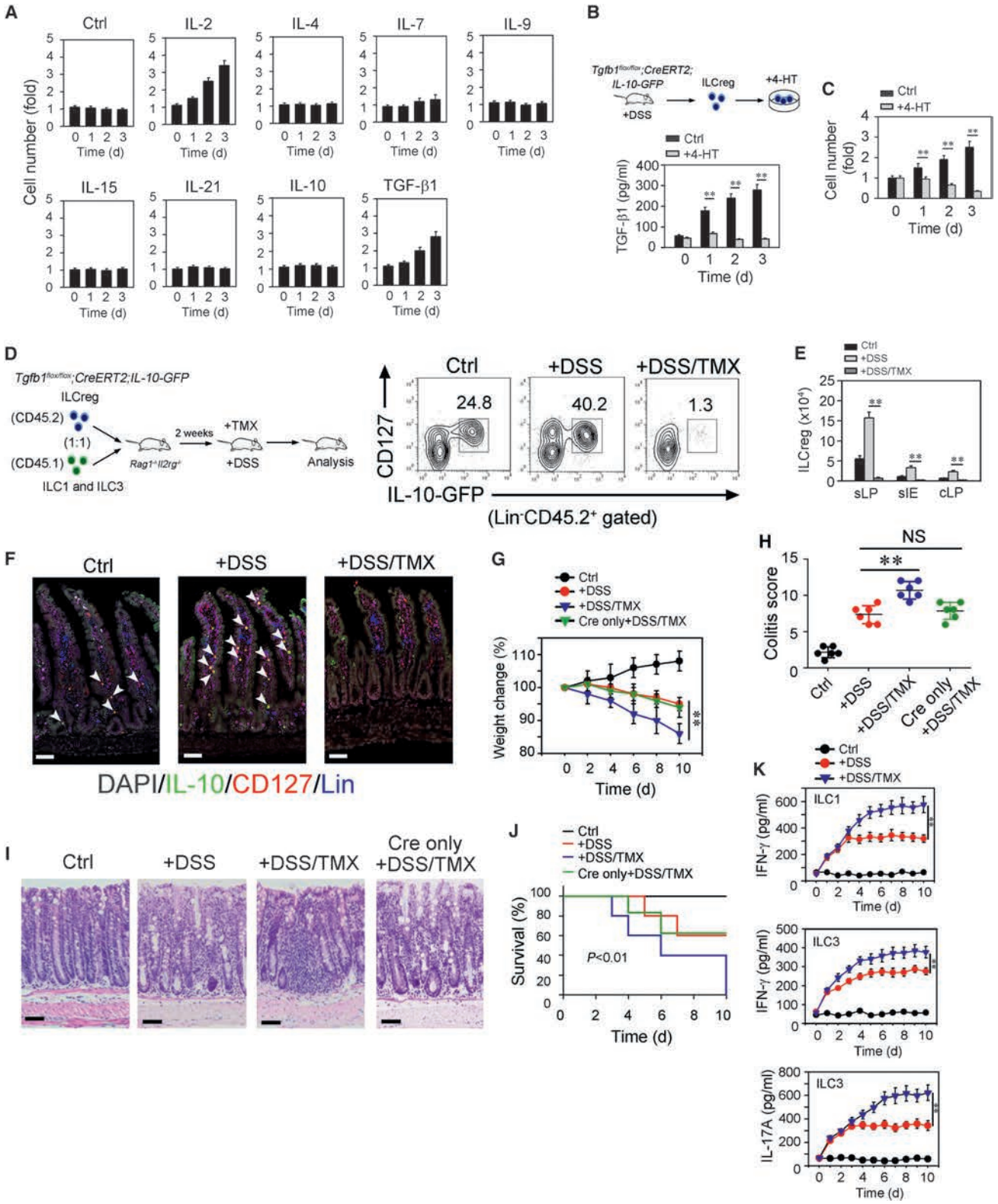
(D) Indicated mice were treated with TMX followed by flow cytometry analysis. WT ILCregs were transferred into *Id3$^{flox/flox}$;Id2-CreERT2* mice for rescue of ILCregs.

(E and F) H&E staining (E) and colitis scores (F) of colons of indicated mice treated for 7 days of 3% DSS in drinking water (+DSS). Scale bar, 50 μm. **p < 0.01 by one-way RM ANOVA.

(G) Secreted cytokines of ILCs from the indicated mice like in (D) were analyzed by ELISA and are shown as means ± SD. **p < 0.01 by One Way RM ANOVA. All data are representative of at least three independent experiments.

noticed that IL-2 and TGF-β1 were able to promote expansion of ILCregs (Figure 7A), suggesting that IL-2 and TGF-β signaling are required for the proliferation and/or maintenance of ILCregs.

Above, we demonstrated that TGF-β1 was induced in ILCregs during innate immune responses (Figures 4A–4C). Next, we sought to test whether autocrine TGF-β1 exerted its function on the development of ILCregs. We then generated *Tgfb1$^{flox/flox}$;Lck-Cre* mice to delete TGF-β1 expression in ILCregs. We mixed bone marrow (BM) of WT (CD45.1) and *Tgfb1$^{flox/flox}$;Lck-Cre* (CD45.2) mice to reconstitute sublethally irradiated *Rag1$^{-/-}$Il2rg$^{-/-}$* recipients. Interestingly, ILCregs from WT (CD45.1) and *Tgfb1$^{flox/flox}$;Lck-Cre* (CD45.2) were successfully reconstituted in BM transplanted recipient mice (Figure S7A). Of note, numbers of CD45.1 and CD45.2 hematopoietic progenitors were comparable (Figure S7B). These data suggest that autocrine TGF-β1 is dispensable for the



(legend on next page)

development of ILCregs. After treatment with DSS, the number of WT ILCregs (CD45.1) increased 2-fold in transplanted recipient mice compared to that with non-DSS treatment (Figure S7A). However, the number of TGF-β1-deleted ILCregs (CD45.2) dramatically reduced post-DSS treatment (Figure S7A). These data indicate that autocrine TGF-β1 is required for the expansion and survival of ILCregs during innate immune responses.

Next, we generated *Tgfb1*$^{flox/flox}$*;CreERT2;IL-10-GFP* mice to determine the role of TGF-β1 secreted by ILCregs in their expansion. We then isolated ILCregs from *Tgfb1*$^{flox/flox}$*;CreERT2;IL-10-GFP* mice and incubated them in media containing IL-2 following treatment with 4-hydroxytamoxifen (4-HT). With administration of 4-HT, ILCregs abrogated TGF-β1 production (Figure 7B). Consequently, the numbers of ILCregs were dramatically reduced (Figure 7C), suggesting autocrine TGF-β1 is required for the expansion of ILCregs.

To further verify TGF-β1 acted as an autocrine manner in vivo, we transferred *Tgfb1*$^{flox/flox}$*;CreERT2;IL-10-GFP* ILCregs, together with ILC1s and ILC3s, into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice (Figure 7D). After administration of TMX, ILCregs were almost abolished in the intestine with DSS stimulation (Figures 7D, 7E, and S7C). These observations were further validated by in situ immunofluorescence staining (Figure 7F). Moreover, TMX-treated mice lost body weight quickly and displayed more severe colitis with DSS treatment (Figures 7G–7J). Finally, these mice produced substantial amounts of IFN-γ and IFN-γ/IL-17A by respective ILC1s and ILC3s (Figure 7K). Similar results were obtained by *S. typhimurium* infection (Figures S7D and S7E). Altogether, autocrine TGF-β1 by ILCregs sustains the maintenance/survival and expansion of inflammation-activated ILCregs.

The TGF-β receptor complex is a tetrameric structure, consisting of two type I TGF-β receptors (TGF-βRI) and two type II TGF-β receptors (TGF-βRII) (Travis and Sheppard, 2014). Deletion either of them is able to abrogate TGF-β signaling (Kang et al., 2009). We generated *Tgfbr2*$^{flox/flox}$*;CreERT2;IL-10-GFP* mice and performed adoptive transfer assays (Figure S7F). After TMX treatment, ILCregs almost lost in the intestine over DSS treatment (Figure S7G). Consequently, *Tgfbr2* abrogation in ILCregs led to more severe colitis and more persistent intestinal damage (Figures S7H and S7I). In parallel, these *Tgfbr2*-deficient mice exhibited active ILC1s and ILC3s (Figure S7J). Therefore, TGF-β1 signaling is required for the expansion and survival of ILCregs.

DISCUSSION

It has been reported that innate lymphoid cells (ILCs) play important roles in innate immunity (Spits and Cupedo, 2012). ILCs are able to cause inflammation independent of T and B cells (Buonocore et al., 2010; Coccia et al., 2012; Powell et al., 2012). However, how the functions of ILCs in intestinal inflammation are regulated still remains elusive. Here, we show that a regulatory subset of ILCs (we called ILCregs) exist in mouse and human intestines. During inflammatory stimulation, ILCregs can be induced in the intestine and suppress the activation of ILC1s and ILC3s via secretion of IL-10, leading to protection against innate intestinal inflammation. Moreover, ILCregs also produce large amounts of TGF-β1, and this autocrine TGF-β1 is required for their expansion and survival during intestinal inflammation. Therefore, ILCregs play a regulatory role in favoring the resolution of innate intestinal inflammation.

ILCs are rapidly activated at the early stage of innate immune response and implicated in eliminating pathogens-infected cells or provoking profound immune response (Eberl et al., 2015). Many studies reported that lack or dysfunction of ILCs by deletion of key transcription factors or signaling factors causes apparent pathologic disorders (Guo et al., 2015; Rankin et al., 2013; Spooner et al., 2013). Like all the immune responses, the activation of these cells also needs to be balanced by stringent regulations because the excessive activation may result in tissue

Figure 7. Autocrine TGF-β1 Is Required for the Expansion of ILCregs

(A) ILCregs from IL-10-GFP mice were isolated, treated with indicated cytokines (2 ng/ml murine IL-2, 5 ng/ml murine IL-4, 5 ng/ml murine IL-7, 5 ng/ml murine IL-9, 5 ng/ml murine IL-15, 5 ng/ml murine IL-21, 5 ng/ml murine IL-10, and 5 ng/ml murine TGF-β1), and analyzed by flow cytometry. Fold changes of cell numbers were analyzed and are shown as means ± SD.

(B) ILCregs from *Tgfb1*$^{flox/flox}$*;CreERT2;IL-10-GFP* mice were isolated and cultured in media containing 2 ng/ml IL-2 with or without of 0.5 μm 4-hydroxytamoxinfen (4-HT) treatment for the indicated days. TGF-β1 amounts were analyzed by ELISA and are shown as means ± SD. **$p < 0.01$ by a two-tailed unpaired Student's t test.

(C) Cell numbers of ILCregs in (B) were calculated and are shown as means ± SD. **$p < 0.01$ by a two-tailed unpaired Student's t test.

(D) ILCregs were isolated from *Tgfb1*$^{flox/flox}$*;CreERT2;IL-10-GFP* mice (CD45.2) and transferred with ILC1s (CD45.1) and ILC3s (CD45.1) into *Rag1*$^{-/-}$*Il2rg*$^{-/-}$ mice. After 2 weeks, mice were treated with water (Ctrl) or 3% DSS (+DSS) or injection of tamoxinfen (TMX) with oral gavage of 3% DSS (+DSS/TMX) for 8 days. ILCregs were analyzed by flow cytometry.

(E) Cell numbers of ILCregs in different tissues as treated in (D) are shown as means ± SD. **$p < 0.01$ by a two-tailed unpaired Student's t test.

(F) Immunofluorescence staining of ILCregs in the indicated mouse intestine sections. Scale bar, 50 μm. The white arrowhead indicates Lin^{-}CD127^{+}IL-10^{+} ILCregs.

(G) Body weight changes of indicated mice were calculated and are shown as means ± SD. n = 6 for each group. **$p < 0.01$ by a non-parametric Mann-Whitney U-test. Mice transferred with ILCregs (from *IL-10-GFP;CreERT2* mice), together with ILC1s and ILC3s, were treated with DSS and TMX and used as negative controls (Cre only+DSS/TMX).

(H and I) Colitis score (H) and histology (I) of the indicated mouse colon were analyzed and are shown as means ± SD. **$p < 0.01$ by one-way ANOVA. n = 6 for each group. Scale bar, 50 μm.

(J) Survival rates of the indicated adoptively transferred mice with administration of DSS and/or TMX. n = 10 for each group. p values (+DSS/TMX versus Cre only+DSS/TMX) were analyzed by Kaplan-Meier method.

(K) ILC1s or ILC3s were isolated from indicated mice and secreted cytokines were analyzed by ELISA. Data are shown as means ± SD. **$p < 0.01$ by a non-parametric Mann-Whitney U-test. All data represent at least three independent experiments.

See also Figure S7.

damage. For example, ILCs are implicated in many inflammatory disorders, including inflammatory bowel diseases, allergic asthma, tissue fibrosis, and oncogenesis (Buonocore et al., 2010; Chang et al., 2011; Kirchberger et al., 2013; McHedlidze et al., 2013). Here, we have shown that ILCregs can inhibit the activation of ILC1s and ILC3s via secretion of IL-10 during the innate intestinal colitis. However, we found that ILCregs do not suppress the function of ILC2s in the process of intestinal inflammation. Of note, we noticed that *Il10rb* was highly expressed in ILC1s and ILC3s, but undetectable in ILC2s (data not shown). A recent study reported that ILC2s are involved in lung inflammation (Monticelli et al., 2016). Whether ILCregs exert tissue-specific regulation needs to be further investigated.

Tolerance in the intestine relies on a wide array of independent immunosuppressive mechanisms. Accumulating evidence indicates that Tregs are able to utilize multiple mechanisms to inhibit immune responses in intestinal inflammation (Asseman et al., 1999; Izcue et al., 2009). Tregs can suppress the activation of other T cells either directly via a cell-contact-dependent fashion or indirectly through downregulating the activity of antigen-presenting cells (APCs) (Thornton and Shevach, 1998; Vignali et al., 2008). Tregs secreted effector cytokines, such as IL-10 and transforming growth factor (TGF)-β play a critical role in the regulation of intestinal inflammation (Maloy et al., 2003). Indeed, IL-10-deficient mice exhibit severe intestinal inflammation and elevated proinflammatory cytokines secretion (Kühn et al., 1993). Moreover, polymorphisms at the *Il10* locus confer risk for ulcerative colitis and Crohn's disease (Franke et al., 2008, 2010), suggesting that IL-10 exerts a central role in the regulation of intestinal mucosal homeostasis. Here, we show that ILCregs uniquely secrete large amounts of IL-10 to downregulate the effector functions of ILC1s and ILC3s during the innate intestinal inflammation. However, Tregs have no apparent inhibitory roles in the suppression of ILC1s and ILC3s during the innate intestinal inflammation. More importantly, ILCregs were located very closely to ILCs in the small intestine (data not shown). The physical closer distance between ILCregs and ILCs might contribute to the regulatory activity of ILCregs to ILC1s and ILC3s.

Of note, IL-22 exerts a protective role in mucosal immunity (Wang et al., 2014; Sugimoto et al., 2008). Here, we found that ILCregs inhibit the generation of IL-17A/IFN-γ, but do not impact the production of IL-22 in ILC3s, which is beneficial to the protective function of ILCregs against innate inflammation. Thus, secretion of IL-22 and IL-17A/IFN-γ by ILC3s might be separately regulated by ILCregs during the process of inflammation as Th17 cells do (Zheng et al., 2007). Accumulating evidence indicates that IL-27/IL-27R signaling triggers IL-10 production in a variety of immune cells (Saraiva and O'Garra, 2010). Here, we demonstrated that ILCregs express low levels of IL-27R, suggesting IL-27/IL-27R signaling may not participate in the regulation of IL-10 production in ILCregs. We observed that ILCregs are activated by IL-2 and TGF-β signaling. It has been reported that IL-2 and TGF-β signaling induces IL-10 generation in Tregs (Brandenburg et al., 2008; Maynard et al., 2007). Therefore, IL-2 and TGF-β signaling might be involved in the regulation of IL-10 production in ILCregs. How ILCregs induces IL-10 secretion needs to be further investigated.

All ILC lineages are originally derived from common lymphoid progenitors (CLPs) in the bone marrow, which also differentiate into T and B cells (Diefenbach et al., 2014). More recently, the progenitors that restrictedly differentiate into ILCs have been defined (Constantinides et al., 2014; Klose et al., 2014; Yang et al., 2015). ILCs are developed from the earliest α-lymphoid progenitor cells (αLPs, $CXCR6^+$ integrin α4β7-expressing CLPs) (De Obaldia and Bhandoola, 2015), which differentiate into restricted CHILP cells (Klose et al., 2014). CHILPs generate all ILCs, including LTi cells, but they fail to give rise to conventional NK cells. Subsequently, their downstream precursor ILCPs (common precursor of ILCs), characterized by expression of the transcription factor PLZF, lose the ability to generate LTi cells and give rise to all ILC1, ILC2, and ILC3 subsets (Constantinides et al., 2014). RORγt (encoded by *Rorc*) drives differentiation of ILC3s from their precursor ILCPs (Montaldo et al., 2014). RORγt deletion causes a complete loss of ILC3s but not ILC1s or ILC2s. These observations indicate that the specifications of different ILC subsets are controlled by transcription factor networks. Here, we show that Id2 is required for the development of ILCregs as other ILCs, suggesting that ILCregs are derived from CHILPs. However, deletion of PLZF or RORγt did not affect the development of ILCregs, indicating that ILCregs were not derived from the all ILC precursor ILCPs. Our data suggest that CHILPs, but not from ILCPs, differentiate into ILCregs, validating that ILCregs are a distinct cell lineage. A recent study clustered all Lin^-CD127^+ cells into 23 transcriptionally homogeneous subgroups based on 6,637 differential genes (Gury-BenAri et al., 2016). Among these genes they used, however, several ILCreg feature genes were not included. We then chose 112 feature genes for ILCregs and ILCs and re-analyzed the single-cell RNA sequencing (RNA-seq) data (GEO: GSE85152). We noticed that ILCreg subpopulation was virtually clustered and distinct from the three existing ILC subpopulations (data not shown).

Transforming growth factor-β (TGF-β) is a multifunctional cytokine that modulates cell survival, proliferation, and differentiation in a variety of settings ranging from embryogenesis to adult tissue homeostasis (Wu and Hill, 2009). TGF-β needs to bind its respective cell surface receptor to initiate TGF-β signaling for its effector functions. The TGF-β receptor complex is a tetrameric structure, consisting of two type I TGF-β receptors (TGF-βRI) and two type II TGF-β receptors (TGF-βRII) (Travis and Sheppard, 2014), deletion either of which can abrogate TGF-β signaling (Kang et al., 2009). For instance, TGF-β signaling plays critical roles in preventing uncontrolled immune responses via inducing development of Tregs (Curotto de Lafaille and Lafaille, 2009; Sakaguchi et al., 2008). A recent study showed that TGF-β signaling is required for the differentiation of salivary gland ILC1s through suppression of Eomes expression (Cortez et al. 2016). In this study, we show that ILCregs produce large amounts of TGF-β, and this autocrine TGF-β is required for the expansion and maintenance of inflammation-induced ILCregs. More importantly, autocrine TGF-β is dispensable for the development of ILCregs. Whether paracrine TGF-β affects the development and differentiation of ILCregs in the intestine needs to be further investigated.

In sum, we have identified a new distinct regulatory subset of ILC cells (called ILCregs), and ILCregs exert inhibitory functions during the innate immune response, which contributes to resolution of innate intestinal inflammation. Hence, we believe that ILCregs may be used to develop new and potential therapies to restore immune tolerance in chronic inflammatory and autoimmune diseases for their effective treatment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice and human tissues
  - Generation of inflammation mouse models
- METHOD DETAILS
  - Isolation of ILCreg cells
  - Flow cytometry assay
  - Imaging flow cytometry
  - Immunofluorescence assay
  - Adoptive transfer assays
  - ELISA assay
  - Histology and immunohistochemistry assay
  - Colitis scores and histologic analysis
  - Generation of IL-10 deficient ILCregs
  - Electron microscopy
  - Real-time PCR assay
  - Microarray analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.07.027.

AUTHOR CONTRIBUTIONS

S.W. designed and performed the experiments, analyzed the data, and wrote the paper; P.X. performed experiments and analyzed data; Y.C., Y.Q., Z.X., B.Y., and Y.D. performed the experiments; Y.T, Z.Y., and Z.X. provided the genetic tool mice. Z.F. initiated the study and organized, designed, and wrote the paper.

ACKNOWLEDGMENTS

We thank Junying Jia, Lei Sun, Yan Teng, Shu Meng, Zhonglin Fu, Yapu Zhao, and Feng Zhang for technical support. We thank Drs. Flavell (Yale University), Yuan Zhuang (Duke University), and Chen Dong (Tsinghua University) for providing the genetic mouse stains. We thank Jing Li (Cnkingbio Company) for technical support. This work was supported by the National Natural Science Foundation of China (91640203, 31530093, 81330047, 81601361, 31471386, 31570872, and 31671531), the Strategic Priority Research Programs of the Chinese Academy of Sciences (XDB19030203 and XDA12020219), the Youth Innovation Promotion Association of CAS (2015073) (to S.W.), and the China Postdoctoral Science Foundation (2015M571141) (to P.X.).

Received: December 7, 2016
Revised: June 16, 2017
Accepted: July 20, 2017
Published: August 24, 2017

REFERENCES

Artis, D., and Spits, H. (2015). The biology of innate lymphoid cells. Nature *517*, 293–301.

Asseman, C., Mauze, S., Leach, M.W., Coffman, R.L., and Powrie, F. (1999). An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. J. Exp. Med. *190*, 995–1004.

Bedoui, S., Gebhardt, T., Gasteiger, G., and Kastenmüller, W. (2016). Parallels and differences between innate and adaptive lymphocytes. Nat. Immunol. *17*, 490–494.

Brandenburg, S., Takahashi, T., de la Rosa, M., Janke, M., Karsten, G., Muzzulini, T., Orinska, Z., Bulfone-Paus, S., and Scheffold, A. (2008). IL-2 induces in vivo suppression by CD4(+)CD25(+)Foxp3(+) regulatory T cells. Eur. J. Immunol. *38*, 1643–1653.

Brestoff, J.R., Kim, B.S., Saenz, S.A., Stine, R.R., Monticelli, L.A., Sonnenberg, G.F., Thome, J.J., Farber, D.L., Lutfy, K., Seale, P., and Artis, D. (2015). Group 2 innate lymphoid cells promote beiging of white adipose tissue and limit obesity. Nature *519*, 242–246.

Buonocore, S., Ahern, P.P., Uhlig, H.H., Ivanov, I.I., Littman, D.R., Maloy, K.J., and Powrie, F. (2010). Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. Nature *464*, 1371–1375.

Chang, Y.J., Kim, H.Y., Albacker, L.A., Baumgarth, N., McKenzie, A.N., Smith, D.E., Dekruyff, R.H., and Umetsu, D.T. (2011). Innate lymphoid cells mediate influenza-induced airway hyper-reactivity independently of adaptive immunity. Nat. Immunol. *12*, 631–638.

Coccia, M., Harrison, O.J., Schiering, C., Asquith, M.J., Becher, B., Powrie, F., and Maloy, K.J. (2012). IL-1β mediates chronic intestinal inflammation by promoting the accumulation of IL-17A secreting innate lymphoid cells and CD4(+) Th17 cells. J. Exp. Med. *209*, 1595–1609.

Constantinides, M.G., McDonald, B.D., Verhoef, P.A., and Bendelac, A. (2014). A committed precursor to innate lymphoid cells. Nature *508*, 397–401.

Cortez, V.S., Cervantes-Barragan, L., Robinette, M.L., Bando, J.K., Wang, Y., Geiger, T.L., Gilfillan, S., Fuchs, A., Vivier, E., Sun, J.C., et al. (2016). Transforming growth factor-beta signaling guides the differentiation of innate lymphoid cells in salivary glands. Immunity *44*, 1127–1139.

Curotto de Lafaille, M.A., and Lafaille, J.J. (2009). Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor? Immunity *30*, 626–635.

De Obaldia, M.E., and Bhandoola, A. (2015). Transcriptional regulation of innate and adaptive lymphocyte lineages. Annu. Rev. Immunol. *33*, 607–642.

Diefenbach, A., Colonna, M., and Koyasu, S. (2014). Development, differentiation, and diversity of innate lymphoid cells. Immunity *41*, 354–365.

Eberl, G., Colonna, M., Di Santo, J.P., and McKenzie, A.N. (2015). Innate lymphoid cells. Innate lymphoid cells: a new paradigm in immunology. Science *348*, aaa6566.

Franke, A., Balschun, T., Karlsen, T.H., Hedderich, J., May, S., Lu, T., Schuldt, D., Nikolaus, S., Rosenstiel, P., Krawczak, M., and Schreiber, S. (2008). Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat. Genet. *40*, 713–715.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. *42*, 1118–1125.

Fuchs, A., Vermi, W., Lee, J.S., Lonardi, S., Gilfillan, S., Newberry, R.D., Cella, M., and Colonna, M. (2013). Intraepithelial type 1 innate lymphoid cells are a unique subset of IL-12- and IL-15-responsive IFN-γ-producing cells. Immunity *38*, 769–781.

Gagliani, N., Amezcua Vesely, M.C., Iseppon, A., Brockmann, L., Xu, H., Palm, N.W., de Zoete, M.R., Licona-Limón, P., Paiva, R.S., Ching, T., et al. (2015). Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. Nature *523*, 221–225.

Garrett, W.S., Lord, G.M., Punit, S., Lugo-Villarino, G., Mazmanian, S.K., Ito, S., Glickman, J.N., and Glimcher, L.H. (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. Cell *131*, 33–45.

Gasteiger, G., and Rudensky, A.Y. (2014). Interactions between innate and adaptive lymphocytes. Nat. Rev. Immunol. *14*, 631–639.

Gordon, S.M., Chaix, J., Rupp, L.J., Wu, J., Madera, S., Sun, J.C., Lindsten, T., and Reiner, S.L. (2012). The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. Immunity *36*, 55–67.

Guo, X., Liang, Y., Zhang, Y., Lasorella, A., Kee, B.L., and Fu, Y.X. (2015). Innate lymphoid cells control early colonization resistance against intestinal pathogens through ID2-dependent regulation of the microbiota. Immunity *42*, 731–743.

Gury-BenAri, M., Thaiss, C.A., Serafini, N., Winter, D.R., Giladi, A., Lara-Astiaso, D., Levy, M., Salame, T.M., Weiner, A., David, E., et al. (2016). The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. Cell *166*, 1231–1246.

Ishizuka, I.E., Constantinides, M.G., Gudjonson, H., and Bendelac, A. (2016). The innate lymphoid cell precursor. Annu. Rev. Immunol. *34*, 299–316.

Izcue, A., Coombes, J.L., and Powrie, F. (2009). Regulatory lymphocytes and intestinal inflammation. Annu. Rev. Immunol. *27*, 313–338.

Josefowicz, S.Z., Lu, L.F., and Rudensky, A.Y. (2012). Regulatory T cells: mechanisms of differentiation and function. Annu. Rev. Immunol. *30*, 531–564.

Kamanaka, M., Kim, S.T., Wan, Y.Y., Sutterwala, F.S., Lara-Tejero, M., Galán, J.E., Harhaj, E., and Flavell, R.A. (2006). Expression of interleukin-10 in intestinal lymphocytes detected by an interleukin-10 reporter knockin tiger mouse. Immunity *25*, 941–952.

Kang, J.S., Liu, C., and Derynck, R. (2009). New regulatory mechanisms of TGF-beta receptor function. Trends Cell Biol. *19*, 385–394.

Kaser, A., Zeissig, S., and Blumberg, R.S. (2010). Inflammatory bowel disease. Annu. Rev. Immunol. *28*, 573–621.

Kirchberger, S., Royston, D.J., Boulard, O., Thornton, E., Franchini, F., Szabady, R.L., Harrison, O., and Powrie, F. (2013). Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. J. Exp. Med. *210*, 917–931.

Klose, C.S., Kiss, E.A., Schwierzeck, V., Ebert, K., Hoyler, T., d'Hargues, Y., Göppert, N., Croxford, A.L., Waisman, A., Tanriver, Y., and Diefenbach, A. (2013). A T-bet gradient controls the fate and function of CCR6-RORγt+ innate lymphoid cells. Nature *494*, 261–265.

Klose, C.S.N., Flach, M., Möhle, L., Rogell, L., Hoyler, T., Ebert, K., Fabiunke, C., Pfeifer, D., Sexl, V., Fonseca-Pereira, D., et al. (2014). Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. Cell *157*, 340–356.

Kühn, R., Löhler, J., Rennick, D., Rajewsky, K., and Müller, W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. Cell *75*, 263–274.

Li, H., Dai, M., and Zhuang, Y. (2004). A T cell intrinsic role of Id3 in a mouse model for primary Sjogren's syndrome. Immunity *21*, 551–560.

Maloy, K.J., Salaun, L., Cahill, R., Dougan, G., Saunders, N.J., and Powrie, F. (2003). CD4+CD25+ T(R) cells suppress innate immune pathology through cytokine-dependent mechanisms. J. Exp. Med. *197*, 111–119.

Maynard, C.L., Harrington, L.E., Janowski, K.M., Oliver, J.R., Zindl, C.L., Rudensky, A.Y., and Weaver, C.T. (2007). Regulatory T cells expressing interleukin 10 develop from Foxp3+ and Foxp3- precursor cells in the absence of interleukin 10. Nat. Immunol. *8*, 931–941.

McHedlidze, T., Waldner, M., Zopf, S., Walker, J., Rankin, A.L., Schuchmann, M., Voehringer, D., McKenzie, A.N., Neurath, M.F., Pflanz, S., and Wirtz, S. (2013). Interleukin-33-dependent innate lymphoid cells mediate hepatic fibrosis. Immunity *39*, 357–371.

Miyazaki, M., Miyazaki, K., Chen, S., Itoi, M., Miller, M., Lu, L.F., Varki, N., Chang, A.N., Broide, D.H., and Murre, C. (2014). Id2 and Id3 maintain the regulatory T cell pool to suppress inflammatory disease. Nat. Immunol. *15*, 767–776.

Montaldo, E., Teixeira-Alves, L.G., Glatzer, T., Durek, P., Stervbo, U., Hamann, W., Babic, M., Paclik, D., Stölzel, K., Gröne, J., et al. (2014). Human RORγt(+) CD34(+) cells are lineage-specified progenitors of group 3 RORγt(+) innate lymphoid cells. Immunity *41*, 988–1000.

Monticelli, L.A., Buck, M.D., Flamar, A.L., Saenz, S.A., Tait Wojno, E.D., Yudanin, N.A., Osborne, L.C., Hepworth, M.R., Tran, S.V., Rodewald, H.R., et al. (2016). Arginase 1 is an innate lymphoid-cell-intrinsic metabolic checkpoint controlling type 2 inflammation. Nat. Immunol. *17*, 656–665.

Moro, K., Yamada, T., Tanabe, M., Takeuchi, T., Ikawa, T., Kawamoto, H., Furusawa, J., Ohtani, M., Fujii, H., and Koyasu, S. (2010). Innate production of T(H)2 cytokines by adipose tissue-associated c-Kit(+)Sca-1(+) lymphoid cells. Nature *463*, 540–544.

Munneke, J.M., Björklund, A.T., Mjösberg, J.M., Garming-Legert, K., Bernink, J.H., Blom, B., Huisman, C., van Oers, M.H., Spits, H., Malmberg, K.J., and Hazenberg, M.D. (2014). Activated innate lymphoid cells are associated with a reduced susceptibility to graft-versus-host disease. Blood *124*, 812–821.

Powell, N., Walker, A.W., Stolarczyk, E., Canavan, J.B., Gökmen, M.R., Marks, E., Jackson, I., Hashim, A., Curtis, M.A., Jenner, R.G., et al. (2012). The transcription factor T-bet regulates intestinal inflammation mediated by interleukin-7 receptor+ innate lymphoid cells. Immunity *37*, 674–684.

Rankin, L.C., Groom, J.R., Chopin, M., Herold, M.J., Walker, J.A., Mielke, L.A., McKenzie, A.N., Carotta, S., Nutt, S.L., and Belz, G.T. (2013). The transcription factor T-bet is essential for the development of NKp46+ innate lymphocytes via the Notch pathway. Nat. Immunol. *14*, 389–395.

Rivera, R.R., Johns, C.P., Quan, J., Johnson, R.S., and Murre, C. (2000). Thymocyte selection is regulated by the helix-loop-helix inhibitor protein, Id3. Immunity *12*, 17–26.

Robinette, M.L., Fuchs, A., Cortez, V.S., Lee, J.S., Wang, Y., Durum, S.K., Gilfillan, S., and Colonna, M.; Immunological Genome Consortium (2015). Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. Nat. Immunol. *16*, 306–317.

Sakaguchi, S. (2005). Naturally arising Foxp3-expressing CD25+CD4+ regulatory T cells in immunological tolerance to self and non-self. Nat. Immunol. *6*, 345–352.

Sakaguchi, S., Yamaguchi, T., Nomura, T., and Ono, M. (2008). Regulatory T cells and immune tolerance. Cell *133*, 775–787.

Saraiva, M., and O'Garra, A. (2010). The regulation of IL-10 production by immune cells. Nat. Rev. Immunol. *10*, 170–181.

Serafini, N., Vosshenrich, C.A., and Di Santo, J.P. (2015). Transcriptional regulation of innate lymphoid cell fate. Nat. Rev. Immunol. *15*, 415–428.

Shih, H.Y., Sciumè, G., Mikami, Y., Guo, L., Sun, H.W., Brooks, S.R., Urban, J.F., Jr., Davis, F.P., Kanno, Y., and O'Shea, J.J. (2016). Developmental acquisition of regulomes underlies innate lymphoid cell functionality. Cell *165*, 1120–1133.

Sonnenberg, G.F., and Artis, D. (2015). Innate lymphoid cells in the initiation, regulation and resolution of inflammation. Nat. Med. *21*, 698–708.

Spits, H., and Cupedo, T. (2012). Innate lymphoid cells: emerging insights in development, lineage relationships, and function. Annu. Rev. Immunol. *30*, 647–675.

Spooner, C.J., Lesch, J., Yan, D., Khan, A.A., Abbas, A., Ramirez-Carrozzi, V., Zhou, M., Soriano, R., Eastham-Anderson, J., Diehl, L., et al. (2013). Specification of type 2 innate lymphocytes by the transcriptional determinant Gfi1. Nat. Immunol. *14*, 1229–1236.

Strober, W., Fuss, I.J., and Blumberg, R.S. (2002). The immunology of mucosal models of inflammation. Annu. Rev. Immunol. *20*, 495–549.

Sugimoto, K., Ogawa, A., Mizoguchi, E., Shimomura, Y., Andoh, A., Bhan, A.K., Blumberg, R.S., Xavier, R.J., and Mizoguchi, A. (2008). IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. J. Clin. Invest. *118*, 534–544.

Thornton, A.M., and Shevach, E.M. (1998). CD4+CD25+ immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. J. Exp. Med. *188*, 287–296.

Travis, M.A., and Sheppard, D. (2014). TGF-β activation and function in immunity. Annu. Rev. Immunol. *32*, 51–82.

Vignali, D.A., Collison, L.W., and Workman, C.J. (2008). How regulatory T cells work. Nat. Rev. Immunol. *8*, 523–532.

Wang, X., Ota, N., Manzanillo, P., Kates, L., Zavala-Solorio, J., Eidenschenk, C., Zhang, J., Lesch, J., Lee, W.P., Ross, J., et al. (2014). Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes. Nature *514*, 237–241.

Wang, S., Xia, P., Chen, Y., Huang, G., Xiong, Z., Liu, J., Li, C., Ye, B., Du, Y., and Fan, Z. (2016a). Natural killer-like B cells prime innate lymphocytes against microbial infection. Immunity *45*, 131–144.

