AAAS

Science

23 JULY 2026

Hope and tragedy in gene-editing trial in China **p. 346**

The rise and fall of human languages **p. 387**

The magnetic field of an exoplanet **p. 403**



4D NUCLEOME



Apply for the BII & *Science Translational Medicine* Prize for Innovations in Women's Health

Submit an essay describing your journey from basic science to commercialization.

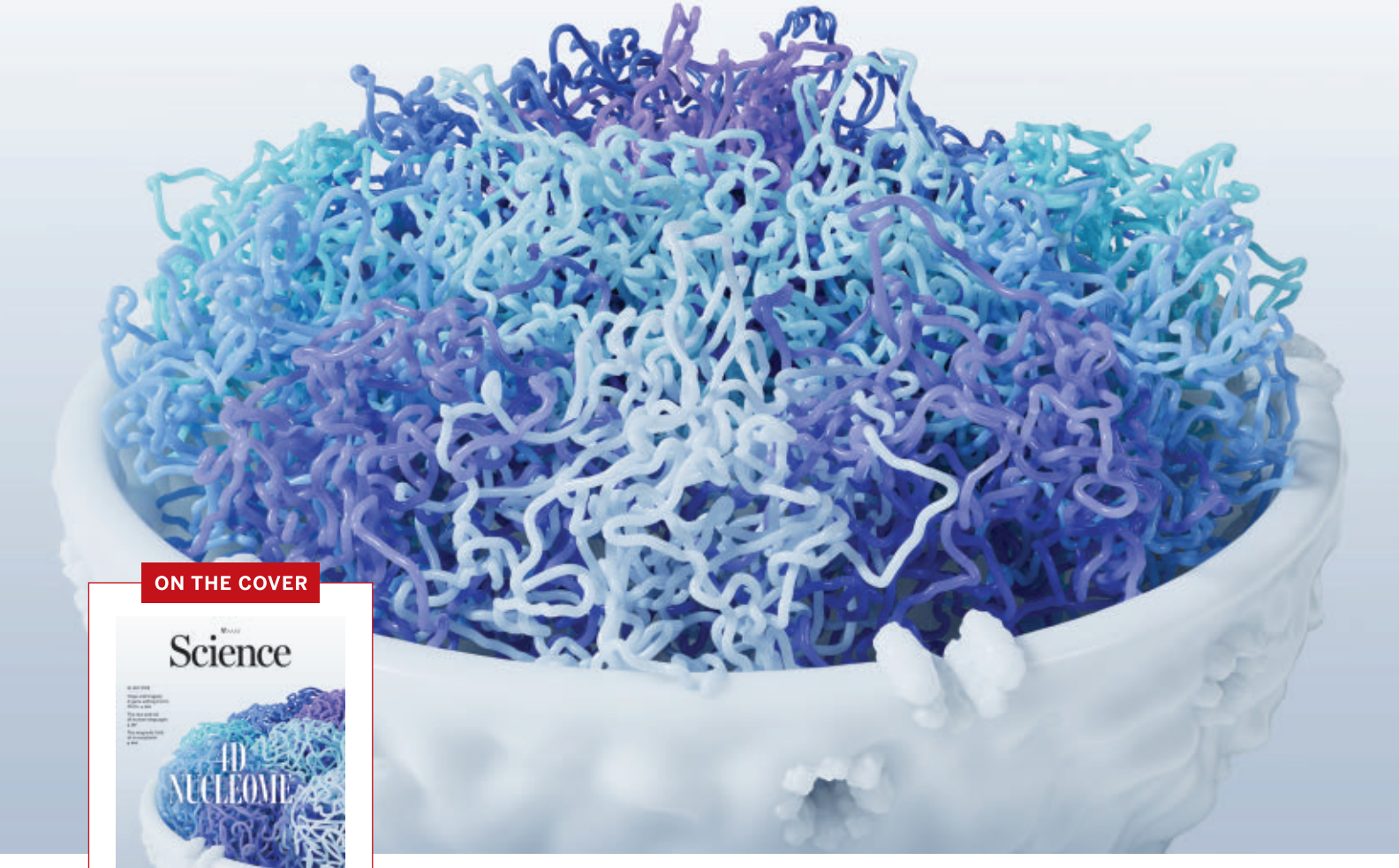
Grand Prize of USD 25,000
Finalist prizes of USD 10,000

Winners have their essays published in *Science Translational Medicine* and are celebrated at Innovation Day at the BioInnovation Institute in Copenhagen.

Submit your application by November 1st 2026

CONTENTS

23 JULY 2026 | VOLUME 393 | ISSUE 6809



ON THE COVER



Strands of DNA-like thread coil, loop, and spill upward, evoking the genome's dynamic three-dimensional architecture within the nucleus of a living cell. This special issue, with papers in *Science* and *Science Advances*, features single-cell and multiomic studies from the National Institutes of Health Common Fund's 4D Nucleome Program, which examines a fourth dimension of nuclear organization—how this folded structure shapes cell identity, shifts across development and aging, and goes awry in disease. See page 364 and 10.1126/sciadv.aek6870. Illustration: A. Fisher/*Science*

ILLUSTRATION: A. FISHER/*SCIENCE*

SPECIAL ISSUE

4D NUCLEOME

INTRODUCTION

364 Genome in motion

RESEARCH SUMMARIES

366 A 3D genome atlas of human tonsil and the role of loop extrusion in B cell somatic hypermutation —Y. Cheng *et al.*

367 Dose-dependent sensitivity of human three-dimensional chromatin to a heart disease–linked transcription factor —Z. L. Grant *et al.*

368 Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure —Y. Xie *et al.*

369 Single-cell multiomics connects 3D genome and transcriptome alterations in Alzheimer's disease —Y. Zhang *et al.*

370 Epigenetic and 3D genome reprogramming during the aging of human hippocampus —N. R. Zemke *et al.*

371 Human body single-cell atlas of three-dimensional genome organization and DNA methylation —J. Zhou *et al.*

SEE ALSO *SCIENCE ADVANCES* 4D NUCLEOME SECTION 10.1126/SCIADV.AEK6870

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**



ILLUSTRATION: DARIA LADA

EDITORIAL

335 University science in the US needs a coherent plan
—H. H. Thorp

NEWS

336 Prominent Salk Institute biologist to resign following sexual misconduct investigation
Staffers complain that Salk allowed circadian rhythms researcher Satchidananda Panda to depart quietly, while the allegations and findings remain under wraps
—S. Reardon

338 Ebola's rapid spread spurs new drug and vaccine trials
Pioneering studies are underway amid difficult conditions to address Bundibugyo's threat
—K. Kupferschmidt

339 Contagious fish cancer overruns New England lake
Genetic studies reveal rare example of identical transmissible tumors
—T. Appenzeller

341 Climate scientists sharpen tools for linking global warming to extreme weather
National Academies report says more rigorous results will boost confidence in fast-growing field
—J. Vaz

342 Oldest known Mars rock offers glimpse of planet's watery youth
Teghaza meteorite hints at surprisingly Earth-like crust and shows Mars was already losing its water 4.1 billion years ago
—P. Voosen

343 Alzheimer's trial results lend momentum to drugs targeting tau
First drug to lower the protein and slow cognitive decline created buzz despite puzzling data —J. E. Smith

344 Detention of U.S. seismologist by China alarms researchers
Youlin Chen analyzed earthquake data that can also aid nuclear test monitoring —R. Stone

FEATURES

346 A fatal reaction
A cutting-edge gene-editing trial in China went disastrously wrong. A family wants accountability —B. Borrell
PODCAST

COMMENTARY

PERSPECTIVES

354 Birdsong diversity across the world
Birdsong motifs are shaped by evolutionary history, biological traits, and environmental constraints
—E. F. Briefer
RESEARCH ARTICLE p. 381

356 Ballooning under water
Insect larval buoyancy is controlled by material properties of their air sacs
—J. F. Harrison and H. A. Woods
RESEARCH ARTICLE p. 398

357 No strain, no gain
A modular approach could expand the repertoire of spring-loaded covalent drugs —F. B. Kraft and M. Gehringer
RESEARCH ARTICLE p. 408

359 A tiny cone of reactivity
Colliding molecules react only when their orientations and point of impact fall within a narrow range
—J. Björk
RESEARCH ARTICLE p. 417

BOOKS ET AL.

360 The tangled world of snake taxonomy
Fringe figures in herpetology advance unlikely (and unchecked) claims in an offbeat taxonomical history —C. Kemp

361 Dispatches from the front lines of science
Young scientists' passion and commitment take center stage in a moving portrait of research in action
—V. Venkatraman

LETTERS

362 Brazil's environmental governance at risk
—M. V. Cianciaruso *et al.*

363 Seafood biosecurity gaps threaten biodiversity
—E. A. Maboloc and J. K.-H. Fang

363 Protect and expand ocean observation systems
—R. Blasiak and A. V. Norström

RESEARCH

HIGHLIGHTS

372 From *Science* and other journals

RESEARCH ARTICLES

376 Invasive species
Invasive species' environmental impacts are more severe in the Global South
—S. Bacher *et al.*

381 Birdsong
The global biogeography of passerine songs
—Q. Bacquelé *et al.*
PERSPECTIVE p. 354

387 Language diversity
The rise and fall of language diversity through the Holocene
—D. E. Blasi *et al.*

392 Physics
Dynamic asymmetric strain imprinted into substrates by an oxide thin film —E. Kisiel *et al.*

398 Comparative physiology
Crush-resistant air sacs allow insect larvae to exploit aquatic habitats at extreme depth —E. K. G. McKenzie *et al.*
PERSPECTIVE p. 356; PODCAST

403 Exoplanets
Constraining an exoplanet's magnetic field using star-planet interactions
—D. Revilla *et al.*

408 Drug design
Late-stage functionalization with strain-release warheads enables tunable covalent inhibition —Z. P. Shultz *et al.*
PERSPECTIVE p. 357

417 Surface chemistry
Spatially resolving the cone of reaction for a single molecule
—M. J. Timm *et al.*
PERSPECTIVE p. 359

422 Nanomaterials
Step geometry–guided growth of rhombohedral graphene
—M. Zhao *et al.*

WORKING LIFE

430 A geologist learns AI
—I. Pope

334 *Science* Staff
429 *Science* Careers

Science **serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news, commentary, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.** *Science* (ISSN 0036-8075) is published weekly on Thursday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2026 by the American Association for the Advancement of Science. The title *Science* is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $3125; Foreign postage extra: Air assist delivery: $135. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A. Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Science AAAS

EDITOR-IN-CHIEF **Holden Thorp**, hthorp@aaas.org

EXECUTIVE EDITOR Valda Vinson

EDITORS, RESEARCH Sacha Vignieri, Jake S. Yeston **EDITOR, COMMENTARY** Lisa D. Chong

DEPUTY EXECUTIVE EDITOR Lauren Kmec

DEPUTY EDITORS Stella M. Hurtley (UK), Phillip D. Szuromi **SENIOR EDITORS** Michael A. Funk, Angela Hessler, Di Jiang, Priscilla N. Kelly, Marc S. Lavine (Canada), Bianca Lopez, Mattia Maroso (UK), Yevgeniya Nusinovich, Ian S. Osborne (UK), Joana Osório (UK), L. Bryan Ray, Cheri Sirois, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Yury V. Suleymanov, Valerie B. Thompson, Brad Wible **ASSOCIATE EDITORS** Jack Huang, Sumin Jin, Sarah Ross (UK), Madeleine Seale (UK), Corinne Simonti, Unnati Sonawala **SENIOR LETTERS EDITOR** Jennifer Sills **SENIOR NEWSLETTER EDITOR** Christie Wilcox **RESEARCH & DATA ANALYST** Jessica L. Slater **LEAD CONTENT PRODUCTION EDITORS** Chris Filiatreau, Harry Jach **SR. CONTENT PRODUCTION EDITORS** Amelia Beyna , Julia Haber-Katris **CONTENT PRODUCTION EDITORS** Anne Abraham, Robert French, Nida Masiulis, Avery Shantha, Suzanne M. White **SENIOR PROGRAM ASSOCIATE** Maryrose Madrid **EDITORIAL MANAGER** Joi S. Granger **EDITORIAL ASSOCIATES** Aneera Dobbins, Lisa Johnson, Jerry Richardson, Anita Wynn **SENIOR EDITORIAL COORDINATORS** Clair Goodhead (UK), Alexander Kief, Kat Kirkman, Ronmel Navas, Isabel Schnaidt, Alice Whaley (UK), Brian White **EDITORIAL COORDINATORS** Samuel Bates, Daniel Young **ADMINISTRATIVE COORDINATOR** Karalee P. Rogers **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI OFFICE MANAGER** Carly Hayward (UK) **ASI SR. OFFICE ADMINISTRATORS** Simon Brignell (UK), Jessica Waldock (UK) **COMMUNICATIONS DIRECTOR** Meagan Phelan **DEPUTY DIRECTOR** Matthew Wright **SENIOR WRITERS** Walter Beckwith, Joseph Cariz, Abigail Eisenstadt **WRITER** Mahathi Ramaswamy **SENIOR COMMUNICATIONS ASSOCIATES** Kiara Brooks, Zachary Graber, Mackenzie Williams, Sarah Woods **NEWSLETTER INTERN** BenjaminHack

NEWS EDITOR Tim Appenzeller

NEWS MANAGING EDITOR John Travis **GLOBAL HEALTH EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Rachel Bernstein, David Grimm, Eric Hand, David Malakoff (Artificial Intelligence), Michael Price, Kelly Servick, Matt Warren (International) **SENIOR CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jocelyn Kaiser, Jeffrey Mervis **ASSOCIATE EDITORS** Michael Greshko, Katie Langin **NEWS REPORTERS** Jeffrey Brainard, Adrian Cho, Phie Jacobs, Sara Reardon, Jennie Erin Smith, Erik Stokstad, Paul Voosen, Celina Zhao **CONSULTING EDITOR** Elizabeth Culotta **CONTRIBUTING CORRESPONDENTS** Vaishnavi Chandrashekhar, Dan Charles, Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Virginia Morell, Dennis Normile (Tokyo), Catherine Offord, Cathleen O'Grady, Elisabeth Pain (Careers), Charles Piller, Zack Savitsky, Richard Stone (Senior Asia Correspondent), Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **INTERNS** Laura Martín Agudelo, Mona Patterson, Julia Vaz **COPY EDITORS** Julia Cole (Senior Copy Editor), Hannah Knighton, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

CREATIVE DIRECTOR Beth Rakouskas

DESIGN MANAGING EDITOR Chrystal Smith **GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** Emily Petersen **MULTIMEDIA MANAGING PRODUCER** Kevin McLean **DIGITAL DIRECTOR** Kara Estelle-Powers **DESIGN EDITOR** Marcy Atarod **DESIGNER** Noelle Jessup **SENIOR SCIENTIFIC ILLUSTRATOR** Noelle Burgess **SCIENTIFIC ILLUSTRATORS** Austin Fisher, Kellie Holoski, Ashley Mastin **SENIOR GRAPHICS EDITOR** Monica Hersher **GRAPHICS EDITOR** Veronica Penney **SENIOR PHOTO EDITOR** Charles Borst **PHOTO EDITOR** Elizabeth Billman **SENIOR PODCAST PRODUCER** Sarah Crespi **SENIOR VIDEO PRODUCER** Meagan Cantwell **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **SOCIAL MEDIA INTERN** Alyssa Trull

CHIEF EXECUTIVE OFFICER AND EXECUTIVE PUBLISHER

Sudip Parikh

PUBLISHER, SCIENCE FAMILY OF JOURNALS Bill Moran

DIRECTOR, BUSINESS OPERATIONS & ANALYSIS Eric Knott **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYSTS** Kurt Ennis, Maggie Clark, Isacco Fusi **BUSINESS OPERATIONS ADMINISTRATOR** Taylor Fisher **DIGITAL SPECIALIST** Marissa Zuckerman **ASSOCIATE DIRECTOR OF PRODUCTION AND PLATFORM** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **CONTENT OPERATIONS MANAGER** Rebecca Doshi **PUBLISHING PLATFORM MANAGER** Jessica Loayza **PUBLISHING SYSTEMS SPECIALIST, PROJECT COORDINATOR** Jacob Hedrick **PRODUCTION LEAD** Kristin Wowk **SENIOR PRODUCTION SPECIALIST** Audrey Diggs **PRODUCTION SPECIALIST** Kelsey Cartelli **SPECIAL PROJECTS ASSOCIATE** Shantel Agnew

MARKETING DIRECTOR Sharice Collins **ASSOCIATE DIRECTOR, MARKETING** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **ASSOCIATE DIRECTOR, MARKETING SYSTEMS & STRATEGY** Aimee Aponte **SENIOR MARKETING MANAGER** Shawana Arnold **MARKETING MANAGER** Ashley Evans **MARKETING ASSOCIATES** Hugues Beaulieu, Ashley Hylton, Lorena Chirinos Rodriguez, Jenna Voris **MARKETING ASSISTANT** Courtney Ford **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Erika Gebel Berg **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SENIOR TRAFFIC ASSOCIATE** Christine Hall

DIRECTOR, PRODUCT MANAGEMENT Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **ASSOCIATE DIRECTOR, PUBLISHING INTELLIGENCE** Rasmus Andersen **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATES** Caroline Breul, Anne Mason

ASSOCIATE DIRECTOR, INSTITUTIONAL LICENSING MARKETING Kess Knight **ASSOCIATE DIRECTOR, INSTITUTIONAL LICENSING SALES** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Nazim Mohammedi, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **MANAGER, RENEWAL & RETENTION** Lana Guz **SYSTEMS & OPERATIONS ANALYST** Ben Teincuff **FULFILLMENT ANALYST** Aminta Reyes

ASSOCIATE DIRECTOR, INTERNATIONAL Roger Goncalves **ASSOCIATE DIRECTOR, US ADVERTISING** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES MANAGER** Chris Hoag **DIRECTOR, OUTREACH AND STRATEGIC PARTNERSHIPS, ASIA** Shoupeng Liu **ASI EMEA SALES ADVERTISING MANAGER** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Rie Rambelli (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren **RIGHTS AND LICENSING COORDINATOR** Dana James **CONTRACT SUPPORT SPECIALIST** Michael Wheeler

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

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

MEDIA CONTACTS
scipak@aaas.org

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.science.org
science_advertising@aaas.org

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/
benefits

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

AAAS BOARD OF DIRECTORS

CHAIR Joseph S. Francisco

IMMEDIATE PAST PRESIDENT
Theresa A. Maldonado

PRESIDENT
Marina Picciotto

PRESIDENT-ELECT
Eric D. Isaacs

TREASURER Mark F. Dingfield

COUNCIL CHAIR
Ichiro Nishimura

CHIEF EXECUTIVE OFFICER
Sudip Parikh

BOARD
Carolyn N. Ainslie
Morton Ann Gernsbacher
Kathleen Hall Jamieson
Jennifer Lodge
Babak Parviz
Gabriela Popescu
Juan S. Ramírez Lugo
Susan M. Rosenberg
Vassiliki Betty Smocovitis
Roger Wakimoto