Wang, S., Xia, P., Huang, G., Zhu, P., Liu, J., Ye, B., Du, Y., and Fan, Z. (2016b). FoxO1-mediated autophagy is required for NK cell development and innate immunity. Nat. Commun. *7*, 11023.

Wirtz, S., Neufert, C., Weigmann, B., and Neurath, M.F. (2007). Chemically induced mouse models of intestinal inflammation. Nat. Protoc. *2*, 541–546.

Wu, M.Y., and Hill, C.S. (2009). Tgf-beta superfamily signaling in embryonic development and homeostasis. Dev. Cell *16*, 329–343.

Xia, P., Wang, S., Huang, G., Zhu, P., Li, M., Ye, B., Du, Y., and Fan, Z. (2014). WASH is required for the differentiation commitment of hematopoietic stem cells in a c-Myc-dependent manner. J. Exp. Med. *211*, 2119–2134.

Xia, P., Ye, B., Wang, S., Zhu, X., Du, Y., Xiong, Z., Tian, Y., and Fan, Z. (2016). Glutamylation of the DNA sensor cGAS regulates its binding and synthase activity in antiviral immunity. Nat. Immunol. *17*, 369–378.

Yang, Q., Li, F., Harly, C., Xing, S., Ye, L., Xia, X., Wang, H., Wang, X., Yu, S., Zhou, X., et al. (2015). TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. Nat. Immunol. *16*, 1044–1050.

Zheng, Y., Danilenko, D.M., Valdez, P., Kasman, I., Eastham-Anderson, J., Wu, J., and Ouyang, W. (2007). Interleukin-22, a T(H)17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. Nature *445*, 648–651.

Article

Cell

Sensory Neurons Co-opt Classical Immune Signaling Pathways to Mediate Chronic Itch

Landon K. Oetjen,[1,2] Madison R. Mack,[1,2] Jing Feng,[1,3] Timothy M. Whelan,[1,2] Haixia Niu,[1,2] Changxiong J. Guo,[1,3] Sisi Chen,[4] Anna M. Trier,[1,2] Amy Z. Xu,[1,2] Shivani V. Tripathi,[1,2] Jialie Luo,[1,3] Xiaofei Gao,[1,3] Lihua Yang,[5] Samantha L. Hamilton,[5] Peter L. Wang,[6] Jonathan R. Brestoff,[6] M. Laurin Council,[2] Richard Brasington,[7] András Schaffer,[2,6] Frank Brombacher,[8] Chyi-Song Hsieh,[6,7] Robert W. Gereau IV,[3] Mark J. Miller,[5] Zhou-Feng Chen,[1,3] Hongzhen Hu,[1,3] Steve Davidson,[4] Qin Liu,[1,3] and Brian S. Kim[1,2,3,6,9,*]

[1]Center for the Study of Itch, Washington University School of Medicine, St. Louis, MO 63110, USA
[2]Division of Dermatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
[3]Department of Anesthesiology, Washington University School of Medicine, St. Louis, MO 63110, USA
[4]Department of Anesthesiology, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
[5]Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
[6]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA
[7]Division of Rheumatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
[8]International Centre for Genetic Engineering and Biotechnology and Institute of Infectious Disease and Molecular Medicine, Division of Immunology, University of Cape Town, Cape Town 7700, South Africa
[9]Lead Contact
*Correspondence: briankim@wustl.edu
http://dx.doi.org/10.1016/j.cell.2017.08.006

SUMMARY

Mammals have evolved neurophysiologic reflexes, such as coughing and scratching, to expel invading pathogens and noxious environmental stimuli. It is well established that these responses are also associated with chronic inflammatory diseases, including asthma and atopic dermatitis. However, the mechanisms by which inflammatory pathways promote sensations such as itch remain poorly understood. Here, we show that type 2 cytokines directly activate sensory neurons in both mice and humans. Further, we demonstrate that chronic itch is dependent on neuronal IL-4Rα and JAK1 signaling. We also observe that patients with recalcitrant chronic itch that failed other immunosuppressive therapies markedly improve when treated with JAK inhibitors. Thus, signaling mechanisms previously ascribed to the immune system may represent novel therapeutic targets within the nervous system. Collectively, this study reveals an evolutionarily conserved paradigm in which the sensory nervous system employs classical immune signaling pathways to influence mammalian behavior.

INTRODUCTION

The nervous and immune systems work in concert to sense potential pathogens and provide host-protective responses (Ordovas-Montanes et al., 2015; Veiga-Fernandes and Mucida, 2016). Emerging evidence has demonstrated extensive cross-talk between these systems at multiple barrier surfaces including the gut (Gabanyi et al., 2016), lung (Talbot et al., 2015), and skin (Kashem et al., 2015; Riol-Blanco et al., 2014). These interactions orchestrate multifaceted responses including neurophysiologic reflexes, such as coughing and scratching, to expel invading pathogens and noxious environmental stimuli (Bautista et al., 2014; Veiga-Fernandes and Mucida, 2016). Although evidence suggests that neuroimmunologic processes have coevolved to promote host survival, whether dysregulation of these responses can drive sensory dysfunction, particularly as in chronic itch, remains poorly understood.

Chronic itch, defined as symptoms of itch lasting longer than six weeks, affects up to 15% of the population and has a profoundly negative impact on quality of life (Kini et al., 2011; Matterne et al., 2011; Yosipovitch and Bernhard, 2013). While chronic itch is a central feature of many primary inflammatory skin disorders, such as atopic dermatitis (AD), it also presents secondarily to a variety of medical conditions and numerous neuropathies (Mollanazar et al., 2016; Yosipovitch and Bernhard, 2013). Further, chronic itch can present in the absence of any known disease process, and, in this setting, has been described as chronic idiopathic pruritus (CIP) (Berger and Steinhoff, 2011; Xu et al., 2016). Alarmingly, despite its high incidence and enormous burden on quality of life, there are currently no medications specifically indicated for the treatment of chronic itch.

It is widely appreciated that skin inflammation can lead to itch (Bautista et al., 2014). For example, the pruritogen histamine is associated with acute itch in the setting of urticaria (Kaplan, 2004), and interleukin (IL)-31, a previously identified cytokine pruritogen, has been linked to AD-associated itch (Cevikbas et al., 2014; Dillon et al., 2004; Ruzicka et al., 2017; Sonkoly et al., 2006). However, the precise mechanisms by which proinflammatory mediators promote the sensation of itch remain poorly defined. In AD, the type 2 cytokines IL-4, IL-5, and IL-13 are known to drive skin inflammation (Gittler et al., 2012; Weidinger and Novak, 2016). Further, mice that overexpress IL-4 or IL-13 in the skin develop AD-like disease and robust chronic

itch (Chan et al., 2001; Zheng et al., 2009). Notably, dupilumab, a recently approved monoclonal antibody that blocks both IL-4 and IL-13 signaling via their shared receptor subunit IL-4Rα, has demonstrated remarkable efficacy in treating AD (Beck et al., 2014; Simpson et al., 2016; Thaçi et al., 2016). Despite these observations, it remains unclear whether these canonical proinflammatory mediators also function as pruritogens. Further, whether downstream signaling pathways are conserved across the immune and nervous systems and can be targeted to treat chronic itch remains unexplored.

In this study, we demonstrate that sensory neurons are directly activated by the classical immune signaling molecules IL-4 and IL-13 along previously defined itch-sensory pathways in mice. Further, we show that IL-4 can activate human sensory neurons, provoking the hypothesis that neuronal type 2 cytokine signaling contributes to itch in both mice and humans. By employing sensory neuron-specific genetic deletion of IL-4Rα, we found a previously unrecognized mechanism by which neuronal IL-4Rα critically mediates chronic itch. Surprisingly, we found that, rather than triggering acute itch, activation of neuronal IL-4Rα sensitizes sensory neurons to multiple other pruritogens. Based on IL-4Rα signaling biology, we hypothesized that type 2 cytokine-mediated activation of sensory neurons would be dependent on neuronal Janus kinase (JAK) signaling. Consistent with this hypothesis, both pharmacologic JAK inhibition and sensory neuron-specific genetic deletion of JAK1 resulted in abatement of chronic itch in mice. Although JAK inhibitors are well-established anti-inflammatory agents (Schwartz et al., 2016), whether they exhibit additional neuromodulatory properties is not known. In support of this possibility, we observed a dramatic improvement of itch in a small cohort of CIP patients treated off-label with the JAK inhibitor tofacitinib after having failed other anti-inflammatory therapies. Thus, these signaling mechanisms previously ascribed to the immune system may represent new targets within the sensory nervous system for the treatment of chronic itch. Beyond itch, the discovery of these neuroimmunologic pathways may reveal new insights into sensory perception at multiple barrier surfaces and how these pathways modulate behavior.

RESULTS

Sensory Neurons Are Directly Activated by Type 2 Cytokines

Although type 2 cytokines are well-established mediators of skin inflammation in AD (Gittler et al., 2012; Weidinger and Novak, 2016), it is not clear how these cytokines promote chronic itch. When we examined mouse dorsal root ganglia (DRG), which contain sensory neurons that innervate the skin, we found expression of IL-4Rα (*Il4ra*) and IL-13Rα1 (*Il13ra1*) but not IL-5Rα (*Il5ra*) (Figure 1A). As expected, we also observed expression of the receptor for the known cytokine pruritogen IL-31 (*Il31ra*) (Figure 1A) (Cevikbas et al., 2014; Dillon et al., 2004; Sonkoly et al., 2006). Similarly, human DRG also expressed IL-4Rα (*IL4RA*), IL-13Rα1 (*IL13RA1*), and IL-31RA (*IL31RA*), but not IL-5Rα (*IL5RA*) (Figure 1B). Further, we found that murine DRG and trigeminal ganglia (TG) express comparable patterns of *Il4ra*, *Il13ra1*, and *Il31ra* (Figure 1C). Taken together, these findings suggest that the type 2 cytokines IL-4 and IL-13 may directly activate sensory neurons.

Next, we sought to functionally test whether type 2 cytokines can directly stimulate sensory neurons using ratiometric calcium imaging. Stimulation of mouse DRG neurons with IL-4 (Figure 1D) or IL-13 (Figure 1E) resulted in responses similar to IL-31 stimulation (Figure 1F) in a subset of neurons. Overall, IL-4 or IL-13 stimulation led to calcium responses in 4.1% and 3.3% of total DRG neurons, respectively (Figure 1G), and the frequency of sensory neurons responsive to IL-4 and IL-13 increased in a dose-dependent manner (Figures S1A–S1D). Consistent with previous reports (Cevikbas et al., 2014; Liu et al., 2009), approximately 3.3% and 20% of neurons were responsive to IL-31 and histamine, respectively (Figure 1G), demonstrating that IL-4 and IL-13 stimulation results in responses from a proportion of neurons similar to other canonical pruritogens like IL-31. As expected due to the lack of *Il5ra* expression (Figure 1A), IL-5 stimulation did not trigger responses from DRG neurons (Figures S1E and S1F). Collectively, these results demonstrate that type 2 cytokines can directly activate a subset of mouse sensory neurons.

Given the similar expression profile of type 2 cytokine receptors in mouse and human DRG (Figures 1A and 1B), we next sought to examine whether human DRG neurons would also respond directly to type 2 cytokines. Thus, we obtained viable sensory neurons from human cadaveric donors and stimulated these cells with IL-4 (Valtcheva et al., 2016). Similar to our observations in mice (Figure 1D), we found that a subset of human sensory neurons directly responded to IL-4 (Figure 1H). Together, these findings indicate that sensory neuron responses to type 2 cytokines, such as IL-4, are conserved across mice and humans.

Type 2 Cytokines Activate Itch-Sensory Pathways

To thoroughly characterize neurons that express type 2 cytokine receptors, we reanalyzed two previously published datasets of neuronal transcriptional profiles (Chiu et al., 2014; Usoskin et al., 2015). First, we found that *Il4ra* expression was favored on sort-purified non-peptidergic and peptidergic pruriceptive and nociceptive sensory neurons compared to proprioceptive neurons (Figure S2A) (Chiu et al., 2014). Second, using a single-cell RNA-seq dataset (Usoskin et al., 2015), we found that families of small-diameter neurons predicted to mediate itch (NP1, NP2, and NP3) were enriched in neurons that expressed *Il4ra* and *Il13ra1* compared to other families of nociceptive (PEP1 and PEP2) and mechanoreceptive (TH) neurons (Figure 2A). Although previously defined pruriceptors, such as Mas-related G-protein-coupled receptor A3 (MrgprA3) and IL-31RA, were highly specific to defined itch subsets, *Il4ra* expression spanned multiple pruriceptive clusters (Figure 2A). Consistent with these findings, we found that IL-4- and IL-13-responsive neurons were predominantly small-diameter neurons, similar to histamine-responsive neurons, using calcium imaging (Figures 2B and S2B). Next, to examine whether IL-4 or IL-13 directly activates established itch-sensory pathways, we exposed DRG neurons to successive challenges with defined pruritogens. When DRG neurons were serially stimulated with IL-4, and then IL-31, followed by histamine, 89% of IL-4-responsive neurons also responded to IL-31, histamine, or both (Figure 2C). Similarly, 75% of IL-13-responsive neurons

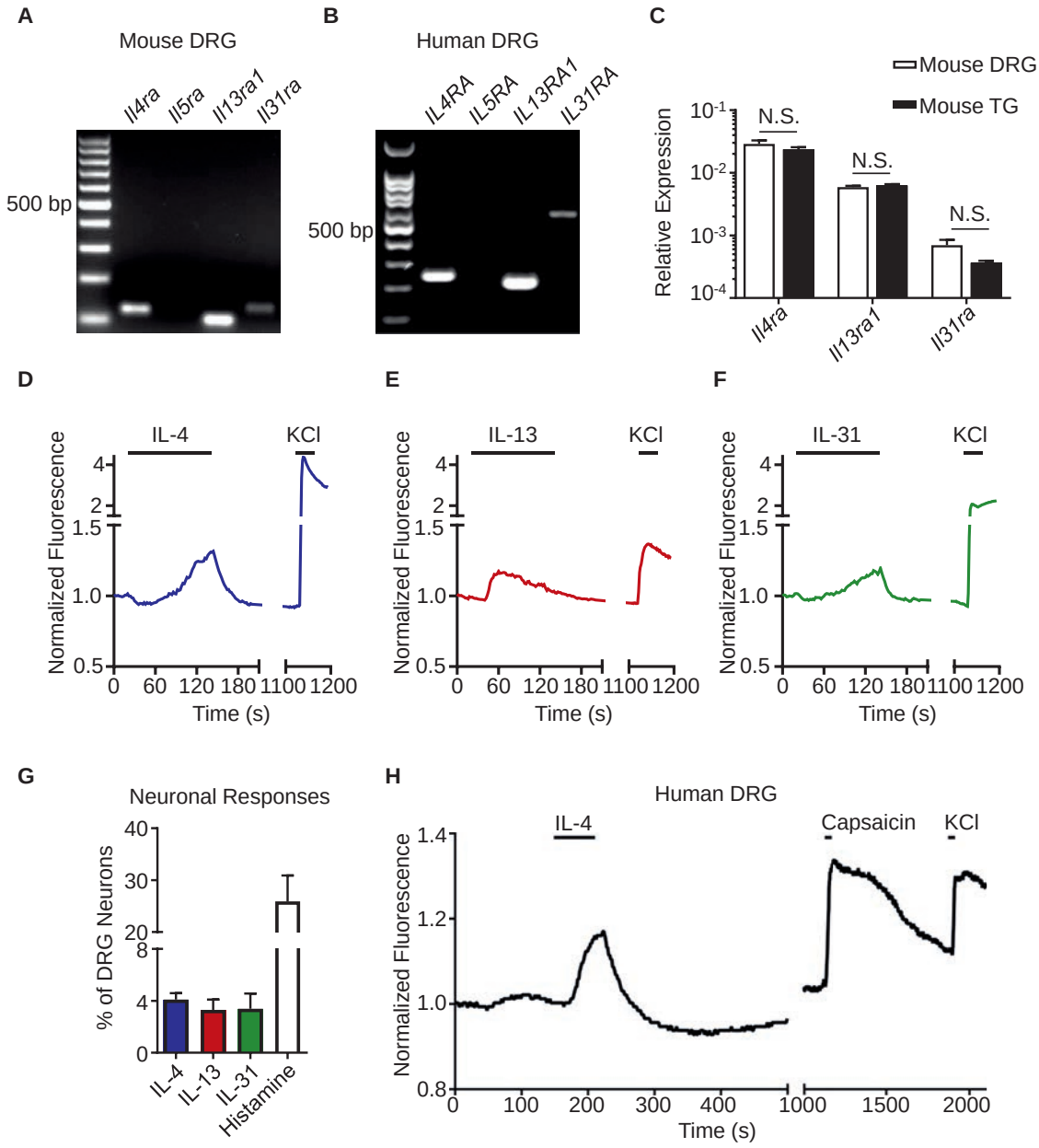


Figure 1. Type 2 Cytokines Activate Mouse and Human Sensory Neurons

(A) Representative gel of RT-PCR product of whole mouse dorsal root ganglia (DRG), n = 4 WT mice.

(B) Representative gel of RT-PCR product of whole human DRG, n = 3 donors.

(C) Comparison of *Il4ra*, *Il13ra1*, and *Il31ra* expression by qRT-PCR in whole mouse DRG and trigeminal ganglia (TG), n = 4 WT mice.

(D–F) Representative calcium imaging trace of mouse DRG neurons in response to (D) recombinant murine (rm)IL-4 (300 nM), (E) rmIL-13 (300 nM), or (F) rmIL-31 (300 nM) and potassium chloride (KCl, 100 mM).

(G) rmIL-4-, rmIL-13-, rmIL-31-, and histamine-responsive DRG neurons as a percentage of total KCl-responsive neurons, n > 500 neurons from 2 WT mice.

(H) Representative calcium imaging trace of human DRG neurons in response to recombinant human IL-4 (300 nM), capsaicin (500 nM), and KCl (50 mM), n > 40 neurons from a male donor.

Data are represented as mean ± SEM. Black bars in calcium imaging traces indicate timing of challenges.

See also Figure S1.

also responded to IL-31, histamine, or both (Figure S2C). Additionally, 63% of neurons that responded to IL-4 also responded to IL-13 (Figure 2D), and 45% of IL-4-responsive neurons also responded to the previously reported cytokine pruritogen thymic stromal lymphopoietin (TSLP) (Wilson et al., 2013a) (Figure S2D). Collectively, these studies indicate that both IL-4 and IL-13 can

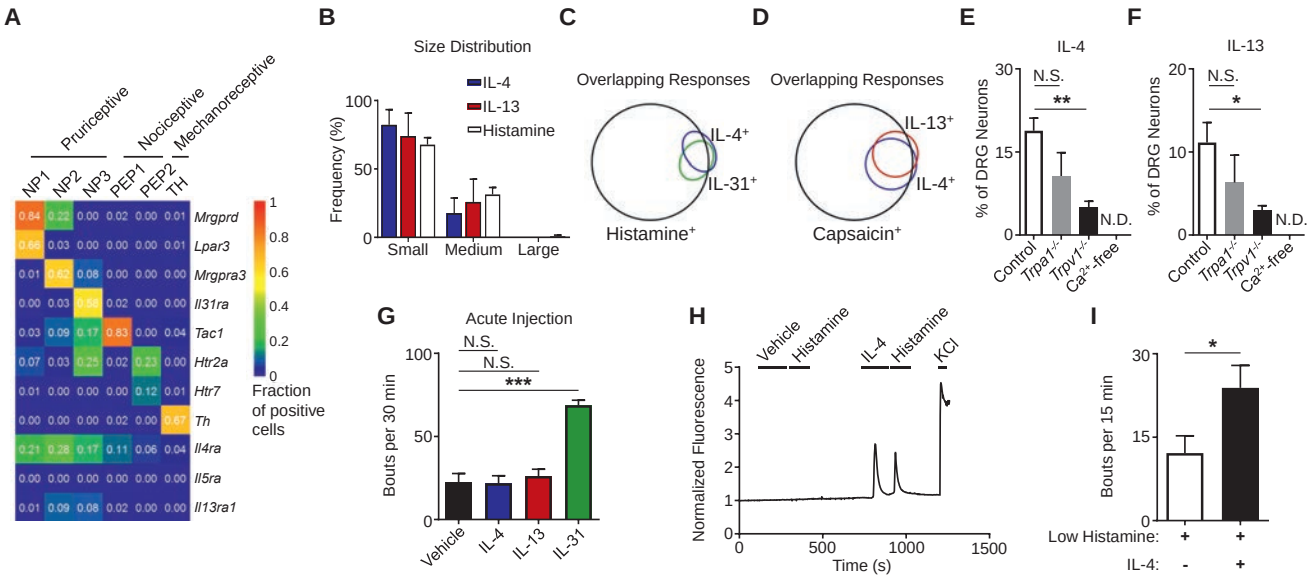


Figure 2. Type 2 Cytokines Activate Itch-Sensory Pathways

(A) Expression of selected genes in mouse DRG neuron populations clustered into functional subsets based on single-cell RNA-seq data. Numbers indicate the fraction of positive cells within individual clusters by thresholding method. Full dataset and methods are available in Usoskin et al. (2015).

(B) Representative size distribution of rmIL-4-, rmIL-13-, and histamine-responsive mouse DRG neurons classified as small-, medium-, and large-diameter neurons (< 18 μm, 18–25 μm, and > 25 μm), n > 500 neurons from a WT mouse.

(C and D) Representative Venn diagrams of overlapping responses of mouse DRG neurons to stimulation with rmIL-4 with subsequent challenge with (C) rmIL-31 (300 nM) and histamine (50 μM) or with (D) rmIL-13 (4.0 μg/mL) and capsaicin (300 nM), n > 300 neurons from a WT mouse.

(E and F) Responses to (E) rmIL-4 (4 μg/mL) and (F) rmIL-13 (4 μg/mL) of mouse DRG neurons from *Trpa1*$^{-/-}$ mice, *Trpv1*$^{-/-}$ mice, and in calcium-free conditions as a percentage of total KCl-responsive neurons compared to WT control mice, n > 300 neurons per group.

(G) Total scratching bouts elicited in response to intradermal (i.d.) cheek injection of vehicle control (0.1% BSA in PBS, 10 μL), rmIL-4, rmIL-13, or rmIL-31 (all 2.5 μg/10 μL), n ≥ 5 mice per group.

(H) Representative calcium imaging trace of mouse DRG neurons responding to low levels of histamine (5 μM) after vehicle control or rmIL-4 (4 μg/mL) challenge, n > 200 neurons from 2 WT mice.

(I) Cheek-directed scratching bouts elicited in response to i.d. cheek injection of a low dose of histamine (4 mM, 10 μL) with or without rmIL-4 (2.5 μg), n ≥ 7 mice per group.

Data are represented as mean ± SEM. Black bars in calcium imaging traces indicate timing of challenges.

See also Figure S2.

activate families of neurons that include previously identified itch-sensory pathways.

Earlier studies have shown that pruritogens, such as histamine and the MrgprA3 agonist chloroquine, are differentially dependent on the transient receptor potential (TRP) channels TRPV1 and TRPA1 for their activity via extracellular calcium influx (Liu et al., 2009; Wilson et al., 2013b). Given the substantial overlap in IL-4- and IL-13-responsive neurons with neurons responsive to the TRPV1 agonist capsaicin (91%; Figure 2D), we sought to test whether neuronal activation by type 2 cytokines is dependent on these TRP channels. Neurons from both *Trpa1*$^{-/-}$ and *Trpv1*$^{-/-}$ mice demonstrated reduced responses to both IL-4 (Figure 2E) and IL-13 (Figure 2F), although only significantly reduced in *Trpv1*$^{-/-}$ mice, and these responses were entirely dependent on extracellular calcium (Figures 2E and 2F). Together, our findings indicate that type 2 cytokines directly activate sensory neurons via TRP channel-dependent calcium influx.

Based on our findings that type 2 cytokines directly activate itch-sensory neurons, we hypothesized that intradermal (i.d.) administration of IL-4 and IL-13 would evoke acute itch. Surprisingly, in contrast to IL-31, high doses of either IL-4 or IL-13 did not elicit acute itch (Figure 2G). Thus, these findings suggest that type 2 cytokines do not function as potent acute pruritogens despite activating itch-sensory neurons.

Several studies have shown that type 2 inflammation-associated cytokines can modulate the transcriptional activity and responses to subsequent stimuli of multiple immune cell populations (Halim et al., 2016; Martinez-Gonzalez et al., 2016). Thus, we hypothesized that sensory neurons would similarly respond to type 2 cytokines by altering their responsiveness to other pruritogens. To test this, we exposed DRG neurons to multiple pruritogens including histamine (Figure 2H), chloroquine (Figure S2E), TSLP (Figure S2F), and IL-31 (Figure S2G) both before and after exposure to IL-4. We observed that IL-4 sensitized subsets of neurons to respond to previously subthreshold levels of histamine (Figure 2H) as well as other pruritogens (Figures S2E–S2G). In support of these findings, i.d. coinjection of a low dose of histamine with IL-4 caused significantly more scratching in mice than histamine alone (Figure 2I). Therefore, our findings demonstrate that although type 2 cytokines are not potent acute pruritogens themselves, they act to sensitize sensory neurons to many different pruritogens. These observations mirror prior

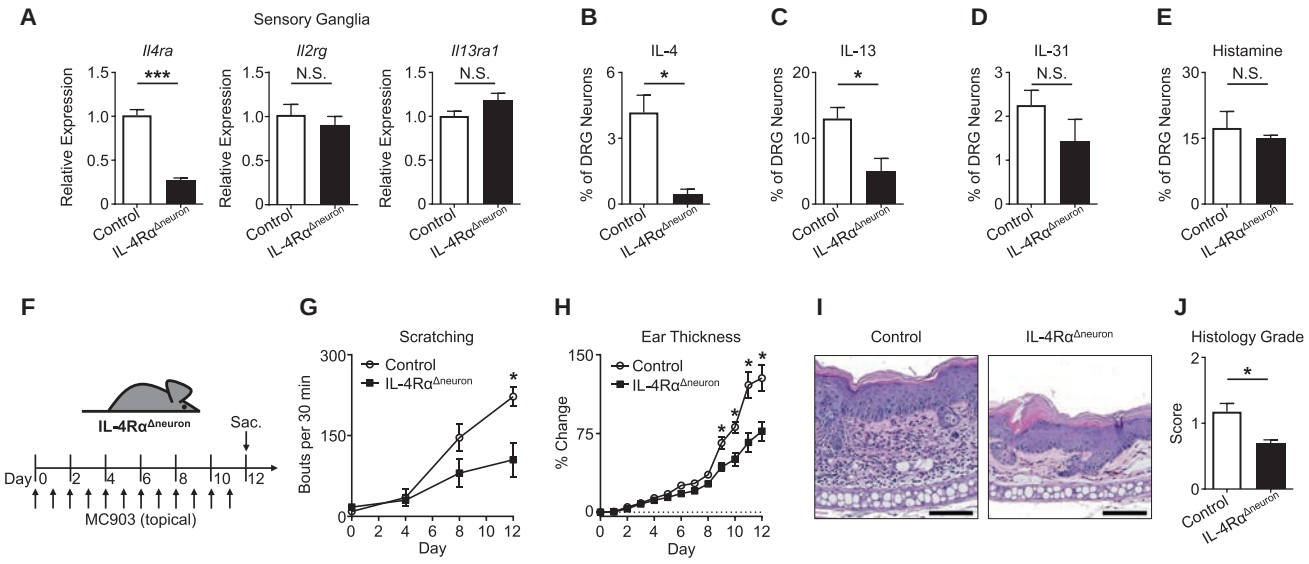


Figure 3. Neuronal IL-4Rα Expression Is Necessary for Chronic Itch

(A) Quantification of *Il4ra*, *Il2rg*, and *Il13ra1* expression by qRT-PCR in whole sensory trigeminal ganglia (TG) of IL-4Rα$^{\Delta neuron}$ mice, n ≥ 4 mice per group.

(B–E) Representative responses of DRG neurons from an IL-4Rα$^{\Delta neuron}$ mouse to (B) rmIL-4 (1 μg/mL), (C) rmIL-13 (4 μg/mL), (D) rmIL-31 (3 μM), and (E) histamine (50 μM) as a percentage of total KCl-responsive neurons compared to neurons from a littermate control mouse, n > 300 neurons per group.

(F) Experimental schematic indicating daily topical treatment with MC903 to the ears of IL-4Rα$^{\Delta neuron}$ and littermate control mice.

(G) Scratching behavior of IL-4Rα$^{\Delta neuron}$ mice compared to littermate control mice over the course of MC903 treatment, n ≥ 7 mice per group.

(H–J) Ear thickness measurements (H), representative H&E histopathology (I), and histology score (J) of MC903-treated IL-4Rα$^{\Delta neuron}$ mice compared to littermate control mice, n ≥ 4 mice per group.

Scale bars indicate 100 μm. Data are represented as mean ± SEM.

See also Figures S3, S4, and S5.

human studies that have shown that challenge with pruritogens in lesional skin of AD patients results in amplified itch responses compared to control subjects (Andersen et al., 2017; Ikoma et al., 2003). Thus, we hypothesized that, rather than acute itch, neuronal type 2 cytokine signaling promotes pathologic chronic itch and that interrupting these signals may represent an effective strategy to target itch.

Neuronal Type 2 Cytokine Signaling Is Necessary for the Development of Chronic Itch

To explore the role of type 2 cytokines in chronic itch, we employed an established model of AD-like skin inflammation in which mice were treated with the topical irritant MC903 (calcipotriol) (Figure S3A) (Kim et al., 2013, 2014; Li et al., 2006). Compared to control mice treated with vehicle only (ethanol, EtOH), MC903-treated mice developed robust skin inflammation as measured by skin thickening (Figure S3B), histologic features of AD (Figures S3C and S3D), and a characteristic AD-associated RNA-seq profile (Figure S3E), as well as marked chronic itch (Figure S3F). To ensure that MC903 does not directly stimulate sensory neurons to elicit itch, we challenged sensory neurons with MC903 and did not observe significant activation (Figure S3G).

To directly investigate the contributions of sensory neuron-intrinsic type 2 cytokine signaling to chronic itch, we generated mice that lack IL-4Rα specifically on sensory neurons ($Na_V1.8$-Cre$^+$ IL-4Rα$^{fl/fl}$, IL-4Rα$^{\Delta neuron}$). Deletion of *Il4ra*, but not *Il2rg* or *Il13ra1*, was confirmed in sensory ganglia of IL-4Rα$^{\Delta neuron}$ mice (Figure 3A). Further, we observed diminished responses to IL-4 and IL-13 in sensory neurons from IL-4Rα$^{\Delta neuron}$ mice (Figures 3B and 3C). In contrast, sensory neurons from IL-4Rα$^{\Delta neuron}$ mice showed normal responses to other pruritogens including IL-31 (Figure 3D) and histamine (Figure 3E) indicating no gross developmental abnormalities in these mice. Strikingly, upon induction of skin inflammation (Figure 3F), IL-4Rα$^{\Delta neuron}$ mice demonstrated a significant reduction in scratching behavior compared to controls (Figure 3G). This decrease was greater than what we observed in *Il31ra*$^{-/-}$ mice treated with MC903 (Figures S4A–S4C) and comparable to decreases in scratching behavior found in other models of itch in which pruriceptors were globally deleted (Liu et al., 2009; Morita et al., 2015). Additionally, skin inflammation in MC903-treated IL-4Rα$^{\Delta neuron}$ mice was reduced as determined by ear thickness measurements (Figure 3H) and histopathologic assessment (Figures 3I and 3J). IL-4Rα$^{\Delta neuron}$ mice also demonstrated a reduction in type 2 immune cells in the skin (Figures S4D–S4G) and a distinct skin transcriptional profile (Figure S4H). Thus, deletion of neuronal IL-4Rα signaling results in attenuation of both itch and skin inflammation.

Given that skin inflammation was reduced, we sought to ensure that the decrease in scratching behavior in these mice was not due to spurious deletion of IL-4Rα in non-neuronal tissues. Although the $Na_V1.8$-Cre system has high specificity for sensory neurons (Agarwal et al., 2004), we elected to investigate

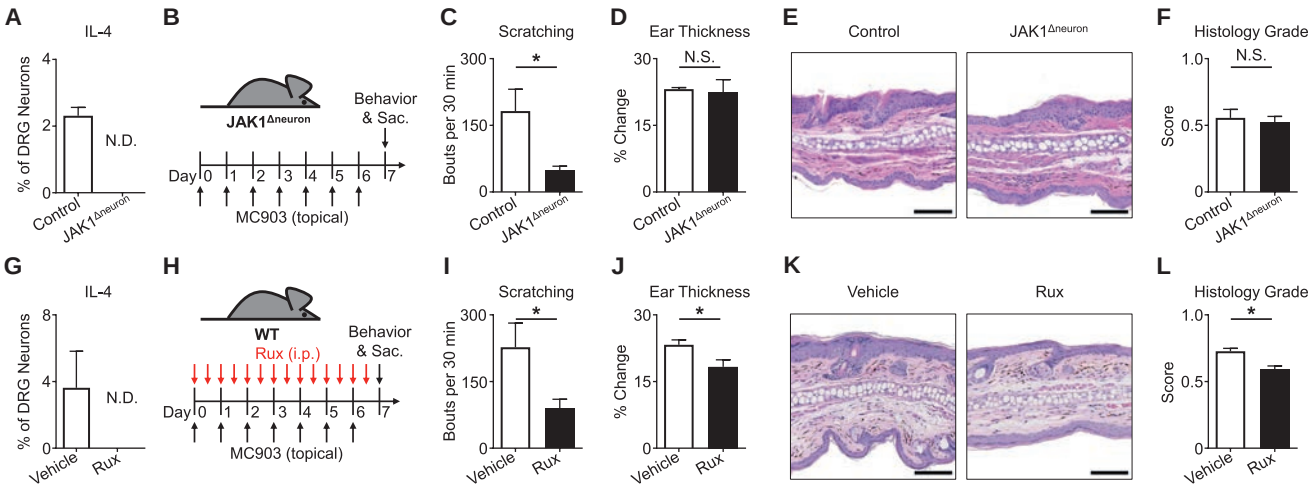


Figure 4. Disruption of Neuronal JAK1 Reduces Chronic Itch

(A) Representative responses to rmIL-4 (1 μg/mL) of DRG neurons from a JAK1$^{\Delta neuron}$ mouse compared to a littermate control mouse as a percentage of total KCl-responsive neurons, n ≥ 200 neurons per group.

(B) Experimental schematic indicating MC903 treatment of JAK1$^{\Delta neuron}$ and littermate control mice.

(C–F) Scratching behavior (C), ear thickness measurement (D), representative H&E histopathology (E), and histology score (F) of JAK1$^{\Delta neuron}$ mice compared to littermate control mice on day 7, n ≥ 8 mice per group.

(G) Representative responses to rmIL-4 (1 μg/mL) of DRG neurons from a WT mouse after incubation with ruxolitinib (Rux, 1 μg/mL) compared to vehicle control as a percentage of total KCl-responsive neurons, n ≥ 200 neurons per group.

(H) Experimental schematic indicating twice daily i.p. injection of vehicle control or Rux (100 μg) during MC903 treatment to the ears of WT mice.

(I–L) Scratching behavior (I), ear thickness measurement (J), representative H&E histopathology (K), and histology score (L) of vehicle control and Rux-treated mice on day 7, n ≥ 10 mice per group.

Scale bars indicate 100 μm. Data are represented as mean ± SEM.

See also Figure S6.

the fidelity of this Cre in the setting of skin inflammation. While we observed $Na_V1.8$ expression in cells from the DRG under steady-state (Figures S5A and S5B) and inflammatory conditions (Figure S5C), no evidence of $Na_V1.8$ expression was found in cells from the skin or skin-draining lymph nodes (Figures S5A–S5C). Additionally, analysis of the skin and skin-draining lymph nodes after MC903 treatment did not yield differences in expression levels of *Il4ra*, *Il2rg*, or *Il13ra1* in IL-4Rα$^{\Delta neuron}$ mice (Figures S5D–S5F). Overall, these findings demonstrate that neuronal IL-4Rα signaling is critical for the elicitation of chronic itch.

Disruption of Neuronal JAK1 Signaling Reduces Chronic Itch

Given that type 2 cytokines are known to signal through JAK-dependent pathways in immune cells (Kelly-Welch et al., 2003), we chose to examine whether itch-sensory neurons also express JAK signaling components. Again using a previously published single-cell RNA-seq dataset (Usoskin et al., 2015), we identified that JAK1 is enriched in pruriceptive neurons (Figure S6A). We then examined whether activation of IL-4Rα on sensory neurons results in increased JAK1 signaling and found that, following stimulation with IL-4, more sensory neurons exhibited JAK1 phosphorylation (Figures S6B–S6D). To functionally assess whether type 2 cytokines activate sensory neurons in a JAK1-dependent manner, we generated mice in which JAK1 is conditionally deleted in sensory neurons ($Na_V1.8$-Cre^{+} JAK1$^{fl/fl}$, JAK1$^{\Delta neuron}$) and found that neuronal IL-4 responses were abrogated in the absence of JAK1 using calcium imaging (Figure 4A). Given the importance of JAK1 in many cellular processes, we sought to ensure that neurons from JAK1$^{\Delta neuron}$ mice were developmentally and functionally intact. We performed a morphological assessment of DRG neurons and found a comparable size distribution between JAK1$^{\Delta neuron}$ mice and controls (Figure S6E). Additionally, no significant differences were observed in neuronal responses across a number of pruritogens (Figures S6F–S6I). Rotarod performance was also found to be equivalent between JAK1$^{\Delta neuron}$ mice and controls (Figures S6J and S6K). Collectively, these data indicate that neuronal development and function are intact in JAK1$^{\Delta neuron}$ mice. Thus, given the enrichment of JAK1 expression in pruriceptive neurons (Figure S6A) and our data demonstrating the dependence of chronic itch on neuronal IL-4Rα (Figures 3F and 3G), we hypothesized that neuronal JAK1 would be required for chronic itch.

To directly investigate the dependence of chronic itch on neuronal JAK1 signaling, we treated JAK1$^{\Delta neuron}$ mice with MC903 (Figure 4B). Upon induction of AD-like disease, JAK1$^{\Delta neuron}$ mice exhibited a marked reduction in scratching behavior (Figure 4C) even in the presence of skin inflammation (Figures 4D–4F). To complement this genetic approach, we employed pharmacologic inhibition of JAK signaling using the JAK inhibitor ruxolitinib. Similar to our findings in JAK1$^{\Delta neuron}$ mice (Figure 4A), we found that application of ruxolitinib to sensory neurons resulted in a reduction in responses to IL-4 (Figure 4G). We then used MC903 treatment to induce AD-associated itch in

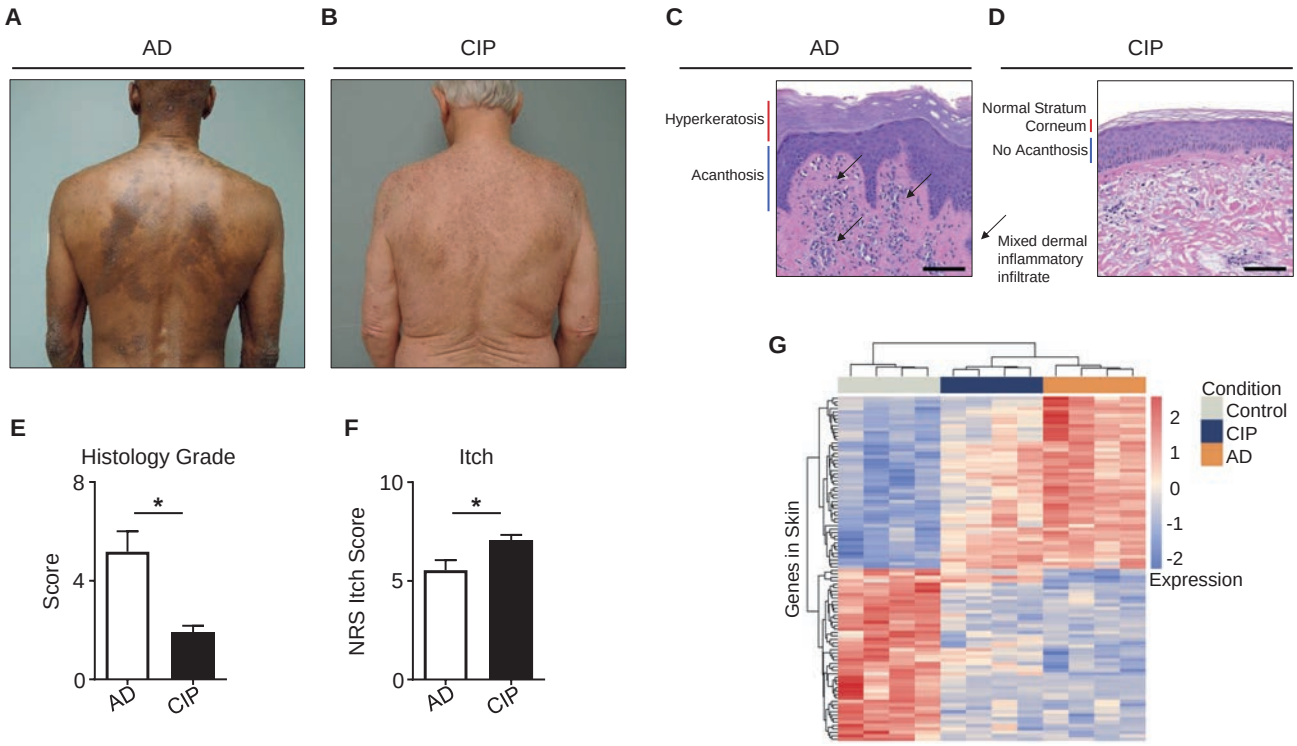


Figure 5. CIP Is a Distinct Chronic Itch Disorder that Exhibits Severe Itch Despite Minimal Skin Inflammation

(A and B) Representative clinical pictures of (A) AD and (B) CIP.

(C and D) Representative H&E histopathology of pruritic sites from (C) AD and (D) CIP.

(E) Histology score of AD and CIP patient biopsies, n $\geq$ 4 biopsies per group.

(F) Numerical rating scale (NRS) itch scores of AD and CIP patients, n $\geq$ 22 patients per group.

(G) Clustering of AD, CIP, and control skin samples by row *Z* scores of the regularized logarithm of gene expression values of the top 100 differentially expressed genes from RNA-seq of AD versus control skin, n = 4 donors per condition.

Scale bars indicate 100 μm. Data are represented as mean ± SEM.

See also Tables S1, S2, and S3.

wild-type (WT) mice while concurrently treating with ruxolitinib (Figure 4H) and observed that systemic intraperitoneal (i.p.) delivery of ruxolitinib significantly reduced scratching behavior (Figure 4I). Surprisingly, skin thickening (Figure 4J) and histologic parameters of AD-like disease (Figures 4K and 4L) demonstrated only a slight reduction in the setting of systemic ruxolitinib treatment, despite its well-established anti-inflammatory properties. Thus, we speculated that the anti-itch effects of ruxolitinib may be mediated predominantly through direct neuronal JAK inhibition rather than through suppression of skin inflammation. To test whether JAK1 signaling is necessary for itch even in the absence of robust skin inflammation, we employed a model of chronic itch not associated with overt inflammation using twice daily topical treatment with a mixture of acetone, ether, and water (AEW) (Miyamoto et al., 2002; Valtcheva et al., 2015; Wilson et al., 2013b). Strikingly, we found that $JAK1^{\Delta neuron}$ mice treated with AEW scratched significantly less than controls (Figures S6L and S6M). Collectively, these findings demonstrate that neuronal JAK1 is a critical mediator of chronic itch even in the absence of robust skin inflammation. Thus, JAK inhibition may represent a broad therapeutic strategy for chronic itch disorders.

CIP Patients Exhibit Severe Itch in the Absence of Overt Skin Inflammation

CIP is an itch disorder of unknown etiology and, in contrast to AD, does not present with overt skin inflammation. However, CIP is strongly associated with aging and believed to be a manifestation of systemic immune senescence (Patel and Yosipovitch, 2010; Reich et al., 2011). As a result, it has been proposed that CIP patients exhibit a mild type 2 immune profile due to the loss of type 1 immunity (Berger and Steinhoff, 2011). In support of this, we recently showed that subsets of CIP patients exhibit features of systemic type 2 inflammation associated with low-grade peripheral eosinophilia and elevation of immunoglobulin E (IgE) (Xu et al., 2016). Although successful elimination of skin inflammation in AD limits itch symptoms, patients with CIP frequently suffer from severe itch even with potent immunosuppression.

Unlike AD, a characteristic inflammatory skin disease (Figure 5A), CIP presents with grossly normal skin findings (Figure 5B). AD histopathology demonstrates features of skin inflammation including hyperkeratosis, acanthosis, and a mixed dermal inflammatory infiltrate (Figure 5C). In stark contrast, histopathology of CIP exhibits only minimal inflammation upon

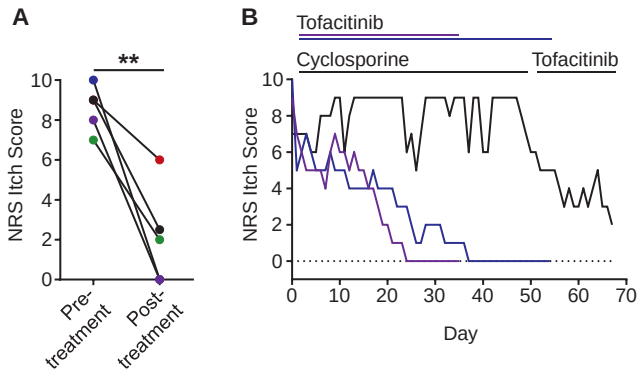


Figure 6. Patients with Refractory CIP Improve when Treated with JAK Inhibition

(A) NRS itch scores for a cohort of CIP patients (n = 5) given the JAK inhibitor tofacitinib. Individual patients are shown by unique colors. Significance was calculated using a paired t test.

(B) Daily NRS itch scores of three CIP patients treated with tofacitinib, including one patient treated with cyclosporine immediately preceding tofacitinib treatment (black).

See also Table S4.

biopsy of the most pruritic cutaneous sites (Figure 5D). Indeed, histologic grading of patient biopsies confirmed that CIP patients exhibit markedly lower levels of skin inflammation compared to AD patients (Figure 5E). Remarkably, despite much milder skin inflammation, CIP patients report more itch than AD patients (Figure 5F). Thus, chronic itch can manifest both in the setting of robust skin inflammation as well as in the absence of notable inflammatory processes.

To characterize CIP at the molecular level, we performed RNA-seq on skin from CIP patients, on lesional skin from AD patients, and on skin from control subjects (Table S1). Clustering of samples by expression of the top 100 most differentially expressed genes between control and AD skin revealed that CIP skin is more closely related to AD skin than control skin (Figure 5G). However, gene set enrichment analysis (GSEA) (Subramanian et al., 2005) comparing CIP and AD skin revealed a number of specific transcriptional programs that distinguish CIP from AD. This analysis showed that AD skin is associated with broad activation of immune responses (Table S2) and that CIP skin is enriched with a variety of transcriptional programs including those for cell morphology and calcium transport (Table S3). Similar studies examining the skin of mice with itch in the setting of minimal inflammation have uncovered homologous transcriptional programs (Wilson et al., 2013b). Collectively, these findings demonstrate that CIP is a chronic itch disorder that shares molecular signatures with AD that distinguish it from control skin despite marked differences in inflammation.

JAK Inhibition Improves Pruritus in Patients with CIP

Recent studies in AD have demonstrated rapid and significant reduction of itch symptoms in response to treatment with JAK inhibitors (Bissonnette et al., 2016; Levy et al., 2015). While these studies sought to investigate JAK inhibition as an anti-inflammatory treatment, our preclinical finding that disruption of neuronal JAK1 signaling limits both inflammatory (Figures 4B and 4C) and non-inflammatory itch (Figures S6L and S6M) suggests that JAK inhibition may represent a novel neuromodulatory approach to target itch. Therefore, despite the absence of overt skin inflammation in CIP, we hypothesized that CIP patients may benefit from JAK blockade.

Because of the lack of approved therapeutics specifically for CIP, we prescribed off-label treatment with the JAK inhibitor tofacitinib to five patients with severe CIP. All of these patients had previously failed multiple other off-label treatments including potent immunosuppression (Table S4). However, one month following treatment with oral tofacitinib, all five patients demonstrated marked improvement in their itch (Figure 6A). Strikingly, patients who closely tracked their itch over time reported rapid onset of itch relief with JAK blockade in contrast to other prior immunosuppressive therapy (Figure 6B). Thus, these findings suggest that JAK inhibition may represent a novel therapeutic strategy for patients with chronic itch disorders that are resistant to conventional anti-inflammatory agents. Taken together, our study reveals that classical cytokine signaling pathways are functional within the sensory nervous system and mediate chronic itch. The discovery of the significance of these pathways may provide new insights into sensory biology and promote the development of new therapies.

DISCUSSION

Chronic itch is an example of how sensory responses that are normally employed by mammals to remove noxious stimuli can become dysregulated and pathologic. While itch is a highly debilitating symptom of multiple distinct medical conditions, chronic itch can also present in the absence of a clearly defined disorder as seen in CIP (Berger and Steinhoff, 2011; Mollanazar et al., 2016; Xu et al., 2016). Although itch is an often overlooked symptom in patient care, clinical studies have established that chronic itch has a profoundly negative impact on quality of life (Kini et al., 2011; Yosipovitch and Bernhard, 2013). However, despite its substantial burden on society, there are currently no specific treatments indicated for chronic itch.

In the setting of AD, the type 2 cytokines IL-4 and IL-13 coordinate an inflammatory response that is critical for disease pathogenesis and the eventual development of debilitating chronic itch. However, the mechanisms by which proinflammatory mediators, such as the type 2 cytokines, elicit itch remain poorly understood. In this study, we provide three conceptual advances that broaden our understanding of chronic itch. First, we demonstrate that activation of the classical immune signaling molecule IL-4Rα directly stimulates both mouse and human sensory neurons. Second, we show that sensory neuron-intrinsic activation of this signaling pathway (IL-4Rα and JAK1) is necessary for chronic itch in inflammatory and non-inflammatory settings. Third, and finally, we show that CIP patients, who are recalcitrant to immunosuppressive therapy, improve when treated with a JAK inhibitor. Taken together, the current study identifies novel functions of immune signaling pathways within the sensory nervous system that represent promising targets for the treatment of chronic itch.

Despite the observation that type 2 cytokines can directly activate itch-sensory neurons *in vitro*, we found that, in contrast to

other cytokine pruritogens, both IL-4 and IL-13 do not function as acute pruritogens *in vivo*. However, in the setting of AD-like skin inflammation where many cytokines are known to be highly upregulated, neuron-specific deletion of IL-4Rα was sufficient to abate chronic itch. To better understand how disruption of neuronal type 2 cytokine signaling could play such a central role in reducing chronic itch, we explored whether neuronal IL-4Rα signaling could alter the responsiveness of sensory neurons to other pruritogens. Indeed, we identified that IL-4 rapidly sensitizes subsets of neurons to a variety of pruritogens. Additionally, we observed that IL-4 significantly amplifies scratching behavior to low doses of known pruritogens like histamine. Thus, we speculate that type 2 cytokines may act as master regulators of chronic itch by intensifying itch responses to multiple pruritogens present in inflamed skin.

In immune cells, cytokine signaling through the JAK-STAT pathway results in transcriptional changes to promote cellular activation. However, although transcriptional alterations remain possible in neurons, our calcium imaging studies indicate that neuronal responses to IL-4Rα stimulation occur within seconds to minutes. Surprisingly, we also observed that these rapid responses were dependent on both JAK1 as well as TRP channels. Thus, we predict that alterations in classical STAT-mediated transcriptional changes alone are insufficient to explain how IL-4Rα modulates sensory neurons. Previous studies have shown that TRP channels harbor several phosphorylation sites and can be regulated by multiple kinases (Sun and Dong, 2016; Zheng, 2013). Although JAKs typically phosphorylate STAT proteins, our current study provokes the hypothesis that JAK proteins may have novel functions in neurons and may regulate TRP channels by unknown pathways. However, we note that such rapid alteration of sensory neuron function does not exclude the role of type 2 cytokines in modulating transcription or otherwise altering expression of pruriceptors or ion channels within neurons. Future studies will be required to better understand how both rapid sensitization and transcriptional changes induced by type 2 cytokines impact sensory function.

Classically, the dominant source of type 2 cytokines in AD lesional skin was believed to be the adaptive immune system via T helper type 2 (Th2) cells (Weidinger and Novak, 2016). Further, recent studies have shown that Th2 cell-derived IL-31 can elicit scratching (Cevikbas et al., 2014; Dillon et al., 2004; Sonkoly et al., 2006). Thus, both AD-associated inflammation and itch were largely believed to be mediated by the adaptive immune system. However, we and others have recently shown that innate immune cell populations, such as group 2 innate lymphoid cells (ILC2s) and basophils, are critical mediators of skin inflammation independent of the adaptive immune system and are essential sources of type 2 cytokines *in vivo* (Kim et al., 2013, 2014; Roediger et al., 2013; Salimi et al., 2013). Based on our new data demonstrating that neuronal IL-4Rα signaling critically mediates chronic itch, we speculate that innate immune cells, such as ILC2s and basophils, play important roles in promoting AD-associated itch, in addition to adaptive Th2 cells. Identification of the precise contributions of both adaptive and innate immune cells to the development of chronic itch requires further investigation.

Recently, the epithelial cell-derived cytokines TSLP and IL-33 have been shown to be master initiators of type 2 inflammation via their effects on a variety of cells including Th2 cells, ILC2s, and basophils (Hammad and Lambrecht, 2015; Oetjen et al., 2016). These cytokines are believed to rapidly signal barrier breach in order to generate innate immune responses and help prime adaptive immune cells. Strikingly, two recent studies have demonstrated that both TSLP and IL-33 can function as pruritogens (Liu et al., 2016; Wilson et al., 2013a). Additionally, we and others have shown that epithelial cell-derived cytokines are highly expressed in the setting of AD-like disease (Kim et al., 2013; Li et al., 2006; Salimi et al., 2013). Despite high expression of these cytokines and robust activation of numerous other downstream cytokine pathways in AD, our study identifies that sensory neuron-specific deletion of IL-4Rα or JAK1 is sufficient to abate chronic itch. Thus, it appears that both neuronal IL-4Rα and JAK1 signaling are exquisitely robust targets for chronic itch.

Clinical trials for dupilumab, an anti-IL-4Rα monoclonal antibody, have repeatedly demonstrated rapid and remarkable improvement of itch symptoms in AD patients (Beck et al., 2014; Simpson et al., 2016; Thaçi et al., 2016). Additionally, significant improvement of itch in AD patients has been observed in early clinical studies employing JAK inhibitors (Bissonnette et al., 2016; Levy et al., 2015). Previously, the changes in itch observed with these therapies have been attributed to the anti-inflammatory properties of either IL-4Rα or JAK blockade. However, our study demonstrates that sensory neurons co-opt classical immune signaling pathways to mediate chronic itch in multiple settings. Thus, we speculate that the rapid improvement of pruritus observed in patients treated with IL-4Rα or JAK blockade may be mediated, at least in part, by disruption of these signals in sensory neurons and that these therapies may abate chronic itch in diseases beyond overt inflammatory skin disorders. Strikingly, a recently published clinical trial for nemolizumab, an anti-IL-31RA monoclonal antibody, also described itch improvement in patients with moderate-to-severe AD (Ruzicka et al., 2017). Given our current study demonstrating a direct role of neuronal IL-4Rα and JAK1 in chronic itch, whether combinations of IL-4Rα-, JAK1-, and IL-31RA-blockade can lead to synergistic therapeutic improvements of itch in patients demands deeper investigation. However, larger prospective, randomized controlled trials will be required to fully determine the efficacy of such treatments.

Immunity at barrier surfaces is believed to have evolved to mediate rapid expulsion of pathogens and potential toxins by employing the specialized functions of the epithelial, immune, and nervous systems. Whereas the epithelium was previously thought to be solely a physical barrier, recent advances have highlighted its direct role in a variety of immune functions, demonstrating that crosstalk between multiple cellular systems is essential for proper host defense (Artis and Spits, 2015; Hammad and Lambrecht, 2015). In addition to the immune system and the epithelial barrier itself, the mammalian host also employs behavioral responses initiated by the sensory nervous system, such as coughing and scratching, to mechanically remove noxious stimuli. Our current study demonstrates that classical immune mediators are an essential part of sensory circuits that modulate host behavior by activating and sensitizing the sensory nervous system. Given that the innate immune system

employs type 2 cytokines to amplify immune responses at barrier surfaces, we speculate that the sensory nervous system may have evolved the ability to use these signals to similarly intensify sensory responses. Thus, direct neuronal priming by type 2 cytokines may be a key first step in the host response to promote protective behavioral responses like scratching. By identifying previously unrecognized neuroimmunologic pathways, we have uncovered novel therapeutic approaches to combat conditions in which these sensory processes become pathologic as in chronic itch.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Ethics statement
  - Animal studies
  - Human studies
  - Primary neuron cultures
- METHOD DETAILS
  - RNA isolation from mouse DRG and TG and qRT-PCR
  - RNA isolation from human DRG and RT-PCR
  - Overnight mouse DRG neuron culture
  - Calcium imaging of mouse DRG neurons
  - Human DRG neuron culture and calcium imaging
  - Scratching behavior assessment
  - MC903 treatment and pathology assessment
  - Mouse and human skin RNA extraction
  - Library preparation and RNA-seq
  - Differential gene expression and gene set enrichment analysis
  - Flow cytometry
  - Immunofluorescence of DRG neurons
  - Rotarod testing
  - Pharmacologic JAK inhibition
  - Mouse model of chronic itch without robust inflammation
  - Human histopathology
  - Numerical rating scale (NRS) itch scoring
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.006.

AUTHOR CONTRIBUTIONS

M.R.M., J.F., T.M.W., H.N., C.J.G., S.C., A.M.T., A.Z.X, S.V.T., J.L., X.G., L.Y., S.L.H., P.L.W., M.L.C., R.B., A.S., and M.J.M. performed the experiments and analyzed the data. R.W.G. and S.D. provided data from human cadaveric dorsal root ganglia. J.R.B., C.-S.H., Z.-F.C., H.H., and Q.L. helped design the study and interpret results. F.B. generated the IL-4Rα^{flox} mice. L.K.O. and B.S.K. performed the experiments, designed the study, interpreted the results, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank all our collaborators at Washington University School of Medicine (WUSM) for their advice and discussion. We also thank members of the Division of Dermatology at WUSM for their assistance with patient sample collection. Specifically, we thank Ms. Mary Tabacchi, Ms. Nancy Bodet, and Dr. Lynn Cornelius for their support. This publication was supported, in part, by a grant from LEO Pharma and by the WUSM Institute of Clinical and Translational Sciences (ICTS), which is supported by the NIH National Center for Advancing Translational Sciences (NCATS) Clinical and Translational Science Award (CTSA) (UL1TR000448). We also thank the Genome Technology Access Center (GTAC) at WUSM for assistance with RNA sequencing. GTAC is partially supported by an NCI grant (P30CA91842) and by an ICTS CTSA grant (UL1TR000448) from the National Center for Research Resources (NCRR). We also thank the WUSM Alafi Neuroimaging Laboratory, supported by an NCRR Shared Instrumentation grant (1S10RR027552). We thank the WUSM Digestive Disease Research Core Center for histology assistance, supported by an NIDDK grant (P30DK052574). We also thank the Immunomonitoring Laboratory (IML) at the Bursky Center for Human Immunology & Immunotherapy Programs (ChiiPs) for flow cytometry support. Research using human DRG in the Gereau Lab is supported by an NINDS grant (R01NS042595). The Davidson Lab is supported by an NIAMS grant (R21AR068012). The Hu Lab is supported by an NIGMS grant (R01GM101218) and an NIDDK grant (R01DK103901). The Kim Lab is supported by NIAMS grants (K08AR065577 and R01AR070116), the American Skin Association, and the Doris Duke Charitable Foundation. L.K.O. is supported by an NHLBI grant (T32HL007317). A.Z.X. is supported by the Howard Hughes Medical Institute (HHMI) Medical Fellows Program. IL-4Rα^{flox} mice were kindly provided by Dr. Herbert (Skip) Virgin at WUSM and Dr. Ajay Chawla at University of California, San Francisco. We also thank Dr. Steven Saenz for critical discussions.

Received: December 4, 2016
Revised: May 18, 2017
Accepted: August 3, 2017
Published: September 7, 2017

REFERENCES

Agarwal, N., Offermanns, S., and Kuner, R. (2004). Conditional gene deletion in primary nociceptive neurons of trigeminal ganglia and dorsal root ganglia. Genesis *38*, 122–129.

Andersen, H.H., Elberling, J., Sølvsten, H., Yosipovitch, G., and Arendt-Nielsen, L. (2017). Non-histaminergic and mechanical itch sensitization in atopic dermatitis. Pain *158*, 1780–1791.

Artis, D., and Spits, H. (2015). The biology of innate lymphoid cells. Nature *517*, 293–301.

Bautista, D.M., Wilson, S.R., and Hoon, M.A. (2014). Why we scratch an itch: the molecules, cells and circuits of itch. Nat. Neurosci. *17*, 175–182.

Beck, L.A., Thaçi, D., Hamilton, J.D., Graham, N.M., Bieber, T., Rocklin, R., Ming, J.E., Ren, H., Kao, R., Simpson, E., et al. (2014). Dupilumab treatment in adults with moderate-to-severe atopic dermatitis. N. Engl. J. Med. *371*, 130–139.

Berger, T.G., and Steinhoff, M. (2011). Pruritus in elderly patients–eruptions of senescence. Semin. Cutan. Med. Surg. *30*, 113–117.

Bissonnette, R., Papp, K.A., Poulin, Y., Gooderham, M., Raman, M., Mallbris, L., Wang, C., Purohit, V., Mamolo, C., Papacharalambous, J., and Ports, W.C. (2016). Topical tofacitinib for atopic dermatitis: a phase IIa randomized trial. Br. J. Dermatol. *175*, 902–911.

Cevikbas, F., Wang, X., Akiyama, T., Kempkes, C., Savinko, T., Antal, A., Kukova, G., Buhl, T., Ikoma, A., Buddenkotte, J., et al. (2014). A sensory neuron-expressed IL-31 receptor mediates T helper cell-dependent itch: Involvement of TRPV1 and TRPA1. J. Allergy Clin. Immunol. *133*, 448–460.

Chan, L.S., Robinson, N., and Xu, L. (2001). Expression of interleukin-4 in the epidermis of transgenic mice results in a pruritic inflammatory skin disease: an experimental animal model to study atopic dermatitis. J. Invest. Dermatol. *117*, 977–983.

Chiu, I.M., Barrett, L.B., Williams, E.K., Strochlic, D.E., Lee, S., Weyer, A.D., Lou, S., Bryman, G.S., Roberson, D.P., Ghasemlou, N., et al. (2014). Transcriptional profiling at whole population and single cell levels reveals somatosensory neuron molecular diversity. eLife *3*, 3.

Dillon, S.R., Sprecher, C., Hammond, A., Bilsborough, J., Rosenfeld-Franklin, M., Presnell, S.R., Haugen, H.S., Maurer, M., Harder, B., Johnston, J., et al. (2004). Interleukin 31, a cytokine produced by activated T cells, induces dermatitis in mice. Nat. Immunol. *5*, 752–760.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Eichenfield, L.F., Tom, W.L., Chamlin, S.L., Feldman, S.R., Hanifin, J.M., Simpson, E.L., Berger, T.G., Bergman, J.N., Cohen, D.E., Cooper, K.D., et al. (2014). Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. J. Am. Acad. Dermatol. *70*, 338–351.

Gabanyi, I., Muller, P.A., Feighery, L., Oliveira, T.Y., Costa-Pinto, F.A., and Mucida, D. (2016). Neuro-immune interactions drive tissue programming in intestinal macrophages. Cell *164*, 378–391.

Gittler, J.K., Shemer, A., Suárez-Fariñas, M., Fuentes-Duculan, J., Gulewicz, K.J., Wang, C.Q.F., Mitsui, H., Cardinale, I., de Guzman Strong, C., Krueger, J.G., and Guttman-Yassky, E. (2012). Progressive activation of T(H)2/T(H)22 cytokines and selective epidermal proteins characterizes acute and chronic atopic dermatitis. J. Allergy Clin. Immunol. *130*, 1344–1354.

Halim, T.Y.F., Hwang, Y.Y., Scanlon, S.T., Zaghouani, H., Garbi, N., Fallon, P.G., and McKenzie, A.N.J. (2016). Group 2 innate lymphoid cells license dendritic cells to potentiate memory TH2 cell responses. Nat. Immunol. *17*, 57–64.

Hammad, H., and Lambrecht, B.N. (2015). Barrier epithelial cells and the control of type 2 immunity. Immunity *43*, 29–40.

Ikoma, A., Rukwied, R., Ständer, S., Steinhoff, M., Miyachi, Y., and Schmelz, M. (2003). Neuronal sensitization for histamine-induced itch in lesional skin of patients with atopic dermatitis. Arch. Dermatol. *139*, 1455–1458.

Kaplan, A.P. (2004). Chronic urticaria: pathogenesis and treatment. J. Allergy Clin. Immunol. *114*, 465–474.

Kashem, S.W., Riedl, M.S., Yao, C., Honda, C.N., Vulchanova, L., and Kaplan, D.H. (2015). Nociceptive sensory fibers drive Interleukin-23 production from CD301b+ dermal dendritic cells and drive protective cutaneous immunity. Immunity *43*, 515–526.

Kelly-Welch, A.E., Hanson, E.M., Boothby, M.R., and Keegan, A.D. (2003). Interleukin-4 and interleukin-13 signaling connections maps. Science *300*, 1527–1528.

Kim, B.S., Siracusa, M.C., Saenz, S.A., Noti, M., Monticelli, L.A., Sonnenberg, G.F., Hepworth, M.R., Voorhees, A.S.V., Comeau, M.R., and Artis, D. (2013). TSLP elicits IL-33–independent innate lymphoid cell responses to promote skin inflammation. Sci. Transl. Med. *5*, 170ra16.

Kim, B.S., Wang, K., Siracusa, M.C., Saenz, S.A., Brestoff, J.R., Monticelli, L.A., Noti, M., Tait Wojno, E.D., Fung, T.C., Kubo, M., and Artis, D. (2014). Basophils promote innate lymphoid cell responses in inflamed skin. J. Immunol. *193*, 3717–3725.

Kini, S.P., DeLong, L.K., Veledar, E., McKenzie-Brown, A.M., Schaufele, M., and Chen, S.C. (2011). The impact of pruritus on quality of life: the skin equivalent of pain. Arch. Dermatol. *147*, 1153–1156.

Levy, L.L., Urban, J., and King, B.A. (2015). Treatment of recalcitrant atopic dermatitis with the oral Janus kinase inhibitor tofacitinib citrate. J. Am. Acad. Dermatol. *73*, 395–399.

Li, M., Hener, P., Zhang, Z., Kato, S., Metzger, D., and Chambon, P. (2006). Topical vitamin D3 and low-calcemic analogs induce thymic stromal lymphopoietin in mouse keratinocytes and trigger an atopic dermatitis. Proc. Natl. Acad. Sci. USA *103*, 11736–11741.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liu, Q., Tang, Z., Surdenikova, L., Kim, S., Patel, K.N., Kim, A., Ru, F., Guan, Y., Weng, H.-J., Geng, Y., et al. (2009). Sensory neuron-specific GPCR Mrgprs are itch receptors mediating chloroquine-induced pruritus. Cell *139*, 1353–1365.

Liu, B., Tai, Y., Achanta, S., Kaelberer, M.M., Caceres, A.I., Shao, X., Fang, J., and Jordt, S.-E. (2016). IL-33/ST2 signaling excites sensory neurons and mediates itch response in a mouse model of poison ivy contact allergy. Proc. Natl. Acad. Sci. USA *113*, E7572–E7579.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Malin, S.A., Davis, B.M., and Molliver, D.C. (2007). Production of dissociated sensory neuron cultures and considerations for their use in studying neuronal function and plasticity. Nat. Protoc. *2*, 152–160.

Martinez-Gonzalez, I., Mathä, L., Steer, C.A., Ghaedi, M., Poon, G.F.T., and Takei, F. (2016). Allergen-experienced group 2 innate lymphoid cells acquire memory-like properties and enhance allergic lung inflammation. Immunity *45*, 198–208.

Matterne, U., Apfelbacher, C.J., Loerbroks, A., Schwarzer, T., Büttner, M., Ofenloch, R., Diepgen, T.L., and Weisshaar, E. (2011). Prevalence, correlates and characteristics of chronic pruritus: a population-based cross-sectional study. Acta Derm. Venereol. *91*, 674–679.

Miyamoto, T., Nojima, H., Shinkado, T., Nakahashi, T., and Kuraishi, Y. (2002). Itch-associated response induced by experimental dry skin in mice. Jpn. J. Pharmacol. *88*, 285–292.

Mollanazar, N.K., Sethi, M., Rodriguez, R.V., Nattkemper, L.A., Ramsey, F.V., Zhao, H., and Yosipovitch, G. (2016). Retrospective analysis of data from an itch center: Integrating validated tools in the electronic health record. J. Am. Acad. Dermatol. *75*, 842–844.

Morita, T., McClain, S.P., Batia, L.M., Pellegrino, M., Wilson, S.R., Kienzler, M.A., Lyman, K., Olsen, A.S.B., Wong, J.F., Stucky, C.L., et al. (2015). HTR7 mediates serotonergic acute and chronic itch. Neuron *87*, 124–138.

Oetjen, L.K., Noti, M., and Kim, B.S. (2016). New insights into basophil heterogeneity. Semin. Immunopathol. *38*, 549–561.

Ordovas-Montanes, J., Rakoff-Nahoum, S., Huang, S., Riol-Blanco, L., Barreiro, O., and von Andrian, U.H. (2015). The regulation of immunological processes by peripheral neurons in homeostasis and disease. Trends Immunol. *36*, 578–604.

Patel, T., and Yosipovitch, G. (2010). The management of chronic pruritus in the elderly. Skin Therapy Lett. *15*, 5–9.

Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat. Biotechnol. *32*, 462–464.

Reich, A., Ständer, S., and Szepietowski, J.C. (2011). Pruritus in the elderly. Clin. Dermatol. *29*, 15–23.

Riol-Blanco, L., Ordovas-Montanes, J., Perro, M., Naval, E., Thiriot, A., Alvarez, D., Paust, S., Wood, J.N., and von Andrian, U.H. (2014). Nociceptive sensory neurons drive interleukin-23-mediated psoriasiform skin inflammation. Nature *510*, 157–161.

Roediger, B., Kyle, R., Yip, K.H., Sumaria, N., Guy, T.V., Kim, B.S., Mitchell, A.J., Tay, S.S., Jain, R., Forbes-Blom, E., et al. (2013). Cutaneous immunosurveillance and regulation of inflammation by group 2 innate lymphoid cells. Nat. Immunol. *14*, 564–573.