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

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Anastassia Alexandrova, *UCLA*
James Analytis, *UC Berkeley*
Amy Ando, *The Ohio State U.*
Paola Arlotta, *Harvard U.*
Madan Babu, *St. Jude*
Jennifer Balch, *U. of Colorado*
Nenad Ban, *ETH Zürich*
Carolina Barillas-Mury, *NIH, NIAID*
Christopher Barratt, *U. of Dundee*
François Barthelat, *U. of Colorado Boulder*
Joseph Bass, *Northwestern U.*
Franz Bauer, *Universidad de Tarapacá*
Andreas Baumler, *UC Davis*
Carlo Beenakker, *Leiden U.*
Camilla Bellone, *Geneva U.*
Sarah Bergbreiter, *Carnegie Mellon U.*
Kiros T. Berhane, *Columbia U.*
Aude Bernheim, *Inst. Pasteur*
Joseph J. Berry, *NREL*
Aditi Bhargav, *UCSF*
Dominique Bonnet, *Francis Crick Inst.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Malcolm Brenner, *Baylor College of Med.*
Ron Brookmeyer, *UCLA* (**S**)
Christian Büchel, *UKE Hamburg*
Johannes Buchner, *TUM*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
Erin Calipari, *Vanderbilt U.*
Annmarie Carlton, *UC Irvine*
Jane Carlton, *Johns Hopkins U.*
Pedro Carvalho, *U. of Oxford*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
Hilde Cheroutre, *La Jolla Inst.*
Wendy Cho, *UIUC*
Ib Chorkendorff, *Denmark TU*
Chunaram Choudhary, *København U.*
Karlene Cimprich, *Stanford U.*
Laura Colgin, *UT Austin*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Nicolas Dauphas, *U. of Chicago*
Liset de la Prida, *Instituto Cajal CSIC*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Samuel Díaz-Muñoz, *UC Davis*
Ulrike Diebold, *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd A. Ehlers, *U. of Glasgow*
Tobias Erb, *MPS, MPI Terrestrial Microbiology*
Beate Escher, *UFZ & U. of Tübingen*
Vanessa Ezenwa, *U. of Georgia*
Fleur Ferguson, *U. San Dieg*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Natascha Förster Schreiber, *MPI Extraterrestrial Phys.*
Serita Frey, *U. of New Hampshire*
Elaine Fuchs, *Rockefeller U.*
Caixia Gao, *Inst. of Genetics and Developmental Bio., CAS*
Daniel Geschwind, *UCLA*
Lindsey Gillson, *U. of Cape Town*
Alemu Gonsamo Gosa, *McMaster U.*
Simon Greenhill, *U. of Auckland*
Gillian Griffiths, *U. of Cambridge*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Osvaldo Gutierrez, *UCLA*
Taekjip Ha, *Johns Hopkins U.*
Daniel Haber, *Mass. General Hos..*
Hamida Hammad, *VIB IRC*
Brian Hare, *Duke U.*
Kelley Harris, *U. of Wash*
Sheng-Yang He, *Duke U.*
Carl-Philipp Heisenberg, *IST Austria*
Emiel Hensen, *TU Eindhoven*
Christoph Hess, *U. of Basel & U. of Cambridge*
Brian Hie, *Stanford U.*
Heather Hickman, *NIAID, NIH*
Janneke Hille Ris Lambers, *ETH Zürich*
Kai-Uwe Hinrichs, *U. of Bremen*
Pinshane Huang, *UIUC*
Christina Hulbe, *U. of Otago, New Zealand*
Gwyneth Ingram, *ENS Lyon*
Darrell Irvine, *Scripps Res.*
Bharat Jalan, *U. of Minnesota*
Erich Jarvis, *Rockefeller U.*
Sheena Josselyn, *U. of Toronto*
Matt Kaeberlein, *U. of Wash.*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
V. Narry Kim, *Seoul Nat. U.*
Nancy Knowlton, *Smithsonian*
Etienne Koechlin, *École Normale Supérieure*
LaShanda Korley, *U. of Delaware*
Paul Kubes, *U. of Calgary*
Deborah Kurrasch, *U of Calgary*
Laura Lackner, *Northwestern U.*
Hedwig Lee, *Duke U.*
Fei Li, *Xi'an Jiaotong U.*
Jianyu Li, *McGill U.*
Xiaobo Li, *Westlake U.*
Ryan Lively, *Georgia Tech*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *WUSTL*
Mo Lotfollahi, *Wellcome Sanger Institute*
Ke Lu, *Inst. of Metal Res., CAS*
Jean Lynch-Stieglitz, *Georgia Tech*
David Lyons, *U. of Edinburgh*
Vidya Madhavan, *UIUC*
Anne Magurran, *U. of St. Andrews*
Oscar Marín, *King's College London*
Matthew Marinella, *Arizona State U.*
Charles Marshall, *UC Berkeley*
Geraldine Masson, *CNRS*
Jennifer McElwain, *Trinity College Dublin*
Scott McIntosh, *NCAR*
Rodrigo Medellín, *U. Nacional Autónoma de México*
Mayank Mehta, *UCLA*
C. Jessica Metcalf, *Princeton U.*
Tom Misteli, *NCI, NIH*
Jeffery Molkentin, *Cincinnati Children's Hospital Medical Center*
Alison Motsinger-Reif, *NIEHS, NIH* (**S**)
Rosa Moysés, *U. of São Paulo School of Medicine*
Carey Nadell, *Dartmouth College*
Thi Hoang Duong Nguyen, *MRC LMB*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
James Owen, *Imperial College London*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Giovanni Parmigiani, *Dana-Farber* (**S**)
Zak Page, *UT Austin*
Pierre Paoletti, *Ecole Normale Supérieure, Paris*
Sergiu Pasca, *Stanford U.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Suzie Pun, *U. of Wash*
Lei Stanley Qi, *Stanford U.*
Simona Radutoiu, *Aarhus U.*
Maanasa Raghavan, *U. of Chicago*
Trevor Robbins, *U. of Cambridge*
Adrienne Roeder, *Cornell U.*
Joeri Rogelj, *Imperial College London*
John Rubenstein, *SickKids*
Sylvie Roke, *École Polytechnique Fédérale de Lausanne*
Yvette Running Horse Collin, *Toulouse U.*
Mike Ryan, *UT Austin*
Alberto Salleo, *Stanford U.*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Bhaven Sampat, *Johns Hopkins U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Ken Shirasu, *RIKEN CSRS*
Robert Siliciano, *JHU School of Med.*
Emma Slack, *ETH Zürich & U. of Oxford*
Richard Smith, *UNC* (**S**)
Ivan Soltesz, *Stanford U.*
John Speakman, *U. of Aberdeen*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
Sandip Sukhtankar, *U. of Virginia*
Christopher Summerfield, *U. of Oxford*
Tavneet Suri, *MIT*
Naomi Tague, *UC Santa Barbara*
A. Alec Talin, *Sandia Natl. Labs*
Patrick Tan, *Duke-NUS Med. School*
Fuchou Tang, *BIOPIC Beijing*
Leticia Tarruell, *ICFO*
Sarah Teichmann, *Wellcome Sanger Inst.*
Dörthe Tetzlaff, *Leibniz Institute of Freshwater Ecology and Inland Fisheries*
Amanda Thomas, *UC Davis*
Bozhi Tian, *U. of Chicago*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Kimani Toussaint, *Brown U.*
Li-Huei Tsai, *MIT*
Jason Tylianakis, *U. of Canterbury*
Matthew Vander Heiden, *MIT*
Jo Van Ginderachter, *VIB, U. of Ghent*
Ivo Vankelecom, *KU Leuven*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Elizabeth Villa, *UC San Diego*
Bert Vogelstein, *Johns Hopkins U.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* (**S**)
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Alex Webb, *U. of Cambridge*
Chris Wikle, *U. of Missouri* (**S**)
Ian A. Wilson, *Scripps Res.* (**S**)
Marjorie Weber, *U. of Michigan*
Sylvia Wirth, *ISC Marc Jeannerod*
Catherine Wolfram, *MIT*
Amir Yacoby, *Harvard U.*
Benjamin Youngblood, *St. Jude*
Kenneth Zaret, *UPenn School of Med.*
Magdalena Zernicka-Goetz, *Caltech*
Lidong Zhao, *Beihang U.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

University science in the US needs a coherent plan

H. Holden Thorp

The Trump administration continues to degrade America's scientific establishment, causing widespread harm to the country's broader interests and well-being. Last week, the Association of American Universities reported that graduate student enrollment in major research universities, which award half of all doctorate degrees in the United States, is down 15% compared to a year ago. And no wonder, considering the effect of recent immigration policies that deter international students and the financial uncertainties that universities are facing given the unpredictable nature of federal funding. The ability to support new doctoral students depends heavily on research grants, and any responsible administrator would be cautious about making commitments to students they can't keep. This is discouraging to the scientific community but should be alarming to everyone.

The loss of young scientists means a loss of the enthusiasm and creativity that drive progress. It's also a disheartening message to the members of Congress who worked to preserve federal funding for research earlier this year. The grim news comes on the heels of last month's move by the US National Science Foundation to cut grants for basic science by up to 30% or more and move support away from academia and toward entrepreneurial technology-oriented organizations. And just last week, the White House Office of Management and Budget (OMB) ignored backlash from the scientific community, as well as bipartisan pushback, when it declined to extend the public comment period for a proposal to shift the power over federal research grants from long-standing peer review boards to political appointees.

Most discouraging of all may be how actions like these are making other countries increasingly attractive to US scientists who may be looking for greener pastures. Earlier this month, Nobel laureate chemist Omar Yaghi from the University of California, Berkeley, was appointed to Tsinghua University, China, where he will lead a new institute devoted to using artificial intelligence to discover new materials. Organic chemist David Nicewicz also left the University of North Carolina at Chapel Hill this month for the Nanyang Technological University in Singapore, saying that he was excited to move to a "country that truly cares about the development and welfare of its citizens." No doubt stars like Yaghi and Nicewicz have been given abundant resources in their new locales. But it's also possible that they just wanted to go where science is growing, not shrinking.

The universities are mostly quiet...rather than banding together to strengthen the academic ecosystem.

To retain top scientific talent—from graduate students to Nobel laureates—the US urgently needs a practical, forward-looking plan that ensures that academic science can adapt to, and even thrive in, the current political climate. Unfortunately, universities and other associations have been quiet about this and other things. Only a few major organizations, including the American Association for the Advancement of Science (AAAS, publisher of *Science*) and the Association of American Medical Colleges, immediately issued firm condemnations of the OMB proposal. With a few exceptions, those institutions that issued objections did so at the last minute and said little beyond their registered comments. The reason for this behavior is obvious—they don't want to invite punishment from the Trump administration. But the cautious silence leaves faculty and students on the campus—or those considering going there—struggling to see how they can advance their careers and feel hopeful.

This leaves science faculty and students confronting a messy patchwork of guidance. The clearest statements about concrete actions are from political activism organizations like Stand Up for Science (SUFS; for more, see my interview with SUFS founder, Colette Delawalla). The universities are mostly quiet or spending more time competing with each other for the spotlight rather than banding together to strengthen the academic ecosystem. What's missing is a cogent plan that articulates the steps by which university science in America can succeed under the principles it followed over the past 80 years, while uniting support across the scientific community.

Fighting for the future of science requires a broad coalition. As Delawalla and I discuss, if the leadership of academia disagrees with the approaches espoused by SUFS and related organizations and advocates, then it needs to be more aggressive and express a different strategy. If universities oppose the actions of the administration, they need to say so quickly so that their constituents aren't left wondering whether there's a plan. The alternative is a community of university scientists adrift and grasping for a lifeline. □

H. Holden Thorp is Editor-in-Chief of the *Science* journals. hthorp@aaas.org

10.1126/science.aek7500



The Salk Institute for Biological Studies campus in San Diego.

PHOTO: ABLOKHIN/GETTY

SCIENTIFIC COMMUNITY

Prominent Salk Institute biologist to resign following sexual misconduct investigation

Staffers complain that Salk allowed circadian rhythms researcher Satchidananda Panda to depart quietly, while the allegations and findings remain under wraps **SARA REARDON**

Satchidananda Panda, a prominent biologist known for his work on circadian rhythms, light sensing, and intermittent fasting, will resign next month after 22 years at the Salk Institute for Biological Studies. His planned departure, which has not been officially announced outside of Salk, comes after the institute contracted an employment law firm to investigate allegations of sexual misconduct by Panda. Panda denies the allegations, which were made earlier this year.

Current and former lab members, and other Salk employees, told *Science* the investigation stemmed from a complaint by a researcher in Panda's lab. *Science* has learned it is at least the second time in recent years that Salk has investigated Panda following harassment allegations. Salk told lab members on 1 July that the investigation had concluded but would not share its findings. Salk employees say Panda has not been on campus for several weeks.

In an email sent to faculty and seen by *Science*, Salk President Gerald Joyce said, "Salk leadership has been informed that Dr. Panda intends to resign from the Institute, effective August 13, 2026." Panda, who is in his mid-50s, told lab members and faculty in a separate email that he plans to write a book about how recent advances in circadian rhythms research will shape human health. "This is the right moment to pause, take fresh stock of these shifts, and redefine the goals for the next phase of my career," Panda wrote. He has authored two books on circadian biology and is one of the most cited researchers in the world, according to the journal analysis firm Clarivate. In 2023, he was elected as a fellow at AAAS, which publishes *Science*.

Multiple Salk employees, including lab members, told *Science* they were worried Salk's decision to let Panda resign without making the investigation public could allow him to join another institution where colleagues have no knowledge of the past concerns about his behavior.

In an emailed statement, Salk confirmed that an independent investigation had taken place after a research assistant raised "workplace concerns." "Because Salk is committed to protecting the privacy of its personnel and encouraging an environment where all employees are empowered to speak up and raise concerns, we will not comment on the nature of the allegations, the workplace investigation, or the investigative findings," the statement said.

In 2018, Salk suspended famed cancer biologist Inder Verma after *Science* reported allegations from eight women—six of them at Salk—of forcible sexual contact and other inappropriate behavior. Verma resigned later that year. The same year, Salk settled three

lawsuits brought by senior female faculty who alleged widespread gender discrimination at the institute.

Science spoke with six researchers with knowledge of the situation, both in Panda's group and other labs at Salk and at the University of California San Diego, which is affiliated with Salk. Most spoke on condition of anonymity out of fear of career repercussions. They say the latest investigation concerned allegations of sexual misconduct brought against Panda by a research assistant in his lab.

Tiffany Chin, a graduate student at Salk who left Panda's lab in April, said the research assistant approached her in late February. She told Chin that Panda had been harassing her, including giving her massages in his office. She said on one occasion she turned down a foot massage, and he later screamed at her that she would never make it in science, which "scared her," Chin told *Science* in a text. "It made her realize that unless she complies with him that she'd lose her letter of recommendation, might get fired, etc." The research assistant also showed Chin texts Panda sent her that she found inappropriate. She told Chin she planned to make a formal complaint against Panda to Salk's human resources (HR) department.

Other people who spoke with the research assistant said she reported that Panda had drunk alcohol with her in his office and made sexually explicit references in conversation. One person *Science* spoke with witnessed Panda regularly meeting with the research assistant for hours in his office—far more time than he usually spent with his postdocs and graduate students. "She was the favorite among the lab personnel—[the favorites] were mostly women," the person says.

A source with knowledge of the situation says in recent years Salk has conducted at least one other formal investigation into allegations of sexual misconduct by Panda. The source says an employee then in Panda's lab alleged that while the two were traveling to a conference, he suggested she share a hotel room with him. The source says Salk took disciplinary action against Panda at the time.

In an email, Panda denies the "allegations and accusations. … I am deeply committed to maintaining a safe, respectful, and professional environment for everyone in my research team and across the Institute. These values are central to my leadership."

Laura van Rosmalen, a postdoc in Panda's lab, reached out to *Science* and said she finds Panda to be a good mentor to all lab members. "I see that he's very supportive to the people in the lab and caring for everyone," she says. The lab "really feels like a family," she says.

After the February incidents, the research assistant went on extended leave and then resigned from the lab, according to multiple people.

According to emails seen by *Science*, Salk contracted attorney Renée Schor to conduct an independent investigation into the allegations. Beginning in March, Schor interviewed lab members in an inquiry that stretched over several months.

On 1 July, Salk's HR department met with Panda's lab to tell them the investigation had concluded and Panda was resigning, but that they would not share the outcome of the investigation, according to multiple lab members present at the meeting. All the sources *Science* spoke with who were at the meeting say they came away with the impression that they were not to discuss the investigation.

Two trainees at Salk independently reached out to *Science* afterward, concerned about the lack of transparency. "I believe that a senior scientist that abuses his position of power should not be able to move on to the next thing with his reputation fully protected—all while his victim(s) and lab members suffer the consequences," Adam Rochussen, a postdoc in Panda's lab who attended the meeting, wrote in an email. According to Rochussen, Panda told him it would be career "suicide" if lab members didn't go along with his official story.

A second trainee at Salk expressed concern that vulnerable employees, especially international ones who rely on Salk to sponsor their visas, felt silenced by the decision to allow Panda to resign without revealing the investigation's nature or outcome. "This dynamic can create an environment where people feel unable to ask questions or share information," they said.

Jennifer Freyd, a psychologist and president of the Center for Institutional Courage who studies sexual harassment, echoes this sentiment. Allowing a person who commits harassment to resign quietly, she says, "is really harmful to the people who've been victims of this person but also bad for the institution and the next institution where he shows up. It's a very destructive pattern to do that."

Freyd notes it is sometimes necessary to keep information private to protect alleged victims. But she says institutions can work with lawyers to be transparent about an investigation's results, at least internally or by reporting the outcome to funders. In a statement, Salk said, "The Institute complies with its reporting obligations."

Rochussen says at least nine of the 26 employees in Panda's lab have left since the beginning of the year, and many cited their discomfort after learning about the allegations as their reason. Multiple sources told *Science* some lab members—especially staff scientists and graduate students—have struggled to find new placements on short notice, in part because research and training funds are tight across the institute amid federal budget cuts.

Salk did not answer questions about what will happen to Panda's grants, but said it is "arranging a transition plan to ensure those in the lab, and their scientific research, are disrupted as little as possible." □

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

IN OTHER NEWS

GRANT-TERMINATION CLAUSE TERMINATED

The White House Office of Management and Budget (OMB) cannot cancel active research grants solely because of a shift in the administration's priorities, a federal judge ruled last week. That removes a tool President Donald Trump has used to terminate billions of dollars in research grants and other funding. Last year, 23 states sued OMB and 11 federal agencies over their use of what is called the termination clause, which allows the government to pull the plug on any award that "does not effectuate program goals, agency priorities, or the national interest as they exist at the time of the termination." But U.S. District Court Judge Indira Talwani said on 17 July that the clause "does not permit agencies to terminate grants based on program goals and agency priorities identified after grants were awarded." The states did not ask for any grants to be reinstated. But legal experts say the ruling makes it much easier for universities to ask a federal claims court to restore canceled grants. The Trump administration has 60 days to appeal last week's decision. The clause underlying the government's claim was added in 2020 to rules governing federal financial assistance. It remains in the most recent revision, proposed by OMB in late May, which has attracted extensive criticism from scientists and professional societies.
—Jeffrey Mervis



Medical workers bury a deceased Ebola patient in Bunia, Democratic Republic of the Congo, on 9 July.

PHOTO: XINHUA VIA GETTY IMAGES

INFECTIOUS DISEASES

Ebola's rapid spread spurs new drug and vaccine trials

Pioneering studies are underway amid difficult conditions to address Bundibugyo's threat **KAI KUPFERSCHMIDT**

More than 2 months after it was discovered, the latest Ebola outbreak in the Democratic Republic of the Congo (DRC) is still out of control. The country had reported 2423 cases by 19 July, including 967 deaths, already making this the third largest Ebola outbreak ever recorded. Dozens of new cases are confirmed every day.

Those grim statistics also mean the outbreak is a chance for scientists to learn more about Bundibugyo, the rare Ebola species responsible, and test vaccines and drugs to fight it. On 15 July, the first ever trial of an Ebola drug to protect people who aren't sick but have been exposed to the virus was launched. Two weeks earlier, a trial of drugs to treat patients had begun. And a phase 1 study of a vaccine candidate kicked off in Oxford, U.K., on 13 July. "It's remarkable to see these clinical trials up and running this quickly," says Isaac Bogoch, an infectious disease expert at Toronto General Hospital.

But the factors that are fueling the rapid spread of the virus in the DRC also complicate the trials. Craig Spencer, a public health expert at Brown University and an Ebola survivor himself, says, "The chaos of the current outbreak" means lots of cases for researchers to draw on. On the other hand, he says, many patients don't reach hospitals and their contacts can't be found, and the population is distrustful—all of which makes trials harder to do. "You can't do these [studies] well if you don't have the basic things underpinning them all in place," Spencer says.

The outbreak was likely brewing for months before the first case was detected; by the time the response started, the virus had already spread widely, says Placide Mbala, head of epidemiology and global health at the DRC's National Institute of Biomedical Research. "Many health zones are affected, and many of them are in a place where access is very difficult because of security." People in the outbreak region, which is plagued by armed conflict, are highly mobile and have little reason to trust public health officials. Because Bundibugyo appears to be milder than other Ebola species, patients may seek care later, or not at all. The poor infrastructure makes it harder to reach them.

One worrisome sign, Ebola experts say, is that more than 80% of newly confirmed cases are not on any patient's list of contacts, indicating that many chains of transmission are still being missed. Most new cases never receive any care in a health facility. Patients often die in their community and are confirmed as Ebola cases only after their death.

That's a particular problem for the prevention trial, which will include people who were in close contact with an Ebola patient who was bleeding or vomiting, or with the body of a deceased patient. They will receive obeldesivir, a drug developed by Gilead that resembles a better known antiviral known as remdesivir but can be administered as a pill instead of injections. Current treatments are often given late in the disease course, when the virus has already overrun the body's defenses; earlier intervention might work better. But for the trial to work, contact tracing has to improve, says Mbala, one of the principal investigators: "Without that, we won't be able to identify high-risk contacts eligible for the study."

The trial won't enroll pregnant people or children younger than 12 years old because it's unknown whether obeldesivir is safe for them; instead they will receive remdesivir as part of a "compassionate use" program to see whether it, too, can

work as a prophylactic. Given the high mortality from Ebola in these groups, it's important to produce those safety data as soon as possible, says Yazdan Yazdanpanah, who runs ANRS Emerging Infectious Diseases, a French government agency that's involved in the trial. ANRS and several other institutions hope to launch a phase 1 study "quite quickly," he says.

The treatment trial started on 2 July in a facility close to the outbreak's center in Bunia, the capital of Ituri province. Initially, researchers are testing two potential drugs. One, named MPB-134, is made up of two monoclonal antibodies isolated from a survivor of the Ebola epidemic in West Africa more than a decade ago. The second one is remdesivir, which has been used with mixed results against SARS-CoV-2 and several other viruses. Participants will be randomly selected to receive one of the drugs, both, or neither. So far, patients have only been enrolled at one location, but researchers hope to scale up to multiple sites soon and include about 1200 participants, says clinical scientist Amanda Rojek of the University of Oxford, one of the trial's principal investigators.

Even if the drugs work, it might take a vaccine to end the outbreak, Yazdanpanah says. But that will take time. There is a licensed vaccine against Zaire, another Ebola species, as well as an investigational vaccine against a third species, named Sudan, that was used in a vaccination trial in Uganda last year. Combining these two could give some protection against Bundibugyo, and a small trial could soon test that approach in health care workers in the DRC. But most researchers are betting on quickly developing a vaccine specific to Bundibugyo.

The first of them is now being tested in a phase 1 safety trial at Oxford. It uses a surface protein of Bundibugyo stitched into a chimpanzee adenovirus. A messenger RNA vaccine developed by Moderna is likely to start a phase 1 trial next month, says Richard Hatchett, who leads the Coalition for Epidemic Preparedness Innovations (CEPI), which is funding both vaccine candidates. A third vaccine, also funded by CEPI, uses a livestock virus as a vector, like the licensed Ebola vaccine against the Zaire species. Even under the best of circumstances, a vaccine is unlikely to be available for efficacy trials in the DRC before October.

Ebola vaccines are typically used not to immunize entire populations, but "rings" of contacts around known cases. For that strategy, too, it will be crucial to find Ebola cases early and track the people they have been in contact with. In a modeling study posted as a preprint last week, Bogoch and others reported that even a highly protective vaccine will not have much impact if contact tracing is struggling.

Given all the obstacles, controlling this epidemic will be "like turning a cruise ship," Bogoch says. "It's going to turn, but it's going to turn slowly, with sustained pressure over time." □

ECOLOGY

Contagious fish cancer overruns New England lake

Genetic studies reveal rare example of identical transmissible tumors

TIM APPENZELLER

Lake Memphremagog stretches an idyllic 50 kilometers through woods and farmland from northern Vermont into southern Quebec. "That lake has great fishing … big lake trout, big salmon, big rainbow trout, a great perch fishery," says Peter Emerson of the Vermont Fish & Wildlife Department. Its abundant brown bullhead catfish (*Ameiurus nebulosus*), however, are less enticing: Nearly one-third are spotted with an inky-black cancer.

It's a rare and strange form of the disease, a team of geneticists and fisheries biologists including Emerson shows in a study this week in *Nature*. Unlike most cancers, this one—a melanoma—is not an abnormal growth of each fish's own tissue. Instead, it appears to be a single cancer that has somehow spread from fish to fish: a giant malignant clone.

What the researchers are calling brown bullhead melanoma joins only a handful of other known transmissible cancers. The most notorious is a ghastly face-eating tumor in Tasmanian devils, which transfer the cancer cells when they bite each other. Dogs can catch a transmissible venereal tumor during mating, and some clams, cockles, and mussels have a transmissible leukemia. But the melanoma is the first such cancer to be found in a fish, and it brings a host of mysteries, including when and where it originated, how it spreads, why the catfish are susceptible—and what it bodes about other contagious cancers yet to be discovered. Answering those questions, says Elizabeth Murchison, a transmissible cancer expert at the University of Cambridge, "is going to be so exciting."

The cancer may have a long pedigree. In 1858, writer Henry David Thoreau described a brown bullhead in a Massachusetts stream whose head was "covered with an inky-black kind of leprosy, like a crustaceous lichen." In 1925, a zoologist who visited Cape Cod reported a "black tumor of the catfish" that was widespread in a pond. But both reports were overlooked until 2012, when alarmed anglers in Vermont began to contact the Fish & Wildlife Department about catches pulled from Lake Memphremagog. It didn't take long to confirm the problem. "It was remarkable how common it was," state fish pathologist Tom Jones recalls.

Jones got in touch with Vicki Blazer, a fish pathologist then at the U.S. Geological Survey, who looked at samples under the microscope



A transmissible cancer begins as superficial spots, but can invade internal organs of catfish.

PHOTO: VERMONT FISH & WILDLIFE DEPARTMENT

and saw the chaotic pigment-rich cells that signal melanoma. As word of the cancerous catfish spread, a local residents' group called DUMP—Don't Undermine Memphremagog's Purity—raised the alarm about a landfill near the southern end of the lake, which it feared was leaking contaminants into the water. But the cancers were no more common near the dump than in more distant parts of the lake, and Blazer only detected low levels of per- and polyfluoroalkyl substances (PFAS), the carcinogenic "forever chemicals" that were an early suspect, in the blood of the affected fish.

Catfish DNA points to a different cause, says data scientist Julie Dragon of the University of Vermont, senior author of the *Nature* study. The team analyzed variations in mitochondrial and nuclear DNA from cancerous and healthy tissue in the affected fish, as well as tissue from unaffected bullheads around New England. Because tissues with similar genetic variations presumably have a common origin, the researchers could build family trees of the tumors and the fish.

Typical cancers are more closely related to their host than to tumors in other individuals. Not this one, Dragon says. "We built these trees, and it turned out that all the tumors clustered together, and all the hosts clustered together." The unexpected pattern pointed to a single, separate origin for the cancer, as if it was a distinct organism that had colonized the fish.

Blazer, who was originally part of the research team, isn't convinced. Before the fish develop full-blown cancers, she says, their tissues show inflammation and cellular changes that suggest the effects of a contaminant such as arsenic, which she found in unusually high levels in affected catfish. It's possible, Blazer suggests, that pollution simply abets the transmissible cancer by lowering the fish's resistance, but she thinks it is more likely to be the primary cause. "I am not on board yet with the transmissible tumor. That's why I took my name off the paper," she says. Local residents, too, are not ready to exonerate water pollution.

But Murchison, who was not part of the research team, says the genetics leave "no question. ... It is a clone, a single clone that's spreading through the population." Michael Metzger, a biologist at the Pacific Northwest Research Institute who discovered the transmissible leukemia in shellfish, agrees. "There isn't really any other explanation for the genetic results."

Since its discovery the cancer has turned up in other lakes and ponds in the Northeast and Canada. Wherever it occurs, its genome looks more like that of healthy brown bullheads from New Hampshire and Maine than from Lake Memphremagog, Dragon says. That suggests the original cancer might have arisen in another water body, perhaps hundreds of years ago, and invaded the Vermont lake not long before it was detected. But how the melanoma traveled is a mystery, as is how it spread so quickly through the long, deep lake.

The age of the afflicted fish offers one clue. "You look at younger fish, and they've never got cancer," Emerson says, but it is rampant after sexual maturity at 3 years or so. Mating is one of the rare times the generally solitary catfish encounter one another. "They're aggregated in these clusters, and they have little spines behind their pectoral fins," Dragon says. "So it's possible they scratch each other."

Or perhaps cells from the cancerous fish spread like a waterborne pathogen. By putting afflicted and healthy fish together in laboratory tanks, Blazer and others hope to investigate those possibilities—and, she says, test whether the cancer really is transmissible.

Brown bullheads remain plentiful in Lake Memphremagog, suggesting the cancer doesn't quickly kill the fish, even though it invades internal organs. At the same time, the cancer's frequency is a sign that the fish don't mount an effective immune defense.

"Every new transmissible cancer provides a whole new perspective," Murchison says. To her, the catfish cancer suggests nature has many similar surprises in store. "I think it's really, really likely that there are more out there." □

IN FOCUS



PHOTO: TAVIPHOTO/ISTOCK

Headbutting gives goats brain damage

Goats may suffer from the perils of headbutting more than assumed, researchers report in a preprint on bioRxiv this month, which finds early signs of neurodegeneration in the animals at just 1 year old. Goats clash with each other to assert dominance and to play, but their skull shape and beefy necks were thought to protect them from long-term neurological damage. Scientists tracked three goats that butted heads over 6 months, finding that their brains at the end of the study contained abnormal tau proteins and evidence of neuroinflammation, both associated with neurodegenerative disease in humans. This damage may accumulate with time, mirroring what happens to athletes exposed to regular head impacts. The researchers hope goats could be a better animal model for studying brain injury than rodents, because their brains are closer in size and structure to humans. —Ailie McWhinnie



PHOTO: VALERY HACHE/AFP VIA GETTY IMAGES

GLOBAL WARMING

Climate scientists sharpen tools for linking global warming to extreme weather

National Academies report says more rigorous results will boost confidence in fast-growing field **JULIA VAZ**

June's European heat wave, in which temperatures above 40°C led to the deaths of more than 10,000 people, made headlines not just for its severity, but for what it said about climate change. Earlier this month, a group called World Weather Attribution (WWA) linked it firmly to global warming, saying such a heat wave would have been "virtually impossible" 50 years ago. It was a high-profile claim from a field called extreme event attribution, which seeks to gauge how global warming impacted a particular heat wave, flood, or storm. The field is still evolving, according to a report released last week by the National Academies of Sciences, Engineering, and Medicine (NASEM).

"We need to make sure that we are clearly communicating the uncertainties of any given study," says James Hurrell, an atmospheric scientist at Colorado State University and chair of the NASEM committee that wrote the report. "But I think the report is also trying to make the point that we can reduce some of the uncertainties." It recommends researchers apply multiple attribution methods, adopt common standards, and communicate inherent uncertainties.

In the past decade, attribution science has advanced rapidly, with the annual number of studies more than doubling since 2012, according to the report. Better climate models, longer observational records, and more sophisticated statistical techniques have all improved researchers' abilities to detect climate change's influence. "The first attribution studies that come to my mind appeared in the early 2000s," Hurrell says. "So, it's a pretty new endeavor."

It has moved well beyond academic journals. Groups such as WWA now routinely release assessments within days of disasters, without peer review. Attribution studies are increasingly cited in lawsuits seeking to hold fossil fuel companies financially responsible for climate-related damages. Some researchers are attempting to trace portions of those damages to emissions from specific sources and companies.

The field also faces political scrutiny, which dogged the NASEM committee. Republican lawmakers and groups allied with the fossil fuel industry questioned whether some members had conflicts of interest because of ties to climate advocacy organizations and sought their internal communications through public records requests. Hurrell dismisses the criticisms, saying the report reflects the consensus opinions of the committee as a whole. "We really didn't let the outside external chatter influence us in any way," he says.

Because every weather event arises from a combination of many influences, including natural variability, assessing the role of global warm-

A café in Nice, France, during June's European heat wave. Attribution studies found the event would have been virtually impossible 50 years earlier when greenhouse gas emissions were lower.

ing is challenging. Most attribution studies compare climate model simulations of today's greenhouse-warmed world with simulations of a counterfactual world without human warming. Scientists then compare how often a given type of event occurs in both models, and how similar it looks.

A more computationally demanding method feeds data from a specific event into a weather model while removing the effects of greenhouse gases. It generates the atmospheric conditions surrounding the event in this hypothetical world, which can then be compared with actual observations.

Other methods forgo modeling entirely. Davide Faranda, a climate scientist at Paris-Saclay University, compares data on weather patterns leading up to a recent extreme weather event with historical analogs from the mid–20th century, predating most warming. The challenge is the quality of the available historical data, which is especially limited in the Global South, the report finds.

Confidence in the outcome of these methods depends on the type of event in question, the report concludes. Heat waves and heavy rainfall are closely linked to rising temperatures and yield the most confident attribution results. But hurricanes and severe storms are shaped not only by warming, but also by complex atmospheric dynamics that climate models still struggle to capture. And disasters such as wildfires can defy modeling, because they depend on human behavior, such as land management and accidental or deliberate ignitions.

Different attribution studies can also produce substantially different results for the same event. A 2025 study comparing analyses of a record-breaking 2021 heat wave in the Pacific Northwest found that some methods concluded the event was 340 times more likely now than in a world without global warming, whereas others found it was only eight times more likely. "The sort of simple declarations are always suspicious because they're likely just one of many valid approaches," says Philip Mote, a climate scientist at Oregon State University who served on a 2016 NASEM review of attribution science.

The committee argues that confidence increases when studies combine different approaches to validate their results. It also recommends investment in higher resolution climate models and "a common overarching framework with recommended best practices." And it urges groups that produce rapid attribution studies to routinely submit their analyses for peer review.

No one should expect certainty, says Friederike Otto, a climatologist and WWA's founder. The purpose of attribution science is not to claim that climate change alone caused an extreme event, but to calculate how it increased its odds or intensity, she says. "If you say smoking causes cancer, you also mean that smoking dramatically increases the likelihood of you developing cancer." □



PHOTO: DARRYL PITT/MAINE MINERAL & GEM MUSEUM

PLANETARY SCIENCE

Oldest known Mars rock offers glimpse of planet's watery youth

Teghaza meteorite hints at surprisingly Earth-like crust and shows Mars was already losing its water 4.1 billion years ago **PAUL VOOSEN**, in Montreal

A meteorite discovered 4 years ago in the Sahara Desert by Sahrawi prospectors may be the oldest ever identified from Mars, preserving a fragment of the planet's crust from more than 4.1 billion years ago. Although analyses are still unfolding, the nearly 800-gram rock, which split into eight pieces, is already offering an unprecedented glimpse of the planet's watery youth—and challenging traditional views about how its crust formed.

The meteorite, called Teghaza 001, "will revolutionize the way that we think about early Mars," said Lee Saper, a geochemist at NASA's Jet Propulsion Laboratory (JPL), in a talk last week here at the Goldschmidt geochemistry conference. One study, examining hydrogen trapped in the rock's minerals, appeared last month in *Science Advances*; another, establishing its age, is under review at *Science* and is available as a preprint, and more are underway.

Chunks of Mars blasted into space by impacts occasionally make it to Earth as meteorites. Because the martian atmosphere has a distinctive fingerprint of noble gases, researchers can distinguish martian meteorites from other space rocks by tiny amounts of gas trapped in them. Although most of the planet's surface is likely more than 3.5 billion years old, nearly all of the several hundred known martian meteorites are relatively young volcanic basalts—hard, durable rocks that better withstood their violent ejection from Mars and passage through Earth's atmosphere.

Until now, only one known martian meteorite, 4.1-billion-year-old Allan Hills 84001, discovered in Antarctica in 1984, preserved crust from that first chapter of martian history. (Another meteorite, nicknamed Black Beauty, contains individual mineral grains as old as 4.4 billion years, but assembled into a rock much more recently.) Allan Hills provided critical insight into early Mars, but relying on one sample has always been uncomfortable, says Lydia Hallis, a planetary scientist at the University of Glasgow. Teghaza

Rich in silicon, Teghaza 001 looks more like a granite than a basalt—unexpected for such an ancient Mars rock.

changes that. "We've doubled our old meteorites," she says. "There will be a lot of people, including me, who want to get a piece of this." (The main mass of the meteorite is owned by the Maine Mineral & Gem Museum, which provided the small sample used in the studies.)

Teghaza is rich in zircons, mineral crystals prized by geologists because tiny amounts of uranium trapped within them decay into lead at a known rate, providing a precise clock. Work led by Yang Liu, a planetary scientist at JPL, dates the meteorite to more than 4.1 billion years ago. The rock may actually be older, however. When the mineral grains melt or are exposed to fluids, their clocks are reset to zero. "There's more of a story to this rock than most other martian meteorites," says Christopher Herd, a geologist at the University of Alberta who also works on the rock sampling efforts of NASA's Perseverance rover.

Perhaps Teghaza's biggest surprise is its composition: unusually rich in silicon. "It's very strange, we don't expect that," says Eva Scheller, a planetary scientist at Stanford University. Most of Mars's surface is made up of basalts, the dark volcanic rocks produced by direct melting of the mantle, much like Earth's ocean crust. By contrast, lighter, silica-rich rocks such as granite generally require repeated episodes of melting and chemical separation to evolve from basalt. On Earth, those processes are closely associated with plate tectonics, leading many researchers to assume that Mars, which lacks active tectonics, has little granitelike crust.

That picture has begun to change. Mars orbiters have spotted hints of granitelike rocks exposed by impacts. The density of the crust, inferred from marsquakes captured by NASA's Insight lander, is lower than that of pure basalt. Other martian meteorites contain bits of quartz and related silica-rich minerals. And Perseverance, exploring terrain similar in age to Teghaza, last year discovered a quartz-rich outcrop, dubbed Plankeholmane, that looks almost exactly like granite, Liu reported earlier this year at the Lunar and Planetary Science Conference.

Taken together, the findings suggest that even without plate tectonics, Mars—and perhaps distant planets—can nevertheless generate complex magmatic systems capable of producing compositionally diverse crust. The repeated recycling of crust into magma and back again could have helped create the rich chemical assemblages that are necessary for life, says Tobermory Mackay-Champion, a geophysicist at the University of Bristol who led a study published last month in *Nature Astronomy* proposing such a system. "Compositionally diverse crust may be far more common on rocky planets than previously thought."

Teghaza also strengthens evidence that Mars began to dry up very early in its history. Hydrogen, the key ingredient in water, provides one of the best tracers of the planet's water. The meteorite's ratio of hydrogen to its heavier isotope, deuterium, closely matches measurements from Allan Hills. Both indicate early Mars contained a higher ratio of hydrogen to deuterium than it does today. Because lighter hydrogen escapes to space more readily than deuterium, the changing ratio implies a large loss of hydrogen, and hence of water.

Mars's ancient magnetic field would have shielded the atmosphere from charged particles in the solar wind, reducing the escape of hydrogen. But after the field disappeared about 3.9 billion years ago, researchers believe water loss accelerated. The new measurements, which also include the hydrogen-deuterium ratio the planet formed with, recorded in parts of the meteorite derived from unaltered mantle rock, show this loss was already well underway 4.1 billion years ago. "This is the product of rapid hydrogen loss from the juvenile martian atmosphere," Saper said during his talk.

Researchers expect Teghaza to yield more surprises. The rock clearly records alteration by water or other fluids in the crust, but the list of possible sources is long. More scrutiny of this piece of early Mars could reveal whether the planet resembled the volcanically active, water-altered landscape that many geologists envision occurred on Earth during its first several hundred million years, says James Day, a geochemist at the Scripps Institution of Oceanography. So far, he adds, "that's what it appears to be." □

BIOMEDICINE

Alzheimer's trial results lend momentum to drugs targeting tau

First drug to lower the protein and slow cognitive decline created buzz despite puzzling data **JENNIE ERIN SMITH**

Alzheimer's disease has long been seen as a tale of two proteins: beta amyloid, which forms sticky plaques in the brain, and tau, which in its diseased state creates tangles inside neurons. Antibodies that clear beta amyloid have been approved to treat Alzheimer's, but it was tau that made headlines last week, as researchers presented key new details about a drug that reduced production of the protein and slowed cognitive decline in a recent clinical trial.

At the Alzheimer's Association International Conference, neurologist Catherine Mummery of University College London presented results from a phase 2 trial testing diranersen, a drug developed by Biogen to lower the body's production of tau, in more than 400 patients with early-stage Alzheimer's. On the study's main measure of cognition, participants getting diranersen saw as much as a 26% slowing of decline—about on par with the effect seen in earlier trials of approved antiamyloid drugs. (Biogen had announced in May that the drug slowed cognitive decline but did not say by how much.)

Although the presentation sparked enthusiasm from Alzheimer's researchers, it also drew attention to puzzling aspects of the trial results. A potentially worrisome side effect emerged at high doses, and contrary to expectations, patients taking the lowest of three possible doses saw the greatest benefit. For these reasons, the findings represent "a double, not a home run," says neurologist Adam Boxer of the University of California San Francisco, who this month launched a clinical trials platform to try different antitau therapies in Alzheimer's.

Diranersen belongs to a class of drugs known as antisense oligonucleotides, strands of RNA that dampen activity of specific genes. It interrupts production of tau by binding to the messenger RNA that encodes instructions for producing the protein, and must be injected directly into the cerebrospinal fluid to reach the brain. After 18 months, participants in all three dose groups had between one-third and one-half as much tau in their cerebrospinal fluid as when they started the trial, whereas levels increased slightly among people receiving placebo. Imaging done on a subgroup of participants showed the drug also reduced tau tangles in the brain.

Glenn Harris, who directs research partnerships for the Rainwater Charitable Foundation, a group that supports research on brain diseases involving tau, says many in the field had higher expectations for the drug's cognitive benefit. But the findings create momentum for tau therapies at a time when several other candidates—mostly antibodies targeting abnormal tau—have failed in Alzheimer's trials. "I think the enthusiasm is highest with this one because [it works by] shutting off the faucet of [tau] production or at least slowing it," he says, apparently allowing the brain to clean out tangles.

The puzzle is that cognitive benefits decreased at higher doses, which caused the trial to fall short of the dose-dependent effect the investigators had chosen as its primary endpoint. The low dose appears to be "showing the most consistent positive clinical effect" across neuropsychological tests, Mummery said at the meeting. Biogen is moving forward with a larger clinical trial, but has not revealed which doses it will investigate.

About one-quarter of people in the higher dose groups experienced a "confusional state" after diranersen was administered, a side effect that lasted up to 1 week. This was unexpected, Boxer says, and might be an off-target effect of the antisense molecules. "But a troubling possibility," he says, is that it is related to tau reduction. "Maybe we can't knock down tau without some effects."

The cognitive benefit seen with diranersen is not enough to argue it should replace the approved antiamyloid antibodies, says Washington University in St. Louis neurologist Eric McDade, who co-directs a research group that tests Alzheimer's drugs in families with inherited forms of the disease. Combination treatment with both drug types appears to be "the strongest way forward," he suggests. "Maybe there is a synergistic effect—but that needs to be tested now."

A few trials are testing antitau and antiamyloid drugs in combination, though none involve diranersen. One, by McDade's group, uses an antitau antibody, and Boxer's group is working with a vaccine that prompts the immune system to attack abnormal tau.

The modest cognitive effects in the trial may speak to the complexity of the disease process, Boxer says. "There's been a long-standing hypothesis among Alzheimer's researchers that amyloid lights the fire of the disease, but that it's tau [fueling neurodegeneration]," Boxer says. These data "provide pretty clear support for that." But Boxer thinks other abnormal proteins seen frequently in Alzheimer's brains—among them alpha synuclein and TDP-43—might be stoking the flames, too, possibly lessening the beneficial effects of tau lowering.

And the disease involves a lot more than proteins, researchers have increasingly come to acknowledge. "We know that there's a huge vascular component [in Alzheimer's]," Harris says. "There's a huge neuroinflammation component and neuroimmune component and synaptic health component." Fortunately, he says, "There are pipelines for those therapies as well." □

THEY SAID IT

This method is misleading and [the U.S. government] should fix it.

Data analysts Abigail Haddad and Christopher Marcum, on their Substack, explaining why the number of public comments on a controversial plan to give political appointees greater control over grantmaking is likely about 54,000—not the 496,775 reported on a U.S. government website. A quirk enables website users to artificially boost the tally, they wrote.

SCIENCE & SECURITY

Detention of U.S. seismologist by China alarms researchers

Youlin Chen analyzed earthquake data that can also aid nuclear test monitoring **RICHARD STONE**

On 5 November 2024, Chinese-born U.S. seismologist Youlin Chen was preparing to fly home to Boston after a visit to China, where he celebrated his mother's 80th birthday and gave talks at two universities. For years, Chen, who became a U.S. citizen in 2011, had taken annual trips to China to see family and meet with collaborators on projects for his employer, EMR Solutions & Technology, a Florida-based consulting firm.

But that day state security officers arrested Chen at Beijing's international airport. Over the following weeks, officials interrogated him more than 100 times about his research, says Chen's wife, Yufang Rong, also a seismologist. He was often forced to sit on a hard stool for hours on end, she adds. Chinese authorities charged Chen with espionage on 1 May 2025 and he remains in detention without a trial, as Reuters first reported earlier this month.

One strand of Chen's work used Chinese data to analyze North Korea's six underground nuclear tests for the U.S. Department of State. But none of his work is or was classified, says EMR Senior Vice President Hafidh Ghalib. "He never had a security clearance," so the company had "no concerns for his safety."

"This is a really tragic case and one of great concern," says nuclear nonproliferation expert Siegfried Hecker, a former director of Los Alamos National Laboratory now at Stanford University.

Chen's detention has heightened alarm over a separate clampdown on the kind of geophysical data he analyzed. Under a policy posted in 2016, seismic networks and other government-funded projects in China were expected to deposit raw observations and processed data in a national platform operated by the China Earthquake Networks Center. Although CENC restricted access to some sensitive data, other records could be downloaded freely or were available to domestic or foreign users after registration. But on 26 March, CENC's data platform announced it would no longer provide raw seismic observations or basic information about monitoring stations, except in specially approved cases. Citing data security and national security, it said users would instead receive processed data or customized analyses.

That move, and Chen's detention, have sent a chill through international seismology. Three U.S. seismologists who have worked with Chinese data declined to be interviewed, citing concerns that their collaborators in China could get in trouble. "This must put tremendous fear into Chinese seismologists who have been collaborating with foreign colleagues," says Frank von Hippel, a senior research physicist with Princeton University's Program on Science and Global Security. As one U.S. seismologist put it, "We're petrified for them."

Placing China's raw data off-limits could impede studies of earthquake hazards, fault systems, and Earth's interior, and make Chinese findings harder for outsiders to reproduce, say the seismologist and others. But Chen's own research demonstrates how difficult it is to separate civilian and security applications of such work.

Chen had openly collaborated with Chinese government and university seismologists for well over a decade. A collaborative study he led, published in *Geophysical Journal International* in July 2024, used more than 31,000 waveforms recorded by 346 stations across eastern Asia, including China's seismic network, to produce maps of how seismic energy is absorbed by the crust. Such analyses can help estimate the size of seismic sources and improve predictions of ground motion and earthquake hazards. The paper disclosed funding from the U.S. Air Force Research Laboratory and the State Department and thanked CENC staff for providing data and technical support.