Ruzicka, T., Hanifin, J.M., Furue, M., Pulka, G., Mlynarczyk, I., Wollenberg, A., Galus, R., Etoh, T., Mihara, R., Yoshida, H., et al.; XCIMA Study Group (2017). Anti-interleukin-31 receptor A antibody for atopic dermatitis. N. Engl. J. Med. *376*, 826–835.

Salimi, M., Barlow, J.L., Saunders, S.P., Xue, L., Gutowska-Owsiak, D., Wang, X., Huang, L.-C., Johnson, D., Scanlon, S.T., McKenzie, A.N.J., et al. (2013). A role for IL-25 and IL-33-driven type-2 innate lymphoid cells in atopic dermatitis. J. Exp. Med. *210*, 2939–2950.

Schwartz, D.M., Bonelli, M., Gadina, M., and O'Shea, J.J. (2016). Type I/II cytokines, JAKs, and new strategies for treating autoimmune diseases. Nat. Rev. Rheumatol. *12*, 25–36.

Simpson, E.L., Bieber, T., Guttman-Yassky, E., Beck, L.A., Blauvelt, A., Cork, M.J., Silverberg, J.I., Deleuran, M., Kataoka, Y., Lacour, J.-P., et al.; SOLO 1 and SOLO 2 Investigators (2016). Two phase 3 trials of dupilumab versus placebo in atopic dermatitis. N. Engl. J. Med. *375*, 2335–2348.

Sonkoly, E., Muller, A., Lauerma, A.I., Pivarcsi, A., Soto, H., Kemeny, L., Alenius, H., Dieu-Nosjean, M.-C., Meller, S., Rieker, J., et al. (2006). IL-31: a new link between T cells and pruritus in atopic skin inflammation. J. Allergy Clin. Immunol. *117*, 411–417.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sun, S., and Dong, X. (2016). Trp channels and itch. Semin. Immunopathol. *38*, 293–307.

Talbot, S., Abdulnour, R.-E.E., Burkett, P.R., Lee, S., Cronin, S.J.F., Pascal, M.A., Laedermann, C., Foster, S.L., Tran, J.V., Lai, N., et al. (2015). Silencing nociceptor neurons reduces allergic airway inflammation. Neuron *87*, 341–354.

Thaçi, D., Simpson, E.L., Beck, L.A., Bieber, T., Blauvelt, A., Papp, K., Soong, W., Worm, M., Szepietowski, J.C., Sofen, H., et al. (2016). Efficacy and safety of dupilumab in adults with moderate-to-severe atopic dermatitis inadequately controlled by topical treatments: a randomised, placebo-controlled, dose-ranging phase 2b trial. Lancet *387*, 40–52.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.

Valtcheva, M.V., Samineni, V.K., Golden, J.P., Gereau, R.W., 4th, and Davidson, S. (2015). Enhanced nonpeptidergic intraepidermal fiber density and an expanded subset of chloroquine-responsive trigeminal neurons in a mouse model of dry skin itch. J. Pain *16*, 346–356.

Valtcheva, M.V., Copits, B.A., Davidson, S., Sheahan, T.D., Pullen, M.Y., McCall, J.G., Dikranian, K., and Gereau, R.W., 4th. (2016). Surgical extraction of human dorsal root ganglia from organ donors and preparation of primary sensory neuron cultures. Nat. Protoc. *11*, 1877–1888.

Veiga-Fernandes, H., and Mucida, D. (2016). Neuro-immune interactions at barrier surfaces. Cell *165*, 801–811.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Weidinger, S., and Novak, N. (2016). Atopic dermatitis. Lancet *387*, 1109–1122.

Wilson, S.R., Thé, L., Batia, L.M., Beattie, K., Katibah, G.E., McClain, S.P., Pellegrino, M., Estandian, D.M., and Bautista, D.M. (2013a). The epithelial cell-derived atopic dermatitis cytokine TSLP activates neurons to induce itch. Cell *155*, 285–295.

Wilson, S.R., Nelson, A.M., Batia, L., Morita, T., Estandian, D., Owens, D.M., Lumpkin, E.A., and Bautista, D.M. (2013b). The ion channel TRPA1 is required for chronic itch. J. Neurosci. *33*, 9283–9294.

Xu, A.Z., Tripathi, S.V., Kau, A.L., Schaffer, A., and Kim, B.S. (2016). Immune dysregulation underlies a subset of patients with chronic idiopathic pruritus. J. Am. Acad. Dermatol. *74*, 1017–1020.

Yosipovitch, G., and Bernhard, J.D. (2013). Clinical practice. Chronic pruritus. N. Engl. J. Med. *368*, 1625–1634.

Zhao, Z.-Q., Huo, F.-Q., Jeffry, J., Hampton, L., Demehri, S., Kim, S., Liu, X.-Y., Barry, D.M., Wan, L., Liu, Z.-C., et al. (2013). Chronic itch development in sensory neurons requires BRAF signaling pathways. J. Clin. Invest. *123*, 4769–4780.

Zheng, J. (2013). Molecular mechanism of TRP channels. Compr. Physiol. *3*, 221–242.

Zheng, T., Oh, M.H., Oh, S.Y., Schroeder, J.T., Glick, A.B., and Zhu, Z. (2009). Transgenic expression of interleukin-13 in the skin induces a pruritic dermatitis and skin remodeling. J. Invest. Dermatol. *129*, 742–751.

Article

Cell

A Human Bi-specific Antibody against Zika Virus with High Therapeutic Potential

Jiaqi Wang,[1,2,11] Marco Bardelli,[3,11] Diego A. Espinosa,[4] Mattia Pedotti,[3] Thiam-Seng Ng,[1,2] Siro Bianchi,[5] Luca Simonelli,[3] Elisa X.Y. Lim,[1,2] Mathilde Foglierini,[3] Fabrizia Zatta,[5] Stefano Jaconi,[5] Martina Beltramello,[5] Elisabetta Cameroni,[5] Guntur Fibriansah,[1,2] Jian Shi,[2,6] Taylor Barca,[4] Isabel Pagani,[7] Alicia Rubio,[7] Vania Broccoli,[7,8] Elisa Vicenzi,[7] Victoria Graham,[9] Steven Pullan,[9] Stuart Dowall,[9] Roger Hewson,[9] Simon Jurt,[10] Oliver Zerbe,[10] Karin Stettler,[5] Antonio Lanzavecchia,[3] Federica Sallusto,[3] Andrea Cavalli,[3] Eva Harris,[4] Shee-Mei Lok,[1,2,12,*] Luca Varani,[3,12,*] and Davide Corti[5,12,13,*]

[1]Program in Emerging Infectious Diseases, Duke–National University of Singapore Medical School, Singapore 169857, Singapore
[2]Centre for BioImaging Sciences, National University of Singapore, Singapore 117557, Singapore
[3]Insitute for Research in Biomedicine, Università della Svizzera italiana, Via Vincenzo Vela 6, 6500 Bellinzona, Switzerland
[4]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, 185 Li Ka Shing Center, 1951 Oxford Street, Berkeley, California, 94720-3370, USA
[5]Humabs BioMed SA a subsidiary of Vir Biotechnology, Inc., Via Mirasole 1, 6500 Bellinzona, Switzerland
[6]CryoEM unit, Department of Biological Sciences, National University of Singapore, Singapore 117557
[7]Viral Pathogens and Biosafety Unit, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy
[8]CNR-Institute of Neuroscience, Via Vanvitelli 32, 20129, Milan, Italy
[9]National Infection Service, Public Health England, Porton Down, Salisbury, Wiltshire, UK
[10]Department of Chemistry, University of Zurich, Zurich, Switzerland
[11]These authors contributed equally
[12]These authors contributed equally
[13]Lead Contact
*Correspondence: sheemei.lok@duke-nus.edu.sg (S.-M.L.), luca.varani@irb.usi.ch (L.V.), davide.corti@humabs.ch (D.C.)
http://dx.doi.org/10.1016/j.cell.2017.09.002

SUMMARY

Zika virus (ZIKV), a mosquito-borne flavivirus, causes devastating congenital birth defects. We isolated a human monoclonal antibody (mAb), ZKA190, that potently cross-neutralizes multi-lineage ZIKV strains. ZKA190 is highly effective in vivo in preventing morbidity and mortality of ZIKV-infected mice. NMR and cryo-electron microscopy show its binding to an exposed epitope on DIII of the E protein. ZKA190 Fab binds all 180 E protein copies, altering the virus quaternary arrangement and surface curvature. However, ZIKV escape mutants emerged in vitro and in vivo in the presence of ZKA190, as well as of other neutralizing mAbs. To counter this problem, we developed a bispecific antibody (FIT-1) comprising ZKA190 and a second mAb specific for DII of E protein. In addition to retaining high in vitro and in vivo potencies, FIT-1 robustly prevented viral escape, warranting its development as a ZIKV immunotherapy.

INTRODUCTION

Effective therapeutics and vaccines against Zika virus (ZIKV) are urgently needed. Although ZIKV generally causes asymptomatic or mild disease, it can lead to neurological complications both in adulthood and during fetal development (Honein et al., 2017), likely due to infection of neural progenitor cells (Frontera and da Silva, 2016; Mlakar et al., 2016). Further, infection of the testes has been shown to cause infertility in mice (Govero et al., 2016).

Cryo-electron microscopy (cryoEM) of ZIKV revealed high structural similarity to other flaviviruses (Kostyuchenko et al., 2016; Sirohi et al., 2016). The smooth-surfaced mature ZIKV particle comprises 180 copies of E protein organized in icosahedral symmetry, each asymmetric unit containing three E proteins. Three E protein dimers lie parallel to each other building a raft, and thirty rafts form a herringbone pattern. The E protein consists of three domains – DI, DII, and DIII (Barba-Spaeth et al., 2016; Kostyuchenko et al., 2016; Rey et al., 1995; Zhao et al., 2016). DIII contains the receptor-binding site and mediates virus attachment (Crill and Roehrig, 2001), whereas the fusion loop in DII facilitates virus-endosomal membrane fusion (Allison et al., 2001). After cell entry, the low endosomal pH triggers dramatic structural rearrangement of E protein dimers into trimers, leading to membrane fusion (Bressanelli et al., 2004; Modis et al., 2004; Stiasny et al., 1996). The post-fusion structure shows DIII rotated ~70° relative to DI, which requires movement of the DI-DIII linker (Bressanelli et al., 2004; Modis et al., 2004).

We recently isolated a panel of anti-ZIKV neutralizing human mAbs and identified a DIII-binding mAb, ZKA190, that showed strong neutralizing activity across ZIKV strains (Stettler et al., 2016). Here, we further characterized ZKA190 in vitro and in vivo. Using a combination of NMR, cryoEM and molecular simulations, we mapped the ZKA190 epitope to an exposed region consisting of the DI-DIII linker and the lateral ridge (LR) region of DIII. Different from other DIII antibodies, ZKA190 Fab binds to all 180 E proteins on the virus surface, altering the E protein quaternary interactions, causing separation of the rafts and flattening

their curvature. In vivo studies of ZIKV infection in lethal mouse models demonstrated strong protection of ZKA190 from multiple strains. To minimize the risks of viral escape and antibody-mediated disease enhancement, we engineered a bispecific antibody, which combines ZKA190 with the DII-binding mAb ZKA185 and also includes a mutation in the Fc region (LALA) abolishing binding to Fcγ receptor and complement (Hessell et al., 2007). FIT-1 retains the breadth and potencies of both parental antibodies and was not observed to generate escape mutants, encouraging FIT-1 development as a ZIKV therapeutic.

RESULTS

In Vitro Characterization of the DIII-Specific mAb ZKA190

Recombinant ZKA190 synthesized in the IgG1 wild-type (WT) and IgG Fc-LALA format potently neutralized African, Asian, and American ZIKV strains, with IC_{50} ranging from 0.004 to 0.05 nM (Figure 1A; 0.6 to 8 ng/mL). Although both the ZKA190 IgG and its Fab bound to E protein with similar affinities (Figure 1C; K_D values of 0.3 and 0.2 nM, respectively), the Fab neutralized ZIKV infection less potently with a 300-fold higher IC_{50} (Figure 1B). Since ZIKV has been shown to infect human neural progenitor cells (hNPC) (Garcez et al., 2016; Nowakowski et al., 2016; Onorati et al., 2016; Tang et al., 2016), we tested ZKA190 WT and LALA, and both fully abolished infection and replication of ZIKV in hNPCs (Figure 1D).

Antibody-dependent enhancement (ADE) of infection may contribute to disease severity in heterologous secondary DENV infections (Halstead, 2014), caused by the binding of non-neutralizing or sub-neutralizing concentrations of E- and prM-specific antibodies to mature, partially mature and fully immature viruses. ZKA190 supports ADE in K562 cells from 0.0001 to 1 nM, while ZKA190-LALA did not show any ADE activity (Figure 1E) and completely inhibited ADE induced by plasma antibodies from four ZIKV-immune donors (Figure 1F).

Anti-prM antibodies are elicited during the human immune response against flaviviruses, and they enhance in vitro infection of immature or partially immature virions (Dejnirattisai et al., 2010). ZKA190-LALA can fully block ADE of ZIKV infection induced by a flavivirus cross-reactive anti-prM mAb, DV62, isolated from a dengue patient (Beltramello et al., 2010) (Figure 1G). Finally, we compared the ability of ZKA190 WT, LALA and Fab to cause or block ADE of ZIKV by human anti-DENV2 plasma or DV62 mAb (Figure 1H). Low concentrations of ZKA190 increased DV62-mediated ADE of ZIKV infection, consistent with its ability to promote the entry of both immature and mature virions, while at concentrations above 1.3 nM (i.e., 200 ng/mL), ZKA190 blocked ADE induced by both anti-DENV plasma and DV62. At concentrations above 0.06 nM, ZKA190 Fab was equally effective as ZKA190-LALA in reducing ADE induced by DV62, indicating that inhibition occurs at a post-attachment step, likely fusion.

ZKA190 Binds to a Conserved and Highly Accessible Region of DIII

The ZKA190 epitope was determined by solution NMR spectroscopy (Bardelli et al., 2015; Simonelli et al., 2010; 2013). Differences in the NMR spectra of free and mAb-bound DIII identified DIII residues affected by ZKA190 binding, either directly or through allosteric effects (Figure S1A). The ZKA190 epitope consists of the LR of DIII (the BC, DE and FG loops), as well as part of the DI-DIII linker (Figure 2A). These residues are conserved among 217 known ZIKV strains, with the exception of V341I and E393D substitutions in the Uganda 1947 MR766 strain (Figure S1D), which ZKA190 was shown to neutralize (Figure 1A). The ZKA190 epitope is highly accessible on the ZIKV surface, except for the FG loop in the 5-fold vertex (Figures 2B and S1C, molecule A).

Computational docking followed by molecular dynamics simulation, guided and validated by both NMR- and cryoEM-derived epitope information as well as DIII mutagenesis, showed that ZKA190 binds through an interface characterized by shape and charge complementarity (Figures 2B and S1E). The computational model fits well to the cryoEM density map of ZIKV-ZKA190 Fab complex (Figure 2C and 3D). Docking suggests that there are no direct contacts between ZKA190 and the FG loop on DIII, thus changes in NMR signals likely derive from allosteric effects. Consistently, mutations of FG loop residues, but not of other NMR-affected residues in the recombinant DIII, did not affect the binding affinity of ZKA190 (Figures 2B and S2).

CryoEM Structure of ZKA190 Complexed with ZIKV

CryoEM micrographs showed that ZKA190 binds to the virus at 4°C, 37°C and 40°C (Figure S3A). Structures of ZIKV complexed with ZKA190 Fab at 4°C and 37°C were determined to ~22 Å resolution and found to be similar (Figures 2D, S3B and S3C). At this resolution, the Fabs and E protein rafts can only be roughly fitted into the density; the footprint on DIII cannot be accurately evaluated using cryoEM results alone. In contrast to other mAbs against DIII LR of other flaviviruses, Fab ZKA190 bound to all 180 copies of E protein on the viral surface (Figures 2D and S3), achieving full occupancy.

Superposition of the structure of the ZKA190 Fab:E protein complex onto the uncomplexed ZIKV structure showed that Fabs bound to the E protein molecule around the 5-fold vertex clashed with each other (Figure 2G). On the contrary, cryoEM data showed equally strong densities of the variable region of the Fabs at all three individual E proteins in the asymmetric unit (Figures 2F and S4). To accommodate binding of all five Fabs without clashes, the E proteins have to separate from each other in the 5-fold vertex (Figure 2G). Comparison of the cryoEM structures of uncomplexed and ZKA190-bound ZIKV showed that in the latter, the E protein rafts move to a higher radius (Figure 2E) and are separated from each other (Figure 2I); the curvature of the rafts is also flattened (Figure 2H). In the uncomplexed ZIKV cryoEM structure, the E protein ectodomain interacts with the membrane-associated stem regions of E and M proteins on the virus, and this interaction might be disrupted by ZKA190 Fab binding, as suggested by the spatial gap observed between the E protein and the lipid bilayer (Figure 2E). The dissociation may increase the motion of the rafts, explaining the inability to obtain a higher resolution cryoEM map.

Mechanisms of ZKA190 Neutralization

The ability of ZKA190 to efficiently neutralize the virus may involve inhibition of either cell attachment or membrane fusion.

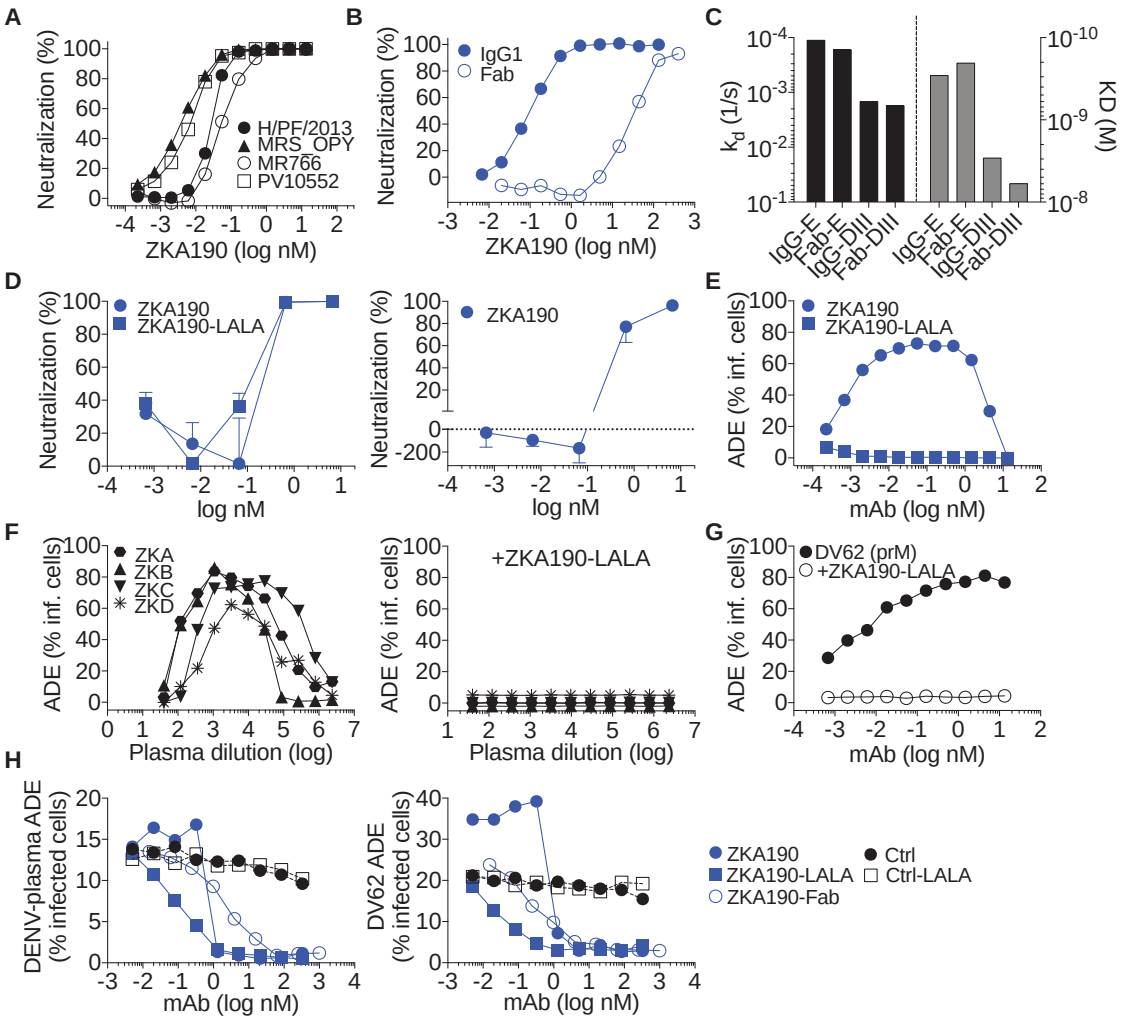


Figure 1. Neutralization and Enhancement of ZIKV Infection by ZKA190 mAb

(A) ZKA190 was tested for neutralization of four strains of ZIKV, as determined by the percentage of infected Vero cells in the presence of increasing amounts of mAbs. Data are representative of at least two independent experiments.

(B) Neutralization of ZIKV H/PF/2013 strain by ZKA190 IgG and Fab determined as in (A).

(C) Binding of ZKA190 IgG and Fab to recombinant ZIKV E and DIII proteins as assessed by SPR.

(D) Neutralization of ZIKV strain PRVABC59 infection of hNPCs by ZKA190, ZKA190-LALA and a control mAb as determined by plaque assay on Vero cells (left panel) and indirect immunofluorescence of infected hNPCs using fluorophore-labeled anti-E antibody (right panel).

(E) ADE of ZIKV infection of non-permissive K562 cells by ZKA190 and ZKA190-LALA.

(F) ADE induced in K562 cells when ZIKV is pre-incubated with serial dilutions of plasma or serum from different ZIKV-positive patients (left panel). When ZKA190 LALA is added to the ZIKV-serum complexes, ADE is inhibited (right panel).

(G) ADE induced in K562 cells when ZIKV is pre-incubated with serial dilutions of a prM cross-reactive mAb (DV62) derived from a DENV-immune donor. ZKA190-LALA inhibits ADE of ZIKV when complexed with prM-reactive mAb DV62.

(H) Effect on ADE induced by peak enhancing dilution of a DENV2 plasma sample (left panel) or anti-prM DV62 mAb (right panel) by serial dilutions of indicated mAbs.

Error bars represent the SD of the mean for each determination.

Another mechanism might involve virus inactivation through aggregation. Virus aggregation due to simultaneously engaging epitopes on different particles by ZKA190 was assessed by dynamic light scattering (DLS). Particles with size similar to single virus particle were detected at low mAb concentration, whereas large viral aggregates appeared at higher concentration ($\geq$33 antibodies per virus) and correlated with neutralizing activity (Figure 3A). One caveat is that the contribution of aggregation to neutralization may be different in vivo, as the amount of virus could differ.

IgG ZKA190 may also engage neighboring E proteins on the viral surface. As shown by cryo-EM, the distance between the Fabs' CH1 regions bound on the viral surface is compatible with that of a full IgG (Figure 3B). However, bivalent binding of a full IgG depends not only on the distance between two Fab molecules, but also on their relative orientation. Full atomistic

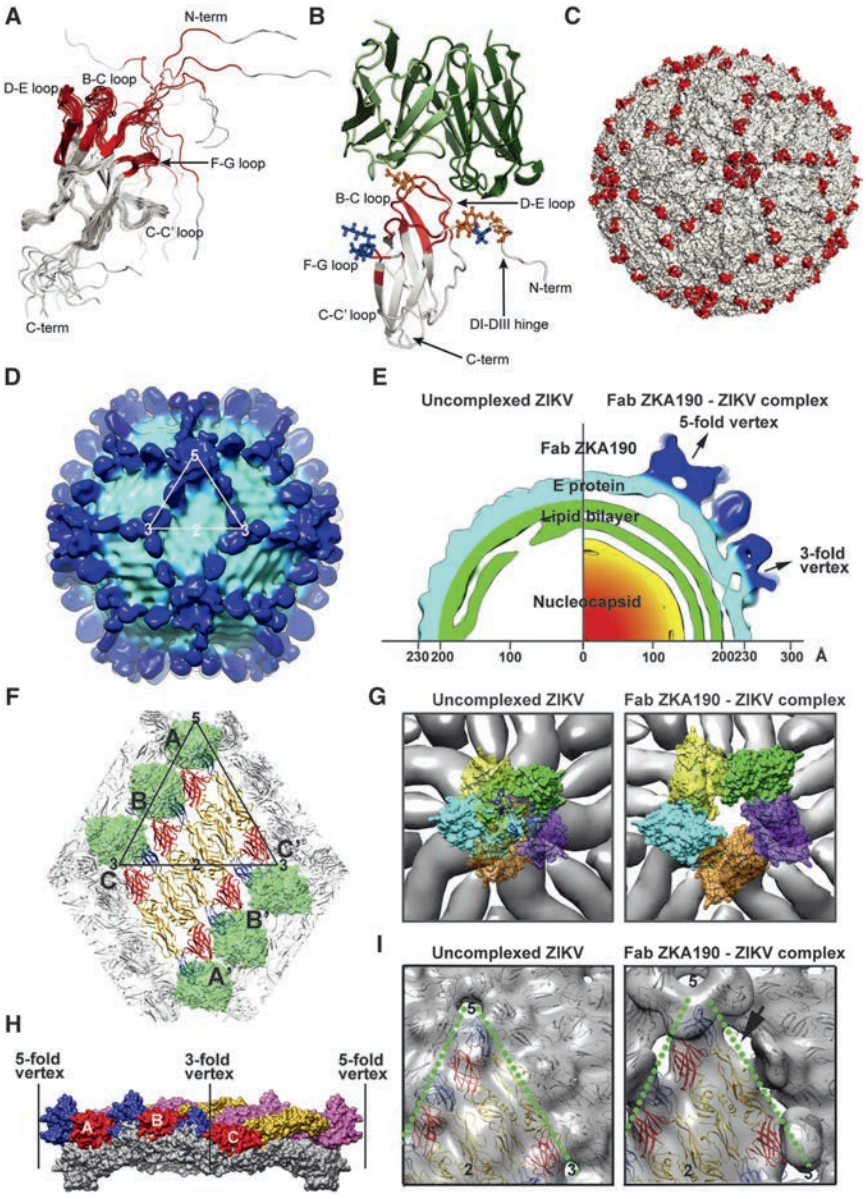


Figure 2. ZKA190 Epitope Identified By NMR, CryoEM, and Docking

(A) Cartoon representation of the 12 lowest-energy NMR structures of ZIKV DIII, with residues affected by ZKA190 binding in red. Flexibility in the N terminus of the construct is apparent.

(B) Model of the ZKA190:DIII complex derived by computational docking and molecular simulation validated by cryoEM and NMR results. The NMR-identified epitope on DIII (gray) is indicated in red. The ZKA190 heavy and light chains are colored in dark and light green, respectively. DIII residues that affect or do not affect antibody binding when mutated are shown as orange and blue sticks, respectively.

(C) NMR-identified ZKA190 epitope (red) is accessible on the virus surface (white).

(D–I) The ~22Å resolution cryoEM map of the ZIKV:ZKA190 Fab complex at 37°C. (D) Surface of the cryoEM map. One icosahedral asymmetric unit is indicated by a white triangle. Light and dark blue correspond to the E protein shell and the Fab densities, respectively. (E) A quarter of the central section of the uncomplexed (left) and complexed (right) cryoEM maps showing that the E protein layer in the latter moved outward with the highest protrusion at the 5-fold vertex. (F) CryoEM structure of ZIKV:ZKA190 Fab complex. The DIIIs of all E proteins are bound by Fabs (green). E proteins in an asymmetric unit are labeled A, B and C, and the neighboring asymmetric unit within a raft A', B' and C'. DI, DII and DIII are colored in red, yellow and blue, respectively. (G) E proteins at the 5-fold vertices move apart upon Fab ZKA190 binding. (Left) Superposition of the Fab ZKA190:E complex onto the uncomplexed ZIKV shows clashes between the Fabs. (Right) In contrast, no clashes are observed in the cryoEM complex structure. E proteins are in gray and Fabs are colored. (H) Superposition of the E protein raft in the free (gray) and ZKA190-complexed (colors) ZIKV; the curvature flattens in the complex structure. E proteins of one asymmetric unit (A, B and C) are in red, yellow and blue; those from another unit are in pink. (I) The rafts separate from each other in the complex (right) compared to the uncomplexed ZIKV structure (left). The inter-raft interface is indicated with a green dotted line, and the gap between rafts is indicated with an arrow.

See also Figure S1, S2, S3 and S4.

molecular dynamics simulations of a homology model of ZKA190 IgG showed that a single ZKA190 molecule can bridge adjacent E proteins around the 5- and 3-fold vertices, but not the 2-fold vertex (Figures 3C and 3D). The Fab regions of the modeled ZKA190 IgG correlated well with the cryo-EM structure (Figure 3D). This suggests that ZKA190 cross-linking of neighboring E proteins might prevent the E dimers from rearrangement into trimers that is required for membrane fusion. Both proposed mechanisms imply bivalent binding of ZKA190 IgG. Since ZKA190 Fab also neutralizes ZIKV, albeit less efficiently than the full IgG, additional mechanisms may be involved. ZKA190 might prevent the attachment of ZIKV to target cells or viral fusion to the endosomal membrane.

Neutralization at a post-attachment step is supported by data showing that ZKA190 inhibits in vitro ZIKV infection with similar potency regardless of whether the mAb was added to the virus pre- or post- attachment to cells (Figure 3E). This result is further supported by the qRT-PCR experiments that detect the amount of virus on the cell surface, which does not decrease at increasing antibody concentrations (Figure 3F), but rather increased compared to the virus-only control, likely due to antibody-mediated virus aggregation (Figure 3A). Moreover, confocal microscopy analysis showed that ZIKV-complexed ZKA190 Fab can enter cells at neutralizing concentrations exceeding the IC_{50} by 10,000-fold (Figure 3G).

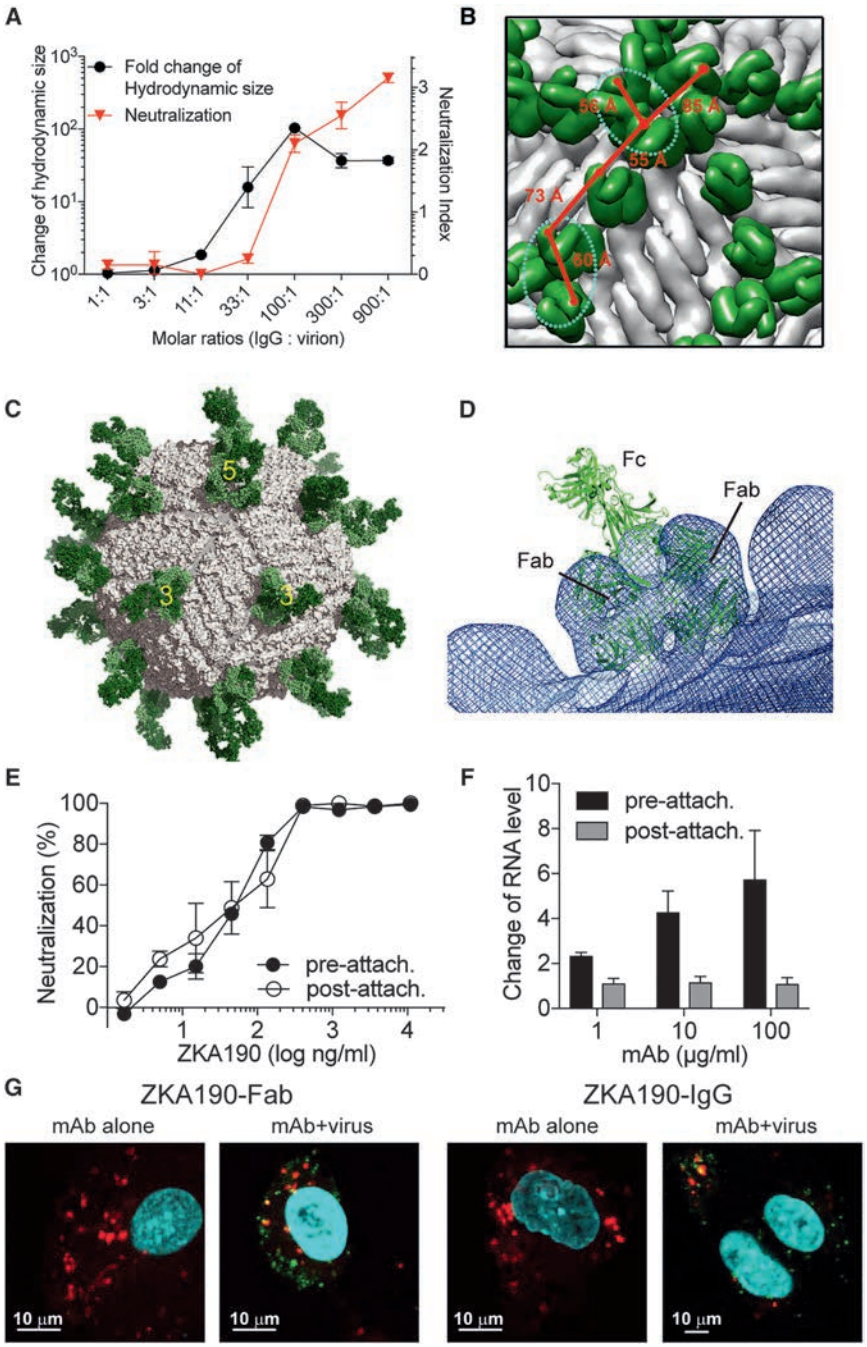


Figure 3. ZKA190 Aggregates Virus at High Antibody:Virus Molar Ratios and Inhibits a Post-Attachment Stage of Virus Infection

(A) Dynamic light scattering experiments measuring the hydrodynamic size of the virus complexes shows ZKA190 IgG aggregates ZIKV at IgG:virion molar ratio of ≥ 33:1. The neutralization activity of antibody correlates with the formation of virus-antibody aggregates. The neutralization is expressed as the $\log_{10}$-transformed fold-reduction of the virus titer compared to the virus-only control.

(B) ZKA190 Fab (green) is shown on the viral surface (gray). The close proximity (< 86Å) between CH1 domains of two Fabs (red lines) in the cryoEM ZIKV:ZKA190 Fab structure indicates possible bivalent binding of IgG on the same virus particle. However, due to the orientation of the Fabs, only the sites (cyan dotted circles) around the 5-fold and the 3-fold vertices can accommodate bivalent binding of a full IgG.

(C) Molecular Dynamics (MD)-based atomic model of the full ZKA190 IgG (green) bridging across two neighboring E proteins on the viral surface (gray). For clarity, only one IgG binding is shown. Bivalent binding to other sites within these vertices is possible.

(D) Superposition of the MD-based full IgG model (green cartoon) on the cryoEM map (blue mesh) near the 3-fold vertex shows good fitting of the modeled Fabs to the cryoEM density.

(E) Neutralization test of ZIKV when ZKA190 is added pre- or post-attachment of virus to cells that were incubated at 4°C. Results show no difference in the neutralization profile.

(F) RT-PCR detecting the amount of virus remaining on cells when ZKA190 was added pre- and post- virus attachment to cells. When increasing concentrations of mAb were mixed with virus before attaching to cells ("pre-attach."), there is a gradual increase in the amount of virus on cell surface, likely due to aggregation of virus. Addition of the mAb to virus that had already bound to cells ("post-attach.") showed that the mAb is unable to strip the virus off the cell surface. Therefore, the mAb can also inhibit a post-attachment stage of the infection cycle.