The same methods, however, can aid nuclear test monitoring. A December 2020 technical report Chen authored under a contract from the State Department's arms control bureau sought to improve estimates of the magnitude and yield of North Korea's nuclear tests and distinguish their signals from natural earthquakes. Ghalib says that, adhering to Chinese rules then in force, Chen analyzed the restricted waveforms in China and did not take the raw records out of the country.

No public evidence links Chen's detention to CENC's subsequent data restriction. But the strategic sensitivity of China's seismic records is illustrated by two small signals detected near China's Lop Nur nuclear test site in June 2020. In February, U.S. officials alleged that they resulted from a concealed low-yield nuclear explosion; China denies the accusation. International monitoring data were insufficient to determine whether the signals were natural or explosive, and seismologists have said finer grained records from China's much denser network might resolve the question.

"There are no doubts that seismologists worldwide, including us at EMR, have examined seismic waveforms from events near Lop Nur recorded by stations outside China," says Ghalib, also a seismologist. *Science* found no evidence that Chen analyzed the signals using waveforms from Chinese seismometers.

In March, the State Department determined Chen is wrongfully detained. Last week, a spokesperson for the Ministry of Foreign Affairs of the People's Republic of China rejected that designation, stating, "There is no instance of wrongful detention." A 16 July commentary on the Chinese portal 163.com claimed that in November 2024 Chen "surreptitiously deployed temporary monitoring instruments to collect extensive, classified geological data firsthand" in "strategic mountainous regions."

Rong says, "These are crazy stories. He never went to those places and never touched any instruments."

After news broke of Chen's detention, Rong's own collaborators in China sent her frantic messages asking her to remove her name from a paper to appear in the October issue of *Engineering Geology*. Rong says she tried, but it was too late—the article had already posted online on 26 June.

Chen's detention has prompted the State Department to consider discouraging U.S. citizens from traveling to China, says Eric Lebson, chief strategy officer at Global Reach, a nonprofit that works to free U.S. citizens held abroad. The department, he says, is also mulling labeling China a State Sponsor of Wrongful Detention. Created by executive order last year, the designation has so far been applied to Iran and Afghanistan, and can lead to sanctions, travel restrictions, export controls, and other penalties.

In the meantime, U.S. and Chinese cooperation in seismology hangs in the balance. "Many of our members, including those in China, depend on data collected and shared by regional and global seismic networks for their research," the Seismological Society of America, of which Chen is a member, said in a statement.

U.S. embassy staff who have visited Chen and spoken with his lawyer—who was not allowed to see Chen until early this year and is not permitted to discuss the case with anyone other than Chen—have told Rong her husband has lost about 18 kilograms. "We tried to give him shoes for exercise," she says, "but that was not allowed." □



Seismologists Youlin Chen (right) and Yufang Rong, who are married.

PHOTO: CHEN FAMILY



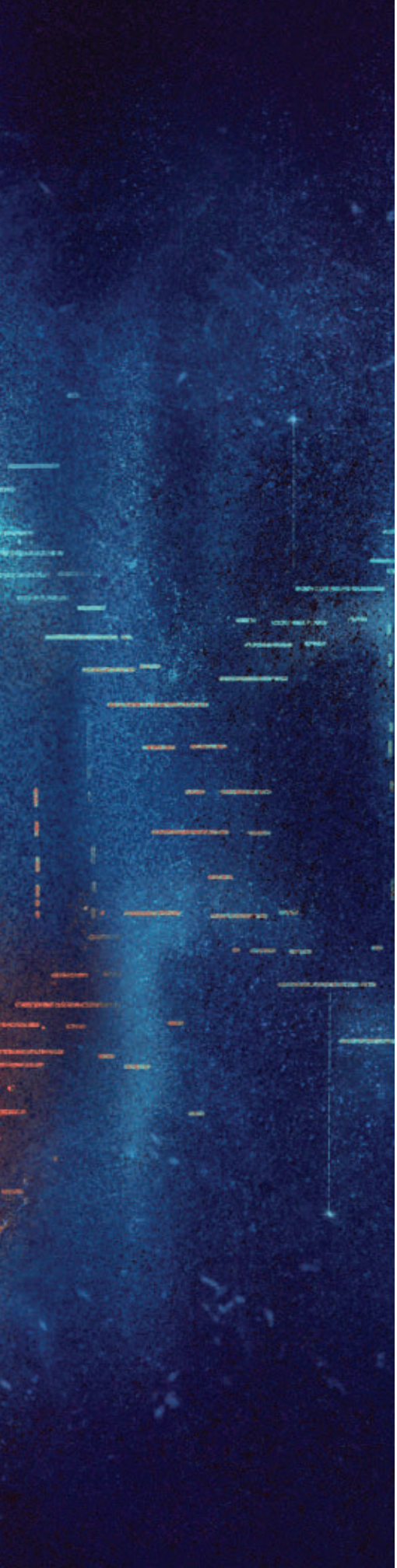


ILLUSTRATION: DARIA LADA

A FATAL REACTION

A cutting-edge gene-editing trial in China went disastrously wrong. A family wants accountability

BRENDAN BORRELL, Retraction Watch

The young girl tugged on her mother's hand as they pressed through the doors of the hospital in Shanghai. She was 6 years old, bouncing along in a pink jacket and blue pants decorated with cartoon bears. Behind them, her father rolled a large suitcase with everything the child needed for the weeklong stay: stuffed animals, Play-Doh, an iPad loaded with episodes of *Peppa Pig*.

She told her parents it felt like they were going on vacation. In fact, they brought her here for an experimental gene therapy. The girl was slipping behind her peers in kindergarten. She still spoke in simple sentences and ate with training chopsticks. Underneath it all was a single mutated DNA base, a T that should have been a C.

"Your 'book' has a small mistake, which has caused you to have a disease that affects your growth," read the children's version of the informed consent form from the hospital. "Over time, it can get more serious." Doctors hoped to repair that mistake while her brain was still building itself. It would be a clinical trial of one, funded in part by $860,000 the parents had scraped together from their own savings and from relatives.

The parents felt they were in good hands. Xinhua Hospital, which is affiliated with the Shanghai Jiao Tong University School of Medicine, was acclaimed for its pediatrics department. It was the first Chinese institution to perform open heart surgery on infants, and the first in the country to separate conjoined twins. If all went well, the hospital would make history once again. The girl would be the first person in the world to receive a gene-editing therapy directed at the brain. It would rewrite the mutated gene in her neurons, restoring the needed DNA base so she could make a vital protein.

Leading the effort was Zilong Qiu, a neuroscientist at the university's brain center, the Songjiang Research Institute. At the time, in late March 2025, Qiu was one of several researchers around the world vying to push base editors—a more precise form of the powerful gene editor CRISPR—into custom treatments for children with rare diseases. One month earlier, KJ Muldoon, an infant with a life-threatening metabolic disorder, had quietly received his first intravenous infusion of one such treatment at the Children's Hospital of Philadelphia.

Although news that base editing saved "Baby KJ" would soon rocket around the world—*Science* named the feat one of the runners-up for its 2025 Breakthrough of the Year—the story of what happened at Xinhua Hospital has remained hidden. An entry for the study posted to ClinicalTrials.gov has not been updated for more than a year. And when Qiu and his colleagues published proof-of-concept animal studies related to the trial in *Nature* early this year, they stripped the paper of references to the family and its financial contributions, noting only that "bridging the gap between preclinical research and clinical translation remains a significant challenge."

That vague language glossed over tragedy: Seven days after the girl's medical team infused trillions of viruses carrying the recipe for the base editor into her spinal fluid, she died of a severe immune reaction linked to the therapy, *Science* and Retraction Watch can now reveal.

According to official documents and accounts provided by the girl's parents, the hospital had allowed Qiu's experimental treatment to proceed under a regulatory provision that does not require approval from national regulators. After the child's death, the hospital paid a modest fine to a local health authority but Qiu was not publicly sanctioned.

The troubling death represents a new stumble in China's push to rival the United States as a biotech power. Three years ago, the government strengthened its biomedical research rules in response to the 2018 scandal around He Jiankui, a biophysicist who violated ethical norms to secretly produce gene-edited babies. The lax oversight of this recent trial and the failure to publicly report the fatality "shows the gap between what is intended and what has been put in place," says Joy Zhang, a sociologist at the University of Kent who has written about the pervasive culture of secrecy in Chinese scientific institutions.



Seven experts in fields including genetics, virology, and bioethics who reviewed details of the *Nature* study and the clinical trial for *Science* and Retraction Watch expressed concern that Qiu and his team downplayed the trial's risks in describing them to the parents, overlooked safety signals in animal studies, and proceeded even though success was unlikely. "This shouldn't have gone to trial," says Steven Gray of the University of Texas Southwestern Medical Center, who develops viruses for gene therapy.

Gray and several of the other experts are calling for a full review of the images and other data in submitted and published versions of the *Nature* paper and full disclosure of the study's funding. Some of the issues might warrant a retraction, they say. Neither Qiu nor his university or the hospital responded to multiple requests for comment for this story, and *Nature* says it was not aware of the issues surrounding the clinical trial before it published the group's paper.

The girl's parents, who requested that *Science* use pseudonyms for them and their daughter for privacy reasons, have decided to tell her story now because they are angry about what they feel is a lack of accountability by the researchers and the institutions. "Learning the reality of these missing safeguards has fundamentally changed how we now view the entire project," says the father, a software engineer. He asked that he be called Jason, his wife Linda, and their daughter Mei (Chinese for "beautiful"). "We did not realize how unusual and dangerous many of the arrangements were."

As with Jesse Gelsinger, the 18-year-old whose death during an early U.S. gene therapy trial in 1999 chilled research in the field for nearly a decade, Mei's story raises questions about whether such risky treatments should be initially tested on patients with nonfatal conditions. "I don't think these kinds of experiments should stop, but we need to make sure we are appropriately careful about them," says bioethicist Hank Greely, director of the Center for Law and the Biosciences at Stanford University. "It's so easy to be blinded by hope—whether it's hope for your kid, hope for your research, or hope for your company."

JASON AND LINDA had been trying to have a child for 4 years before Mei was born in early 2019. They marveled at every milestone: her first words, her first steps. She did seem a little clumsier than other toddlers. But within a few years, she was attending Parkour classes and bouncing on a neighborhood trampoline.

When Mei was 4, one of her kindergarten teachers pulled Linda aside: Mei didn't draw or write as well as the other kids and her language skills weren't developing normally. Her mother might want to get her evaluated, the teacher said. In March 2023, Mei was diagnosed with global developmental delay, a broad label with many causes. Specialists explained that some of Mei's behaviors—the funny sounds she liked to make, for instance—were associated with autism.

The family followed up with a battery of tests at the Shanghai Children's Medical Center. A brain MRI, chromosome analysis, and a test panel for genes known to influence neurodevelopment all came back normal. Detailed sequencing of more of Mei's DNA finally pointed toward the delay's root cause: a defect in a gene, called *CHD3*, that influenced how strands of DNA were wound into chromosomes inside Mei's cells.

This packing affects how well other genes are expressed, including those involved in the early development of the brain. By altering gene activity, *CHD3* mutations produce a condition called Snijders Blok-Campeau syndrome that is known to afflict just 237 individuals around the world.

Philippe Campeau, a medical geneticist at the University of Montréal who helped identify the syndrome's genetic basis in 2018, says people with the mutation often have a normal life expectancy, but their symptoms vary widely. Most have slightly larger than normal heads, and about two-thirds have intellectual deficits. Moderate to severe cases may be nonverbal, suffer from seizures and heart problems, and have fluid-filled voids in their heads. For those patients, gene-editing therapies could be revolutionary, according to Campeau. "For somebody who is just mildly affected … that's a bit more debatable," he says.

Mei sat on the milder end of the gradient, but Linda soon quit her job to devote her energy to their only child. She and Jason took Mei to speech and occupational therapists and got her special education support. It worked,

ILLUSTRATION: DARIA LADA

to a degree. After they transferred Mei to another kindergarten to repeat a year, most of the other parents weren't even aware of her condition, they said.

Whereas Linda was an optimist by nature, Jason feared the worst. The condition is not degenerative, but the gap between Mei and her peers in school would likely widen as courses grew harder. She was also skinnier than other children and had poor muscle tone. She might not ever be able to live independently. "We were worried about her future," Jason says.

After Mei's diagnosis in 2023, the family joined a private WeChat group in which parents of children with autism and similar disorders trade advice. It was there that they first heard about Qiu.

An expert in neurodevelopmental disorders, he did his Ph.D. at the Shanghai Institute of Biochemistry and Cell Biology and, like many promising Chinese students of his generation, went to the U.S. for a postdoctoral position. In 2003, he joined the laboratory of Anirvan Ghosh, a neuroscientist then at the University of California (UC) San Diego, who found Qiu to be "talented and motivated."

In 2009, as China was ramping up efforts to lure its scientists back home, Qiu took a post at his undergraduate alma mater. He published in all the right places—*Nature*, *Neuron*, and the *Proceedings of the National Academy of Sciences*. He also strove to demystify genetics for the Chinese public with a book, *The Gene Enlightenment*, and through public lectures, including a TEDx talk titled "Defying the destiny dictated by our genes."

Qiu was among more than 100 Chinese scientists who condemned He's rogue embryo gene editing in an open letter. "The project completely ignored the principles of biomedical ethics, conducting experiments on humans without proving it's safe," he told the *South China Morning Post*.

Qiu still saw potential in tackling autismlike conditions that have a clear genetic link, such as Rett syndrome. "He was taken by Rett syndrome and wanted to make a difference," says Monica Coenraads, CEO of the Rett Syndrome Research Trust, which funded some of Qiu's research in the U.S.

For parents in the WeChat group, Qiu was someone to watch. His gene therapy for Rett syndrome had recently been patented in China, and he founded a company, Lanqi Xintu Gene Technology, which would soon bring it into the clinic for testing. In July 2023, he gave a lecture about this work and other efforts. A parent from the group attended, and posted slides.

The following week, Jason sent Qiu an email describing Mei's condition. "We understand that gene therapy requires complex experiments and can cost tens of millions [Chinese yuan]. We are willing and able to bear these costs," he wrote. "Watching our child suffer every day is the most painful ordeal for us as parents." He attached Mei's genetic results and hit send.

Thirteen minutes later, Qiu replied with interest. He soon invited Jason and Linda to meet at the institute. Seated in his office with a leafy window view, the parents told him they weren't interested in funding research for the sake of research. "We are really coming at this with the goal of an actual treatment," Jason said.

Qiu was encouraging. "To be honest, if you had come to me last year, I wouldn't have dared to say we could do this," he said during the meeting. "We've reached a tipping point now." (The parents recorded many of their conversations with Qiu and the other parties related to Mei's trial because they were overwhelmed by the technical information. They shared the recordings with *Science* and Retraction Watch.)

Editing the brain, Qiu explained, would require delivering viruses into the fluid in the base of Mei's spine. Because her condition was not life-threatening, however, the bar would be high to get the trial approved by the hospital's ethics board, he added. It would require up to 2 years of work to develop the treatment and test it in mice and monkeys, with no guarantee it would be effective. But a top-tier institution such as Xinhua Hospital, Qiu said in one recording, could ensure there would be no safety catastrophes.

The parents had heard about serious side effects, including deaths, caused by other gene therapies, and knew the greatest risk would be Mei's immune response to the massive dose of virus. Qiu said getting the dose right was critical, but infusing the viruses directly into Mei's spinal fluid, rather than the blood, would minimize the threat of a reaction because it would bypass the kidneys and liver.

The parents were starting to feel at ease. "The psychological cost of doing this is very high for me," Linda told Qiu. "But I am willing to try this because you have proven to be very trustworthy."

JASON AND LINDA say they soon wired their first payments to a foundation affiliated with another university. This payment would support one of Qiu's former students who was designing the base editor. More of their money, according to documents reviewed by *Science* and Retraction Watch, flowed to a Chinese affiliate of a U.S.-based firm that would start to make the viruses, and to PriMed, a contract research organization in Sichuan province that would run the monkey studies.



Neuroscientist Zilong Qiu, speaking here in 2015 at a science fiction convention, was confident in the safety of a gene-editing treatment he developed, but it led to a death that was not made public.

Jeremy Sugarman, a medical doctor and bioethicist at Johns Hopkins University, says it's not unusual for a family to bear the costs of developing a personalized treatment. But, according to text messages shared by Jason and Linda, Qiu also asked the couple to pay other members of the research team directly, through informal arrangements they found increasingly troubling.

In April 2024, at a Starbucks café near the university, Jason met with Kan Yang, the researcher overseeing the lab work and a co-inventor on Qiu's Rett Syndrome patent, to discuss a fee schedule. Jason told Yang he was at a financial breaking point after learning it would cost more than

PHOTO: LI YIBO/XINHUA/ALAMY

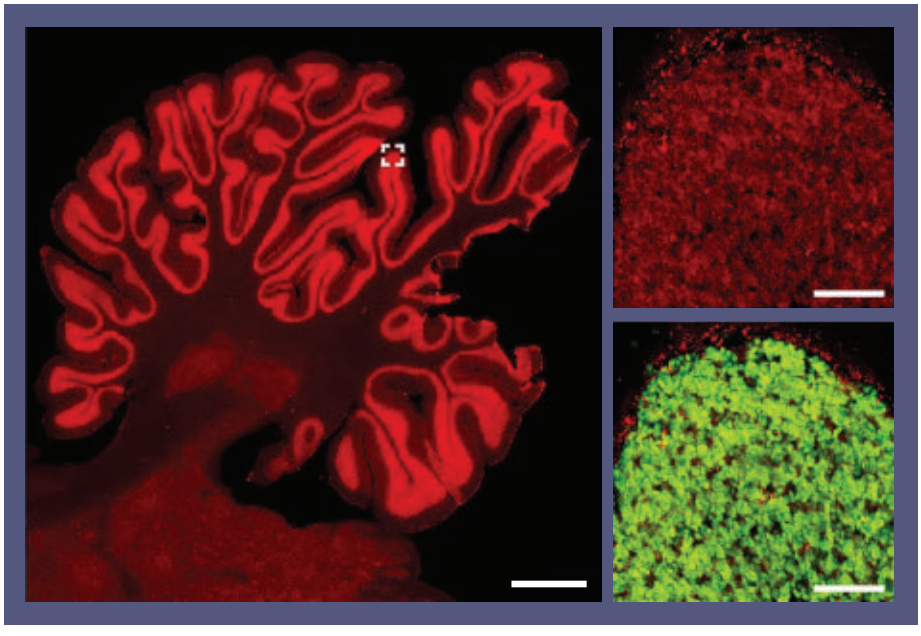
$250,000 to produce clinical-grade viruses. "I'm feeling a bit of pressure," Jason said, according to a recording.

Yang outlined why he needed even more money than Qiu had initially suggested. "The thing is, if you finish those viral injections and don't have a proper person to analyze the results, it's all for nothing," Yang replied. He said the same would be true when it came time for the monkey studies. "It requires extremely meticulous work to get results." (Yang did not respond to multiple requests for comment.)

They haggled over terms, according to the recording. The family ultimately sent $130,000 directly to Yang's personal account over the course of the project, according to bank records. In the U.S., similar payments by desperate parents seeking a treatment may cross a line, according to Sugarman, who says he would be concerned about any "undue inducement" to provide excessive gifts or money.

Qiu never asked for anything himself. Every few months, however, Jason visited him at his office, a nearby restaurant, or his home a couple hours away. Jason always picked up the tab and gave Qiu multiple iPhones, an iPad, and a case of Maotai, a Chinese liquor made of sorghum and wheat, according to family records.

By the end of 2024, the lab's efforts were paying off. Qiu's team had engineered mice to have a human version of the *CHD3* gene with their daughter's mutation, R1025W, which results in a protein with the amino acid tryptophan where there should be an arginine. The mutant pups developed autismlike traits and didn't squeak as much as normal mice when separated from their mothers. When the researchers repaired that mutation, the pups developed normally.



To confirm that viruses carrying the gene editor designed for Mei would reach her brain, scientists injected the treatment into monkeys and stained brain slices from them with a fluorescent antibody targeting part of the editor (left). Comparisons of magnified views (right) of tissue labeled for the editor (red) and for all neurons (green) suggested the treatment was reaching most neurons, but some researchers are skeptical.

Such a surgical edit was only possible because of the base-editing techniques pioneered nearly a decade ago by David Liu's lab at Harvard University. The original gene editor, CRISPR, which had been used by He Jiankui and others, often damaged the DNA it was meant to fix. "Chaos" is how Qiu described its effects to the family during their first meeting. To make an edit, CRISPR breaks both strands of DNA's double helix, which can result in large insertions or deletions in the genome.

Base editors, which use a modified version of CRISPR's DNA-cutting enzyme, only nick a single strand. With the help of an RNA guide, they unwind the DNA at the right spot and chemically convert the base on the intact strand into another base. The cell then repairs the nicked strand with the complementary base. For Mei's mutation, the base editor would transform a specific adenine (A) base in the mutant gene to a guanine (G), which would cause the complementary thymine (T) to change into the correct base, cytosine (C).

Base editing wasn't foolproof or easy to do, though. "Far fewer people have mastered this technique than gene therapy," Qiu said in one recording. He was second only to Liu, he told the family, when it came to taking it to new frontiers.

Unlike the U.S. trial for Baby KJ, which delivered Liu's own base editor into the infant's liver via tiny fatty capsules, the Chinese researchers needed to reach Mei's brain. They chose to pack genes for the editor's enzyme and guide RNA into adeno-associated viruses (AAVs), which have long been used in gene therapy but can trigger inflammatory responses at the doses required. Because only a small fraction of these viruses enter their target cells and spur the production of the base-editing components, fixing enough brain cells would require hundreds of trillions of viruses—thousands of times more than a person receiving a virus-based vaccine gets.

Qiu's team faced a second challenge. Because of the size of the base editor, they needed to split its genetic instructions in half and pack the DNA sequences into two different viruses that would be injected at the same time. Success required both viruses to be expressed inside the same cells throughout the brain. By comparison, Otarmeni, a U.S. Food and Drug Administration (FDA)-approved therapy for hearing loss that uses two vectors infused directly into the tiny, fluid-filled cavity of the inner ear, alters many fewer cells. "I'm skeptical that a dual vector approach would achieve high enough editing" to treat Mei's condition, Gray says.