(G) Confocal microscopy experiments. ZIKV incubated with a concentration exceeding 10,000-fold the IC_{50} value of either ZKA190 Fab or full IgG were added to Vero cells. The ZIKV:antibody complex is detected inside the cells (green) and co-localizes with endosomes (red, yellow overlay). Endosomes and acidic organelles are marked by Lysotracker red; Alexa 488-conjugated ZKA190 is in green. Nuclei are stained with DAPI (blue).

Data in (A) and (F) are represented as mean ± SD of at least two independent experiments. Data in (E) are represented as mean ± SEM of three independent experiments performed in triplicate.

In Vivo Characterization of the DIII-Specific mAb ZKA190

ZKA190 and ZKA190-LALA were tested in A129 mice challenged with a lethal dose of ZIKV strain MP1751 (African lineage). Prophylaxis with ZKA190 and ZKA190-LALA one day before virus challenge protected mice from mortality and morbidity at mAb concentrations of 5, 1 or 0.2 mg/kg (Figures 4A and 4B). ZKA190-LALA, and to a lesser extent ZKA190, delayed morbidity and mortality as compared to the control at 0.04 mg/kg. Viral titers in blood and organs were reduced significantly, even in the presence of serum antibody levels below 1 µg/mL (Figures S5A–S5D). To determine if low doses of ZKA190 could potentially

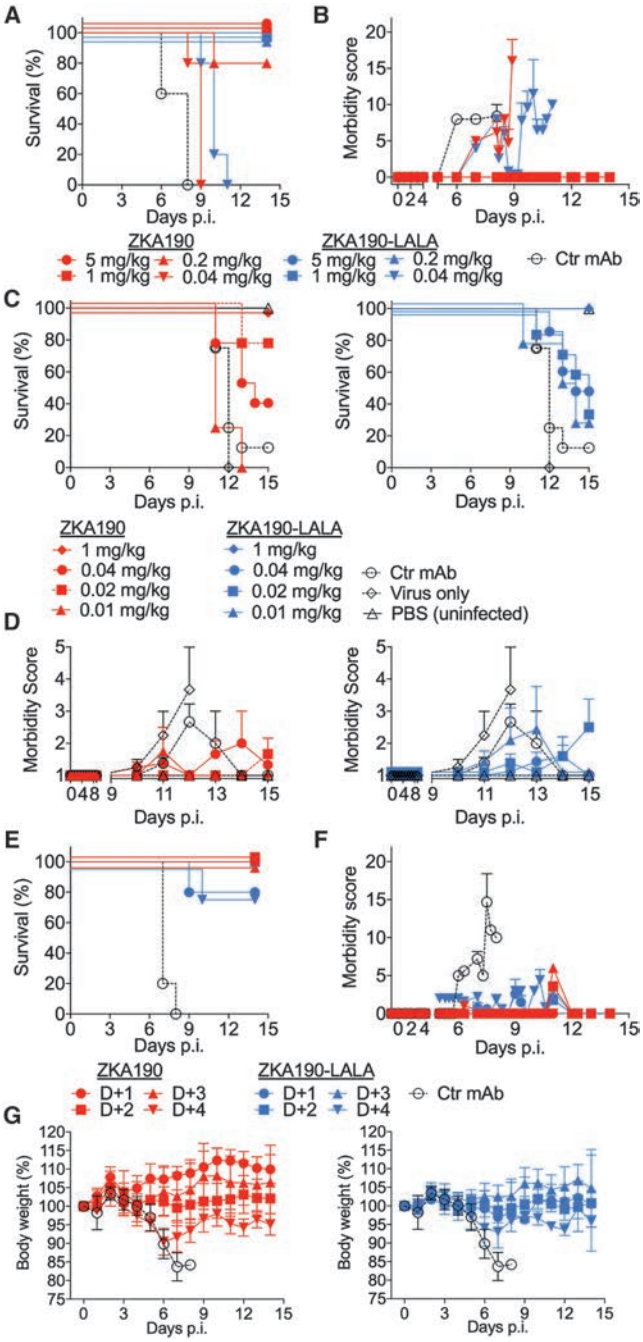


Figure 4. Prophylactic and Therapeutic Efficacy of ZKA190

(A to D) ZKA190 is strongly protective against ZIKV infection when administered prophylactically to mice (A129 in (A) and AG129 in (C, D)) challenged with a lethal dose of ZIKV strain MP1745 (A, B) or Nica 2-16 (C, D). Experiments used N = 4-8 mice per group. (A, C) Kaplan-Meier survival curves are shown. Significance was determined using the Mantel-Cox log-rank test. (A): ZKA190 at 5, 1 and 0.2 mg/kg versus Ctr mAb, p = 0.0031; ZKA190 at 0.04 mg/kg versus Ctr mAb, p = 0.0116; ZKA190-LALA at 5, 1, 0.2 and 0.04 mg/kg versus Ctr mAb, p = 0.0031. Panels C: ZKA190 and ZKA190 LALA at 1 mg/kg versus Ctr mAb, p = 0.0091; ZKA190 at 0.02 mg/kg versus Ctr mAb, p = 0.0035; ZKA190-LALA at 0.04 mg/kg versus Ctr mAb, p = 0.0142. (B, D) Morbidity score of mice monitored over a 14-15 day period (two different scoring methods were used; see (Dowall et al., 2016) for (B) and (Orozco et al., 2012) for (D)). Control mAb is MPE8 mAb (A, specific for RSV F protein (Corti et al., 2013)) and a human IgG polyclonal isotype control (C).

(E–G) ZKA190 or ZKA190-LALA were administered at 15 mg/kg at different time-points after ZIKV infection to A129 mice challenged with a lethal dose of ZIKV strain MP1745. (E) A Kaplan-Meier survival curve is shown. Experiments used N = 5 mice per group. Significance was determined by using the Mantel-Cox log-rank test. ZKA190 and ZKA190-LALA administered on day 1, 2, 3 or 4 versus Ctr., p = 0.0016. (F) Morbidity score of mice monitored over a 14-day period according to (Dowall et al., 2016). (G) Mice were monitored over a 14-day period for body weight loss. Control mAb is MPE8 specific for RSV F protein (Corti et al., 2013).

Data in (B), (D) and (F) are represented as mean ± SEM for each determination. Data in (G) are represented as mean ±SD for each determination. See also Figures S5 and S6.

enhance ZIKV infection, we administered 1 mg/kg, 0.04 mg/kg, 0.02 mg/kg, and 0.01 mg/kg of ZKA190 to AG129 mice and challenged them with a recently isolated ZIKV strain from Nicaragua (Nica 2-16, American lineage). Mice receiving 1 mg/kg of ZKA190 or ZKA190-LALA were completely protected from ZIKV disease (Figures 4C and 4D). ZKA190 or ZKA190-LALA at 0.04 mg/kg and 0.02 mg/kg of partially protected mice from mortality and morbidity, and none of the lower antibody doses resulted in enhanced disease or infection. Blood viral titers on day 3 post-infection were reduced compared to control antibody-treated animals in a dose-dependent fashion (Figure S5E).

To evaluate therapeutic potential, we administered ZKA190 and ZKA190-LALA to A129 mice at different time-points following ZIKV infection. At a dose of 15 mg/kg, survival rates of 80%–100% were achieved, and morbidity was greatly reduced even when treatment was given four days post-infection (Figures 4E–4G). ZKA190 and ZKA190-LALA treatment at all post-infection time-points resulted in significantly reduced viral titers with a clear trend for greater reduction with earlier treatment (Figures S6A–S6C). Of note, ZKA190-LALA showed a reduced antiviral activity on day 5 as compared to ZKA190 when mAbs were given four days post-infection, a result that might relate to the impaired ability of the LALA mAb to facilitate rapid clearance of coated virions.

Of the 16 treated mice, one in vivo escape mutant (Monoclonal Antibody Resistant Mutant 1, MARM1) was isolated containing the amino acid substitution T335R in DIII, which was shown to abrogate ZKA190 binding (Figure S2). Viruses from the other treated mice did not contain any E mutations (Figures S6D and S6E).

In Vitro Selection of ZIKV Escape Mutants

To assess the ability of ZKA190 to select for resistant mutants (MARMs) in vitro, ZIKV (H/PF/2013) was passaged in the presence of sub-neutralizing concentrations of ZKA190. MARM2, carrying the DIII E370K mutation, was isolated after three rounds of selection in Vero cells. The mutation abolished neutralization by ZKA190, but not its binding to DIII (Figure S2). The mutations in the in vivo (T335R) and in vitro (E370K) MARMs are located on the BC and DE loops of DIII, respectively, and are consistent with the epitope identified by NMR.

Development of the Bispecific FIT-1 Antibody

Generation of viral escape mutants can render antibodies useless for treatment. This may be prevented by making a bispecific

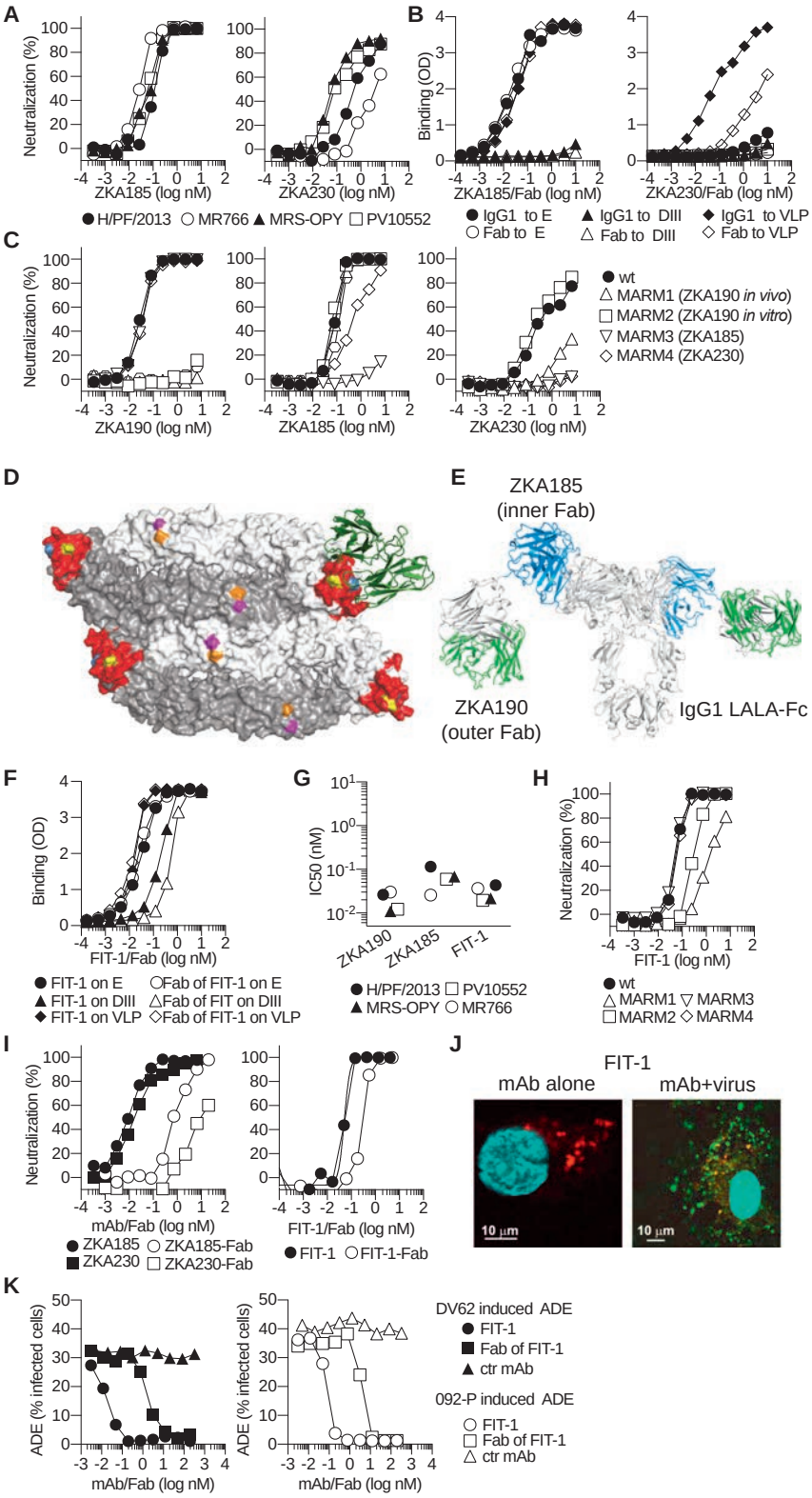


Figure 5. Engineering of ZKA190 into the Bispecific FIT-1 mAb

(A) ZKA185 and ZKA230 mAbs were tested for neutralization of four strains of ZIKV, as determined by the percentage of infected Vero cells in the presence of increasing amounts of mAbs. Data are representative of at least two independent experiments.

(B) Binding of ZKA185 and ZKA230 IgG and Fab to recombinant ZIKV VLP, E and DIII antigens as assessed by ELISA.

(C) ZKA190, ZKA185 and ZKA230 were tested for neutralization of H/PF/2013 (WT) and MARMs 1-4.

(D) Surface representation of two E protein dimers bound by ZKA190 (green); the ZKA190 NMR derived epitope is in red; positions mutated in MARMs are indicated in yellow (E370), blue (T335), orange (D67) and magenta (K84).

(E) Model of FIT-1. The natural linkers between inner and outer Fabs allow flexible movement of Fabs in the FIT-1 antibody. The variable regions of ZKA185 and ZKA190 are highlighted in blue and green, respectively.

(F) Binding of FIT-1 IgG and Fab to recombinant ZIKV VLP, E and DIII antigens as assessed by ELISA.

(G and H) ZKA190, ZKA185 and FIT-1 mAbs were tested for neutralization of four strains (IC_{50} values, (G) and four MARMs (H) of ZIKV).

(I) Neutralization of ZIKV H/PF/2013 strain by ZKA185, ZKA230 and FIT-1 IgG and Fab determined as in (A).

(J) Confocal microscopy experiments as shown in Figure 3G.

(K) Effect on ADE induced by peak enhancing dilution of anti-prM DV62 mAb or DENV2 plasma by serial dilutions of FIT-1 IgG and Fab.

antibody combining two highly neutralizing mAbs, each specific for a distinct epitope. We selected ZKA185 and ZKA230, which potently neutralize African, Asian, and American strains with IC_{50} ranging from 0.02 to 0.62 nM (Figure 5A). ZKA185 binds with high affinity to recombinant ZIKV E protein and to Zika virus-like particles (VLP) but not to the isolated DIII (Figure 5B). ZKA230 bound to ZIKV VLPs, but not to recombinant E or DIII, suggesting that it recognizes a quaternary epitope displayed only on the viral surface (Figure 5B). ZKA185 IgG and Fab were shown to bind to E and VLP antigens with similar high affinity by ELISA. Conversely, ZKA230 Fab bound to VLPs with markedly reduced affinity compared to its IgG.

We isolated MARMs of ZKA185 (MARM3) and ZKA230 (MARM4) by passaging virus in the presence of subneutralizing mAb concentrations. MARM3 contained substitutions at K84E and D67H on DII (Figure 5D). MARM4 showed a mixture of different amino acid substitutions at position 84 (from K to G, E or R). ZKA190 neutralized ZKA185 and ZKA230 MARMs as well as the parental virus (Figure 5C). ZKA185

neutralized both ZKA190 and ZKA230 MARMs. In contrast, ZKA230 neutralized only ZKA190 MARM2.

To gain insight into the development of MARMs capable of escaping from the pressure of multiple mAbs, we serially passaged ZKA190 MARM2 (E370K) in the presence of ZKA185 or ZK230 and found that double MARMs emerged after three to four passages. ZKA230 introduced the K84E mutation while ZKA185 selected for the D76G mutation. These findings indicate that ZIKV can escape multiple mAbs targeting distinct sites when the selection is performed in a stepwise fashion.

MAb ZKA185 was selected to combine with ZKA190 for the development of a bispecific antibody since it potently cross-neutralizes ZIKV strains and does not compete with ZKA190. We produced a bispecific antibody in the tetravalent symmetric format Fabs-in-tandem-Ig (FIT-Ig) (Gong et al., 2017). ZKA190 was placed in the outer and ZKA185 in the inner Fab position (Figure 5E), using the backbone IgG1 antibody in LALA format (here designated FIT-1). FIT-1 bound DIII, E and VLPs in ELISA (Figure 5F) and retained high neutralizing potency against ZIKV strains, with IC_{50} values largely similar to those of the parental mAbs (Figure 5G). FIT-1 bound to E protein with higher affinity than either of the parental mAbs (K_D values assessed by bio-layer interferometry: ZKA185 1.8 nM, ZKA190 9.3 nM and FIT-1 $K_D < 1$ pM due to slower dissociation rate, presumably through avidity effects) and effectively neutralized all the ZKA190, ZKA185 and ZKA230 MARMs (Figure 5H). In addition, the neutralizing activity of the Fab fragment of FIT-1 (comprising one ZKA190 and one ZKA185) was reduced only by about 6-fold (Figure 5I, right panel), whereas those of the individual ZKA190 and ZKA185 Fabs were reduced by about 100-fold compared to their full-length IgG1 (Figure 1B and Figure 5I). Taken together, these findings suggest the ZKA190 and ZKA185 moieties in FIT-1 are both active.

No MARMs were isolated despite 8 rounds of in vitro serial passages in the presence of FIT-1. In contrast, MARMs appeared after 3 to 4 passages with individual mAbs. Simultaneous mutations in both DIII and DII are thus unlikely to occur, suggesting that FIT-1 is a safer therapeutic. Confocal microscopy studies showed that FIT-1, like ZKA190, inhibits virus infection at a post-attachment step, likely fusion (Figure 5J). Finally, FIT-1and its Fab blocked ADE of human anti-DENV2 plasma or DV62 at concentrations above 0.1 and 10 nM, respectively (Figure 5K).

We evaluated the therapeutic potential of FIT-1 by administering three different doses (15, 5 and 1 mg/kg) at three different time-points to mice following ZIKV infection. At a dose of 15 mg/kg, survival rates were 100% without signs of morbidity even when treatment was given three days post-infection (Figure 6); viral titers were abrogated and no escape mutants were detected, indicating high in vivo efficacy. Administration of 5 mg/kg resulted in survival rates of 70%–100%; again, no escape mutants were detected on day 5 after infection. The lowest dose tested (1 mg/kg) protected when administered on day 1 or 2, but not day 3, after infection.

DISCUSSION

The highly potent ZKA190 mAb binds to to the lateral ridge in DIII of the E protein in a unique orientation compared to other flavivirus-neutralizing LR-binding mAbs (Edeling et al., 2014; Nybakken et al., 2005; Zhao et al., 2016) (Figure 7A), resulting in full occupancy of all 180 E proteins. Unlike ZKA190, the cryoEM structures of E16:WNV (Kaufmann et al., 2006) and ZV23:ZIKV (Wang et al., 2016) showed that their Fabs are unable to bind to DIII around the 5-fold vertex. Superposition of the E106:DIII (Edeling et al., 2014), ZV23:DIII and ZV67:DIII (Zhao et al., 2016) structures onto an uncomplexed ZIKV cryoEM structure showed clashes with the neighboring E proteins around the 5-fold vertex (Figure 7B).

Our previous structure of uncomplexed ZIKV revealed that the E protein shell is more compact than that of DENV (Kostyuchenko et al., 2016). Binding of ZKA190 disrupts the tight interactions between E proteins in the 5-fold vertices at both 4°C and 37°C, thus allowing full occupancy. ZKA190 may either select a transient viral conformation where the 5-fold vertex is ''open'' (conformational selection) or cause surface structural changes (allosteric effects). The latter appears more likely since binding occurs even at 4°C, where the thermal motions required for E protein conformational changes are limited. NMR studies and DIII mutants indicated that binding of ZKA190 causes an allosteric effect on the DIII FG loop, which points toward the center of the 5-fold vertex. ZKA190 binding to one DIII near the 5-fold vertex may alter the FG loop conformation, opening the vertex and exposing additional ZKA190 binding sites. Alternatively, the antibody could first bind to the 3- and/or 2-fold vertices, causing allosteric effects throughout the viral surface resulting in opening of the 5-fold vertex, to which other antibodies could then bind; however, experimental evidence for this is lacking.

Multiple mechanisms may contribute to ZKA190 activity. We showed here that high concentrations of ZKA190 induced aggregation of virus particles by crosslinking different particles with the two Fab arms. This correlated with increased neutralization activity, suggesting a role for aggregation in viral neutralization. Molecular dynamics showed that the two Fab arms of ZKA190 can bind within a viral particle, possibly preventing the dramatic rearrangement of the E proteins on the virus surface required for fusion.

RT-PCR and confocal microscopy data indicate that the virus remains attached to cells in the presence of ZKA190 and is able to enter cells in the presence of neutralizing concentrations of ZKA190. Furthermore, similar neutralization profiles of ZKA190 (and FIT-1) in the post- and pre-attachment assays, as well as inhibition by the ZKA190 Fab fragment of DENV plasma and anti-prM antibody-mediated ADE, support the notion that ZKA190 inhibits a post-attachment step, likely fusion. The epitope of ZKA190 encompasses the LR of DIII and the DI-DIII linker region. In a normal infection, the low-pH endosomal environment provokes the E proteins to rearrange from a dimeric structure to a trimeric post-fusion structure. Binding of ZKA190 to the DI-DIII linker and DIII may hinder the movement of DIII, preventing rearrangement and thus membrane fusion, as previously suggested in structural studies of E16:WNV (Kaufmann et al., 2006; Zhang et al., 2015).

Although antibody effector functions can enhance the efficacy of anti-flavivirus antibodies (Chung et al., 2006), comparison of ZKA190 and ZKA190-LALA in vivo efficacy indicates that Fc-dependent mechanisms, such as ADCC or CDC, likely do not play a major role in prophylaxis or therapy. Despite clear evidence of in vitro ADE by ZKA190, and in contrast to a recent

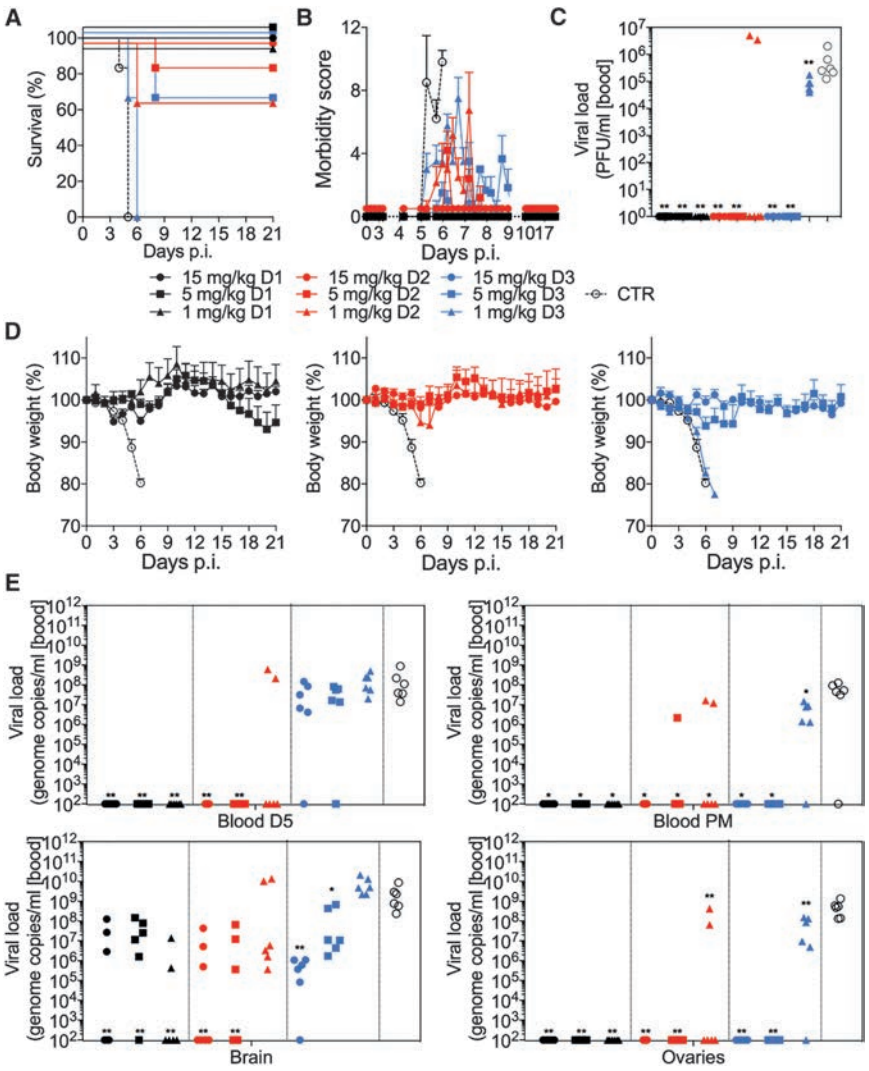


Figure 6. Therapeutic Efficacy of FIT-1

FIT-1 is strongly effective against ZIKV infection when administered therapeutically at different time-points to mice (A129) challenged with a lethal dose of ZIKV strain MP17451. Experiments used N = 5-6 mice per group.

(A) Kaplan-Meier survival curves are shown. Significance was determined by using the Mantel-Cox log-rank test. FIT-1 at 15, 5 and 1 mg/kg given either on day 1 or 2 versus Ctr mAb, p = 0.0012; ZKA190 at 15 and 5 mg/kg given on day 3 versus Ctr mAb, p = 0.0012; ZKA190 at 1 mg/kg given on day 3 versus Ctr mAb, p = 0.0170.

(B) Morbidity score of mice monitored over a 21-day period (Dowall et al., 2016).

(C) Viral load was measured as PFUs on day 5 in blood of all animals.

(D) Mice were monitored over a 21-day period for body weight loss. Control mAb in (A) is MPE8 mAb (specific for RSV F protein (Corti et al., 2013)).

(E) Viral load was measured as genomic copies by qPCR on day 5 in blood of all animals and in blood and indicated tissues when animals were culled at the end of the study or when the humane endpoints were met. Significance was determined compared to control antibody treatment by nonparametric unpaired Mann-Whitney U test. *p < 0.05; **p < 0.01.

Data in (B) are represented as mean ± SEM for each determination. Data in (D) are represented as mean ± SD for each determination.

report (Bardina et al., 2017), our in vivo results indicate that ZKA190-mediated disease enhancement does not occur even when the mAb is administered at doses expected to provide only partial neutralization.

Numerous flavivirus MARMs are readily isolated in in vitro systems (de Alwis et al., 2012; Zou et al., 2012), similarly to what we observed with ZKA190, ZKA185 and ZKA230 mAbs. However, only a few studies described isolation of MARMs in vivo (Lai et al., 2007). Here, we isolated a ZKA190 MARM from one animal treated therapeutically, although no MARMs were detected in animals treated prophylactically. To reduce the risk of emergence of escape mutants, we produced a bispecific antibody coupling ZKA190 to the anti-DII ZKA185 mAb. The FIT-1 bispecific antibody is highly potent in vitro and in vivo and did not generate escape mutants even after 8 passages.

Flavivirus particles in in vitro preparations represent different maturation states (mature, partially immature, and fully immature) (Lok, 2016). Only partially and fully immature virions have prM on their surfaces, and anti-prM antibodies can confer infectivity to the otherwise non-infectious immature virions (Rodenhuis-Zybert et al., 2010). Since anti-prM antibodies are elicited in anti-flavivirus immune responses (Dejnirattisai et al., 2010), they can contribute significantly to pathogenesis (Dejnirattisai et al., 2010). We showed here that ZKA190 and FIT-1 can block in vitro ADE mediated by the cross-reactive prM mAb DV62, indicating that ZKA190 can effectively block prM antibody-mediated infection by at least partially immature ZIKV particles.

Our current studies in vitro and in lethal adult mouse models showed that the FIT-1 bispecific antibody is a good candidate for development of antibody-based therapy for ZIKV infection. Further studies in pregnant rhesus macaques (Dudley et al., 2016; Nguyen et al., 2017) would enable testing FIT-1 efficacy in blocking fetal infection. The in vivo protective activity of FIT-1 indicates that alternative delivery mechanisms, such as DNA or RNA vectors, or half-life-extending modification of the Fc, might be achievable to make it cost-effective for the prevention of congenital ZIKV disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENTS AND RESOURCE SHARING

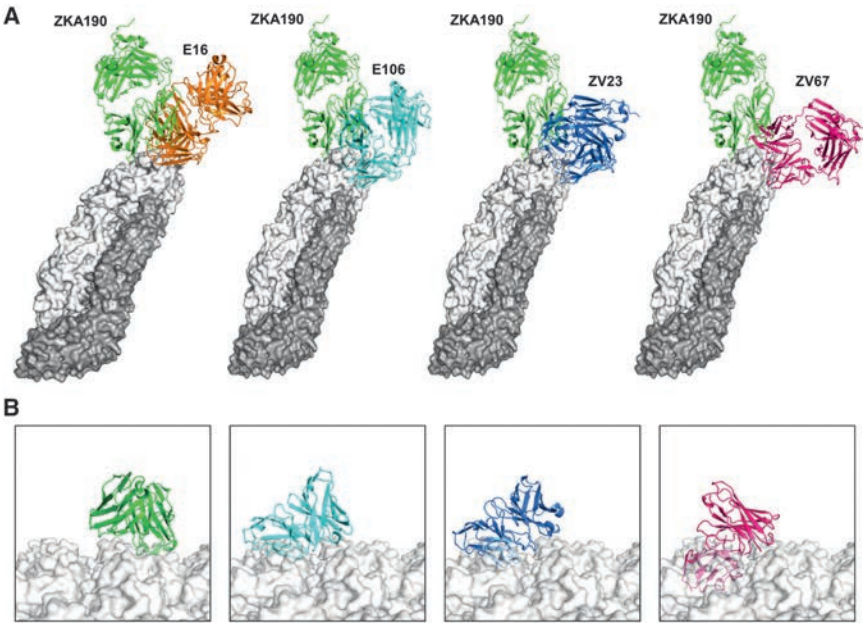


Figure 7. Comparison of ZKA190 to Other DIII Lateral Ridge Antibodies

(A) Superposition of the structures of the Fabs E16 (orange, PDB 1ZTX), E106 (cyan, PDB 4L5F), ZV23 (blue, PDB 5GZR) and ZK67 (purple, PDB 5KVG) onto DIII of ZKA190-E protein cryoEM structure. (B) Fv regions of ZKA190, E106, ZV23 and ZV67 (colors as indicated) over the ZIKV surface, obtained by superimposing their structural complex with DIII to E protein on the viral surface (white surface, the 5-fold vertex is shown). All mAbs, but not ZKA190, clashes with adjacent E proteins when superimposed onto molecule A near the 5-fold vertex.

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Reagents
- METHOD DETAILS
  - Production of recombinant IgG and LALA variant
  - Production of recombinant FIT bispecific antibodies
  - Infection of human neural progenitor cells (hNPCs)
  - Virus neutralization and antibody-dependent enhancement
  - Size measurement of ZKA190 IgG – ZIKV complex by dynamic light scattering
  - Pre- and postattachment neutralization assays
  - RT–PCR to quantitate virus on cell surface
  - Effect of LALA antibodies on ADE
  - ELISA
  - Surface Plasmon Resonance
  - NMR spectroscopy
  - Antibody modeling, computational docking and Molecular Dynamics simulations
  - Modeling of ZKA190 full IgG bound to the viral surface
  - Confocal Microscopy
  - Sequence analysis from ex vivo ZIKV samples
  - MARMs selection
  - Sequencing of MARMs virus
  - ZIKV infection of A129 mice
  - Assessment of prophylactic and therapeutic activity of ZKA190 and FIT-1 in A129 mice
  - Assessment of prophylactic activity of ZKA190 in AG129 mice
  - Viral load quantification from AG129 ZIKV infected mice
  - Viral load quantification from A129 ZIKV infected mice
  - DIII coding region sequencing from biological samples
  - Virus sample preparation for cryoEM studies
  - CryoEM sample preparation
  - CryoEM data collection and image processing
  - Model fitting
  - Structure analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.09.002.

AUTHOR CONTRIBUTIONS

F.Z. and S. Jaconi produced and purified antibodies. M.F. produced figures and carried out bioinformatic analysis. S.B. performed in vitro cellular neutralization, ADE assays and MARMs selection. D.A.E., T.B., V.G., S.P., S.D. and R.H. designed and carried out in vivo studies. J.W., T.-S.N., E.X.Y.L., G.F., J.S. designed and performed structural cryoEM research. M.B., M.P., L.S. designed and performed NMR structural research, docking, molecular dynamics, mutagenesis, binding and immunofluorescence studies; A.C. designed and performed molecular dynamics simulations; S. Jurt and O.Z. performed NMR studies. I.P., A.R., V.B. and E.V. designed and performed in vitro studies on hNPCs infection models. K.S. carried out binding studies, in vitro cellular neutralization, MARMs isolation and sequencing, analyzed the data and coordinated in vitro characterization activities. E.C. and M.B. provided supervision and contributed analyzing data. F.S. and A.L. edited the paper and provided supervision. J.W., E.H., S.-M.L., L.V., and D.C. designed research, analyzed data, and wrote the paper.

ACKNOWLEDGMENTS

We thank Victor A. Kostyuchenko and P. Robert Beatty for scientific discussion and Kuiama Lewandowski at PHE for the technical support in viral sequencing and Rocco D'Antuono for microscope assistance. We thank Fabio Benigni at Humabs for critical reading of the manuscript and Antonio Piralla from San Matteo Hospital in Pavia for helping in the sequencing of MARMs. We thank the European Virus Archive for consenting to the use of the ZIKV

H/PF/2013 strain. The work was supported by Singapore Ministry of Education Tier 3 grant (MOE2012-T3-1-008), National Research Foundation Investigatorship award (NRF-NRFI2016-01), National Research Foundation Competitive Research Project grant (NRF2016NRF-CRP001-063) and the Duke-NUS Signature Research Programme funded by the Ministry of Health, Singapore, awarded to S-M.L, and grant 5U19AI109761 from the US National Institutes of Health to Ian Lipkin and E.H. L.V. is grateful for support by G. Noseda, SNF grant 310030_166445, 157699, KFS-3728-08-2015 and Lions Club Montecenerі. A.L. is the scientific founder of Humabs BioMed SA. A.L. holds shares in Humabs BioMed. K.S., F.Z., S.J., E.C., M.B. and D.C. are employees of Humabs Biomed.

Received: January 17, 2017
Revised: June 14, 2017
Accepted: August 31, 2017
Published: September 21, 2017

REFERENCES

Allison, S.L., Schalich, J., Stiasny, K., Mandl, C.W., and Heinz, F.X. (2001). Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. J. Virol. *75*, 4268–4275.

Barba-Spaeth, G., Dejnirattisai, W., Rouvinski, A., Vaney, M.-C., Medits, I., Sharma, A., Simon-Lorière, E., Sakuntabhai, A., Cao-Lormeau, V.-M., Haouz, A., et al. (2016). Structural basis of potent Zika–dengue virus antibody cross-neutralization. Nature *536*, 1–23.

Bardelli, M., Livoti, E., Simonelli, L., Pedotti, M., Moraes, A., Valente, A.P., and Varani, L. (2015). Epitope mapping by solution NMR spectroscopy. J. Mol. Recognit. *28*, 393–400.

Bardina, S.V., Bunduc, P., Tripathi, S., Duehr, J., Frere, J.J., Brown, J.A., Nachbagauer, R., Foster, G.A., Krysztof, D., Tortorella, D., et al. (2017). Enhancement of Zika virus pathogenesis by preexisting antiflavivirus immunity. Science *356*, 175–180.

Baronti, C., Piorkowski, G., Charrel, R.N., Boubis, L., Leparc-Goffart, I., and de Lamballerie, X. (2014). Complete coding sequence of zika virus from a French polynesia outbreak in 2013. Genome Announc *2*, e00500–14–e00500–14.