The experimental therapy had apparently worked in the mice. But the virus Qiu and colleagues used in those experiments, AAV-PHP, can be injected into a mouse vein and crosses the rodent's blood-brain barrier. Because of species differences in certain cellular receptors, that virus would not cross

IMAGES: K. YANG, WK. LI, YX. GENG *ET AL*., *NATURE* 651, 785–795 (2026)



An ethics panel at Xinhua Hospital in China approved a gene-editing trial, but had not reviewed primate data suggesting safety concerns. Another hospital panel later concluded the experimental treatment was "definitely related" to Mei's death.

PHOTO: WANG GANG/COSTFOTO/BARCROFT MEDIA VIA GETTY IMAGES

the barrier as easily in humans or other primates. So Qiu's team shifted to another common gene therapy "vector," AAV9, which would be injected directly into the spinal canal, bypassing the barrier. Before Mei would be treated, however, the researchers intended to show the method could safely deliver the editing instructions to the brain cells of monkeys.

IN JANUARY 2025, Qiu shared some good news with Jason and Linda. In a WeChat message reviewed by *Science* and Retraction Watch, he sent them a manuscript he had recently submitted to *Nature*. It compiled all his lab's preclinical work on Mei's mutation over the previous 18 months. The first figure in the paper illustrated the gene sequences of Jason, Linda, and Mei, along with a laboratory study demonstrating that the R1025W mutation damages the protein encoded by *CHD3*.

The submitted paper also included striking images that Yang and colleagues had captured from brain slices from two macaques that had received Mei's potential therapy via AAV9. Yang had stained the neurons green, and the gene editor red to estimate what fraction of neurons it had reached. Under fluorescence imaging, a zoomed-in view of the cerebellum appeared almost entirely red, suggesting the editor had made it into nearly every neuron in the region where *CHD3* is most strongly expressed.

This was welcome proof that the clinical trial should go ahead, Qiu told the family over WeChat. Indeed, that result apparently helped convince the hospital ethics committee to approve the family's clinical trial on 2 January 2025.

But four of the experts *Science* asked to assess this figure, including one of the original reviewers of the *Nature* manuscript, expressed concerns about how well the primate proof-of-concept study was conducted. The submitted paper had no control samples to show the stain was selectively highlighting the gene editor. "It's completely unconvincing," says David Sanders, a biochemist at Purdue University who has studied gene therapy. "One can't have the confidence that one isn't mostly looking at background staining."

An accompanying graph suggested AAV9 only altered 30% of targeted neurons, but that was several times more than achieved with similar vectors in other primate studies. A fifth expert, a neuroscientist who asked not to be named, said the criticisms raised may be valid but there was no "obvious data manipulation or image doctoring."

Another potential concern, however, emerged about 1 month before Mei was treated: the final conclusions of a primate toxicology study conducted at PriMed. The company's report, dated 17 February 2025, concluded all four monkeys that received the therapy, whether at low or high doses, developed moderate to severe liver damage.

That data alone "should have triggered additional studies at lower doses to determine the maximally tolerated dose," says James Wilson, the former University of Pennsylvania researcher who led the ill-fated Gelsinger trial and who is now president of Gemma Biotherapeutics, a gene therapy company.

One monkey, which received a high dose of the viruses, also showed kidney damage. This could have been a result of a well-known reaction to AAV gene therapy: thrombotic microangiopathy, a pattern of microscopic blood clots that can be fatal. Because the team at PriMed did not collect data during the key window in the first month after the monkey's infusion, it is impossible to be certain of the cause.

Under China's dual-track biomedical regulatory system, noncommercial, investigator-initiated trials at major hospitals can test novel gene therapies without the kind of rigorous review from national drug regulators that is required in the U.S. The rules have varied over time, but investigators typically register the study in a national clinical trial database and undergo a scientific and ethical review from the local institution, which can vary in their attention to detail.

The hospital ethics committee did not review the final report from the primate safety study before it approved Mei's trial, according to documents from the Shanghai Yangpu District Health Commission. Jason and Linda say Qiu described the findings to them but seemed unconcerned about their significance.

The trial would move forward. "The contract signing with Xinhua Hospital is basically complete," Qiu wrote Jason in a WeChat message that January. "Next we'll have the kickoff meeting, sign the informed consent, and then administer the drug within 30 days."

On 19 February, the family arrived at Xinhua Hospital to sign the consent document and run some tests on Mei's blood. The form described throm-

botic microangiopathy as a possible outcome that would require "prompt treatment." The children's version—the one with the "book" metaphor—asked Mei herself to check a box marked yes or no. "If you don't want to write," it told her, "you can draw a smiley face instead of your name."

According to the parents' recording, Yongguo Yu, the medical doctor overseeing the clinical side of the study, told them he had selected a doctor who performed 3000 spinal injections a year to deliver the gene therapy. Two days earlier, as part of a clinical trial testing Qiu's Rett syndrome gene therapy, this doctor had performed the same procedure in a child without incident.

Yu also indicated the biggest risk for Mei was the possibility that she had antibodies to the viral vector. "Other than that," he said, "the data shows it's relatively safe." (Yu did not respond to multiple requests for comment.)

A test before the trial started found Mei had no antibodies. Still, to blunt the body's immune response to the viral infusion, she would get a short course of prednisone, a steroid, according to the informed consent document. Although some AAV-based gene therapies approved by FDA are still delivered with prednisone alone, gene therapists are increasingly erring on the side of caution and incorporating more powerful immunosuppressant drugs, such as rapamycin and rituximab.

That decision, and others, might have gotten more scrutiny if the trial had been reviewed as a commercially sponsored study by the National Medical Products Administration—China's version of FDA—rather than an investigator-initiated trial.

In either type of trial, charging for an unproven therapy is forbidden under Chinese law. The informed consent noted that "Participation in this study will not incur any additional costs to you. All expenses will be borne by the sponsor, Lanqi Xintu Gene Technology."

But the family knew better. Yu seemed to recognize the contradiction during his conversation with them. "On the surface, it's not money you're paying, but actually it's all being paid by the company," he said at one point. "If you've contributed, you have stake in it."

On 23 March 2025, the day before the procedure, the family settled into a shared room inside a crowded pediatric ward. When the other girl in the room broke into tears, Mei handed her one of her favorite stuffed animals: a blue bird. "She liked to give things to other kids," Jason said.

The next day Mei put on a brave face as she was taken to the procedure room. Yu's handpicked doctor inserted a needle between two vertebrae on her lower back and withdrew just over a teaspoon of spinal fluid to create space for the drug. Then, he swapped syringes and gently pressed down on the plunger, filling her spinal canal with viruses.

After 3 days, Mei developed a fever, a common reaction to AAVs that was expected to last for a couple days. Qiu swung by as her fever began to wane. He assured Jason and Linda that their daughter's condition would soon improve. As they walked him to the elevator, he announced—excitedly—that the *Nature* paper had been accepted with revisions. Then he added that Mei's treatment would make an even bigger splash: the first brain-directed gene-editing therapy.

Mei did not improve. Because she hadn't urinated since the infusion—a sign that her kidneys had been seriously injured—her doctors had limited her fluid intake, and she became desperately thirsty. The platelets in her blood also dropped to dangerous levels. It was the exact sequence of symptoms that the consent form had warned the family about.

But nowhere did the form explicitly indicate that any of this could end in death, nor did that come up during any conversations with Qiu or the other doctors, Jason and Linda say. "Death should always be mentioned in a first-in-human trial," says Greely, the bioethicist.

Doctors took Mei for a full-body CT scan and then rushed her to the intensive care unit. Her brave face had begun to crack. "Mom, I want to go home," she told her mother, before vanishing into the ward.

Over the next 24 hours, Jason and Linda sat in the waiting room with their elderly parents. Qiu, who had returned upon hearing of Mei's problems, seemed unable to process it. He was still with the family the next morning when a doctor emerged to tell them Mei had died. Qiu was in shock, Jason recalled.

So was the family. They stood by Mei's body and caressed her one last time.

TWO DAYS LATER, the hospital's ethics board convened an emergency meeting and concluded Mei's death was "definitely related" to the treatment, according to its report. The cause of death: thrombotic microangiopathy.

Over the next few weeks, Linda and Jason say Qiu expressed his sadness to the family, but he was unwilling to speculate on what went wrong. The parents say they asked him to withdraw the *Nature* paper out of concern that it could mislead other families

> **I'm feeling a bit of pressure.**
> **Jason**, Mei's father, responding to a request for payment
>
> **The thing is, if you finish those viral injections and don't have a proper person to analyze the results, it's all for nothing. It requires extremely meticulous work to get results.**
> **Kan Yang**, Songjiang Research Institute

ILLUSTRATION: DARIA LADA

with a similarly affected child, as well as researchers. Qiu initially agreed. Eventually, they say, he stopped responding to their queries. Yang returned the full payment he received for the work.

In September 2025, the district health department fined the hospital approximately $3600 for failing to properly oversee the trial and not registering it as commercially sponsored research on the national database. As the responsible physician, Yu was named in the report, and the hospital required him to receive verbal counseling, according to the family.

Although Qiu's company had taken out insurance for the trial, the family said they were offered no compensation. They had not asked for it, they said; instead, they blamed themselves for not asking more questions and for trusting Qiu's authority so completely.

THE MUTED RESPONSE to Mei's death sharply contrasts with the Gelsinger tragedy decades ago. His death drew widespread headlines and prompted a major self-examination by the gene therapy field. The university where the trial was conducted paid a fine of more than a half-million dollars and had further gene therapies suspended. FDA barred Wilson, the principal investigator, from human research studies for 5 years for violations that included failing to disclose the adverse reactions observed in the primate toxicology studies on Gelsinger's informed consent form. "Public disclosure … is essential in safely moving these fields forward," Wilson says.

Within days of their daughter's death, Jason and Linda moved into a hotel and, later, a new apartment. Locked away in the old apartment was all their furniture and their daughter's possessions. They also cut off contact with some of their closest friends. "I don't know how to reply to them, so I pretend not to see their messages," Linda says. "I am not so brave to face everything."

Jason and Linda were still mourning the loss of Mei when they learned in February that a version of the *Nature* study had been published after a review process that had stretched over a year. The family's genetic data had been removed and a sentence thanking them for their "participation and support" had been eliminated.

The paper still described a base-editing therapy directed at their daughter's mutation, but it featured entirely new data on its ability to work in monkey brains. "I remain very skeptical of the data in this figure," Gray says. Qiu's team had added control samples at the request of peer reviewers. But the controls appear to have been processed at a different time from the treated samples, using different microscope settings, according to Mike Rossner, an image integrity expert. That makes any comparison challenging.

All the original mouse experiments, which were funded by the family, also appear to have been redone. Campeau, who reviewed the original manuscript for *Nature*, said he hadn't noticed the extent of the changes when the revised version landed in his inbox, and he had approved it. Notably, the paper also did not mention any of the kidney and liver troubles from the primate safety study of the therapy.

The paper had an enthusiastic reception. "These promising results might pave the way for the development of an effective clinical treatment," Kevin Bender, a neuroscientist at UC San Francisco, wrote in an accompanying commentary. At the time, Bender had no idea that a girl had received it and was already dead. Meanwhile, Chinese state media, CCTV, called the work "the first ray of hope" for "countless families suffering such diseases." Some families in Jason and Linda's WeChat group said they planned to reach out to Qiu about their own children.

When Jason and Linda saw that, they decided to act. They submitted a complaint to the dean of Qiu's university, demanding an investigation and a retraction of the *Nature* paper "to prevent more children from suffering and our blood and tears flowing in vain."

Gray, Sanders, and several others who reviewed the paper for this story believe *Nature* should request and analyze the raw data and image files from the authors. They also say the study's sources of funding should be fully disclosed. "To not acknowledge one of the sources of funding should be enough for retraction," Sanders says.

Greely observes that it would be unreasonable for the family to have veto power over the publication, but he would like to know whether Qiu had told *Nature*'s editors about the outcome of the clinical trial and why the family was removed from the paper. In a June response to an email from the parents noting their child's death and other concerns, *Nature*'s deputy editor, Victoria Aranda, wrote that the ethical issues they raised "fall outside our purview in terms of data integrity" and that the matter should be handled by the university.

In a more recent statement responding to queries from *Science* and Retraction Watch, the journal said it was not informed of these issues or the sanctions the hospital had received before it published the paper. "Had information relevant to the editorial assessment of the manuscript been disclosed during consideration of the paper, it would have been carefully evaluated as part of our decision-making process," noted Francesca Cesari, *Nature*'s chief biological, clinical, and social sciences editor.

On 30 April, Jason and Linda received an emailed reply from the university, followed by a phone call: The dean's office was taking no action against Qiu. The payments made to Yang were simply "research service fees" for the clinical trial, though he was given a formal admonitory talk for his behavior. The *Nature* paper, the university wrote, "bears no relation to the experiments you funded."

DURING A ZOOM CALL last month, framed by the bare, white walls of their home, Jason and Linda described how each had coped with Mei's death. At times during the conversation, Linda became overwhelmed and turned away from the camera, as Jason pressed on. They recalled that when Mei's birthday came around earlier this year, they said nothing to each other, though they both assumed it was on the other's mind.

"We do not discuss that," Linda said.

During the call, Linda finally told her husband how she had spent that day. While he was at work, she finally returned to the old apartment and sat down on Mei's bed surrounded by all her things. She had purchased a chocolate cupcake on the way and now picked at it as she imagined that her daughter was still at her side.

By the time she finished telling her story, Jason was welling up. Linda glanced over at him tenderly. "My husband didn't know this," she said. "I cannot cry every day, but that day, I think, I can cry in the room." □

This story was produced in collaboration with Retraction Watch and supported by the *Science* Fund for Investigative Reporting. Brendan Borrell is a journalist in Los Angeles. Additional reporting was provided by Bian Huihui.



COMMENTARY

PERSPECTIVES

ECOLOGY

Birdsong diversity across the world

Birdsong motifs are shaped by evolutionary history, biological traits, and environmental constraints

Elodie F. Briefer

Many tropical birds, such as the Guianan warbling antbird (*Hypocnemis cantator*), use song motifs that propagate well across rainforests.

Why do skylarks or blackbirds sing such diverse and melodious songs (*1*, *2*), whereas great tits repeat the same two motifs over and over? Differences in sexual selection processes, evolutionary history (phylogeny), or morphological traits could underlie this variation. On page 381 of this issue, Bacquelé *et al.* (*3*) report on the evolutionary drivers and distribution of the songs of passerines—the largest bird order—using birdsong recordings from a large open-source database. They reveal that birdsong diversity can be mapped into eight core motifs, whose use and geographical distribution are determined by a trade-off between environmental constraints, evolutionary history, and biological traits.

Passerines are present on all Earth continents except Antarctica. This order includes songbirds, which emit vocalizations that are long (lasting from a few seconds to several minutes) and complex (made of several acoustic motifs that are organized according to a specific order) (see the figure). These vocalizations, called songs, are typically used to defend territories and attract mates, but some species also use them for strenghening pair bonds (e.g., duets) (*4*). The factors that determine the acoustic features (duration, frequency, and amplitude) of passerine songs have been debated (*5*). These features are mainly constrained by the sender's sound-producing organ (the syrinx), its vocal tract, and its cognitive skills (learning and memory) (*6*). After propagation in the environment, the song can be attenuated; distorted; or masked by abiotic (nonbiological), biotic, and human-generated noises (*7*). Later, other birds decode the information contained in the song and might respond. The efficiency of this decoding process depends on the anatomy of the hearing system and the cognition of the receiver (*6*). The acoustic features of songs are also strongly shaped by their function. Songs aimed at close neighbors are more quiet, whereas those aimed at attracting females located at a distance are much louder (*8*). The capacity of oscine passerines ("true songbirds") for song learning, and hence cultural transmission, makes the search for song feature drivers within this subgroup even more complex (*4*).

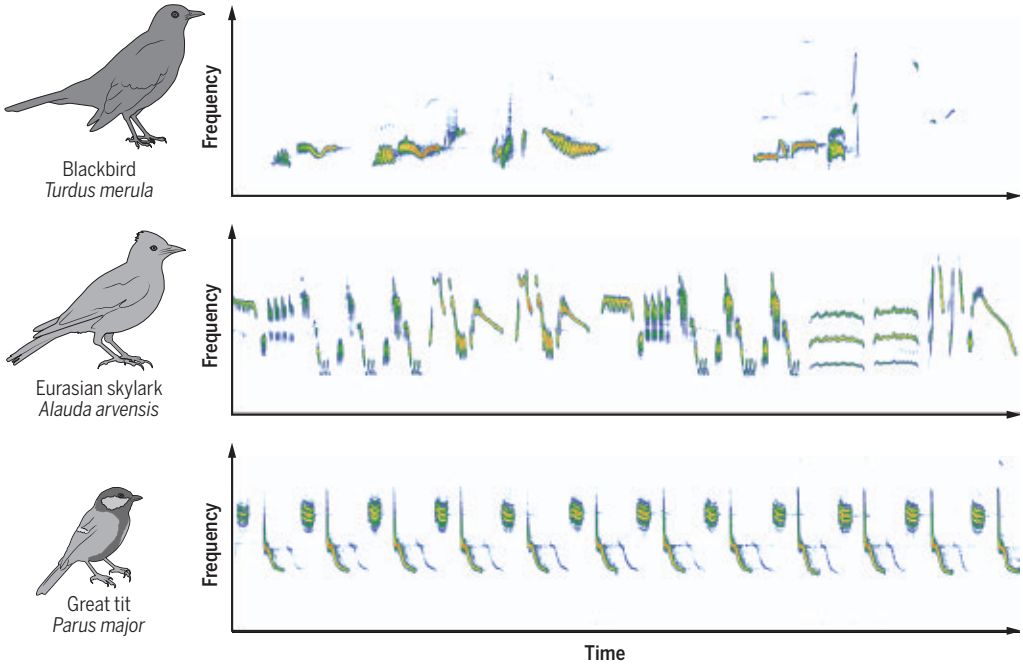
A song's acoustic features therefore result from a trade-off between multiple factors that optimizes the fulfillment of the song's purpose despite environmental constraints. This might lead to acoustic similarities between closely related birds that share anatomical features or between birds living in the same type of environment. However, the large diversity of songbird species across the world makes it challenging to identify the main factors that shape their songs. So far, studies on the evolution of birdsong have mainly focused on single parameters [e.g., frequency (*9*)], specific climate regions or communities (*10*), or a restricted number of factors [e.g., body size and bill size (*11*)].

Bacquelé *et al.* investigated how various factors previously proposed to drive birdsongs influence the shape of song motifs rather than single acoustic parameters. Using data from the citizen science

PHOTO: LUIZ CLAUDIO MARIGO/NPL/MINDEN PICTURES

Bird music

Sound recordings from three bird species living in temperate regions illustrate the variety and complexity of acoustic motifs found in passerine birdsongs.



project Xeno-canto (*12*), an online resource containing song recordings from most bird species across the world, the authors analyzed the acoustic structure of song bouts from more than 3000 passerine species. Unsupervised classification revealed eight main types of motifs, including various kinds of trills and whistles, chaotic notes, and harmonic stacks. The majority of the songs analyzed (65%) were composed of two to eight different motifs. The types and number of motifs varied between species, and 82% of this variation was related to bird phylogeny. The remaining differences were mostly explained by social organization (whether the birds were solitary or formed short-term or long-term pairs or groups and whether they engaged in duets or choruses) followed by morphology (body mass and beak dimensions) and mating system (ranging from strict monogamy to extreme polygamy).

The authors then mapped the geographical distribution of song motifs globally and investigated how this distribution reflects the boundaries of terrestrial biomes (geographical regions with similar climate, plants, and animals). They also modeled the transmission efficiency of each motif—how quicky the information about species identity that the motif conveys is degraded—in different types of habitat (forest or open land). Bacquelé *et al.* then tested the relationships between the transmission efficiency of motifs used in each portion of the world map with the same area and environmental conditions (temperature, humidity, terrain ruggedness, human influence, and vegetation). The geographical distribution of acoustic motifs aligned with that of terrestrial biomes. Notably, flat whistles and slow trills, which are highly resilient to information degradation, were overrepresented in tropical rainforests. This finding suggests that these two motifs are ideal for long-distance communication in densely vegetated environments, which impede sound transmission. By contrast, in temperate latitudes, more complex motifs were favored despite their shorter communication distances. These motifs probably emerged as a result of more intense sexual selection associated with short breeding seasons in these regions.

The use of citizen science data has advantages and disadvantages (*13*). Although bird species can be misidentified when people upload sounds and attribute them incorrectly, such projects enable the collection of very large amounts of data. Indeed, Xeno-canto (*12*) contains the recordings of more than 12,900 species, which are mostly birds but also include bats, frogs, and insects. This resource allowed Bacquelé *et al.* to cover 80% of all existing passerine families in their study despite restricting their analyses to high-quality recordings in the database.

Whether the eight "universal" acoustic motifs identified by Bacquelé *et al.* are also perceived as such by the birds themselves is unknown. Playback experiments would allow researchers to test whether birds classify these motifs in the same way as the unsupervised clustering method used in the study or whether their perception breaks them down into more refined categories (*14*). Future studies could also investigate the order in which these motifs are produced (i.e., their "syntax") to see whether their organization within songs also follows a universal pattern and, if so, which factors drive this pattern.

Through combining large-scale citizen science data with a motif-based framework, Bacquelé *et al.* provide one of the most comprehensive analyses of passerine song evolution to date. Their results reveal that birdsong diversity emerges from the interplay between deep phylogenetic constraints and adaptive responses to social and environmental pressures. In addition to advancing our understanding of avian communication, their approach could be used to study the evolution of complex acoustic systems across other biological groups. □

REFERENCES AND NOTES

1. E. Briefer, T. S. Osiejuk, F. Rybak, T. Aubin, *J. Theor. Biol.* **262**, 151 (2010).
2. T. Dabelsteen, *Ornis Scand.* **15**, 227 (1984).
3. Q. Bacquelé *et al.*, *Science* **393**, 381 (2026).
4. C. K. Catchpole, P. J. B. Slater, *Bird Song: Biological Themes and Variations* (Cambridge Univ. Press, 1995).
5. C. K. Catchpole, *Perspect. Biol. Med.* **30**, 47 (1986).
6. O. N. Larsen *et al.*, in *Exploring Animal Behavior Through Sound: Volume 2: Applications*, C. Erbe, J. A. Thomas, Eds. (Springer, 2025), pp. 285–359.
7. M. Wahlberg, O. N. Larsen, in *Comparative Bioacoustics: An Overview*, C. Brown, T. Riede, Eds. (Bentham Science Publishers, 2017), pp. 62–119.
8. T. Dabelsteen, in *Animal Communication Networks*, P. K. McGregor, Ed. (Cambridge Univ. Press, 2005), pp. 38–58.
9. P. Mikula *et al.*, *Ecol. Lett.* **24**, 477 (2021).
10. G. C. Cardoso, T. D. Price, *Evol. Ecol.* **24**, 447 (2010).
11. J. I. Friis, J. Sabino, P. Santos, T. Dabelsteen, G. C. Cardoso, *Behav. Ecol.* **33**, 798 (2022).
12. Xeno-canto Foundation, Xeno-canto: Sharing wildlife sounds from around the world (2005); www.xeno-canto.org.
13. J. L. Dickinson, B. Zuckerberg, D. N. Bonter, *Annu. Rev. Ecol. Evol. Syst.* **41**, 149 (2010).
14. J. E. Elie, A. de Witasse-Thézy, L. Thomas, B. Malit, F. E. Theunissen, *Science* **389**, 1210 (2025).