Beltramello, M., Williams, K.L., Simmons, C.P., Macagno, A., Simonelli, L., Quyen, N.T.H., Sukupolvi-Petty, S., Navarro-Sanchez, E., Young, P.R., de Silva, A.M., et al. (2010). The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. Cell Host Microbe *8*, 271–283.

Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. EMBO J. *23*, 728–738.

Brister, J.R., Bao, Y., Zhdanov, S.A., Ostapchuck, Y., Chetvernin, V., Kiryutin, B., Zaslavsky, L., Kimelman, M., and Tatusova, T.A. (2014). Virus Variation Resource–recent updates and future directions. Nucleic Acids Res. *42*, D660–D665.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., and Reilein, A. (2000). Leginon: an automated system for acquisition of images from vitreous ice specimens. J. Struct. Biol. *132*, 33–45.

Case, D.A., Cerutti, D.S., Cheatham, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Greene, D., Homeyer, N., et al. (2017). AMBER 2017 University of California, San Francisco.

Chung, K.M., Nybakken, G.E., Thompson, B.S., Engle, M.J., Marri, A., Fremont, D.H., and Diamond, M.S. (2006). Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms. J. Virol. *80*, 1340–1351.

Corti, D., Bianchi, S., Vanzetta, F., Minola, A., Perez, L., Agatic, G., Guarino, B., Silacci, C., Marcandalli, J., Marsland, B.J., et al. (2013). Cross-neutralization of four paramyxoviruses by a human monoclonal antibody. Nature *501*, 439–443.

Crill, W.D., and Roehrig, J.T. (2001). Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. J. Virol. *75*, 7769–7773.

de Alwis, R., Smith, S.A., Olivarez, N.P., Messer, W.B., Huynh, J.P., Wahala, W.M.P.B., White, L.J., Diamond, M.S., Baric, R.S., Crowe, J.E., Jr., and de Silva, A.M. (2012). Identification of human neutralizing antibodies that bind to complex epitopes on dengue virions. Proc. Natl. Acad. Sci. USA *109*, 7439–7444.

Dejnirattisai, W., Jumnainsong, A., Onsirisakul, N., Fitton, P., Vasanawathana, S., Limpitikul, W., Puttikhunt, C., Edwards, C., Duangchinda, T., Supasa, S., et al. (2010). Cross-reacting antibodies enhance dengue virus infection in humans. Science *328*, 745–748.

Dowall, S.D., Graham, V.A., Rayner, E., Atkinson, B., Hall, G., Watson, R.J., Bosworth, A., Bonney, L.C., Kitchen, S., and Hewson, R. (2016). A Susceptible Mouse Model for Zika Virus Infection. PLoS Negl. Trop. Dis. *10*, e0004658.

Dudley, D.M., Aliota, M.T., Mohr, E.L., Weiler, A.M., Lehrer-Brey, G., Weisgrau, K.L., Mohns, M.S., Breitbach, M.E., Rasheed, M.N., Newman, C.M., et al. (2016). A rhesus macaque model of Asian-lineage Zika virus infection. Nat. Commun. *7*, 12204.

Edeling, M.A., Austin, S.K., Shrestha, B., Dowd, K.A., Mukherjee, S., Nelson, C.A., Johnson, S., Mabila, M.N., Christian, E.A., Rucker, J., et al. (2014). Potent dengue virus neutralization by a therapeutic antibody with low monovalent affinity requires bivalent engagement. PLoS Pathog. *10*, e1004072.

Faria, N.R., Sabino, E.C., Nunes, M.R.T., Alcantara, L.C.J., Loman, N.J., and Pybus, O.G. (2016). Mobile real-time surveillance of Zika virus in Brazil. Genome Med. *8*, 97.

Fibriansah, G., Tan, J.L., Smith, S.A., de Alwis, R., Ng, T.-S., Kostyuchenko, V.A., Jadi, R.S., Kukkaro, P., de Silva, A.M., Crowe, J.E., and Lok, S.M. (2015). A highly potent human antibody neutralizes dengue virus serotype 3 by binding across three surface proteins. Nat. Commun. *6*, 6341.

Frontera, J.A., and da Silva, I.R.F. (2016). Zika Getting on Your Nerves? The Association with the Guillain–Barré Syndrome. N. Engl. J. Med., NEJMe1611840–NEJMe1611842.

Garcez, P.P., Loiola, E.C., Madeiro da Costa, R., Higa, L.M., Trindade, P., Delvecchio, R., Nascimento, J.M., Brindeiro, R., Tanuri, A., and Rehen, S.K. (2016). Zika virus impairs growth in human neurospheres and brain organoids. Science *352*, 816–818.

Gong, S., Ren, F., Wu, D., Wu, X., and Wu, C. (2017). Fabs-in-tandem immunoglobulin is a novel and versatile bispecific design for engaging multiple therapeutic targets. MAbs *165*,. http://dx.doi.org/10.1080/19420862.2017.1345401.

Govero, J., Esakky, P., Scheaffer, S.M., Fernandez, E., Drury, A., Platt, D.J., Gorman, M.J., Richner, J.M., Caine, E.A., Salazar, V., et al. (2016). Zika virus infection damages the testes in mice. Nature *540*, 438–442.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., and Baker, D. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J. Mol. Biol. *331*, 281–299.

Güntert, P., and Buchner, L. (2015). Combined automated NOE assignment and structure calculation with CYANA. J. Biomol. NMR *62*, 453–471.

Halstead, S.B. (2014). Dengue Antibody-Dependent Enhancement: Knowns and Unknowns. Microbiol. Spectr. *2*, 249–271.

Hessell, A.J., Hangartner, L., Hunter, M., Havenith, C.E., Beurskens, F.J., Bakker, J.M., Lanigan, C.M., Landucci, G., Forthal, D.N., Parren, P.W., et al. (2007). Fc receptor but not complement binding is important in antibody protection against HIV. Nature *449*, 101–104.

Honein, M.A., Dawson, A.L., Petersen, E.E., Jones, A.M., Lee, E.H., Yazdy, M.M., Ahmad, N., Macdonald, J., Evert, N., Bingham, A., et al. (2017). Birth Defects Among Fetuses and Infants of US Women With Evidence of Possible Zika Virus Infection During Pregnancy. JAMA *317*, 59–68.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kaufmann, B., Nybakken, G.E., Chipman, P.R., Zhang, W., Diamond, M.S., Fremont, D.H., Kuhn, R.J., and Rossmann, M.G. (2006). West Nile virus in complex with the Fab fragment of a neutralizing monoclonal antibody. Proc. Natl. Acad. Sci. USA *103*, 12400–12404.

Keller, R. (2004). The Computer Aided Resonance Assignment Tutorial. Cantina: Verlag.

Kostyuchenko, V.A., Lim, E.X.Y., Zhang, S., Fibriansah, G., Ng, T.-S., Ooi, J.S.G., Shi, J., and Lok, S.-M. (2016). Structure of the thermally stable Zika virus. Nature *533*, 425–428.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

Lai, C.-J., Goncalvez, A.P., Men, R., Wernly, C., Donau, O., Engle, R.E., and Purcell, R.H. (2007). Epitope determinants of a chimpanzee dengue virus type 4 (DENV-4)-neutralizing antibody and protection against DENV-4 challenge in mice and rhesus monkeys by passively transferred humanized antibody. J. Virol. *81*, 12766–12774.

Lanciotti, R.S., Kosoy, O.L., Laven, J.J., Velez, J.O., Lambert, A.J., Johnson, A.J., Stanfield, S.M., and Duffy, M.R. (2008). Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg. Infect. Dis. *14*, 1232–1239.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Lok, S.-M. (2016). The Interplay of Dengue Virus Morphological Diversity and Human Antibodies. Trends Microbiol. *24*, 284–293.

Loman, N.J., and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics *30*, 3399–3401.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods *12*, 733–735.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell *143*, 527–539.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. *142*, 334–347.

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodušek, V., et al. (2016). Zika Virus Associated with Microcephaly. N. Engl. J. Med. *374*, 951–958.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. Nature *427*, 313–319.

Nguyen, S.M., Antony, K.M., Dudley, D.M., Kohn, S., Simmons, H.A., Wolfe, B., Salamat, M.S., Teixeira, L.B.C., Wiepz, G.J., Thoong, T.H., et al. (2017). Highly efficient maternal-fetal Zika virus transmission in pregnant rhesus macaques. PLoS Pathog. *13*, e1006378.

Nowakowski, T.J., Pollen, A.A., Di Lullo, E., Sandoval-Espinosa, C., Bershteyn, M., and Kriegstein, A.R. (2016). Expression Analysis Highlights AXL as a Candidate Zika Virus Entry Receptor in Neural Stem Cells. Cell Stem Cell *18*, 591–596.

Nybakken, G.E., Oliphant, T., Johnson, S., Burke, S., Diamond, M.S., and Fremont, D.H. (2005). Structural basis of West Nile virus neutralization by a therapeutic antibody. Nature *437*, 764–769.

Onorati, M., Li, Z., Liu, F., Sousa, A.M.M., Nakagawa, N., Li, M., Dell'Anno, M.T., Gulden, F.O., Pochareddy, S., Tebbenkamp, A.T.N., et al. (2016). Zika Virus Disrupts Phospho-TBK1 Localization and Mitosis in Human Neuroepithelial Stem Cells and Radial Glia. Cell Rep. *16*, 2576–2592.

Orozco, S., Schmid, M.A., Parameswaran, P., Lachica, R., Henn, M.R., Beatty, R., and Harris, E. (2012). Characterization of a model of lethal dengue virus 2 infection in C57BL/6 mice deficient in the alpha/beta interferon receptor. J. Gen. Virol. *93*, 2152–2157.

Pedotti, M., Simonelli, L., Livoti, E., and Varani, L. (2011). Computational docking of antibody-antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin. Int. J. Mol. Sci. *12*, 226–251.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature *530*, 228–232.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature *375*, 291–298.

Rodenhuis-Zybert, I.A., van der Schaar, H.M., da Silva Voorham, J.M., van der Ende-Metselaar, H., Lei, H.-Y., Wilschut, J., and Smit, J.M. (2010). Immature dengue virus: a veiled pathogen? PLoS Pathog. *6*, e1000718.

Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Simonelli, L., Beltramello, M., Yudina, Z., Macagno, A., Calzolai, L., and Varani, L. (2010). Rapid structural characterization of human antibody-antigen complexes through experimentally validated computational docking. J. Mol. Biol. *396*, 1491–1507.

Simonelli, L., Pedotti, M., Beltramello, M., Livoti, E., Calzolai, L., Sallusto, F., Lanzavecchia, A., and Varani, L. (2013). Rational engineering of a human anti-dengue antibody through experimentally validated computational docking. PLoS ONE *8*, e55561.

Sircar, A., Kim, E.T., and Gray, J.J. (2009). RosettaAntibody: antibody variable region homology modeling server. Nucleic Acids Res. *37*, W474-9.

Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T.C., Rossmann, M.G., and Kuhn, R.J. (2016). The 3.8 Å resolution cryo-EM structure of Zika virus. Science *352*, 467–470.

Stettler, K., Beltramello, M., Espinosa, D.A., Graham, V., Cassotta, A., Bianchi, S., Vanzetta, F., Minola, A., Jaconi, S., Mele, F., et al. (2016). Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. Science *353*, 823–826.

Stiasny, K., Allison, S.L., Marchler-Bauer, A., Kunz, C., and Heinz, F.X. (1996). Structural requirements for low-pH-induced rearrangements in the envelope glycoprotein of tick-borne encephalitis virus. J. Virol. *70*, 8142–8147.

Tabata, T., Petitt, M., Puerta-Guardo, H., Michlmayr, D., Wang, C., Fang-Hoover, J., Harris, E., and Pereira, L. (2016). Zika Virus Targets Different Primary Human Placental Cells, Suggesting Two Routes for Vertical Transmission. Cell Host Microbe *20*, 155–166.

Tang, H., Hammack, C., Ogden, S.C., Wen, Z., Qian, X., Li, Y., Yao, B., Shin, J., Zhang, F., Lee, E.M., et al. (2016). Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. Stem Cell, 1–5.

van den Broek, M.F., Müller, U., Huang, S., Aguet, M., and Zinkernagel, R.M. (1995). Antiviral defense in mice lacking both alpha/beta and gamma interferon receptors. J. Virol. *69*, 4792–4796.

Wang, Q., Yang, H., Liu, X., Dai, L., Ma, T., Qi, J., Wong, G., Peng, R., Liu, S., Li, J., et al. (2016). Molecular determinants of human neutralizing antibodies isolated from a patient infected with Zika virus. Sci Transl Med *8*, 369ra179–369ra179.

Xu, M., Lee, E.M., Wen, Z., Cheng, Y., Huang, W.-K., Qian, X., Tcw, J., Kouznetsova, J., Ogden, S.C., Hammack, C., et al. (2016). Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. Nat. Med. *22*, 1101–1107.

Zhang, X., Sheng, J., Austin, S.K., Hoornweg, T.E., Smit, J.M., Kuhn, R.J., Diamond, M.S., and Rossmann, M.G. (2015). Structure of acidic pH dengue virus showing the fusogenic glycoprotein trimers. J. Virol. *89*, 743–750.

Zhao, H., Fernandez, E., Dowd, K.A., Speer, S.D., Platt, D.J., Gorman, M.J., Govero, J., Nelson, C.A., Pierson, T.C., Diamond, M.S., and Fremont, D.H. (2016). Structural Basis of Zika Virus-Specific Antibody Protection. Cell *166*, 1016–1027.

Zou, G., Kukkaro, P., Lok, S.-M., Ng, J.K.W., Tan, G.K., Hanson, B.J., Alonso, S., MacAry, P.A., and Shi, P.-Y. (2012). Resistance analysis of an antibody that selectively inhibits dengue virus serotype-1. Antiviral Res. *95*, 216–223.

A Unifying Theory of Branching Morphogenesis

Edouard Hannezo,[1,2,3,8,9,*] Colinda L.G.J. Scheele,[4,8] Mohammad Moad,[5] Nicholas Drogo,[6] Rakesh Heer,[5] Rosemary V. Sampogna,[7] Jacco van Rheenen,[4,*] and Benjamin D. Simons[1,2,3,10,*]

[1]Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK
[2]The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK
[3]The Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge CB2 1QN, UK
[4]Cancer Genomics Netherlands, Hubrecht Institute-KNAW and University Medical Centre Utrecht, Utrecht 3584CT, the Netherlands
[5]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE2 4AD, UK
[6]Department of Biomedical Engineering, University of Rochester, Rochester, NY 14627, USA
[7]Division of Nephrology, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA
[8]These authors contributed equally
[9]Present address: Institute of Science and Technology IST Austria, 3400 Klosterneuburg, Austria
[10]Lead Contact
*Correspondence: edouard.hannezo@ist.ac.at (E.H.), j.vanrheenen@hubrecht.eu (J.v.R.), bds10@cam.ac.uk (B.D.S.)
http://dx.doi.org/10.1016/j.cell.2017.08.026

SUMMARY

The morphogenesis of branched organs remains a subject of abiding interest. Although much is known about the underlying signaling pathways, it remains unclear how macroscopic features of branched organs, including their size, network topology, and spatial patterning, are encoded. Here, we show that, in mouse mammary gland, kidney, and human prostate, these features can be explained quantitatively within a single unifying framework of branching and annihilating random walks. Based on quantitative analyses of large-scale organ reconstructions and proliferation kinetics measurements, we propose that morphogenesis follows from the proliferative activity of equipotent tips that stochastically branch and randomly explore their environment but compete neutrally for space, becoming proliferatively inactive when in proximity with neighboring ducts. These results show that complex branched epithelial structures develop as a self-organized process, reliant upon a strikingly simple but generic rule, without recourse to a rigid and deterministic sequence of genetically programmed events.

INTRODUCTION

Branching morphogenesis has fascinated biologists and mathematicians for centuries, both because of its complexity and ubiquity (Hogan 1999; Lu and Werb, 2008; Metzger et al., 2008). In higher organisms, many organs are organized into ductal tree-like structures comprising tens of thousands of branches, which typically function to maximize the surface of exchange between the epithelium and its lumen. Examples include lung, kidney, prostate, liver, pancreas, the circulatory system and the mammary gland epithelium. Alongside metazoa, tree crowns and root systems as well as coral reefs often display a similar branched organization (Harrison, 2010) raising the question of whether common mechanisms could underlie their formation. Extensive investigations have identified features shared by all branched organs, which are formed by repeated cycles of branching (either through side-branching or tip-splitting), together with phases of ductal elongation (Iber and Menshykau, 2013).

Attempts to resolve the regulatory basis of branching morphogenesis have been targeted at different length scales, offering contrasting perspectives. First, at the molecular scale, key regulatory signaling pathways, for instance controlling proliferation and cell fate, have been resolved in multiple organs (Iber and Menshykau, 2013). Second, at the cellular and mesoscopic scale, measurements of gene expression patterns and branching shape have implicated Turing-like mechanisms in the regulation of the first rounds of repetitive branching in the lung and kidney (Iber and Menshykau, 2013; Guo et al., 2014). Alternative, potentially overlapping, explanations based on mechanical (Gjorevski and Nelson, 2011) or viscous (Lubkin and Murray, 1995) models have been proposed, and processes such as oriented cell divisions (Yu et al., 2009), collective cell migration (Huebner et al., 2016; Riccio et al., 2016), and cytoskeleton-driven cell shape changes (Elliott et al., 2015; Kim et al., 2015) have been shown to play a role. Yet, even a perfect understanding of how single branching events occur would not explain how thousands of tips and branches become coordinated at the organ scale, to specify a complex ductal network.

Therefore, here, we adopt an alternative "non-reductionist" approach and test whether the statistical properties of branched networks can be predicted without extensively addressing the detailed underlying molecular and cellular regulatory processes. Historically, the development of such statistical approaches has been limited by the lack of high resolution biological data on the complete organ structure. However, this problem is becoming alleviated by advances in imaging techniques (Metzger et al., 2008; Sampogna et al., 2015; Short et al., 2014), which provide an ideal platform to question how a complex 3D organ structure is encoded. Does it form from the unfolding of an intrinsic deterministic program or is it shaped by extrinsic influences and stochastic processes?

 © 2017 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

In the following, we use detailed whole-organ imaging and 3D reconstructions of the mouse mammary gland epithelium, mouse kidney, and human prostate to address the spatiotemporal dynamics of branching morphogenesis. We show that the detailed statistical properties of these organs share key underlying features, which can be explained quantitatively through a remarkably simple and conserved design principle, based on the theory of branching and annihilating random walks (BARWs). In this model, growing ductal tips follow the same, time-invariant, statistical rules based on stochastic ductal branching and random exploration of space. However, when an active tip comes into proximity with a neighboring duct, it becomes irreversibly inactive (differentiating and exiting cell cycle), leading to the termination of the duct. We show that, together, these simple local rules are enough to allow the epithelium to grow in a self-organized manner, into a complex ductal network with conserved statistical properties that are quantitatively predicted by the model. Notably, these isotropic rules predict the emergence of directional bias in the expansion of the ductal network, in the absence of any external guiding signaling gradients. Finally, to further challenge the model, we predict, and discover experimentally, novel signatures of the inferred dynamics, which are consistent with an out-of-equilibrium "phase transition." Moreover, by adjusting experimentally the microenvironment of the branching tips through local or systemic perturbations, we further test the predictive capacity of the model and gain insight into the molecular regulatory basis of the inferred collective cell dynamics.

RESULTS

Defining the Ductal Network Structure of the Mouse Mammary Gland Epithelium

To develop a model of tip-driven ductal morphogenesis, we began by considering the mammary gland epithelium. At birth, the mouse mammary gland is specified as a small rudimentary tree-like structure (Figure S1A). During puberty, precursors localized at ductal tips (termed terminal end-buds) drive the expansion of a complex network through multiple rounds of tip bifurcation and ductal elongation (Sternlicht, 2006) (Figure S1A). These networks are characterized by the structural heterogeneity of their subtrees (that we defined formally as the parts of the ductal tree sharing a common branch ancestor at branch level $l_e = 6$—that reflects the approximate extent of the rudimentary structure prior to pubertal morphogenesis). Indeed, some ductal subtrees become extremely large, containing as many as 30 generations of consecutive branching events, while neighboring subtrees may terminate precipitously (Scheele et al., 2017).

Recently, we showed through quantitative genetic lineage tracing methods that the complexity of the mammary epithelium does not derive from intrinsic heterogeneity of tip precursor populations, but from the stochastic fate decisions of equipotent tips, which either branch (bifurcate) or terminate (through cell-cycle exit) with near equal-probability (Scheele et al., 2017), suggestive of a local control of tip fate. However, such a focus on spatially averaged models of branching morphogenesis (Zubkov et al., 2015) cannot resolve the spatiotemporal dynamics and mechanistic basis of the underlying regulatory program, nor its potential conservation in other organs.

How can such a balance between tip termination and branching be regulated at the population level? One possibility is that tip branching and termination rates are dependent on the local epithelial density. Indeed, an increase in the termination rate and/or a decrease in the branching rate with local density can ensure that the system reaches a robust steady state, characterized by a balance between tip branching and termination (Method Details). Interestingly, such behavior generically produces glands of uniform spatial density, a feature that we verified experimentally by reconstructions of whole adult mammary glands (n = 14 glands, Figures 1A and 1B). However, to understand whether it is branching or termination events that are actively regulated, we turned to quantitative measurements. In particular, parameterization of the distribution of mammary branch lengths (defined as the distance between consecutive branching points) revealed a strikingly exponential dependence, with an average branch length that remains approximately constant over time (Scheele et al., 2017). This observation suggests that the timing between consecutive branching events is random and statistically uncorrelated, pointing to a stochastic and time-invariant program of tip branching. This behavior stands in stark contrast with the early stages of lung morphogenesis, where branch lengths for a given branching level are tightly controlled (Iber and Menshykau, 2013).

We then examined the spatial organization of the ductal network. As the mammary fat pad constrains growth to a thin pancake-like geometry, ductal morphogenesis of the mammary epithelium takes place in a near 2D setting. Against this background, inspection of whole gland reconstructions revealed a strikingly low frequency of ductal crossovers (Figures 1A, 1C, 1D, and S1) with terminated tips often residing close to an existing duct or the fat pad boundary (Silberstein, 2001). This observation suggests that ductal elongation and branching may proceed as a "default state," with tip termination occurring only when tips come into proximity with existing ducts. Such behavior is consistent with *in vitro* measurements (Nelson et al., 2006), which show that ductal branching only occurs when remote from the other ducts, with tips close to neighbors remaining inactive.

Mammary Morphogenesis Proceeds as a Branching and Annihilating Random Walk

Interestingly, such a model of branching morphogenesis maps directly onto the theory of "branching and annihilating random walks," a class of models studied extensively by physicists (Cardy and Täuber, 1996) showing that the dynamics of binary tip-splitting models converge over time onto a common statistical behavior belonging to the universality class of "directed percolation." Here, we implemented a minimal model of branching morphogenesis, inspired by the theory of BARWs, where tip dynamics involves only three processes (depicted in Figure 2A): (1) ducts elongate from active tips in a random direction with a speed v—"a persistent random walk"—leaving behind a trail of static, non-proliferative ducts; (2) at any instant, ducts can branch through stochastic tip bifurcation with a constant probability r_b; and (3) ducts terminate through tip inactivation when tips come within an annihilation radius R_a of an existing duct.

Significantly, numerical simulations of the model dynamics in the absence of physical boundary constraints shows that the

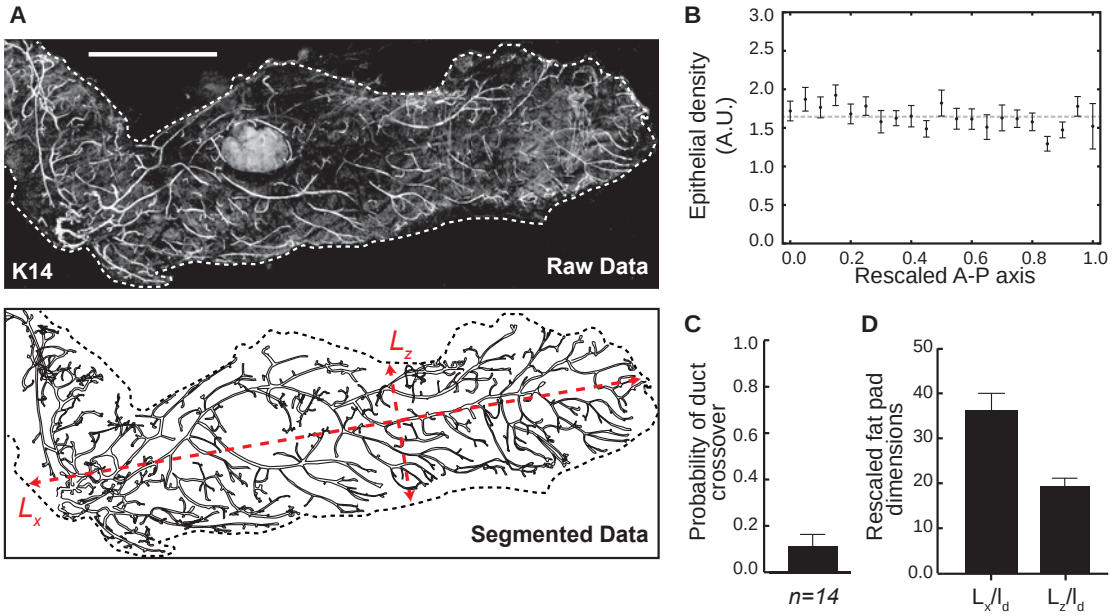


Figure 1. Geometry and Characteristics of Murine Mammary Glands Revealed by Quantitative Reconstructions

(A) Quantitative reconstruction (top) and outline (bottom) of a fourth mammary gland based on Keratin14 (K14) staining (white), reproduced from (Scheele et al., 2017) along with measurements of the fat pad dimensions L_x and L_z (red).

(B) Density profile of ducts along the rescaled antero-posterior axis.

(C) Counting of ductal crossovers (normalized by total number of ductal branches) reveal a low crossing probability.

(D) Experimentally measured ratio between the dimensions of the mammary fat pad (long axis L_x and short axis L_z) and the average length of a branch l_d, used for the simulations of the mammary gland. Error bars represent mean and SEM. Scale bar, 5 mm.

See also Figure S1.

system reaches robustly a non-equilibrium steady state in which the frequency of branching and termination events becomes naturally balanced (Figures S2A and S2B; Method Details). As tips are not observed to cross the boundary of the fat pad and frequently terminate in their proximity (Figure 1A), we further implemented simulations of branching morphogenesis in a rectangular box of length L_x and width L_z to mimic these geometric constraints. Thus, the only key parameter of the model is the ratio between the dimensions of the fat pad and the average branch length l_d (the latter fixed by the ratio v/r_b) (Figure 1D). Indeed, this geometrical parameter was fitted to its measured value (Figure S2; Method Details), so all subsequent comparisons with experiment represent the result of model predictions that do not involve the adjustment of any free parameter.

While a visual inspection of a typical simulation output revealed good qualitative agreement between the experiment and the theoretical predictions of the spatial organization (Figure 2B) and topology (Figure 2C) of the mammary ductal network, can such a simple model dynamics also provide quantitative insights? To address this question, we first quantified how the predicted frequency of tip bifurcation versus termination events evolves with branch level (i.e., the number of generations since the origin). Interestingly, as well as recapitulating long-term balance in the frequency of tip bifurcation and termination, we found that the model faithfully reproduced the dynamics of convergence toward balance, from an initial stage of symmetric branching early in pubertal development, where the ductal density is low (Figure 2D, R^2 = 0.73). Strikingly, the model also predicted with high precision the heterogeneity of subtrees in mammary glands (defined in Figure 2C), quantified both by the subtree size distribution (Figure 2E, R^2 = 0.96) and the subtree persistence to a given level (Figure 2F, R^2 = 0.99).

Importantly, the spatial model accounted more accurately for the abundance of very large subtrees, which appear due to spatial ''priming'' in low density regions, than the previously published ''zero-dimensional'' model (Scheele et al., 2017) in which tip branching and termination events are defined intrinsically and probabilistically. More generally, to explore the specificity of the model, we also considered the quantitative predictions made by eight further classes of models, corresponding to various alternative proposals from the literature. In each case, their applicability to the experimental data was found to be limited (see Figures S3 and S4A–S4D; Method Details).

Branching and Annihilating Random Walks Reproduce the Dynamics of Mammary Morphogenesis

Having established how the final state of the mammary epithelium is specified, we turned to examine whether the full dynamics of growth could also be predicted quantitatively. To gain insight into the nature and parametric dependences of the growth dynamics, we considered the hydrodynamic limit of the model in which the kinetics is captured by a mean-field theory, a

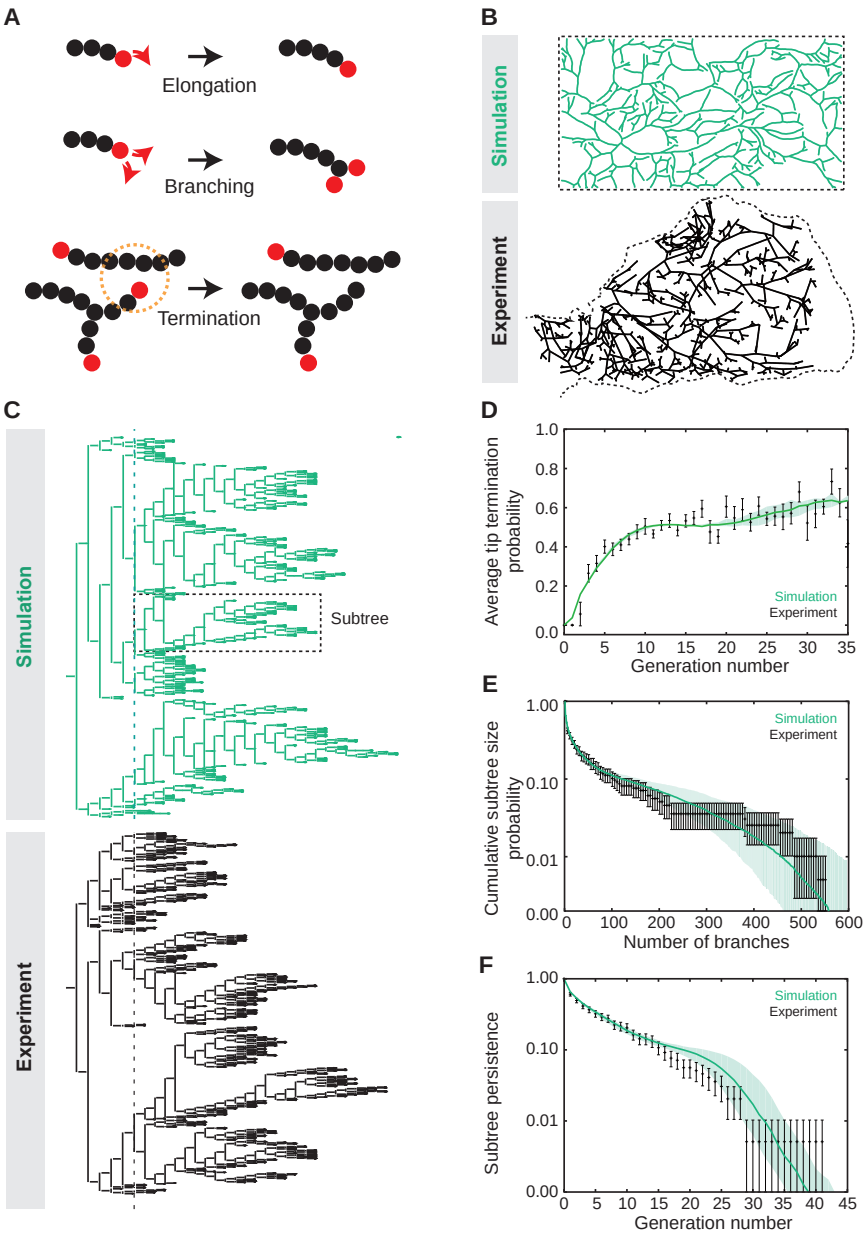


Figure 2. A Model Based on Branching and Annihilating Random Walks Predicts Quantitatively Mammary Branching Morphogenesis

(A) Schematic of the model. Active ductal tips choose between ductal elongation, stochastic branching through tip bifurcation, or termination when in proximity to a neighboring duct.

(B) Comparison between the experimental and theoretical structure of mammary glands.

(C) Comparison between the experimental and theoretical topology of the trees, displaying large heterogeneity, with different subtrees (defined as parts of the tree starting at level 6, delineated as dashed line, with a black box showing an example of a subtree) growing to widely different sizes.

(D–F) The BARW model predicts quantitatively the evolution of the probability for tips to terminate (D), the cumulative distribution of subtree size (E), and the subtree persistence to a given branch generation number (F). Data from Scheele et al. (2017). Shaded area and error bars in (E and F) represent mean ± 1 SD confidence intervals. Error bars in (D) represent mean and SEM. Black represents experiments and green theoretical predictions from simulations.

See also Figure S2 and Movie S1.

manifestation of a "two-species Fisher-KPP equation" (Fisher, 1937) (Method Details):

$$\begin{cases} \partial_t a = D\nabla^2 a + r_b a\left(1 - \dfrac{a+i}{n_0}\right) \\ \partial_t i = r_e a + \dfrac{r_b}{n_0} a(a+i) \end{cases}, \quad (1)$$

where $a(x,t)$ and $i(x,t)$ denote, respectively, the local concentration of active (tip) and inactive (duct) segments or "particles." Referring to the description of the model dynamics above, active particles diffuse with diffusion constant D while producing inactive segments at rate r_e (reflecting the process of ductal elongation), branch at rate r_b, and annihilate when they meet another particle (reflecting the process of tip inactivation), giving rise to a logistic growth term saturating at a total steady-state density, n_0 (for details of how Equation 1 emerges from the stochastic model and can be related to biological signaling pathways, see Method Details). Within this framework, both theory and numerical simulations predict that, during expansion, active tips become self-organized into a narrow pulse at the growing front of the developing epithelium, traveling at constant speed as a solitary wave and leaving in its wake an inactive ductal network of constant density (Figures 3A, 3B, and S4D–S4I; Movie S1).

From a biological perspective, this behavior provides a natural explanation for the constant speed of invasion, a robust feature of mammary morphogenesis (Paine et al., 2016). At the same time, the theory predicts that ducts should be spatially patterned at a constant density (Figures 1A and 1B), while active tips should localize in a predictable pulse-shape distribution at the edge of the invading front. To test these predictions quantitatively, we performed EdU-pulse labeling of mice at 5 weeks of age (approximately the mid-point of branching morphogenesis of the mammary gland) and used whole gland reconstruction to both quantify the morphology of the network (Figures 3C and S4J) and define the regional localization of active tips (defined as proliferative tips with

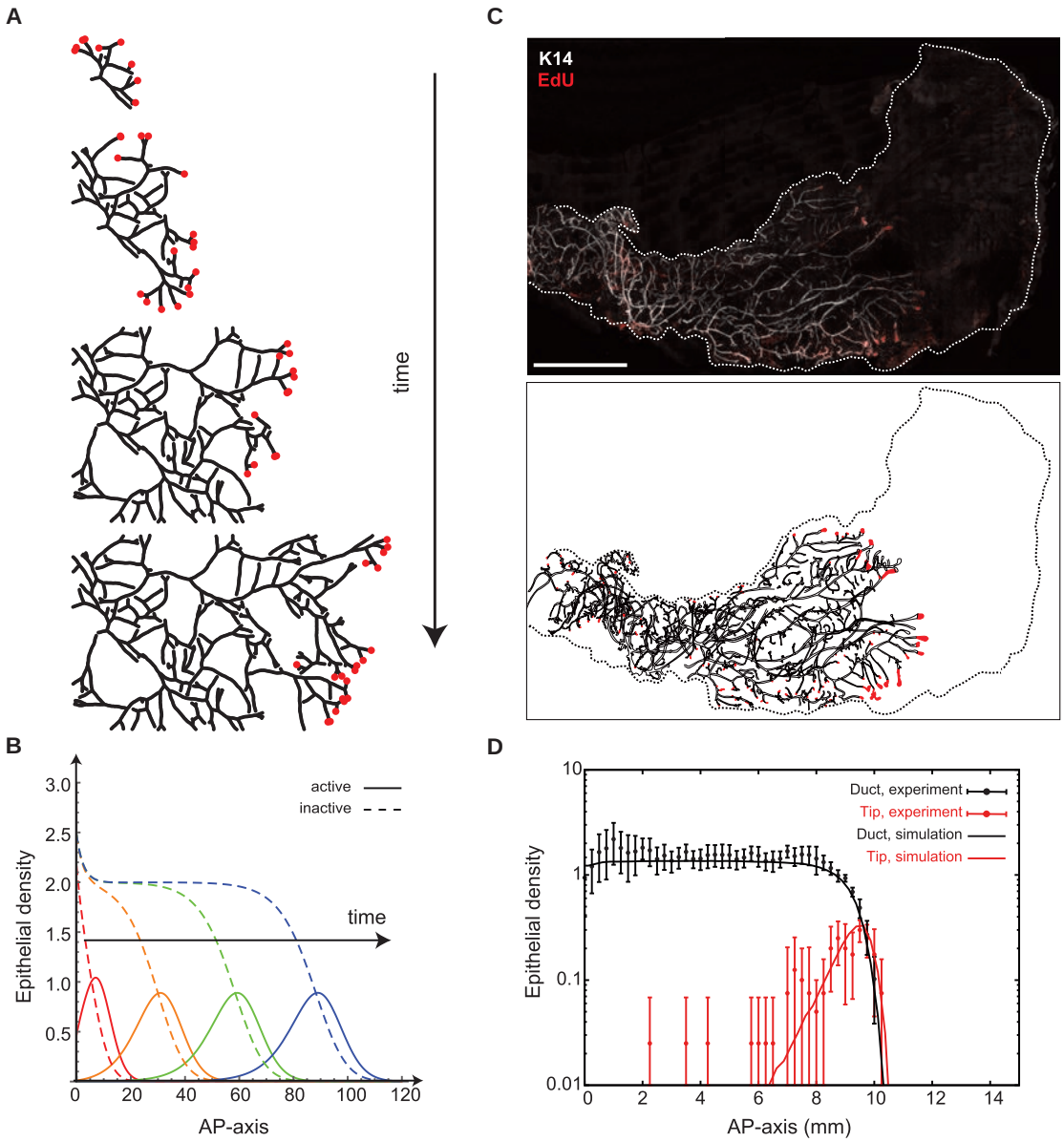


Figure 3. Branching and Annihilating Random Walks Reproduce the Kinetics of Mammary Invasion

(A) Numerical simulation of the model at different developmental time points with ducts shown in black and active tips in red.

(B) Theory predicts a self-organized solitary pulse of active tips positioned at the growing edge of the network, leaving behind a trail of inactive ducts of constant density.

(C) 3D reconstruction of a fourth mammary gland following an EdU pulse at 5 weeks showing the position of active tips. Active tips are localized preferentially at the invasion front, mirroring qualitatively the prediction of the model.

(D) Density profiles of ducts (black) and fully proliferative tips (red), averaged over n = 4 glands, alongside theory (red and black lines, respectively) revealing good quantitative agreement. Error bars represent mean and SEM. Scale bar, 5 mm.

See also Figure S3.

>50% of EdU$^+$ cells, Figures S4J and S4K). Importantly, we found good qualitative agreement between experiment and theory, with active tips present at the edge of the growing front and a remarkably constant density of trailing ducts (Figure 3D). Quantitatively, analysis of the spatial profile at the growing front showed that the density of active tips decayed exponentially both ahead and behind the front, with the decay length of the former larger than the latter by a factor of $(\sqrt{2}-1)$, all key and non-trivial predictions of the Fisher-KPP dynamics (Figures S4F–S4I; Method Details).

Together, these results suggest that the global spatiotemporal dynamics of mammary ductal morphogenesis can be

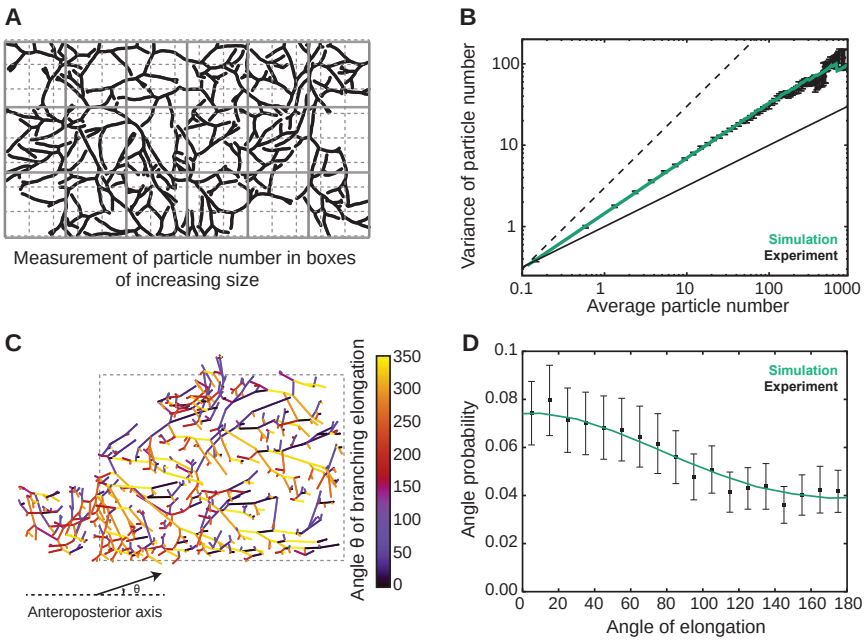


Figure 4. Self-Organized Properties of BARWs Predict Both Giant Density Fluctuations and Emergent Directional Bias of Ducts

(A and B) Experimental variance (y axis) versus average (x axis) of duct volume in boxes of increasing size, *L* (A). The variance in density of the gland at different length scales grows as a power law (B, black bars), with an exponent larger than 1/2 (B, thin and dashed black lines represent exponents of 1/2 and 1, respectively), indicative of giant number fluctuations, and quantitatively predicted by the BARW model (green line).

(C and D) Self-organized directional invasion proceeds from local negative interactions. (C) Representative example of the outline of an 8-week fourth mammary gland (same as Figure 2B), where the angle θ of each branch segment is calculated relative to the AP-axis. (D) Experimental (black bars) and theoretical (green line) distributions showing probabilities of finding a branch growing with a given angle θ. The experimental distribution is predicted quantitatively by the model even in the absence of a large-scale directional gradient. Error bars indicate mean and SEM.

See also Figures S4 and S5.

understood as a process of self-organization following from a program of stochastic tip bifurcation arrested by tip termination at the intersection with neighboring ducts.

Giant Density Fluctuations and Self-Organized Directional Invasion during Mammary Morphogenesis

Although the proposed mechanism of branching morphogenesis can ensure a uniform density of ducts, statistical fluctuations during growth generate large spatial variations in the distribution of active EdU$^+$ tips (Figure 3C). Indeed, the EdU-pulse assay reveals duct-depleted regions formed either by chance mass termination of tips (Figure S5A) or locally ''divergent'' flows of active tips randomly exploring other regions (Figure S5B), both behaviors being well-reproduced in the numerical simulations of the model dynamics. Importantly, according to the rules of the model dynamics, the trailing distribution of newly formed ducts is frozen or ''quenched'' in the fat pad. Therefore, we expect that the statistical fluctuations of epithelial density should persist in the mature network. Thus, in addition to the prediction of the average density profiles of active tips and mature ducts, the model makes further key quantitative predictions on the statistical properties of spatial density fluctuations.

We thus quantified these fluctuations by defining the spatial average, n_L, and SD, $(\Delta n)_L$, of duct volume in boxes of viable size *L* (see Figure 4A for a schematic). For systems at equilibrium (in which each elemental process is equilibrated by its reverse, the property of detailed balance), the central limit theorem requires that $(\Delta n)_L = n_L^{\alpha}$ with the exponent $\alpha = 1/2$. By contrast, in systems characterized by non-equilibrium fluctuations, α can take values larger than 1/2 (Ramaswamy et al., 2003; Narayan et al., 2007)—the phenomenon of giant number fluctuations. Indeed, using the same parameter set as before, model simulations revealed a robust power law dependence of $(\Delta n)_L$ (Figure 4B, green line), with an exponent $\alpha_{theory} \approx 0.66$, that increased with decreasing branching rate (Figure S5C).

Turning to previous mammary gland reconstructions at 8 weeks of age, we found many instances of large spatial density fluctuations that could not be accounted for by boundary effects, or by the presence of obstacles such as lymph nodes. We therefore applied the same statistical approach to determine experimentally the quantitative dependence of $(\Delta n)_L$ and n_L (n = 14 glands from 7 mice). Strikingly, this analysis revealed a robust power law dependence over more than three orders of magnitude (black dots, Figure 4B), with an exponent of $\alpha_{exp.} \approx 0.65 \pm 0.02$ (mean ± SEM), consistent with giant number fluctuations *in vivo*. Moreover, the experimental data collapsed on the theoretical curve with extremely high precision (Figures 4B and S5D), emphasizing the robustness of the model prediction ($R^2 = 0.90$, $R^2_{log} = 0.99$). Overall, this analysis uncovers an unexpected out-of-equilibrium feature of branching morphogenesis *in vivo* and serves as a strong test of the validity and predictive power of the BARW model. In particular, this shows that, while the proposed mechanism enforces (in a self-organized manner) a robust and constant averaged epithelial density, the local density is, as a result, only weakly regulated.

A further ubiquitous feature of mammary gland morphogenesis is the appearance of directional biases in the growth of the ductal network, suggestive of a mechanism that guides tips distally (Figures S5E–S5J). Indeed, quantification of the distribution of angles θ between a given branch and the horizontal proximal-distal axis (Figures 4C and S5K) revealed a 2-fold bias toward a proximal-to-distal orientation (Figure 4D). A puzzle in the field has been the lack of identification of any large-scale gradient that could cause this anisotropy (Gjorevski and Nelson,

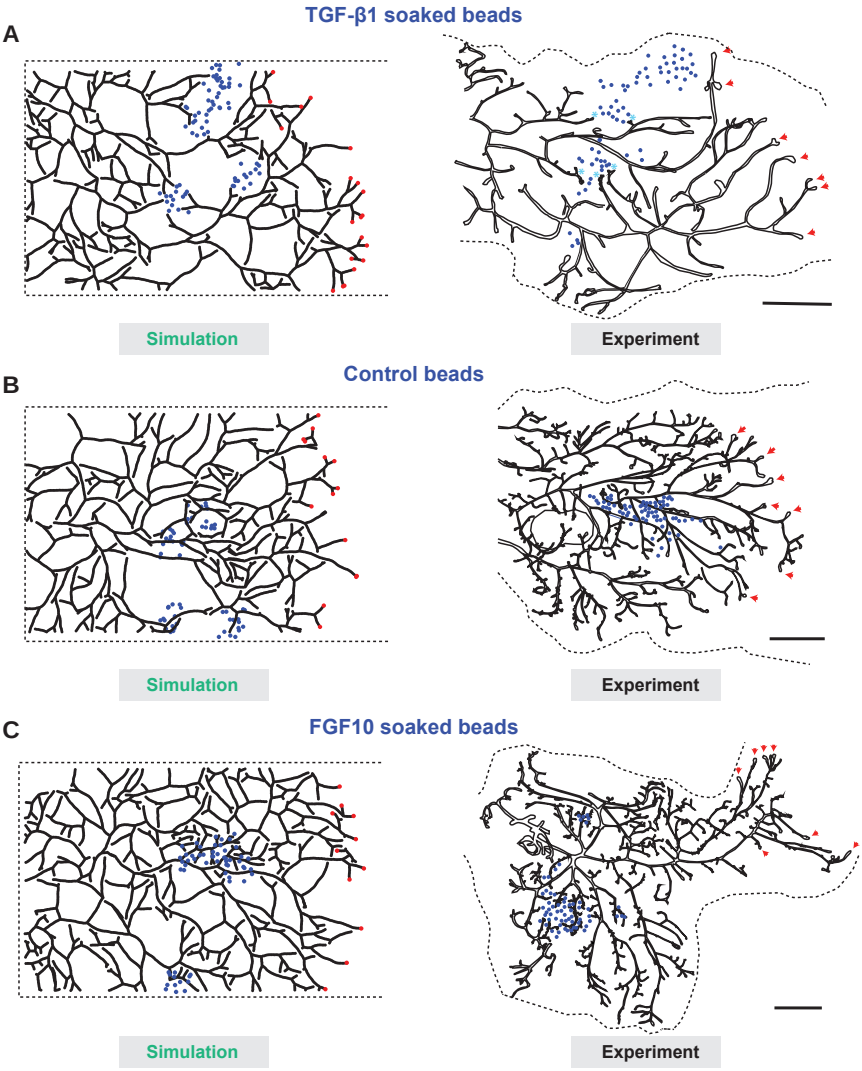


Figure 5. Perturbation Experiments Reveal the Molecular Basis of Termination and Branching

(A) Comparison between a representative reconstructed fourth mammary gland (right) in the presence of TGF-β1 soaked beads (blue spheres) and a simulated theoretical counterpart (left). This confirms tip termination in proximity (light blue asterisks) to TGF-β1 soaked beads.

(B) Comparison between a representative reconstructed fourth mammary gland (right) in the presence of control inactive beads soaked in PBS with 0.1% BSA (blue spheres) and a simulated theoretical counterpart (left).

(C) Comparison between a representative reconstructed third mammary gland (left) in the presence of FGF10 soaked beads (blue spheres) and a simulated theoretical counterpart with local 2-fold increase in branching rate (right). Active TEBs are marked by red arrowheads. Scale bar, 2 mm.

See also Figure S5.

2011). However, we reasoned that such a directional bias could derive naturally from the BARW model, even in the absence of global chemical cues or gradients, because branches growing toward the proximal region are more likely to terminate against existing ducts (i.e., less likely to give rise to progeny), resulting in an "effective" and self-organized bias emerging from isotropic short-range interactions. To test this hypothesis, we computed the theoretical prediction from the same model as above. Strikingly, the model was able to predict quantitatively the experimental profile (R^2 = 0.95, Figure 4D), suggesting that directional bias may simply emerge as a natural consequence of the BARW model.

Molecular Basis of Tip Termination and Branching

Finally, given the importance of tip annihilation in our framework, we sought to test in a more direct way its underlying molecular basis. Ectopic delivery of TGF-β by large pellets has been shown to reversibly inhibit mammary ductal growth (Silberstein and Daniel, 1987). Therefore, to test the local action of TGF-β signaling, we implanted small TGF-β1-soaked agarose beads into the mammary fat pads of 4-week-old mice and waited for 2 weeks before sacrificing the mice (Figure 5A; Method Details). Importantly, as predicted by the theoretical simulations (Figures 5A, S5L, and S5M; Method Details) and as opposed to experiments with control beads soaked in PBS with 0.1% BSA (Figures 5B, S5N, and S5O), we found that mammary ducts never colonized regions rich in TGF-β1 beads, while we could observe numerous events of tips having stopped in their close proximity (100–200 μm, blue asterisks on Figure 5A). By contrast, the branching pattern was unaffected in regions devoid of beads (Figures 5A and S5P). These observations support the hypothesis that chemical signaling from maturing ducts regulate the termination of active terminal end-buds and implicate a role for TGF-β1 in providing the cue in a very local manner.

Next, we wished to assess quantitatively the effect of known positive regulators of branching morphogenesis. We thus performed a similar assay using FGF10-soaked beads (Figure 5C), as FGF10 has been identified as the predominant stromal FGF ligand expressed during pubertal mammary morphogenesis (Zhang et al., 2014). Notably, we found that FGF10 induced a 2-fold increase in branching, consistent with its proposed role in driving branch initiation in *in vitro* studies (Zhang et al., 2014), with a corresponding densification of the network close to beads, which was well-reproduced in model simulations (see Figure 5C; Method Details).

Together, these two sets of experiments provide both an additional test of the branching and annihilating random walk framework and a molecular basis for the regulatory program.

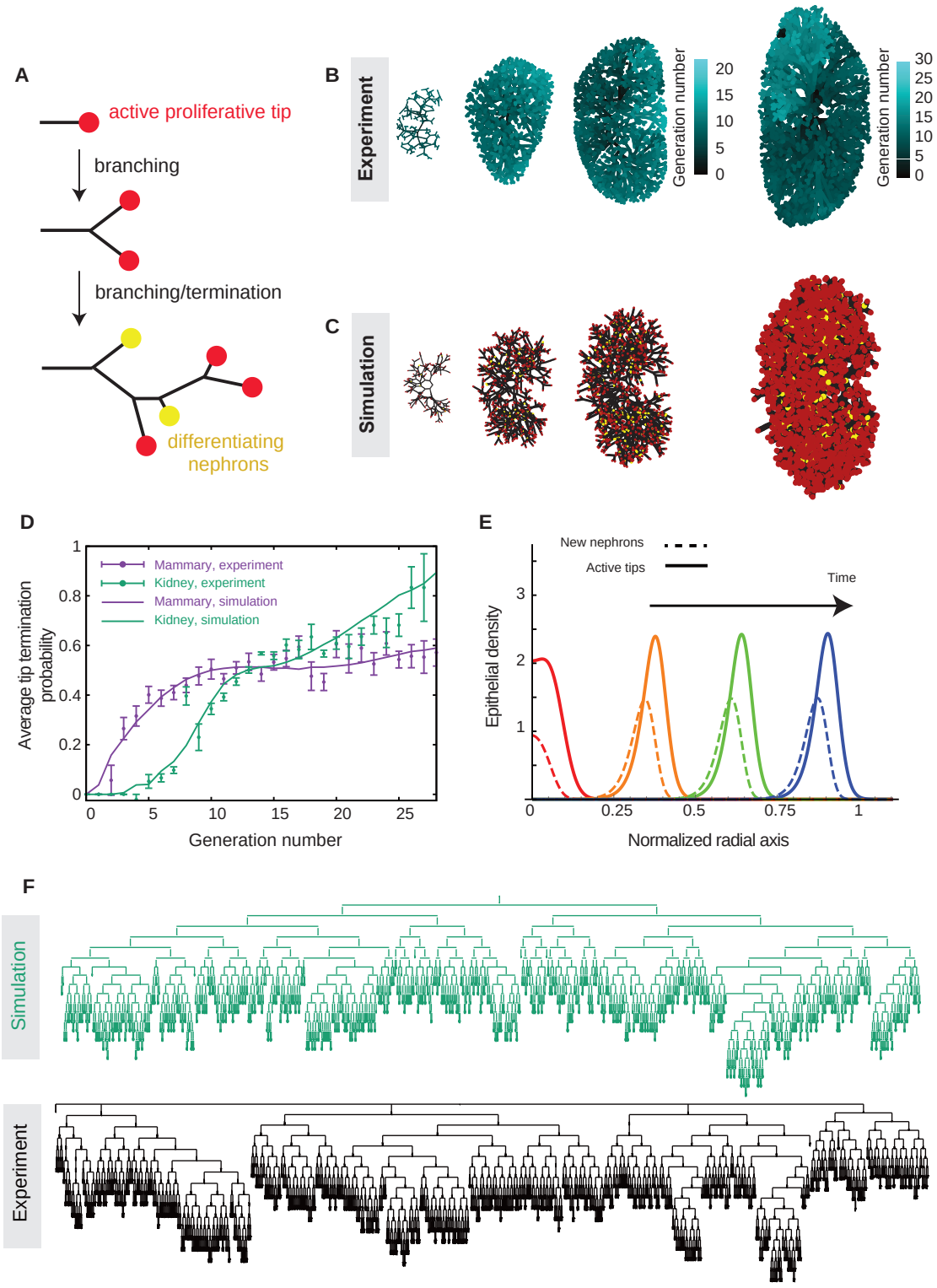


Figure 6. Branching and Annihilating Random Walks Can Reproduce Quantitatively the 3D Kidney Topology

(A) Schematic of kidney morphogenesis as a stochastic branching process where active tips (red) either elongate, branch, or stop contributing to branching via nephron differentiation (yellow).

(B) Reconstructions of murine kidney at E13, E15, E16, and E18 (left to right) with generation number of segments color coded in blue.

(C) Typical output of numerical simulations of BARW model at corresponding time points.

(legend continued on next page)

Kidney Morphogenesis as a 3D Branching-Annihilating Random Walk

So far, we have restricted our analysis to the quasi 2D geometry of the mouse mammary gland. Therefore, to address the potential generality of the model to other organs, we considered the 3D incarnation of the BARW model using the development of kidney as a 3D system. During kidney morphogenesis, the ureteric bud, a single outgrowth that arises around embryonic day 11 (E11) from the nephric duct, arborizes to form the collecting system through a repeating process of mainly dichotomous branching. During the course of this iterative branching process, tips induce an aggregate of adjacent cap mesenchyme to undergo a mesenchymal to epithelial transition, thereby initiating the first steps of nephrogenesis, i.e., the formation of nephrons, the kidney's filtration unit. These aggregates continue to mature while the renal connecting tubule concomitantly forms and joins these nascent nephrons with branching collecting ducts (Short et al., 2014; Sampogna et al., 2015; Cebrián et al., 2004). Crucially, as kidney development progresses, a growing subset of older ureteric tips continue to fuse with adjacent maturing nephrons. Once occupied, these tips are thought to no longer contribute to further branching (Sampogna et al., 2015; Costantini and Kopan, 2010) so that they can be considered to have undergone branching termination (Figure 6A).

Motivated by these findings, we thus considered whether the BARW model could predict kidney morphogenesis. Indeed, the convergence of the BARW model toward balanced ductal bifurcation and termination described above is quite general, applying in all dimensions. However, simulating the model dynamics in 3D (Figures 6B, 6C, and S6) revealed that this convergence, for the same annihilation radius, occurs on much longer timescale (around 3 versus 10 generations, on average, Figure 6D). This behavior can be explained intuitively through differences in the frequency of random collisions between ducts and tips, which become much rarer in 3D as compared to 2D. From a biological perspective, this would mean that the topology of 3D branched organs should appear to be predominantly geometric (deterministic) early in development, displaying serial rounds of symmetric branching events without termination and only later becoming stochastic in character. Interestingly, such behavior is qualitatively consistent with recent reports by several groups using detailed 3D reconstructions (Short et al., 2014; Sampogna et al., 2015), showing structural heterogeneity only at higher branch levels. We therefore analyzed original and more recent data from (Sampogna et al., 2015), involving kidney reconstructions from E12 to E19, to test whether the same framework could apply during the seemingly non-stereotypical later phase of morphogenesis (Figures 6C and S6).

To develop a more precise quantitative comparison, we considered a numerical simulation of the branching dynamics in an unconfined 3D geometry (Figures S6A–S6C; Movie S2). In this case, the dynamics depends only on the ratio of the annihilation radius to the characteristic duct length, $R'_a = R_a/l_d$ (Figures S6D–S6F). In contrast to the 2D setting, this parameter becomes crucial in 3D, where the probability of two branches to cross becomes of measure zero. Moreover, as kidney expands anisotropically, we renormalized all rate constants with respect to growth orientation to match the experimental aspect ratio (Figure S6C; Method Details).

Interestingly, with $R'_a = 0.25$, a value close to that found for mouse mammary gland, we could reproduce with high precision the growth characteristics of the E19 mouse kidney, as exemplified by the evolution of the tip branching versus termination probability as a function of branch level (Figure 6D). As mentioned above, ductal evolution is characterized by a protracted early phase of symmetric branching, converging slowly toward balanced fate. R'_a is thus the key and only fitting parameter, and all subsequent comparisons to experiments do not involve the adjustment of any additional parameter.

We then compared the experimental and theoretical topologies of kidneys (Figure 6F), as well as the distributions of branch number as a function of branch level across a wide range of developmental time points (Figures 7A, S6H, and S6I). Given the simplicity of the model, these results showed remarkably good correspondence, revealing an initial phase of geometric ductal expansion (with the number of branches at level n growing as 2^n), followed by a plateauing and widening of the distributions, a manifestation of increasing ductal termination (R^2 = 0.93 at E13, R^2 = 0.95 at E15, R^2 = 0.94 at E17, and R^2 0.93 at E19, Figures S6I–S6L). Importantly, this behavior does not arise from purely geometric anisotropies, as can be seen in the same simulation without termination events $(R'_a = 0)$ (Figure S6J), or with termination in an isotropic geometry (Figures S7A and S7B).

Stochasticity in Kidney Morphogenesis and Nephron Number Specification

In common with the mammary epithelium, growing tips were also predicted to become self-organized into a pulse of activity at the periphery of the developing kidney, while newly formed nephrons were predicted to form as a secondary pulse behind this front (Figure 6E). Such behavior matched the known organization of the kidney into a nephrogenic zone positioned at the growing periphery of the tissue (Figure S7C) (Sampogna et al., 2015; Short et al., 2014). Moreover, although the time evolution of both the active tip and nephron density can depend on time variations in branching rate, plotting one versus the other provided a robust, time-independent test of model.

Using glomeruli, the capsule of capillaries located at the beginning of a nephron (Sampogna et al., 2015), as a proxy for the

(D) Experimental versus theoretical tip termination probability as a function of generation for mammary gland (purple) and kidney (green), using the radius of termination R'_a = 0.25 as the only fitting parameter.

(E) The model predicts a self-organized zone of active tips growing at the periphery of the kidney, followed spatially by a domain of tip termination, reminiscent of the nephrogenic zone observed *in vivo*.

(F) Tree representation of a E17 kidney branching topology (top, green) and the output of the theory at the corresponding time point (bottom, black). Error bars represent mean and SEM.

See also Figure S6 and Movie S2.

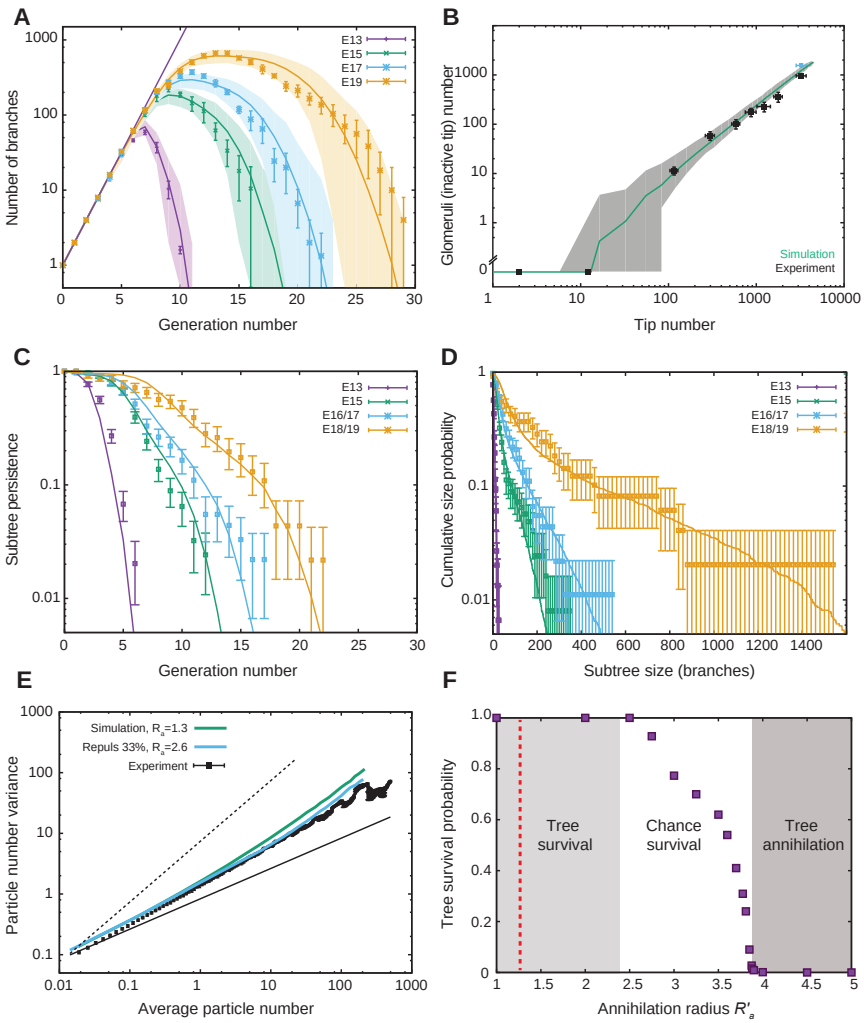


Figure 7. Branching and Annihilating Random Walks Can Reproduce Quantitatively the Detailed Properties of Kidney

(A) Using the radius of termination R_a as the only free parameter, the model predicts well the number of segments per generation at different time points of embryo development (E13, E15, E17, and E19 in purple, green, blue, and orange, respectively).

(B) Inactive tip number (assessed indirectly via glomeruli staining from [Sampogna et al., 2015] in black, or glomeruli counting via a method of acid maturation from [Cebrian et al., 2014] in blue) versus total number of tips, displaying a power law after a phase of purely symmetric branching, predicted by the model (green).

(C) Subtree persistence at different developmental time points (squares) compared to the model (lines).

(D) Cumulative subtree size distribution at different developmental time points (squares) compared to the model (lines).

(E) Variance (y axis) versus the average (x axis) duct volume in a box of size L (experiments in black) in kidney, showing an exponent larger than 1/2 (thin and dashed black line represent exponents of 1/2 and 1, respectively), indicative of giant number fluctuations. The green and blue lines are predictions from default model (no repulsion, R'_a = 0.25), and model with repulsion (f_r = 0.33, R'_a = 0.5).

(F) Tree survival probability versus termination radius, showing a phase transition above which kidney systematically become fully annihilated. Red dashed line shows the best-fit value of R_a used in (A)–(D). Shaded areas represent 95% confidence intervals, and error bars represent mean and SEM.

Lines in (A)–(D) are model predictions, using the parameter R'_a fitted from Figure 6D. See also Figure S7 and Movies S2 and S3.

number of maturing nephrons, we found that, after an initial phase of pure tip production without any nephrons, both quantities robustly scaled experimentally, with the relationship well-fit by a power law (Figures 7B, S6K, and S6L). Such scaling argues for simple time-invariant rules underlying nephron specification, because deviations from this time-invariance would cause deviations from the observed scaling (as simulated in Figure S7D). Crucially, we then compared this observation to our model, using the same parameter set as that used above, and found a good prediction for the scaling relationship throughout the entire time course of embryonic development (R^2 = 0.78, $R^2_{log} = 0.97$).

To determine whether the model could also predict the detailed heterogeneity of branching structures as well as the averages, we examined the size distributions and persistence of subtrees at each time point (defined as in mammary gland). These were consistently broadly distributed, indicative of large-scale heterogeneity, and adopted similar functional dependences to their mammary counterparts, indicative of conserved (or universal) underlying properties. Crucially, the distributions of subtree persistence were consistently very well-fit by the model at all developmental time points (Figure 7C, R^2 = 0.95 at E13, R^2 = 0.99 at E15, R^2 = 0.97 at E17, and R^2 = 0.97 at E19) as were subtree size distributions (Figure 7D, R^2 = 0.89 at E13, R^2 = 0.99 at E15, R^2 = 0.99 at E17, and R^2 = 0.93 at E19).

Finally, although the model captures successfully several nontrivial features of the experimental data, arguing for conserved rules underlying both mammary gland and kidney morphogenesis, we noted that kidney reconstructions were characterized by a rather regular spacing between tips, which was consistently more ordered than the numerical simulations of the model (Figures S6H and S6I). Indeed, although we found again evidence of giant number fluctuations (Figure 7E), the model slightly overestimated the amplitude of fluctuations (Figure 7E), hinting that tips may be partially self-avoiding, as proposed by (Davies et al., 2014). To incorporate this effect into the model, we proposed that tips, in addition to their persistent random motion, are repulsed by neighboring tips and ducts when within a radius R_r. With v_r defining the characteristic speed change induced by

this repulsion, the ratio $f_r = v_r/v$ provides a measure of the strength of the repulsion (Quantification and Statistical Analysis). For large values of f_r, termination becomes extremely rare, yielding a behavior inconsistent with the observed degree of kidney heterogeneity and nephrogenesis kinetics. However, for small values of $f_r = 0.2$ and a correspondingly larger value of the annihilation radius R_a (Quantification and Statistical Analysis), we could still obtain a satisfactory fit to the data (Figures S6J and S6L), while obtaining a more ordered kidney structure (Figure S6I). Interestingly, without further adjustment, this improved the fit to the observed density fluctuations ($R^2 = 0.75$, $R^2_{log} = 0.98$, Figure 7E).

These findings argue that, although self-avoidance is not the dominant characteristic of kidney morphogenesis (Figures S6M–S6O), it may cooperate with termination, i.e., nephron maturation, to produce a partially ordered structure. Notably, both elements of the model arise from purely local rules, maintaining the self-organizing character of branching morphogenesis. Indeed, this might explain why it can proceed robustly *in vitro* in the absence of external chemical or morphogen gradients (Davies et al., 2014).

Branching Defects as a Route to Premature Termination of Branching Morphogenesis

Based on these findings, we then questioned their implications for pathologies of branched organs, such as kidney, which have been linked to defects in branching morphogenesis. For instance, hypertension has been proposed to be at least partially explained by insufficient nephron number (Brenner et al., 1988), whereas renal agenesis is a relatively frequent congenital defect in humans, mirroring the *GDNF* knockout in mice, that results in the formation of tiny rudimentary ductal trees in the kidney (Pichel et al., 1996).

Interestingly, by the stochastic nature of the BARW, chance events may also lead to the premature extinction of active tips, which annihilate against the existing ductal network, inhibiting kidney growth. Although the frequency of such events is negligibly small for the parameters of wild-type tissue (Figures 7F and S7E), higher values of the annihilation radius cause the extinction probability to increase dramatically (Figures 7F, S7F, and S7G; Movie S3). Numerical simulations reveal a critical point above which early extinction always occurs (Figure 7F; Quantification and Statistical Analysis), as well as a continuous transition to a non-zero extinction probability below this critical value, which could explain the variable nature of small ductal trees observed in *GDNF* knock-out mice.

Based on this insight, we examined branching morphogenesis on E15.5 littermate kidneys that developed under the condition of mild maternal-fetal Vitamin A-deficiency as described previously (Sampogna et al., 2015). These kidneys are nearly two times smaller than wild-type (Figures S7H and S7I) and display larger subtree size heterogeneity (n = 4 mice, $p < 0.05$), although the total number of branches remains normal (Sampogna et al., 2015). Noting that vitamin A-deficient kidneys were markedly smaller, we tested whether this measured decrease in branch length was enough to reproduce the enhanced heterogeneity, by producing earlier crowding-induced tip termination (Method Details). We found that a uniformly decreased branch length was indeed sufficient to reproduce quantitatively the changes in branch generation distribution in vitamin A-deficient mice (Figure S7J). Further studies will be needed to address, more generically, whether such mutant conditions can be understood in terms of ratio between branch length and termination radius, i.e., of their proximity to the annihilating critical point.

Balance between Tip Termination and Branching Is Also Observed in Human Prostate

Finally, to further explore the generality of the proposed mechanism of branching morphogenesis, we turned to consider the human prostate, which consists of independent subunits branching independently from the urethra (McNeal, 1968). Organogenesis of the prostate shares key features of tip driven morphogenesis as described above in breast and kidney formation: the adult branching structure derives from epithelial ductal outgrowths into surrounding urogenital mesenchyme during embryogenesis and the immediate postnatal period (Powers and Marker, 2013).