ACKNOWLEDGMENTS

The author thanks R. Mandel, J. H. Ramussen, and A. Ramussen for useful suggestions.

10.1126/science.aej0087

Behavioural Ecology Group, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. Email: elodie.briefer@bio.ku.dk

CREDITS: (GRAPHIC) N. BURGESS/*SCIENCE*; (DATA) E. BRIEFER; BLACKBIRD AND GREAT TIT SONG RECORDINGS BY XENO-CANTO

PHYSIOLOGY

Ballooning under water

Insect larval buoyancy is controlled by material properties of their air sacs

Jon F. Harrison[1] and **H. Arthur Woods**[2]

In lakes and oceans, billions of animals migrate vertically every day, feeding in surface waters at night and descending out of the photic zone during the day to escape predatory fish. Many migrants are crustacean zooplankton, such as copepods, that move vertically by passive sinking, active vertical swimming, and alterations in body lipid composition (*1, 2*). Such transit plays important roles in food webs and nutrient transfer (*3*). On page 398 of this issue, McKenzie *et al.* (*4*) report that insects also migrate vertically. Larvae of the midge *Chaoborus edulis* (family Chaoboridae), which live in Lake Malawi, Africa, dive during the day to depths of up to 200 m using a set of crush-resistant air sacs, the volume of which can be fine-tuned to regulate buoyancy. Surviving the depth of this water column raises the possibility that other insects exploit deep-water habitats, contrary to widespread belief.

McKenzie *et al.* describe how *Chaoborus* larvae and pupae migrate vertically in an aquatic environment using a previously uncharacterized mechanism of buoyancy control. Their buoyancy control organs consist of four cylindrical air sacs, each surrounded by alternating layers of a rubberlike protein called resilin, and chitin, a structural carbohydrate. The resilin layers expand or contract in response to changes in pH, whereas the rigid chitin framework directs those changes so that they alter air sac volume. Changes in pH are driven by proton pumps and ion channels and exchangers in the tracheal epithelium. Alkalinization ionizes resilin's acidic amino acids, which creates more negative charges and thereby renders the protein more hydrophilic, and it therefore absorbs water and expands. Acidification reverses this process and causes shrinkage (*5*). McKenzie *et al.* observed that pH-controlled modifications in air sac volume assist in both descent and ascent of *Chaoborus* larvae through the upper 50 m of the water column. Resilin band expansion in the air sac wall is driven by higher local pH, allowing the air sac to expand and support buoyancy. At greater depths, the capacity of the air sacs to resist further compression prevents a large increase in their density, which would greatly raise the cost of migrating back to the surface.



Aquatic larvae of the insect *C. edulis* modify the stiffness and geometry of their air sacs to tolerate the depths they encounter.

It is unlikely that this resilin-chitin molecular motor evolved de novo, which suggests that other such structures may exist in other insects but perhaps play different roles. The expression of ion channels and transporters in epithelial cells of the *Chaoborus* buoyancy organs is similar to that in tracheal epithelia from related species that have more standard tracheal systems where an air-filled network of tubes transports gases. This suggests that the control system for buoyancy evolved from ancient fluid-clearing processes of insect tracheae (*6*). *C. edulis* relies on cutaneous respiration in which dissolved oxygen diffuses through the insect's exoskeleton; however, their air sacs are referred to as a tracheal system, albeit a nonstandard one. The resilin-chitin laminae surrounding the air sacs can exert force both when lengthening and shortening, suggesting that as a constituent of exoskeletons, this material could play important roles in altering insect dimension, shape, and strength. Molecular details of pH-driven alterations in the dimensions of the resilin-chitin complex remain under investigation, but harnessing this mechanism holds great promise for bio-inspired microdevices (*5*). One could imagine the use of this material as a tiny mechanical switch or pump within an implanted system for delivering daily doses of hormone or medicine or as a key component of micromixing systems for synthetic biology.

The Lake Malawi chironomid larvae have an additional key adaptation that allows them to avoid fish predation. Similar to other chironomids, they can survive long periods of low oxygen by depressing their metabolism and utilizing anaerobic metabolism to convert large stores of malate to succinate. This regenerates nicotinamide adenine dinucleotide (oxidized form), which is required for glycolysis and energy production to continue (*7*). The metabolic switch allows larvae to spend daylight hours in the deepest layer of the lake, where low amounts of dissolved oxygen exclude oxygen-requiring fish.

The Lake Malawi species is not the only *Chaoborus* to use air sacs for buoyancy control. Indeed, air sacs are widespread throughout the genus, though other species dive to shallower depths than *C. edulis*. McKenzie *et al.* show that similar air sacs occur in four different *Chaoborus* species from four widely distributed lakes and that their capacity to resist compression correlates with lake depth. These fly larvae, and perhaps other aquatic insects, have substantial abilities to adapt the capacity of their air-filled tracheal systems to resist compression and likely to regulate buoyancy. The diversity of such capacities will be an important area of future research aimed at understanding insect distributions in different aquatic environments.

The findings of McKenzie *et al.* illuminate one of the major patterns of life on Earth. Although insects play key roles in terrestrial environments and have very successfully invaded fresh water, they

PHOTO: VIRIDIFLAVUS

remain mostly excluded from marine habitats. Nearly 30 years ago, insect physiologist Simon Maddrell proposed that this pattern occurs because compression of the air-filled tracheal system at depth excludes insects from the deep, dark waters necessary to escape fish predation (*8*), an explanation that is generally accepted. The McKenzie *et al.* study refutes this hypothesis by showing that some insects can escape fish by descending into deep waters without complete compression of their air-filled spaces. Another challenge to the Maddrell hypothesis comes from marine seal lice (family Echinophthiriidae). Seal lice survive underwater for long periods of time while attached to their amphibious, deep-diving pinniped hosts (*9*). At shallow depths, they close their spiracles (exoskeletal pores) and use cuticular gas exchange to obtain oxygen at about half the rate they achieve in air (*10*). Seal lice can survive very deep dives, some in excess of 2000 m, but it is not known whether their tracheal system collapses at the high hydrostatic pressures at these depths and whether they can sustain aerobic metabolism throughout a dive (*11*).

The study of McKenzie *et al.* indicates that insects can evolve nonstandard tracheal systems with noncompressible components, but observations of terrestrial insects indicate that compressible tracheal systems are strongly linked to high and flexible rates of gas exchange (*12*–*14*). Therefore, a refined hypothesis could be that insects are unable to sustain vigorous activity, growth, and reproduction in the high hydrostatic pressures of deep marine habitats because they require an air-filled, compressible tracheal system for adequate and flexible aerobic metabolism. In support of this, seal lice stop moving underwater and reproduce only on land (*10*), and *Chaoborus* feed and grow in shallow waters (*4*). This hypothesis predicts that the hydrostatic pressure that must be sustained to avoid fish predation precludes a tracheal system–mediated aerobic metabolism sufficient to support growth and reproduction in aquatic insects. □

REFERENCES AND NOTES

1. N. B. Tack, S. Oliveira Santos, B. J. Gemmell, M. M. Wilhelmus, *Mar. Biol.* **171**, 207 (2024).
2. R. W. Campbell, J. F. Dower, *Mar. Ecol. Prog. Ser.* **263**, 93 (2003).
3. K. Bandara, Ø. Varpe, L. Wijewardene, V. Tverberg, K. Eiane, *Biol. Rev. Camb. Philos. Soc.* **96**, 1547 (2021).
4. E. K. G. McKenzie *et al.*, *Science* **393**, 398 (2026).
5. E. K. G. McKenzie, G. T. Kwan, M. Tresguerres, P. G. D. Matthews, *Curr. Biol.* **32**, 927 (2022).
6. T. Ames, P. G. D. Matthews, B. J. Matthews, *J. Exp. Biol.* **228**, jeb251113 (2025).
7. F. Scholz, I. Zerbst-Boroffka, *J. Insect Physiol.* **44**, 427 (1998).
8. S. H. Maddrell, *J. Exp. Biol.* **201**, 2461 (1998).
9. M. S. Leonardi, J. E. Crespo, F. Soto, C. R. Lazzari, *Insects* **13**, 46 (2021).
10. M. S. Leonardi *et al.*, *Commun. Biol.* **8**, 861 (2025).
11. M. S. Leonardi *et al.*, *J. Exp. Biol.* **223**, jeb226811 (2020).
12. M. W. Westneat *et al.*, *Science* **299**, 558 (2003).
13. J. J. Socha, T. D. Förster, K. J. Greenlee, *Respir. Physiol. Neurobiol.* **173** (suppl.), S65 (2010).
14. J. F. Harrison *et al.*, *J. Exp. Biol.* **226**, jeb245712 (2023).

ACKNOWLEDGMENTS

J.F.H. and H.A.W. were partially supported by fellowships from the Stellenbosch Institute of Advanced Studies, and J.F.H. was partially supported by a sabbatical fellowship from Arizona State University.

10.1126/science.aei9743

[1]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [2]Division of Biological Sciences, University of Montana, Missoula, MT, USA. Email: j.harrison@asu.edu; art.woods@mso.umt.edu

BIOCHEMISTRY

No strain, no gain

A modular approach could expand the repertoire of spring-loaded covalent drugs

Fabian B. Kraft[1,2,3] and **Matthias Gehringer**[1,2,3]

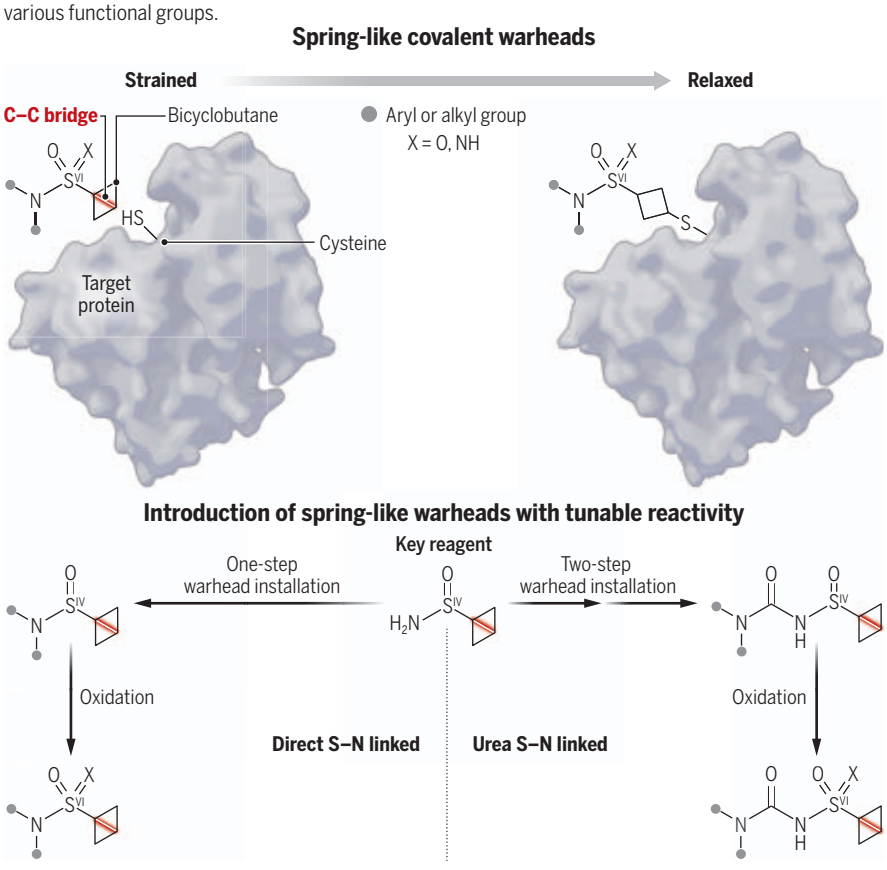
Drug candidates that form covalent bonds with their targets were historically avoided because of toxicity concerns (*1*). However, covalent drugs with controlled reactivity are now recognized as promising agents that combine manageable safety with substantial advantages, including enhanced potency, prolonged target residence times, and access to challenging targets (*2*, *3*). This shift is reflected in recent covalent drug approvals, particularly those targeting protein-phosphorylating enzymes (kinases) (*4*, *5*) and cancer drivers such as the oncogenic protein KRAS (*6*). A central feature of these drugs is a "warhead" reacting with an amino acid residue in the target protein to form a covalent bond. Yet the repertoire of drug-like, chemically accessible reactive moieties remains limited (*7*). On page 408 of this issue, Shultz *et al.* (*8*) report streamlined access to an underexplored warhead type with well-tempered reactivity that can be appended to complex scaffolds late in synthesis, enabling practical use in drug discovery.

Warheads must have preferential reactivity toward a specific functional group (chemoselectivity) and productive engagement with an individual amino acid residue in a target protein (site selectivity). The initial noncovalent interaction between a drug and the protein positions a warhead close to a specific amino acid. This positioning makes the reaction at the exact target residue fast, with similar amino acids elsewhere much less likely to react, resulting in a specific covalent modification (*9*). Among the most widely used warheads are acrylamides, which preferentially target cysteine, a reactive sulfur-containing amino acid that is commonly addressed in contemporary covalent drug discovery. Acrylamides have quickly become an industry standard because they exhibit moderate, tunable reactivity that can help limit nonspecific protein modification, have favorable properties in living organisms, and are easily attached to drug-like molecules at a late stage of synthesis. These advantages may have contributed to the slower adoption of alternative covalent warheads, thereby limiting the chemical diversity of clinically used covalent reactive groups.

A promising new class of covalent warheads for drug discovery are sulfur(VI) bicyclobutanes, in which a highly electron-deficient form of sulfur (oxidation state: +VI) is attached at the bridgehead carbon of two three-membered hydrocarbon rings that share a common bond (*10*). Bicyclobutane is a highly strained structure that exhibits spring-like behavior. The strained bridging C–C bond, which is polarized by the sulfur(VI) group, breaks upon attack by an electron-rich chemical species (nucleophile) through a nucleophile-induced ring-opening reaction. This releases the strain and traps the nucleophile irreversibly (see the figure). Previous work showed that arylsulfone bicyclobutane—a subclass of sulfur(VI) bicyclobutanes where the sulfur atom is additionally attached to an aromatic ring—displays high selectivity toward cysteine, with tunable reactivity (*10*, *11*). However, these motifs are not widely adopted, in part because they have been synthetically difficult to access. Several approaches have been proposed to broaden the toolbox of available bicyclobutane warheads: Carboxamide bicyclobutanes were introduced as more accessible analogs to sulfur(VI) bicyclobutanes (*12*), a one-pot route to sulfur(VI) bicyclobutane analogs was developed (*13*), and a synthesis strategy in which one sulfone oxygen atom is replaced by nitrogen to enable attachment of an additional substituent at the sulfur(VI) center was estab-

Drug-like molecular springs

Sulfur(VI) bicyclobutane is highly strained because of two three-membered hydrocarbon rings that share a common bond. The C–C bridge breaks upon nucleophilic attack by the cysteine side chain, leading to covalent modification of the targeted cysteine residue in the protein. A readily accessible reagent with a less electron-deficient oxidation state (+IV) of sulfur allows formation of S–N bonds in the presence of various functional groups.



lished (*14*). Nevertheless, a simple and versatile means to introduce sulfur(VI) bicyclobutanes onto molecular scaffolds at a late stage of drug synthesis has been lacking.

Shultz *et al.* demonstrate a mild and modular approach for producing nitrogen-linked sulfur(VI) bicyclobutanes, including sulfonamide- and sulfonimidamide-bicyclobutane groups. A readily accessible reagent (bicyclobutane sulfinamide) leverages the low propensity of sulfur in a less electron-deficient oxidation state (+IV), to promote nucleophile-induced bicyclobutane ring opening. This allows for the formation of S–N bonds with various nitrogen-containing chemical species such as amines, anilines, and ureas. Oxidation of the resulting intermediates then generates the corresponding sulfur(VI) bicyclobutanes in a straightforward fashion. This strategy enabled late-stage installation of sulfur(VI) bicyclobutanes onto various complex covalent drug precursors without the need to temporarily protect a specific part of a molecule. Notably, the approach was compatible with different reactive chemical moieties such as free heteroaryl NH groups, amides, phenols, and basic tertiary amines, all of which occur frequently in approved drugs. The method could also introduce nitrogen-15 isotope labels at a late stage of synthesis, which may facilitate pharmacokinetic and pharmacodynamic studies and metabolomics, proteomics, and structural biology analyses. Moreover, Shultz *et al.* established new sulfur(VI) bicyclobutane linkage topologies that may broaden biological target space and enable fine control over the covalent binding process.

The bicyclobutane motifs produced by Shultz *et al.* showed improved cysteine chemoselectivity compared with the gold standard acrylamides, as well as mild and tunable reactivity. When incorporated into covalent kinase inhibitors, the bicyclobutane warheads yielded compounds with bioactivity comparable to that of the acrylamide parent molecules, but with markedly improved selectivity when assessed for off-target activity and through proteomic analyses of reactivity across the proteins expressed in target cells. A bicyclobutane sulfonamide analog of dacomitinib, which is an approved inhibitor of the epidermal growth factor receptor kinase, also showed favorable drug-like properties and pharmacokinetics while displaying antitumor activity in mice. These results further support the potential of sulfur(VI) bicyclobutane warheads for drug discovery.

The study of Shultz *et al.* highlights the importance of the three-dimensional arrangement of substituents around the sulfur(VI) center of sulfonimidamide bicyclobutanes. This emphasizes the need to develop synthetic methods that produce defined geometries around the sulfur(VI) atom. Broadening the strategy to other spring-loaded warhead topologies, such as housanes (bicyclopentanes), could further expand its utility in covalent drug discovery. It will also be important to test whether observed chemoselectivity is preserved across the full proteome of a target cell or tissue. For example, residue-resolved chemical proteomics can comprehensively identify which residues on a given protein are modified (*15*). Although the therapeutic value of bicyclobutane-type covalent warheads remains to be proven, Shultz *et al.* provide a key advance that may help translate the promising chemical properties of these motifs into clinical reality. □

REFERENCES AND NOTES

1. J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discov.* **10**, 307 (2011).
2. L. Boike, N. J. Henning, D. K. Nomura, *Nat. Rev. Drug Discov.* **21**, 881 (2022).
3. J. Dopstadt *et al.*, *J. Med. Chem.* **69**, 9891 (2026).
4. R. Roskoski Jr., *Pharmacol. Res.* **217**, 107805 (2025).
5. S. Qunies, V. Teixeira, 2025 Novel Small Molecule FDA Drug Approvals (2026); https://drughunter.com/articles/2025-novel-small-molecule-fda-drug-approvals.
6. K. Choucair *et al.*, *Signal Transduct. Target. Ther.* **10**, 385 (2025).
7. L. Hillebrand, X. J. Liang, R. A. M. Serafim, M. Gehringer, *J. Med. Chem.* **67**, 7668 (2024).
8. Z. P. Shultz *et al.*, *Science* **393**, 408 (2026).
9. B. Srinivasan, *J. Med. Chem.* **68**, 13137 (2025).
10. R. Gianatassio *et al.*, *Science* **351**, 241 (2016).
11. J. M. Lopchuk *et al.*, *J. Am. Chem. Soc.* **139**, 3209 (2017).
12. K. Tokunaga *et al.*, *J. Am. Chem. Soc.* **142**, 18522 (2020).
13. M. Jung, V. N. G. Lindsay, *J. Am. Chem. Soc.* **144**, 4764 (2022).
14. Z. Zhong *et al.*, *Angew. Chem. Int. Ed.* **64**, e202420028 (2025).
15. P. R. A. Zanon *et al.*, *Nat. Chem.* **17**, 1712 (2025).

ACKNOWLEDGMENTS

The authors thank J. Strumberger and M. Hanne for help with the text and figure and acknowledge support from the German Research Foundation (DFG) under Germany's excellence strategy EXC 2180-390900677.

10.1126/science.aej4566

[1]Department for Medicinal Chemistry, Institute for Biomedical Engineering, Faculty of Medicine, University of Tübingen, Tübingen, Germany. [2]Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies," University of Tübingen, Tübingen, Germany. [3]Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Faculty of Sciences, University of Tübingen, Tübingen, Germany. Email: matthias.gehringer@uni-tuebingen.de

GRAPHIC: A. FISHER/*SCIENCE*

SURFACE CHEMISTRY

A tiny cone of reactivity

Colliding molecules react only when their orientations and point of impact fall within a narrow range **Jonas Björk**

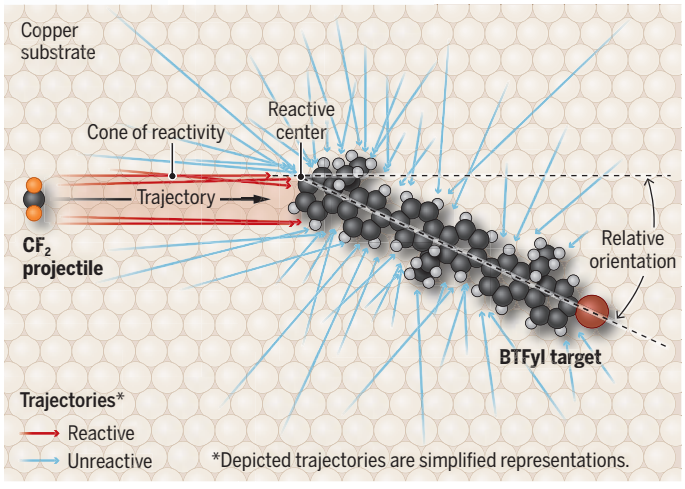
Not every collision causes a reaction. Only a small fraction of molecular collisions brings the reactive sites into a favorable arrangement for bond formation. The reactivity of two colliding molecules therefore depends on both the point of impact (impact parameter) and the relative orientation of the reactants. The effect of molecular orientation on reactivity has been probed by manipulating molecules with a stream of atoms or molecules in a vacuum (*1, 2*). However, examining how the impact parameter influences individual molecular collisions is difficult. On page 417 of this issue, Timm *et al.* (*3*) report an approach to simultaneously manipulate and probe impact parameter and orientation by using scanning tunneling microscopy (STM), offering new opportunities for understanding and controlling chemical reactions at the atomic scale.

The 1981 invention of STM (*4*) not only earned Gerd Binnig and Henrich Rohrer the Nobel Prize in Physics but also transformed how surface science is conducted. By exploiting quantum tunneling of electrons—a quantum-mechanical phenomenon in which particles pass through energy barriers they cannot overcome classically—STM enables real-space imaging with atomic resolution (*5*). When electrons tunnel between a probe tip and surface across the vacuum gap, the resulting tunneling current provides a sensitive measure of the local electronic structure, enabling direct imaging of individual atoms and molecules adsorbed on surfaces. This capability has enabled identification of reactants and products at the single-molecule level, providing insights into important on-surface chemical transformations such as dehalogenative coupling of porphyrins (*6*), which is a prototypical on-surface reaction; synthesis of graphene nanoribbons (*7*) that have tunable electronic properties suitable for organic semiconductor applications; and the formation of biphenylene networks, which establish a new two-dimensional carbon allotrope with unusual electronic properties (*8*). Beyond imaging, STM can manipulate individual atoms and molecules on surfaces through controlled tip–surface interactions. The atomically sharp probe locally injects electrons or applies precise forces, causing atoms or molecules hop or slide across the surface. This has enabled, for example, atom-by-atom assembly of iron atoms into a quantum corral that confines electrons (*9*), the selective breaking of individual chemical bonds (*10*, *11*), and the generation of molecular fragments that travel across surfaces as nanoscale projectiles (*12*).

Timm *et al.* exploited the advanced capabilities of STM to launch a projectile difluorocarbene (CF_2) toward a target radical (BTFyl) to identify chemical reactions between the molecules upon a collision. This design allowed the control of two key parameters governing reactivity (see the figure). The impact parameter was varied by launching projectiles along different atomic rows of a copper surface while the molecular orientation was controlled through rotation of the target molecule around its anchoring site. Of the 79 collisions performed, only six resulted in a chemical reaction, highlighting the strong sensitivity of reactivity to collision geometry. Notably, successful reactions were observed only for a narrow range of impact parameters, corresponding to a spatial tolerance of only a few angstroms. The relative orientation was even more restrictive. Reactions were absent when the molecular alignment deviated by more than about 10° from the optimal geometry. Conversely, when both the impact parameter and molecular orientation were favorable, the probability of reaction was substantially high. Theoretical calculations supported the geometric picture and underscored that reducing the system to simplified reaction coordinates is nontrivial.

Colliding molecules react in a narrow range

A difluorocarbene (CF_2) projectile is aligned along different atomic rows of a copper substrate, then launched toward a target radical (BTFyl). The projectile's trajectory is precisely determined by the position and orientation of the target. Two molecules react and form a bond only when the point of impact and their relative orientation fall within a narrow range.



GRAPHIC: K. HOLOSKI/*SCIENCE*

Timm *et al.*'s approach provides an experimental route toward mapping multidimensional reaction landscapes one collision at a time. Such measurements could reveal how molecular structure, reactive sites, and environment shape reaction pathways, enabling more predictive control of surface chemical transformations. They could also enable single-molecule studies of how chirality influences chemical reactivity by directly controlling the handedness and orientation of reactants. Key open questions include how molecular structure, steric constraints of both target and projectile, and projectile energy influence the width and shape of the cone of reactivity, and how these effects depend on surfaces. Addressing these aspects will be essential for assessing the generality of the observed behavior and for extending collision-based concepts toward predictive control of surface chemical reactions. □

REFERENCES AND NOTES

1. P. R. Brooks, E. M. Jones, *J. Chem. Phys.* **45**, 3449 (1966).
2. R. J. Beuhler Jr., R. B. Bernstein, K. H. Kramer, *J. Am. Chem. Soc.* **88**, 5331 (1966).
3. M. Timm *et al.*, *Science* **393**, 417 (2026).
4. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, *Phys. Rev. Lett.* **49**, 57 (1982).
5. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, *Phys. Rev. Lett.* **50**, 120 (1983).
6. L. Grill *et al.*, *Nat. Nanotechnol.* **2**, 687 (2007).
7. J. Cai *et al.*, *Nature* **466**, 470 (2010).
8. Q. Fan *et al.*, *Science* **372**, 852 (2021).
9. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Science* **262**, 218 (1993).
10. G. Dujardin, R. E. Walkup, P. Avouris, *Science* **255**, 1232 (1992).
11. B. C. Stipe *et al.*, *Phys. Rev. Lett.* **78**, 4410 (1997).
12. K. Anggara, L. Leung, M. J. Timm, Z. Hu, J. C. Polanyi, *Sci. Adv.* **4**, eaau2821 (2018).

ACKNOWLEDGMENTS

The author acknowledges support from the Swedish Research Council.

10.1126/science.aej5890

Materials Design Division, Department of Physics, Chemistry and Biology (IFM), Linköping University, Linköping, Sweden. Email: jonas.bjork@liu.se



Amateur taxonomists such as Raymond Hoser, shown here retrieving a tiger snake from a backyard, often vex professional scientists.

PHOTO: FAIRFAX MEDIA VIA GETTY IMAGES

TAXONOMY

The tangled world of snake taxonomy

Fringe figures in herpetology advance unlikely (and unchecked) claims in an offbeat taxonomical history

Christopher Kemp



Zach St. George
Norton, 2026. 272 pp.

SNAKE MEN | There are around 4249 known snake species worldwide, according to the Reptile Database, which attributes the naming of 255 of those species (or roughly 6% of all known snakes) to Belgian zoologist George Boulenger. The first, described in 1880, was *Atractus duboisi*, the yellow-spotted ground snake. Its habitat, on the eastern slopes of the Ecuadorian Andes, is smaller than a large US cattle ranch. Forty years later, Boulenger described two Indonesian snakes—*Cylindrophis aruensis* and *Calamaria alidae*. They would be his last.

Every species name tells a story: history, exploration, disease, conflict, jungle rot—often disaster, sometimes death. With *Snake Men*, author Zach St. George ventures into the tangled world of snake taxonomy and some of the offbeat characters who populate it. Broadly, the book is about the tension between the slow, scholarly pursuit of taxonomy, performed in institutions and museum archives, and the democratic, less policed version of species description being conducted in basements by upstarts and unaffiliated amateur scientists.

Readers learn about historical figures such as herpetologist Eric Worrell, the rule-breaker who built an Australian snake empire, and Richard Wells, who—as a university student, in collaboration with high school teacher Cliff Wellington—reorganized the taxonomy of Australian herpetology in 1984 in a few mad sleepless days, naming hundreds of new species. The latter pair's work disrupted the field of herpetology for years as others tried to understand what they had done and what it meant.

In an early scene, readers meet Raymond Hoser, the book's main protagonist. Hoser names new species as often as some of us sneeze. Between 1998 and 2025, he named 1573 species—mostly snakes, but not all (he maintains a list at smuggled.com). By the time this publishes, he will have named more. Hoser's species are named for his children, his mother, his friends, and his dead dog. Some of them have scientific names like *Questionthenarrative always* and *Oh yes*. He regularly scoops other researchers who have reported the existence of unnamed species but not yet described them fully—an approach that meets the requirements of species description but also the definition of taxonomic theft.

During his snake work, Hoser says he has been persecuted, followed, singled out, surveilled, set up, victimized, and wrongfully arrested and convicted. The police are corrupt, he tells St. George. The system is rotten. The wildlife protection officers who pursue him are mercenary. His multiple arrests for physical assault are a fiction, he claims. Then there is the global network of respected herpetologists at prestigious institutions who are trying to destroy him.

Over time, Hoser's claims get more outlandish. They cannot all be true. Are any of them true? Does St. George have an obligation to fact-check the wildest ones? Despite hurtling around Melbourne, Australia, with Hoser in his overheated Nissan Micra for hours, the backseat filled with snakes, St. George never does.

Do Hoser's species descriptions, self-published in his own journal from his garden shed, mean anything? Does Hoser have any evidence that respected herpetologists are, as he claims, sex offenders and thieves? St. George never asks him. At times, it is maddening—so many journalistic obligations unmet.

Midway through this often entertaining story, St. George meets Wells, the former student and proto-Hoser who upended Australia's herpetology taxonomy in the 1980s. At one point, he tells St. George that Earth's magnetic poles are reversing and the world is ending soon. Almost all of us will die. Wells says he has collected 17 million books to instruct the people who are left behind how to begin again. St. George writes: "This, admittedly, was not the conversation I had expected, but I found that despite it not conforming to my ordinarily literal understanding of reality, everything Wells was telling me made perfect sense. I sucked my coffee and scribbled notes and thought that Wells was perhaps the most intelligent and charismatic person I'd ever met."

Therein lies the problem at the center of this sometimes uneven book. Writing about people and ideas at the margins of established fields requires a deft touch. It can be done. St. George owes readers a wink or at least some acknowledgment that his subjects are at the fringe. But the wink never comes. □

10.1126/science.aei8559

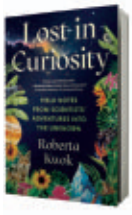
The reviewer is at the Department of Translational Neuroscience, Michigan State University, Grand Rapids, MI, USA, and is the author of *Dark and Magical Places: The Neuroscience of Navigation* (Norton, 2022). Email: cjkemp@gmail.com

SCIENCE LIVES

Dispatches from the front lines of science

Young scientists' passion and commitment take center stage in a moving portrait of research in action

Vijaysree Venkatraman



Roberta Kwok
Sourcebooks, 2026.
368 pp.

LOST IN CURIOSITY | In most known frog species, males advertise their sexual availability with mating calls, while females choose their partners in silence. So when graduate student Johana Goyes Vallejos heard of an obscure frog species with vocalizing females, she wanted to know more. The inch-long nocturnal frogs, known as smooth guardian frogs, were tough to find in the Bornean rainforest. But when Vallejos did find them, she discovered that not only did the females call, they did so before, during, and even after mating. The behavior hinted at an inversion of the Darwinian script, suggesting a rare case of sex role reversal—a phenomenon that had never been reported before in frogs.

In *Lost in Curiosity*, science journalist Roberta Kwok embeds with Vallejos and other young researchers working in fields ranging from paleontology to astronomy to reveal the unglamorous effort that underlies hypothesis testing and the grunt work behind the research stories that typically make the news. The result is nine rigorously reported, character-driven narratives in which the scientific process, rather than its outcomes, takes center stage.

The protagonists in Kwok's narrative share a core persistence, grounded in quiet conviction that their work matters. But small details tell readers about their distinct personalities. Astrophysicist Apurva Oza, for example, names a hypothetical moon after his favorite Bollywood song, revealing that he is unafraid to attach himself emotionally to an object whose existence is difficult to prove. Meanwhile, urban ecologist and Afrofuturist Chris Schell, whose research examines how structural inequities shape the ecology of urban wildlife, frequently quotes Batman—hinting at a preoccupation with justice and hidden systems.

Shadowing researchers, Kwok shows readers what scientists notice, what they infer, and how they recalibrate hypotheses and expectations on the basis of what they find. In Berkeley, California, when a coyote stomach yields a dozen undigested baby rats, Schell is delighted, because it allows him to recast the animal as a human ally rather than a scourge. Rat predation could potentially slow the transmission of certain diseases in cities, a point that he can press with groups intent on culling neighborhood coyotes.

In Greenland, Kwok documents the quick-thinking flexibility that frequently defines work in the field. Here, in the fall of 2023, glaciologist Jessica Mejia has realized she can no longer access her GPS stations, installed earlier that year in a crevasse. A warm summer has melted six feet of snow instead of the expected six inches, making it unsafe to land aircraft near those units. Rather than accepting defeat, Mejia notes that the crevasse has widened—exactly the change the instruments were meant to capture—and decides to return with a rope team the next summer to retrieve the equipment and the data.

Kwok describes how Vallejos—a Colombian citizen—faced a near-decade-long delay in securing the visas and permits required to return to her first field station. Once, when what she thought would be an epic expedition never materialized, the experience left her "flabbergasted…heartbroken and defeated." Fieldwork, it turns out, is the result of a combination of scientific intent and an array of variables—funding, weather, geopolitics, and other factors—that determine what can be accomplished in any given season.

The question arises: Can truth be wrested from nature through valiant but necessarily sporadic efforts? Here, Kwok cites philosopher of science Michael Strevens, who argues in *The Knowledge Machine* that "it is by accounting for the fraction of an inch, the sliver of a degree, that the truth will make itself known."

The protagonists in Kwok's narrative share a core persistence, grounded in quiet conviction that their work matters.

What will all of the labor Kwok recounts—whether in extreme environs, in climate-controlled labs, or late at night in a researcher's home—amount to? Scientific proof offers no guarantee of action. But more than one of Kwok's subjects suggest that hard-won evidence could make inaction indefensible, particularly in the face of existential threats such as climate change and pandemics.

One of the book's strongest chapters unfolds in the Navajo Nation, where a new generation of Diné researchers is tracing the long shadow of uranium mining—a Cold War enterprise that left Indigenous communities exposed to radiation. Cleaning up the abandoned mines would take a "Manhattan Project–like effort," and this may never be forthcoming. Does that mean the damage will never be undone?

Diné environmental toxicologist Richelle Lynn Thomas tests whether medicinal plants absorb heavy metals from the soil, a question her traditional-healer grandfather had wanted answered. She analyzes soil samples and helps people decide where they can safely farm. Scientific agency becomes a form of healing and restoration.

With *Lost in Curiosity*, Kwok brings readers to the wings of a play whose ending has not yet been written. The drama lies in the process of discovery—the almost irrational insistence on continuing to search for answers, and the endurance to persist, despite setbacks. The book reminds readers that science is not only about answers delivered but also about the refusal to stop asking questions. □

10.1126/science.aei1520

The reviewer is a science journalist in Boston, MA, USA. Email: v.vijaysree@gmail.com



Satellite imagery of activities such as deforestation is crucial to Brazil's environmental enforcement efforts.

Brazil's environmental governance at risk

Brazil's global environmental leadership is under threat. Brazil's Congress is advancing legislation that would weaken the scientific and institutional foundations of environmental governance. An agribusiness-backed coalition recently secured approval in the Chamber of Deputies for a package of bills now awaiting Senate endorsement (*1*). If passed, the legislation would constrain environmental enforcement, narrow ecosystem protection, and shift biodiversity decisions away from environmental and scientific institutions. The Senate should reject these bills to preserve the scientific and institutional infrastructure required to translate environmental commitments into enforceable policy.

Bill 2564/2025 (*2*) would bar the use of satellite-based evidence for environmental enforcement by requiring prior notification and on-site inspections by the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA), Brazil's federal environmental agency (*3*). Satellite monitoring has become a cornerstone of Brazilian environmental enforcement, contributing to recent reductions in deforestation (*4*–*6*). Since 2023, Brazil has halved Amazon deforestation to its lowest level in more than a decade (*7*). Slowing the remote detection of illegal land conversion would weaken one of Brazil's most effective evidence-based enforcement tools.

Bill 364/2019 (*8*) would remove legal protection from native nonforest ecosystems despite their disproportionate importance for biodiversity and ecosystem services (*9*). At least 48 million ha of savannas, grasslands, wetlands, and shrublands—roughly twice the size of the UK—could become vulnerable to conversion, including more than half of the Pantanal, 32% of the Pampas, and 7% of the Cerrado (*10*).

Bill 5900/2025 (*11*) would grant the Ministry of Agriculture binding authority over regulations affecting species of economic interest, potentially subordinating environmental decisions to sectoral priorities. For example, the classification of farmed tilapia as a biological risk or potential invader could be delayed or constrained, despite evidence that non-native aquaculture fishes can threaten native fish diversity (*12*). It would also oversee key administrative decisions, including the annual lists of threatened species and pesticides prohibited in Brazil.

Together, these bills would not only threaten biodiversity conservation and ecosystem services essential for agriculture, water security, and climate regulation but also undermine the science-based governance systems that translate environmental data, ecological science, and legal authority into action. Continued scrutiny from scientists, civil society, and the media will be essential to safeguard satellite-based enforcement, protection for native nonforest ecosystems, and independent scientific authority over biodiversity decisions.

Marcus V. Cianciaruso[1], Luis Mauricio Bini[1], José Alexandre Diniz-Filho[1], Rafael Loyola[1,2]

[1]Department of Ecology, National Institute of Science and Technology in Ecology, Evolution, and Biodiversity Conservation (INCT EECBio), Universidade Federal de Goiás, Goiânia, GO, Brazil. [2]Brazilian Foundation for Sustainable Development, Rio de Janeiro, RJ, Brazil. Email: cianciaruso@gmail.com

REFERENCES AND NOTES

1. Instituto Socioambiental (ISA) (2026); https://www.socioambiental.org/en/socio-environmental-news/Agriculture-Day-exposes-rural-offensive-against-the-environment/.
2. Câmara dos Deputados, Projeto de Lei no. 2564/2025 (2025); https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2516808 [in Portuguese].
3. S. Hanbury, "Brazil Congress passes bill to bar use of Amazon deforestation satellite tool," *Mongabay*, 28 May 2026.
4. J. Assunção, C. Gandour, R. Rocha, *Am. Econ. J. Appl. Econ.* **15**, 125 (2023).
5. L. Bragagnolo, R. V. da Silva, J. M. V. Grzybowski, *Ecol. Inform.* **66**, 101454 (2021).
6. M. Finer *et al.*, *Science* **360**, 1303 (2018).
7. World Resources Institute (WRI), "Tropical Rainforest Loss Drops 36% in 2025, but Fires Threaten Global Progress," press release (29 April 2026).
8. Câmara dos Deputados, Projeto de Lei no. 364/2019 (2019); https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2190986 [in Portuguese].
9. G. E. Overbeck *et al.*, *Science* **384**, 168 (2024).
10. F. Wenzel, "Agribusiness bill moves to block grassland protections in Brazilian biomes," *Mongabay*, 28 March 2024.
11. Câmara dos Deputados, Projeto de Lei no. 5900/2025 (2025); https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2586587 [in Portuguese].
12. J. R. Britton, M. L. Orsi, *Rev. Fish Biol. Fish.* **22**, 555 (2012).

10.1126/science.aej1473

PHOTO: USGS/NASA LANDSAT DATA/ORBITAL HORIZON/GALLO IMAGES/GETTY IMAGES

Seafood biosecurity gaps threaten biodiversity

Biological invasions pose serious threats to biodiversity, ecosystem function, economies, and human health (*1*). Live seafood trading is a major pathway for the introduction of non-native and invasive species (*2*–*5*). Although existing local and global frameworks align with international treaties, regulations are often outdated, weakly enforced, and inconsistent across countries (*6*–*9*). To address these gaps, a more robust and coordinated approach is needed to reinforce existing biosecurity protocols.

Recent documentation of invasions demonstrates the risks of the global seafood trade. For example, in 2025, Hong Kong recorded the invasive Atlantic slipper limpet *Crepidula fornicata* and the onyx slipper limpet *C. onyx* as hitchhikers on imported live seafood (*5*). This first local documentation of *C. fornicata* in Hong Kong (*5*, *10*) mirrors the accidental introduction of the species to Europe in the late 19th century (*1*), which caused substantial ecological and economic damages (*1*). Also in 2025, several invasive species, including the beaked barnacle *Austrominius modestus*, the rock barnacle *Semibalanus balanoides*, and the amphipod crustacean *Monocorophium insidiosum*, were reported for the first time in the Tylihul Estuary in Ukraine after arriving as hitchhikers attached to Pacific oyster (*Magallana gigas*) shells (*11*, *12*).

Governments and international organizations should establish a regional and global framework that can facilitate coordinated and enforceable biosecurity measures. The framework should include stricter import controls at entry points, with species-specific and comprehensive shipment declarations. Ships should undergo a prerelease inspection at every port that removes biofouling and hitchhikers, and safe disposal and decontamination practices should be standardized. Improving traceability is essential for the rapid detection of sources and origins and for timely corrective actions. National governments should update and harmonize legislation to combat illegal and unregulated live seafood trading, including species beyond Convention on International Trade in Endangered Species of Wild Fauna and Flora listings. Regular audits and inspections of market suppliers should be institutionalized, with penalties for noncompliance. Collaboration with industry stakeholders is essential for implementing importer certification programs that promote sustainable trading practices. Public awareness campaigns and citizen science monitoring programs are also crucial for educating consumers and improving the detection and reporting of non-native and invasive species. Aligning these principles across countries will help reduce the risk of hitchhikers and high-risk species spreading globally.

Elizaldy Acebu Maboloc[1] and James Kar-Hei Fang[1,2,3]

[1]Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China. [2]State Key Laboratory of Marine Environmental Health, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China. [3]PolyU-BGI Joint Research Centre for Genomics and Synthetic Biology in Global Ocean Resources and Research Institute for Future Food, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China.
Email: james.fang@polyu.edu.hk

REFERENCES AND NOTES

1. J. P. Bray, C. L. R. Hewitt, P. E. Hulme, *BioScience* **74**, 440 (2024).
2. M. Blanchard, *Sci. Mar.* **61**, 109 (1997).
3. J. W. Chapman, T. W. Miller, E. V. Coan, *Conserv. Biol.* **17**, 1386 (2003).
4. A. N. Cohen, "Aquatic Invasive Species Vector Risk Assessments: Live Marine Seafood and the Introduction of Non-native Species into California" (California Ocean Protection Council, 2012).
5. E. A. Maboloc, J. K.-H. Fang, *Glob. Ecol. Conserv.* **61**, e03691 (2025).
6. P. E. Hulme, *One Earth* **4**, 666 (2021).
7. S. A. Simmons, M. De Poorter, Eds., "Best Practices in Pre-Import Risk Screening for Species of Live Animals in International Trade: Proceedings of an Expert Workshop on Preventing Biological Invasions, University of Notre Dame, Indiana, USA, 9–11 April 2008" (Global Invasive Species Programme, 2009); https://www.gisp.org/publications/policy/workshop-riskscreening-pettrade.pdf.
8. W. Lau *et al*., "Hong Kong's Seafood Trade, Port Measures and Import Controls. Parts I, II and III" (ADM Capital Foundation, 2023); https://www.admcf.org/wp-content/uploads/2025/01/HKST_Full-reportdf.pdf.
9. Y. C. Kam, A. Chung, M. Tin, J. D. Gaitan-Espitia, C. Schunter, *Mar. Policy* **165**, 106200 (2024).
10. J. C. Astudillo, *et al*., Eds., Hong Kong Register of Marine Species (2026); https://www.marinespecies.org/hkrms/index.php.
11. H. Gabrielczak, Y. Kvach, M. O. Son, *NeoBiota* **103**, 299 (2025).
12. H. Glenner, J. Lützen, L. C. Pacheco-Riaño, C. Noever, *Aquat. Invasions* **16**, 675 (2021).