From the tracing studies of ductal subtrees, based on large-scale 3D reconstructions of adult human prostates (n = 5 from 5 patients), some of which extended to 70 generations of branching, we found that some regions terminate early, with tips forming differentiated acini structures, while others grow extensively (Figure S7K). From a plot of the relative probability of tip termination versus branching, we found again that the overwhelming majority of subtrees displayed a striking degree of balance between ductal termination and branching (Figure S7L). Additionally, we found that the functional shape of the distribution was again similar to the other organs, with a few subtrees growing to up to 10 times the average subtree size (Figure S7M). These findings suggest that the paradigm uncovered for mammary gland and kidney morphogenesis may be translated to a priori different biological settings.

DISCUSSION

In this study, we have investigated how the branching pattern of the mouse mammary gland epithelium and kidney emerge throughout development. Using a combination of whole-organ 3D reconstruction, proliferation kinetics, and biophysical modeling, we have provided evidence that branching morphogenesis proceeds from the spatial competition of equipotent tips, which randomly explore space through a process of ductal elongation and stochastic branching. If this process occurred without competition between growing tips, branched organs would be characterized by stereotypical rounds of purely symmetric branching, with the number of branches increasing with branch level n as 2^n. Indeed, such behavior would serve to minimize the time required to build a branched structure while filling space efficiently. However, reconstructions of mouse mammary gland, kidney, and human prostate reveal a different scenario, where tip terminations occur even at the earliest stages of branching morphogenesis and rapidly balance tip bifurcations at the population level.

Based on the scarcity of ductal crossovers in mammary gland, we propose that the dominant source of tip termination is the

presence of neighboring ducts inhibiting growth. This provides a density-dependent feedback that naturally balances ductal branching and tip termination. This hypothesis challenges the concept of branching morphogenesis occurring through a rigid and deterministic sequence of genetically programmed events, and replaces it by a stochastic self-organizing model of development. After deducing the branching rate from *in vivo* measurements, our model predicts nearly perfectly and without adjustable parameters the network topology and spatial structures of adult mammary glands, while also making a number of additional non-trivial quantitative predictions.

In particular, it predicts the self-organization of active tips into a spatial domain or pulse, localized at the growing front of the network, which invades into the fat pad at constant speed, leaving behind a constant density of mature ducts. As a consequence, our model suggests that the directional invasion of the mammary gland toward the distal end of the fat pad does not need to be guided by a global chemotactic gradient, but instead can be explained quantitatively in a self-organized manner from the short-range annihilating properties of tips and ducts. As a non-equilibrium process, the BARW model predicts quantitatively the existence and scaling dependence of hallmark giant density fluctuations, which we verify experimentally.

Finally, we have shown that the model applies equally well in the 3D setting of the developing mouse kidney, reproducing accurately the network heterogeneity, with some subtrees colonizing large parts of the kidney while others terminate precipitously, as well as the spatiotemporal pattern of nephrogenesis. Such behavior suggests that this self-organized pattern of growth, consistent with the *in vitro* growth capability of kidney trees, may constitute a conserved (universal) mechanism of branching morphogenesis across different tissues, shifting the focus of future studies to the collective spatiotemporal fate control of branching and termination of entire tips, rather than on individual cells.

From a molecular mechanistic perspective, some of the processes underlying tip termination have been studied individually in several organs. In particular, inhibition of tip growth through TGF-β signaling has been demonstrated in both mammary and prostate glands (Silberstein, 2001; Powers and Marker, 2013). TGF-β is also a good candidate to provide crowding-induced feedback (Silberstein, 2001), as it is known to be diffusible in the stroma, is secreted by mature ducts, and has been shown both *in vitro* and *in vivo* to regulate the branching pattern of pubertal morphogenesis (Silberstein and Daniel, 1987), as confirmed here. Moreover, it was recently reported from *in vitro* culture experiments that the TGF-β superfamily, in particular Bmp7, was also implicated in crossover avoidance in kidney (Davies et al., 2014).

Thus, these findings question the underlying molecular basis of the BARW model. Given the diffusible nature of key underlying regulators, we investigated whether generic reaction-diffusion models could explain the BARW phenomenology. Interestingly, we found that such branching and annihilating dynamics can indeed emerge naturally and robustly from simple Turing-Meinhardt type models (Meinhardt, 1982; Guo et al., 2014) involving only spatial interactions of an activator, an inhibitor, and a consumed substrate (Figure S7N; Method Details). Taken as a whole, our study demonstrates that the morphogenesis of complex ductal tissues can be understood and predicted quantitatively on the basis of a remarkably simple set of local rules that direct the robust self-organization of a large-scale network structure.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Mice Experimental Details
  - Local application of Affi-gel blue beads in mammary glands
  - Whole-mount immunofluorescence staining of mammary glands
  - Whole-mount imaging of mammary glands
  - Models of branching and annihilating random walk
  - Mean-field dynamics of front invasion
  - Revised model of tip termination
  - Mammary gland parameters and numerical simulations
  - Alternative models of mammary branching morphogenesis
  - Theoretical analysis of bead implantation
  - Numerical simulations in 3D
  - Key parameters for the default kidney simulations
  - Mapping between simulated and real embryonic time in kidney
  - Modeling of Vitamin A deficient kidney
  - BARW from reaction diffusion equations
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantifications
  - Goodness-of-fit statistics
  - Statistics for bead experiments
  - Sensitivity analysis on mammary gland simulations
  - Kinetics of mammary gland invasion
  - Giant density fluctuations in kidney
  - Sensitivity analysis on kidney simulations

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and three movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.026.

AUTHOR CONTRIBUTIONS

E.H., C.L.G.J.S., J.v.R., and B.D.S. conceived the study and designed the experiments. C.L.G.J.S., M.M., R.H., N.D., and R.V.S. performed experiments and contributed data. E.H., C.L.G.J.S., M.M., and R.V.S. performed analyses. E.H. and B.D.S. developed the theoretical framework. C.L.G.J.S. and E.H. made the figures. J.v.R. and B.D.S. supervised the study. All authors discussed results and participated in the preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Anko de Graaff and the Hubrecht Imaging Centre for imaging support and the Hubrecht Institute animal caretakers for animal support. This work was supported by an ERC consolidator grant (648804), research grants from the Dutch Organization of Scientific Research (NWO; 823.02.017), the Dutch Cancer Society (KWF; HUBR 2009-4621), the Association for International Cancer Research (AICR; 13-0297) (all J.v.R), the Wellcome Trust (110326/Z/15/Z to E.H. and 098357/Z/12/Z to B.D.S.), and equipment grants from the Dutch Organization of Scientific Research (NWO; 175.010.2007.00 and 834.11.002). E.H. is funded by a JRF from Trinity College and acknowledges the Bettencourt-Schueller Young Researcher Prize for support. C.L.G.J.S. is funded by a Boehringer Ingelheim Fonds PhD Fellowship. R.S. was supported by the Norman S. Coplon Extramural Grant. R.H. and M.M. were funded by a Cancer Research UK Clinician Scientist Fellowship (Ref C10169/A12173).

Received: January 24, 2017
Revised: June 20, 2017
Accepted: August 15, 2017
Published: September 21, 2017

REFERENCES

Brenner, B.M., Garcia, D.L., and Anderson, S. (1988). Glomeruli and blood pressure. Less of one, more the other? Am. J. Hypertens. *1*, 335–347.

Cardy, J., and Täuber, U.C. (1996). Theory of branching and annihilating random walks. Phys. Rev. Lett. *77*, 4780–4783.

Cebrián, C., Borodo, K., Charles, N., and Herzlinger, D.A. (2004). Morphometric index of the developing murine kidney. Dev. Dyn. *231*, 601–608.

Cebrian, C., Asai, N., D'Agati, V., and Costantini, F. (2014). The number of fetal nephron progenitor cells limits ureteric branching and adult nephron endowment. Cell Rep. *7*, 127–137.

Costantini, F., and Kopan, R. (2010). Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. Dev. Cell *18*, 698–712.

Davies, J.A., Hohenstein, P., Chang, C.-H., and Berry, R. (2014). A self-avoidance mechanism in patterning of the urinary collecting duct tree. BMC Dev. Biol. *14*, 35.

Elliott, H., Fischer, R.S., Myers, K.A., Desai, R.A., Gao, L., Chen, C.S., Adelstein, R.S., Waterman, C.M., and Danuser, G. (2015). Myosin II controls cellular branching morphogenesis and migration in three dimensions by minimizing cell-surface curvature. Nat. Cell Biol. *17*, 137–147.

Fisher, R.A. (1937). The wave of advance of advantageous genes. Ann. Eugen. *7*, 355–369.

Gjorevski, N., and Nelson, C.M. (2011). Integrated morphodynamic signalling of the mammary gland. Nat. Rev. Mol. Cell Biol. *12*, 581–593.

Guo, Y., Sun, M., Garfinkel, A., and Zhao, X. (2014). Mechanisms of side branching and tip splitting in a model of branching morphogenesis. PLoS ONE *9*, e102718.

Harrison, L.G. (2010). The Shaping of Life: the Generation of Biological Pattern (Cambridge University Press).

Hogan, B.L. (1999). Morphogenesis. Cell *96*, 225–233.

Honeth, G., Schiavinotto, T., Vaggi, F., Marlow, R., Kanno, T., Shinomiya, I., Lombardi, S., Buchupalli, B., Graham, R., Gazinska, P., et al. (2015). Models of breast morphogenesis based on localization of stem cells in the developing mammary lobule. Stem Cell Reports *4*, 699–711.

Huebner, R.J., and Ewald, A.J. (2014). Cellular foundations of mammary tubulogenesis. Semin. Cell Dev. Biol. *31*, 124–131.

Huebner, R.J., Neumann, N.M., and Ewald, A.J. (2016). Mammary epithelial tubes elongate through MAPK-dependent coordination of cell migration. Development *143*, 983–993.

Iber, D., and Menshykau, D. (2013). The control of branching morphogenesis. Open Biol. *3*, 130088.

Kim, H.Y., Pang, M.-F., Varner, V.D., Kojima, L., Miller, E., Radisky, D.C., and Nelson, C.M. (2015). Localized smooth muscle differentiation is essential for epithelial bifurcation during branching morphogenesis of the mammalian lung. Dev. Cell *34*, 719–726.

Lu, P., and Werb, Z. (2008). Patterning mechanisms of branched organs. Science *322*, 1506–1509.

Lubkin, S.R., and Murray, J.D. (1995). A mechanism for early branching in lung morphogenesis. J. Math. Biol. *34*, 77–94.

McNeal, J.E. (1968). Regional morphology and pathology of the prostate. Am. J. Clin. Pathol. *49*, 347–357.

Meinhardt, H. (1982). Models of Biological Pattern Formation (Citeseer).

Metzger, R.J., Klein, O.D., Martin, G.R., and Krasnow, M.A. (2008). The branching programme of mouse lung development. Nature *453*, 745–750.

Narayan, V., Ramaswamy, S., and Menon, N. (2007). Long-lived giant number fluctuations in a swarming granular nematic. Science *317*, 105–108.

Nelson, C.M., Vanduijn, M.M., Inman, J.L., Fletcher, D.A., and Bissell, M.J. (2006). Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. Science *314*, 298–300.

Odor, G. (2004). Universality classes in non-equilibrium lattice systems. Rev. Mod. Phys. *76*, 663.

Paine, I., Chauviere, A., Landua, J., Sreekumar, A., Cristini, V., Rosen, J., and Lewis, M.T. (2016). A geometrically-constrained mathematical model of mammary gland ductal elongation reveals novel cellular dynamics within the terminal end bud. PLoS Comput. Biol. *12*, e1004839.

Pichel, J.G., Shen, L., Sheng, H.Z., Granholm, A.-C., Drago, J., Grinberg, A., Lee, E.J., Huang, S.P., Saarma, M., Hoffer, B.J., et al. (1996). Defects in enteric innervation and kidney development in mice lacking GDNF. Nature *382*, 73–76.

Powers, G.L., and Marker, P.C. (2013). Recent advances in prostate development and links to prostatic diseases. Wiley Interdiscip. Rev. Syst. Biol. Med. *5*, 243–256.

Ramaswamy, S., Aditi Simha, R.A., and Toner, J. (2003). Active nematics on a substrate: Giant number fluctuations and long-time tails. Europhys. Lett. *62*, 196.

Riccio, P., Cebrian, C., Zong, H., Hippenmeyer, S., and Costantini, F. (2016). Ret and etv4 promote directed movements of progenitor cells during renal branching morphogenesis. PLoS Biol. *14*, e1002382.

Rossi, M., Pastor-Satorras, R., and Vespignani, A. (2000). Universality class of absorbing phase transitions with a conserved field. Phys. Rev. Lett. *85*, 1803–1806.

Sampogna, R.V., Schneider, L., and Al-Awqati, Q. (2015). Developmental programming of branching morphogenesis in the kidney. J. Am. Soc. Nephrol. *26*, 2414–2422.

Scheele, C.L., Hannezo, E., Muraro, M.J., Zomer, A., Langedijk, N.S., van Oudenaarden, A., Simons, B.D., and van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. Nature *542*, 313–317.

Short, K.M., Combes, A.N., Lefevre, J., Ju, A.L., Georgas, K.M., Lamberton, T., Cairncross, O., Rumballe, B.A., McMahon, A.P., Hamilton, N.A., et al. (2014). Global quantification of tissue dynamics in the developing mouse kidney. Dev. Cell *29*, 188–202.

Silberstein, G.B. (2001). Postnatal mammary gland morphogenesis. Microsc. Res. Tech. *52*, 155–162.

Silberstein, G.B., and Daniel, C.W. (1987). Reversible inhibition of mammary gland growth by transforming growth factor-beta. Science *237*, 291–293.

Sternlicht, M.D. (2006). Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res. *8*, 201.

Taketo, M., Schroeder, A.C., Mobraaten, L.E., Gunning, K.B., Hanten, G., Fox, R.R., Roderick, T.H., Stewart, C.L., Lilly, F., Hansen, C.T., et al. (1991). FVB/N: an inbred mouse strain preferable for transgenic analyses. Proc. Natl. Acad. Sci. USA *88*, 2065–2069.

Wagner, K.U., Wall, R.J., St-Onge, L., Gruss, P., Wynshaw-Boris, A., Garrett, L., Li, M., Furth, P.A., and Hennighausen, L. (1997). Cre-mediated gene deletion in the mammary gland. Nucleic Acids Res. *25*, 4323–4330.

Yu, J., Carroll, T.J., Rajagopal, J., Kobayashi, A., Ren, Q., and McMahon, A.P. (2009). A Wnt7b-dependent pathway regulates the orientation of epithelial cell division and establishes the cortico-medullary axis of the mammalian kidney. Development *136*, 161–171.

Zhang, X., Martinez, D., Koledova, Z., Qiao, G., Streuli, C.H., and Lu, P. (2014). FGF ligands of the postnatal mammary stroma regulate distinct aspects of epithelial morphogenesis. Development *141*, 3352–3362.

Zubkov, V.S., Combes, A.N., Short, K.M., Lefevre, J., Hamilton, N.A., Smyth, I.M., Little, M.H., and Byrne, H.M. (2015). A spatially-averaged mathematical model of kidney branching morphogenesis. J. Theor. Biol. *379*, 24–37.

Cell

Retraction Notice to: Calcium and SOL Protease Mediate Temperature Resetting of Circadian Clocks

Ozgur Tataroglu, Xiaohu Zhao, Ania Busza, Jinli Ling, John S. O'Neill, and Patrick Emery*
*Correspondence: patrick.emery@umassmed.edu
http://dx.doi.org/10.1016/j.cell.2017.09.013

(Cell *163*, 1214–1224; November 19, 2015)

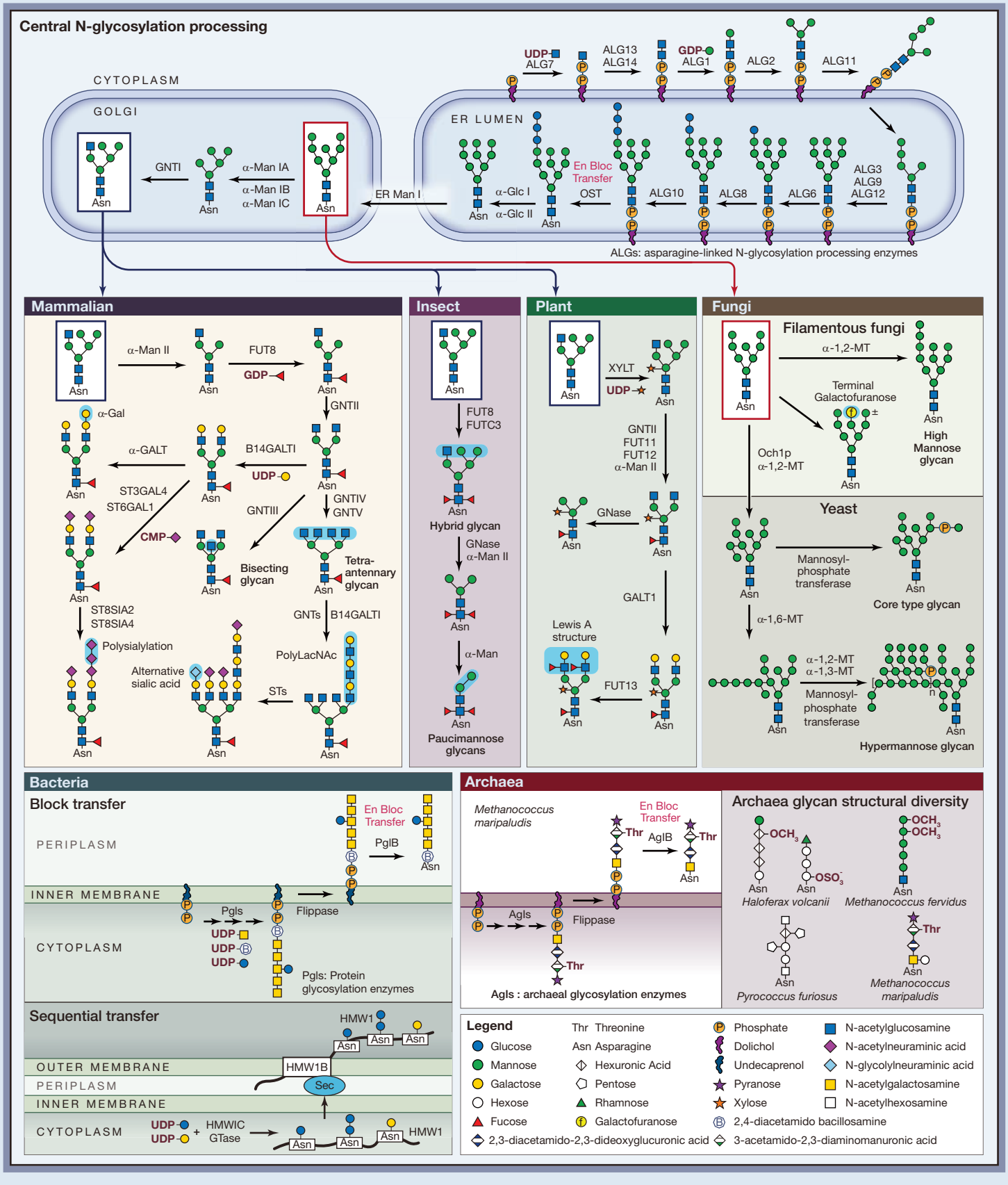
In the above article, we presented evidence that temperature phase shifts animal circadian clocks through regulated degradation of critical circadian pacemaker proteins mediated by calcium and SOL protease. In follow-up experiments, other members of the corresponding author's laboratory were unable to reproduce key observations made in vivo that supported the role of calcium and SOL protease in signaling temperature input to the *Drosophila* circadian clock. An in-depth review of the dataset generated by the first author revealed clear evidence that he had repeatedly misrepresented and altered primary data pertaining to the role of calcium, calmodulin, and SOL protease in *Drosophila* during the study. An independent investigation conducted by the University of Massachusetts Medical School concluded that the first author committed scientific misconduct. Since the manipulated data—which measured the amplitude of circadian phase shifts in response to temperature pulses in various genotypes using luciferase and behavioral assays—were critical to support our conclusions, we are retracting the above article. We sincerely apologize to the scientific community for publishing this erroneous article and profoundly regret any inconvenience and confusion that it might have caused.

The first author, Ozgur Tataroglu, declined to sign this retraction.

CrossMark

SnapShot: N-Glycosylation Processing Pathways across Kingdoms

Cheng-Yu Chung, Natalia I. Majewska, Qiong Wang, Jackson T. Paul, and Michael J. Betenbaugh
Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA



See online version for legends and references

ASHG2017

ORLANDO • OCTOBER 17-21, 2017

SHARING DISCOVERIES. **SHAPING** OUR FUTURE.

ASHG is pleased to welcome two special guests to the

Presidential Symposium at ASHG 2017



Bill Gates
Bill & Melinda Gates Foundation



Francis Collins
National Institutes of Health

Wednesday, October 18, 2017
5:00-6:30 pm
South Hall B, Orange County Convention Center

REGISTER TODAY

ashg.org/2017meeting

University of Nebraska Medical Center

FRED & PAMELA BUFFETT CANCER CENTER

The UNMC Eppley Institute for Research in Cancer seeks candidates for full-time tenure-track faculty positions with open rank (Assistant, Associate or Full Professor). Candidates are expected to 1) develop an independent, extramurally-funded research program in basic/translational cancer research that is complementary to existing strengths and aligned with the overall directions of the Fred & Pamela Buffett Cancer Center, 2) participate in graduate-level teaching in the Cancer Research Doctoral Program and/or other UNMC graduate programs, 3) engage in institutional service based on their expertise.

The Eppley Institute currently has 27 tenured or tenure-track faculty members. In 2017 the NCI-designated Fred & Pamela Buffett Cancer Center opened a state-of-the-art $335 million complex including a 260,000 sq. ft. cancer research tower, a multidisciplinary outpatient clinic (surgical, medical, and radiation oncology), and a 108-bed inpatient cancer hospital. The Cancer Center has over 100 primary members focused on externally funded research.

The Institute and Cancer Center seek talented investigators in any aspect of basic or translational cancer research. Faculty rank and compensation depends on qualifications. EEO/AA individuals from diverse backgrounds are encouraged to apply.

To apply, go to https://jobs.unmc.edu
and reference requisition #2017-210
Additional information is available at:
https://unmc.edu/cancercenter/facultysearch/
Please contact Cindy Plate, cjplate@unmc.edu, with questions.

VANDERBILT UNIVERSITY MEDICAL CENTER

MEDICAL DIRECTOR of DIAGNOSTIC MICROBIOLOGY AND IMMUNOLOGY

Vanderbilt University Medical Center
Department of Pathology Microbiology and Immunology
Vanderbilt Institute for Infection, Immunology, and Inflammation (VI4)

The Department of Pathology, Microbiology and Immunology, and the Vanderbilt Institute for Infection, Immunology, and Inflammation (VI4) at Vanderbilt University Medical Center invite applications for the position of Medical Director of the Diagnostic Microbiology and Immunology laboratories. Responsibilities include clinical oversight and strategic development of the Microbiology, Virology, Molecular Infectious Disease, and Immunology laboratory sections, and supervision of the corresponding associate medical directors. In addition, successful candidates will be expected to establish and maintain an independent research program and participate in teaching of graduate and medical students. Candidates should have substantial post-graduate training highlighted by peer-reviewed publications that demonstrate research productivity. Candidates should hold an M.D. and/or Ph.D. degree, be board certified by the American Board of Medical Microbiology or American Board of Pathology, and have at least 5 years of experience as a laboratory medical director.

Vanderbilt University Medical Center, located on the Vanderbilt University campus, is home to internationally recognized programs in bioinformatics, drug discovery, global health, inflammation, imaging science, pharmacology, proteomics, and vaccine science. The School consistently ranks in the Top 20 US Medical Schools and provides outstanding opportunities for scholarship, collaboration, and teaching.

The Vanderbilt University campus is located in Nashville, the capital of Tennessee, known internationally as "Music City USA", Nashville is home to professional sports teams, the Nashville Symphony, the Frist Center for the Visual Arts, and activities for outdoor enthusiasts, and is a wonderful place to live, work, and raise a family.

Applicants should send a *curriculum vitae*, a statement of current and future research interests, and at least 3 references to: Eric Skaar, Ph.D., Director VI4, Vice Chair for Basic Research, or Adam Seegmiller, M.D., Ph.D., Vice Chair for Clinical Pathology, Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center at this Email: VI4research@vanderbilt.edu

Vanderbilt University is an Affirmative Action/Equal Opportunity Employer.
Women and minority candidates are encouraged to apply.

KOCH INSTITUTE for Integrative Cancer Research at MIT

MIT DEPARTMENT OF BIOLOGY

Faculty Position

The Koch Institute for Integrative Cancer Research at the Massachusetts Institute of Technology (http://ki.mit.edu)invites applications for a junior or senior faculty appointment Appointments are expected to be in the MIT Department of Biology, but other departments will be considered if appropriate. The Koch Institute is an NCI-designated Cancer Center, which features research across a wide range of areas in basic cancer research and cancer-oriented engineering. This is an open search with regard to field of study and specific research focus, and research programs focused on basic science questions are potentially appropriate, but some cancer relevance will be a consideration. The candidate(s) will be expected to develop and lead an internationally competitive research program as well as participate in undergraduate and graduate teaching. The successful candidate(s) will have laboratory space in the Koch Institute.

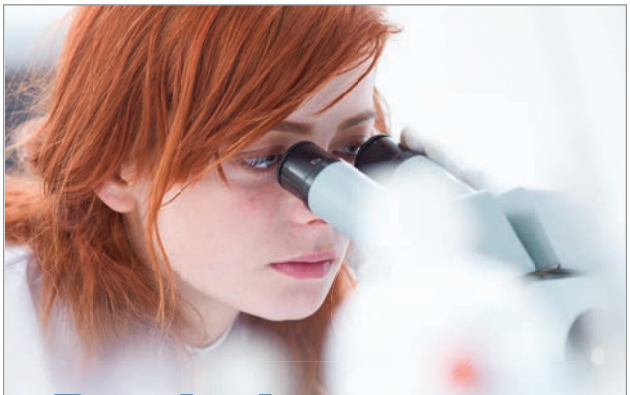
Applicants should include curriculum vitae, brief summaries of past accomplishments and a description of future research plans. Letters of recommendation should be sent separately from three individuals able to evaluate the candidate's accomplishments and future potential for both research and teaching.

Please apply online at
https://academicjobsonline.org/ajo/jobs/9566

Complete applications are due by November 1, 2017.

MIT is an affirmative action employer, and we encourage applications from women and underrepresented minorities.

Massachusetts Institute of Technology



Be their next big discovery.

careers.cell.com

Where top life sciences talent is searching for their next opportunity

Cell Career Network

Discover the best jobs in life science

To showcase your job openings on Cell Career Network, contact:
advertising@cell.com

UK Dementia Research Institute

The UK Dementia Research Institute is open for recruitment

Discovery science applications are invited for programme leaders from across biomedical research

The UK DRI is an exciting new £250 million investment by the UK into dementia research that will transform the way we diagnose, treat and care for people with dementias.

Based across six world-leading UK universities, the UK DRI will focus on innovative early-stage science to advance our understanding of how dementias develop and progress.

We are now seeking ambitious, talented scientists from around the world to bring innovative cross-disciplinary ideas to tackle and beat dementia.

Find out more at ukdri.ac.uk/joinus

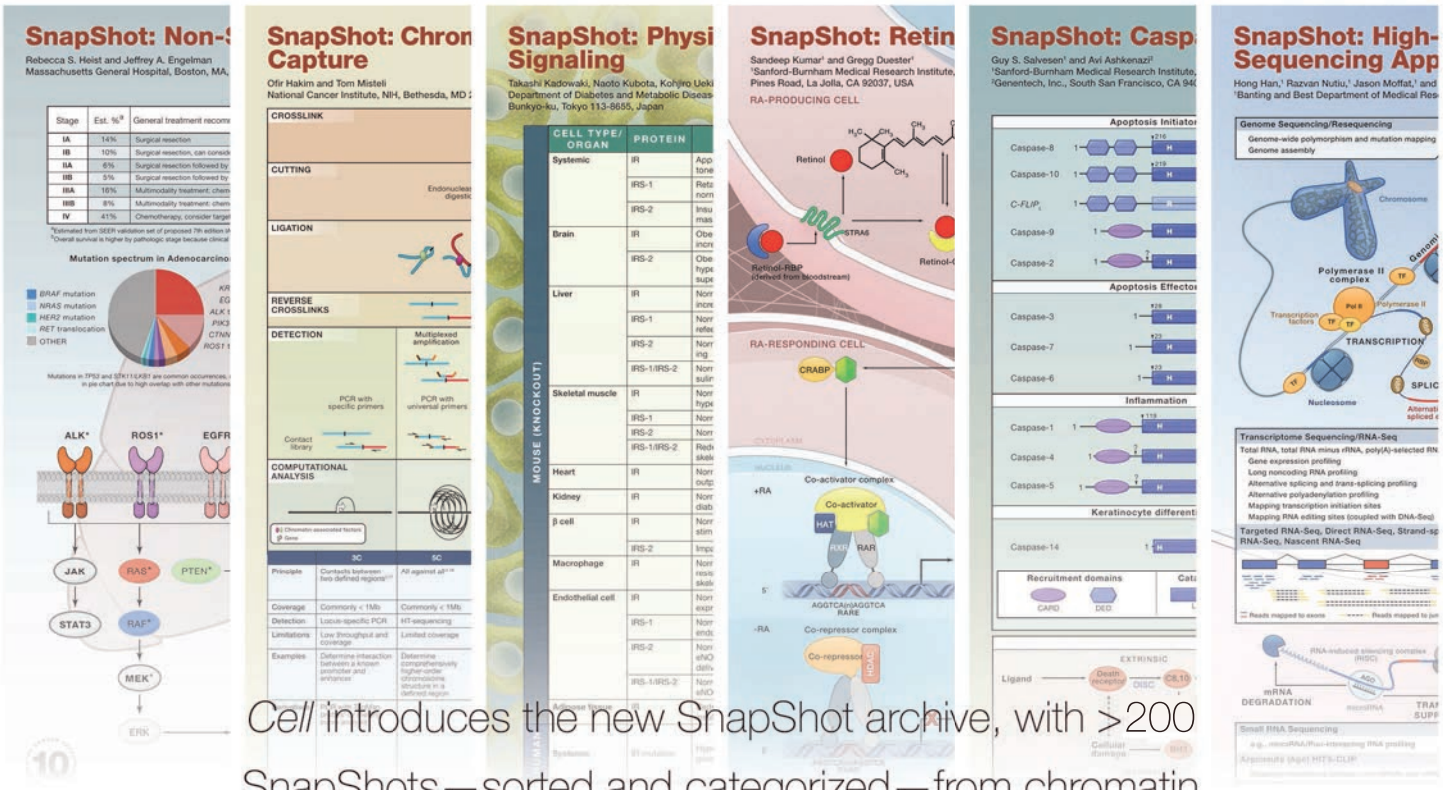
Founding funders:

MRC Medical Research Council

Alzheimer's Society

Alzheimer's Research UK

Cell SnapShot archive



Cell introduces the new SnapShot archive, with >200 SnapShots—sorted and categorized—from chromatin remodelers and autophagy to cancer and autism.

All SnapShots published from a year ago or more are open access and freely available.

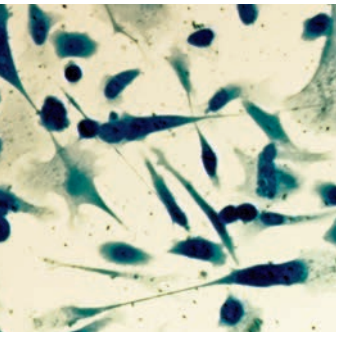
Sponsored By

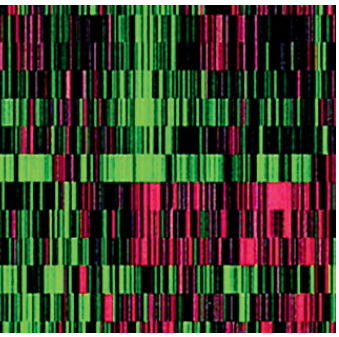
R&D SYSTEMS
a bio-techne brand

view the archive

cell.com/snapshots













2017-2018 SCIENTIFIC CONFERENCES

Presenting the most significant research on cancer etiology, prevention, diagnosis, and treatment

New Horizons in Cancer Research: Research Propelling Cancer Prevention and Cures
Conference Cochairs: Nancy E. Davidson, Kornelia Polyak, Chi Van Dang, Hongyang Wang
November 6-9, 2017 | Shanghai, P.R. China

Prostate Cancer: Advances in Basic, Translational, and Clinical Research
Conference Cochairs: Johann S. de Bono, Karen E. Knudsen, Peter S. Nelson, and Mark A. Rubin
December 2-5, 2017 | Orlando, FL

Pediatric Cancer Research: From Basic Science to the Clinic
Conference Cochairs: Peter C. Adamson, Nada Jabado, and Charles W. M. Roberts
December 3-6, 2017 | Atlanta, GA

San Antonio Breast Cancer Symposium Presented by CTRC-AACR-BCM
Codirectors: Carlos L. Arteaga, Virginia G. Kaklamani, and C. Kent Osborne
December 5-9, 2017 | San Antonio, TX

Fifth AACR-IASLC International Joint Conference: Lung Cancer Translational Science from the Bench to the Clinic
Conference Cochairs: Charles M. Rudin and Charles Swanton
January 8-11, 2018 | San Diego, CA

Obesity and Cancer: Mechanisms Underlying Etiology and Outcomes
Conference Cochairs: Lewis C. Cantley, Michael N. Pollak, and Elizabeth A. Platz
January 27-30, 2018 | Austin, TX

Immunobiology of Primary and Metastatic CNS Cancer: Multidisciplinary Science to Advance Cancer Immunotherapy
Conference Cochairs: Hideho Okada, Robyn S. Klein, Ignacio Melero, and Patricia S. Steeg
February 12-15, 2018 | San Diego, CA

Third AACR-SNMMI Joint Conference on State-of-the-Art Molecular Imaging in Cancer Biology and Therapy
Conference Cochairs: Todd E. Peterson and David R. Piwnica-Worms
February 14-17, 2018 | San Diego, CA

Targeting DNA Methylation and Chromatin for Cancer Therapy
Conference Cochairs: Stephen B. Baylin, Margaret A. Goodell, and Peter A. Jones
March 1-4, 2018 | Atlanta, GA

AACR Annual Meeting 2018
Program Committee Chair: Elaine R. Mardis
April 14-18, 2018 | Chicago, IL

Cancer Dormancy and Residual Disease
Conference Cochairs: Julio Aguirre-Ghiso, Cyrus Ghajar, Christoph Klein, and Dorothy Sipkins
June 19-22, 2018 | Montreal, QC, Canada

Sixth JCA-AACR Special Joint Conference on the Latest Advances in Lung Cancer Research: From Basic Science to Therapeutics
Organizing Committee: Hiroyuki Mano, Seiji Yano, Hiroyoshi Nishikawa, Alice T. Shaw, Roy S. Herbst, and Charles M. Rudin
July 10-12, 2018 | Kyoto, Japan

Intestinal Stem Cells and Colon Cancer: Biology to Therapy
Conference Cochairs: Anil K. Rustgi, Johanna Bendell, Hans Clevers, Christina Curtis, and Owen Sansom
September 27-30, 2018 | Washington, DC

Metabolism and Cancer
Conference Cochairs: Ralph J. Deberardinis, Tak W. Mak, Joshua D. Rabinowitz, and M. Celeste Simon
September 28-October 1, 2018 | New York, NY

Learn more and register at
AACR.org/Calendar

AACR American Association for Cancer Research
FINDING CURES TOGETHER®