10.1126/science.aej1248

Protect and expand ocean observation systems

The United States alone is responsible for roughly half of the world's ocean monitoring platforms (*1*). In May, the US government announced plans to decommission and remove more than 900 deep-sea instruments moored across the Atlantic and Pacific oceans (*2*). Although implementation has since been paused (*3*), the episode exposed the vulnerability of a global ocean observation system that is essential to global risk management infrastructure. The international community must take urgent steps to broaden the coalition of countries investing in ocean observation infrastructure.

The ocean is a central regulator of Earth system stability and climate resilience. Ocean waters have absorbed 90% of excess planetary heat and 25% of human carbon emissions (*4*). This carbon sink function has buffered the effects of increased temperatures on human societies, but it is fundamentally changing ocean conditions (*5*).

Long-term ocean observations are particularly important because some of the most consequential Earth system risks emerge gradually and can be detected only through continuous measurements. These include changes in major circulation systems such as the Atlantic Meridional Overturning Circulation (AMOC), marine heat waves, and ecosystem transitions with potentially far-reaching societal consequences (*6*). Indeed, in 2025, Iceland formally identified a potential AMOC collapse as a national security risk (*7*).

Reducing ocean observation infrastructure has global implications: Roughly 40% of the global population lives along coastlines, and the rapidly expanding ocean economy is projected to outpace global growth for decades (*8*). Satellites, which provide valuable insight into atmospheric change and surface dynamics but reveal little about the structural changes occurring to ocean currents, chemistry, or ecosystems, cannot replace this infrastructure. At a time of accelerating environmental change, ocean observing systems function as a planetary early warning system. Governments should treat them as critical public infrastructure, expanding rather than dismantling collection of the long-term observations needed to detect emerging risks, anticipate change, and navigate an increasingly uncertain future.

Robert Blasiak[1] and Albert V. Norström[1,2]

[1]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. [2]Future Earth, c/o Royal Swedish Academy of Sciences, Stockholm, Sweden. Email: robert.blasiak@su.se

REFERENCES AND NOTES

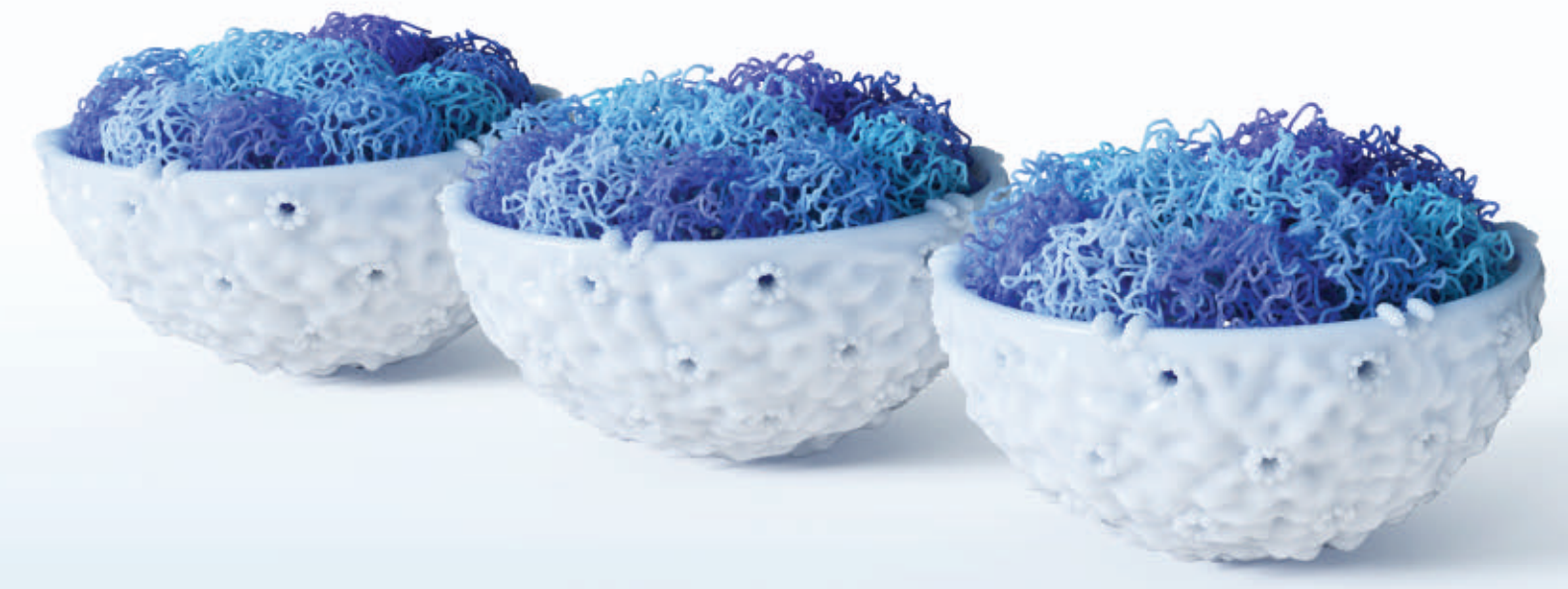
1. Integrated Ocean Observing System (IOOS), Global Ocean Observing System for Climate (2026); https://ioos.noaa.gov/community/global/.
2. Ocean Observatories Initiative (OOI), Announcement on NSF OOI Descoping (21 May 2026); https://oceanobservatories.org/2026/05/announcement-on-ooi-descoping/.
3. US National Science Foundation, Update on Ocean Observatories Initiative (18 June 2026); https://www.nsf.gov/news/update-ocean-observatories-initiative.
4. J. A. Church *et al*., *Geophys. Res. Lett.* **38**, L18601 (2011).
5. Intergovernmental Panel on Climate Change (IPCC), *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner *et al*., Eds. (Cambridge Univ. Press, 2019).
6. D. Nian, M. Willeit, N. Wunderling, A. Ganopolski, J. Rockström, *Commun. Earth Environ.* **7**, 295 (2026).
7. A. Withers, S. Jacobsen, "Iceland deems possible Atlantic current collapse a security risk," *Reuters*, 12 November 2025.
8. J.-B. Jouffray, R. Blasiak, A.-V. Norström, H. Österblom, M. Nyström, *One Earth* **2**, 43 (2020).

10.1126/science.aej4152

SPECIAL ISSUE

GENOME IN MOTION

Mapping the nucleus in health



Like thread gathering into increasingly intricate coils, the genome's linear sequence folds into an elaborate three-dimensional architecture within the nucleus—organization that shifts across development and aging, shapes gene expression, and, when disrupted, contributes to disease.

DI JIANG, YEVGENIYA NUSINOVICH, and SARAH ROSS

The human genome is not a static list of instructions but a dynamic, three-dimensional structure that folds, loops, and reorganizes within the nucleus to control where and when genes are expressed. This spatial and temporal organization—the "4D" nucleome—has emerged as a critical layer of gene regulation, one whose disruption is increasingly linked to disease.

The National Institutes of Health Common Fund's 4D Nucleome (4DN) Program was established to generate the data, tools, and technologies needed to understand this organization at scale. Building on a first phase that produced thousands of public datasets and dozens of software packages, the program's second phase has pushed toward achieving single-cell resolution and analyzing disease-relevant tissue, asking how chromatin architecture shapes—and is reshaped by—human health.

In this collection, published jointly in *Science* and *Science Advances*, researchers applied cutting-edge single-cell and multiomic technologies to map genome architecture across an unusually broad range of biological contexts: the aging and Alzheimer's-affected brain, the failing heart, developing cardiomyocytes, maturing B cells, and diverse tissues spanning the body. Together, these studies reveal how genome folding contributes to cell identity, how it is perturbed in disease, and how epigenetic memory can be encoded in chromatin's physical compaction. Complementing these biological investigations, the computational and imaging tools presented here—for denoising sparse single-cell contact maps and resolving regulatory elements in previously understudied tissues—extend the reach of spatial genomics to systems where, historically, data have been limited.

Collectively, this package illustrates how far single-cell three-dimensional genomics has advanced since the 4DN program's inception and how integrating chromatin architecture with gene expression and epigenetic data is reshaping our understanding of development, aging, and disease. The resources and methods introduced here are poised to accelerate discovery across the broader research community.

10.1126/science.aek3868

This introduction is being simultaneously published in *Science* and *Science Advances*.

ILLUSTRATION: A. FISHER/*SCIENCE*



RESEARCH ARTICLE SUMMARY

A 3D genome atlas of human tonsil and the role of loop extrusion in B cell somatic hypermutation

Yubao Cheng†, Jianshu Wang†, Yuan Zhang†, Anurupa Devi Yadavalli, Miao Liu, Shengyan Jin, Grace Buddle, Ann Haberman, David G. Schatz*, Siyuan Wang*

Full article and list of author affiliations: https://doi.org/10.1126/science.adw4243

INTRODUCTION: Somatic hypermutation (SHM) is a critical mechanism that diversifies immunoglobulin (Ig) genes in germinal center B cells (GCBC) and enables the generation of high-affinity antibodies. SHM is mediated by activation-induced deaminase (AID), which deaminates cytosine in DNA during transcription to initiate mutagenesis. Mistargeting of SHM contributes to oncogenic mutations, chromosomal translocations, and the development of B cell lymphoma. Prior studies suggest a potential linkage between chromatin organization and SHM targeting. However, it is not yet clear how multiscale three-dimensional (3D) genome organization might affect the SHM targeting process.

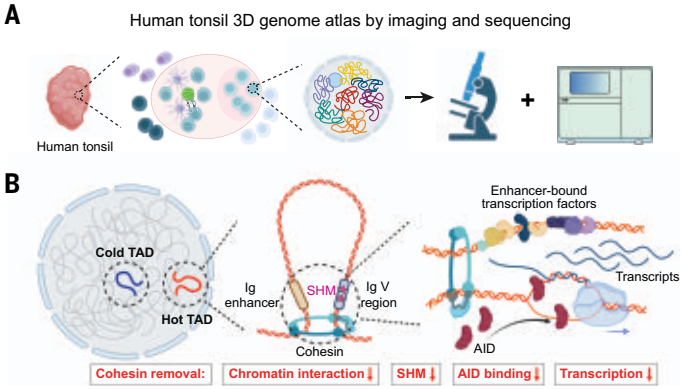
RATIONALE: Characterizing the 3D genome landscape across length scales during GCBC development and active SHM in the native tissue context will lay the foundation for understanding the architectural basis of chromatin-mediated SHM targeting. We previously demonstrated that topologically associating domains (TADs) delineate SHM susceptibility and that the cohesin loader NIPBL, the key regulator of loop extrusion, is enriched in SHM-susceptible (hot) TADs as compared with SHM-resistant (cold) TADs. Direct interrogation of the loop extrusion machinery, integrated with measurements of chromatin architecture, AID binding, and the transcriptional activity in SHM-activated B cells, will elucidate how multiple genomic processes regulate SHM targeting.

RESULTS: In this study, we developed an imaging-based toolkit for multimodal profiling of the 3D genome organization and spatial transcriptome in human clinical tissue samples and integrated it with a sequencing-based, single-cell 3D genomics approach. Using these complementary methods, we constructed a single-cell 3D genome atlas of human tonsil cells and characterized changes in multiscale 3D genome features during GCBC development. The atlas revealed both cell type–specific 3D genome features associated with gene regulation and cell type–invariant features. The SHM-associated 3D genome analysis revealed that cold TADs are located more toward the nuclear interior, whereas hot TADs exhibit stronger intra-TAD chromatin interactions in SHM-active cells.

To further define the relationship between 3D genome organization and SHM, we engineered human Burkitt lymphoma cell line Ramos to allow rapid removal of cohesin through induced RAD21 degradation. Upon cohesin removal, SHM activity was fully abolished between 6 and 12 hours (the earliest interval during which mutations could be detected in untreated cells), coupled with a progressive destabilization of chromatin interactions, some rapidly lost as early as 2 hours. Transcription activities at multiple SHM target genes partially decreased but were still substantial even at 12 hours, and AID binding was only modestly reduced at 6 hours and was completely lost by 18 hours.

CONCLUSION: Our study reveals SHM-associated 3D genome features at different length scales and an essential role for the cohesin complex in SHM targeting. Our data support a model in which cohesin-mediated loop extrusion enables and sustains SHM by maintaining local chromatin architecture, an active transcriptional environment, AID binding, and perhaps other mechanisms as well. These local mechanisms, together with the global chromatin feature of radial nuclear positioning, collectively contribute to SHM susceptibility and resistance. □

*Corresponding author. Email: david.schatz@yale.edu (D.G.S.); siyuan.wang@yale.edu (S.W.) †These authors contributed equally to this work. Cite this article as Y. Cheng *et al.*, *Science* 393, eadw4243 (2026). DOI: 10.1126/science.adw4243



Single-cell 3D genome atlas of human tonsil and the role of chromatin loop extrusion in SHM. (**A**) A single-cell 3D genome atlas of human tonsil generated by integration of sequencing-based and image-based 3D genomics methods. (**B**) Multiscale 3D chromatin features contribute to SHM targeting. Cohesin sustains the chromatin architecture, transcriptional environment, and AID binding necessary for SHM. [Figure created with BioRender.com]

RESEARCH ARTICLE

A 3D genome atlas of human tonsil and the role of loop extrusion in B cell somatic hypermutation

Yubao Cheng[1]†, Jianshu Wang[2]†, Yuan Zhang[1]†, Anurupa Devi Yadavalli[2], Miao Liu[1], Shengyan Jin[1], Grace Buddle[2], Ann Haberman[2], David G. Schatz[2]*, Siyuan Wang[1,3,4,5,6,7,8,9,10]*

B cell maturation within the germinal center tissue microenvironment involves immunoglobulin gene diversification by somatic hypermutation (SHM). How three-dimensional (3D) genome architecture influences SHM is not fully understood. We leveraged sequencing-based and image-based 3D genomics and transcriptomics to map single-cell 3D genome organization and gene expression across cell types and states in human tonsils and in B cell lymphoma cell lines. These analyses revealed trajectories of compartment, looping, and nuclear position changes during the B cell immune response and activation of SHM. Targeted protein degradation of cohesin component RAD21 revealed its contribution to enabling SHM. Our results provide a single-cell 3D genome atlas of human tonsil cells and outline the links between the chromatin loop extrusion machinery and SHM.

Antibody-mediated immunity relies on somatic hypermutation (SHM) to generate point mutations in rearranged immunoglobulin (Ig) genes of germinal center B cells (GCBC) (*1*). SHM is vital for antibody affinity maturation, but mistargeting of the reaction contributes to oncogenic mutations, chromosomal translocations, and the development of B cell lymphoma (*2*–*7*). Previous studies suggest that SHM is intimately linked to three-dimensional (3D) genome organization (*8*, *9*), but how 3D genome architecture and architectural factors regulate the targeting and mistargeting of SHM and how this relates to a healthy versus oncogenic outcome is poorly understood. Defining the 3D genome and nuclear architecture as GCBC develop and activate SHM and how this goes awry in the case of B cell lymphoma will help establish the foundation for understanding the role of nuclear organization in the balance between health and disease.

During SHM, the enzyme activation-induced deaminase (AID) initiates a highly mutagenic process that is regulated by enhancers, super-enhancers (*10*–*14*), transcription factors, and transcriptional pausing (*2*, *11*, *15*, *16*), and as we recently demonstrated, by topologically associating domains (TADs) (*8*). Thus, prior evidence suggests a linkage between 3D chromatin architecture and SHM. Upon immune activation, naïve B cells (NBC) can migrate into the germinal center (GC) and give rise to two populations of GCBC, dark zone (DZ) centroblasts (the primary cell type for SHM), and light zone (LZ) centrocytes. GCBC in the LZ and DZ interconvert and eventually provide the output of the GC in the form of memory B cells (MBC) and plasma cells (PC) (*17*). In the context of cancer, extensive evidence indicates that mistargeted SHM in centroblasts contributes to the generation of B cell lymphoma (*2*–*7*). There is a critical need to uncover the architectural basis of chromatin-mediated SHM targeting to understand how this goes awry in the context of B cell lymphomagenesis.

How 3D chromatin folding affects genomic functions in single cells in their primary tissue microenvironment is not well understood. In this work, we developed a 3D genome and transcriptome imaging toolbox optimized for clinical tonsil tissue, including chromatin tracing (*18*), RNA multiplexed error-robust fluorescence in situ hybridization (MERFISH) (*19*), and multiplexed imaging of nucleome architectures (MINA) (*20*). We integrated this imaging framework with a cutting-edge single-cell Hi-C approach using Droplet Hi-C (*21*) and applied it to primary human tonsil tissue and malignant GC-derived B cell lymphoma cell lines to define 3D genome architecture and nuclear organization in GC B cells undergoing SHM.

Results

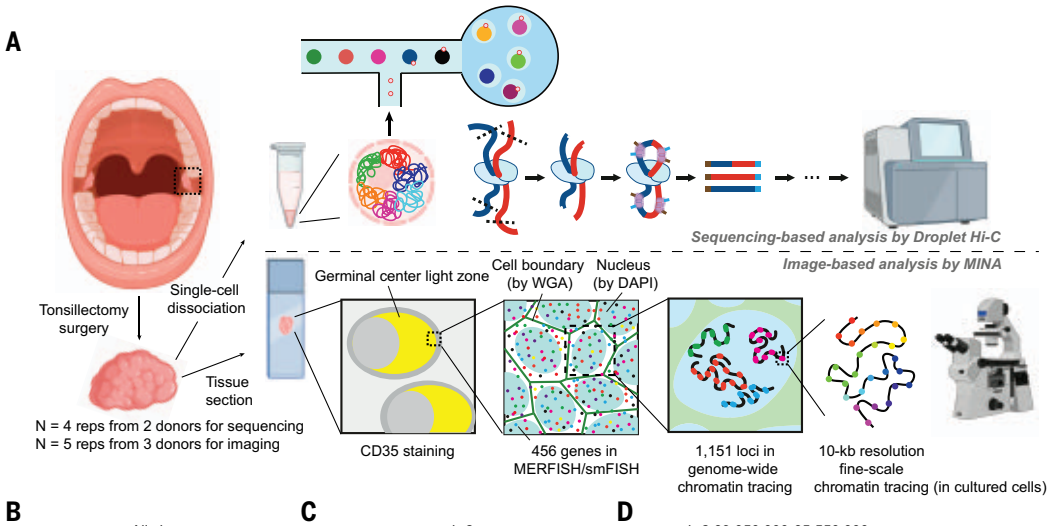
A single-cell 3D genome atlas of human tonsil

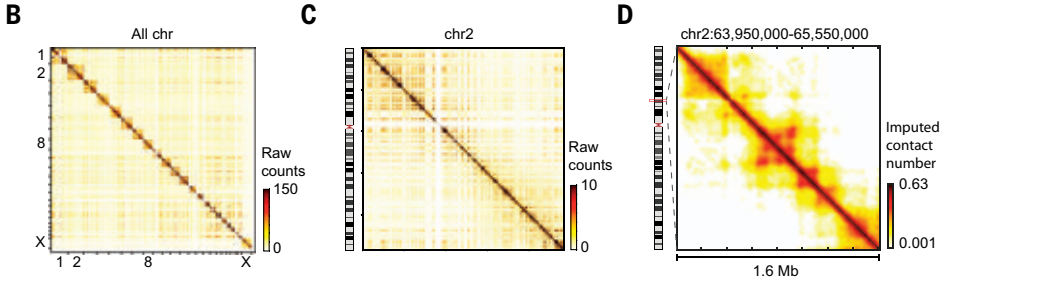
To explore the genome folding patterns of cells in the human tonsil, including the variety of B cells (fig. S1A) and the potential regulatory role of 3D genome in SHM targeting, we integrated sequencing-based analysis by Droplet Hi-C and image-based analysis by MINA (Fig. 1A). In Droplet Hi-C experiments, human tonsil tissues obtained from tonsillectomy were dissociated into single cells and processed following published procedures (*21*). We obtained 12,308 high-quality cells with a median of 145,950 distinct read pairs per cell [with 15,707 cis-long contacts (>1 kb) and 10,145 trans contacts per cell] from four tissue replicates collected from two donors (fig. S1B). Droplet Hi-C revealed chromatin interaction maps at multiple length scales (Fig. 1, B to D). Intrachromosomal contact frequencies were highly correlated among the four replicates (fig. S1, C and D), indicating biological reproducibility.

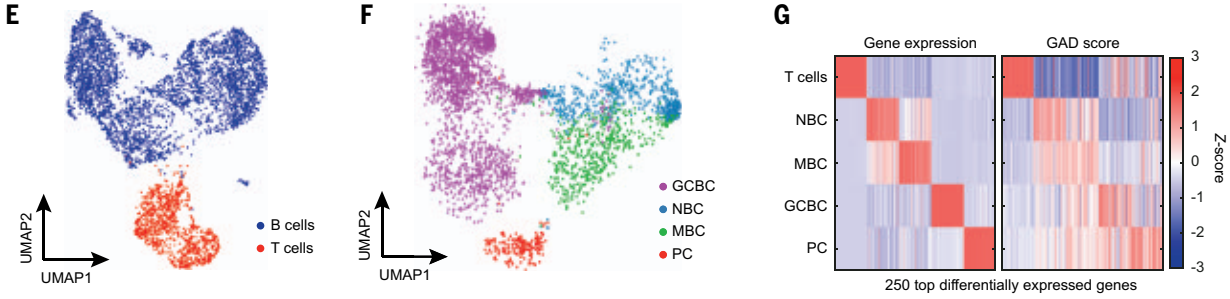
Previous work has demonstrated that the association between chromatin interactions within the gene body and gene expression levels can be leveraged to cluster and annotate single-cell Hi-C data (*21*, *22*). We quantified chromatin contacts within the gene body of each gene to derive gene-body associating domain (GAD) scores and assembled a cell-by-gene GAD score matrix for clustering single cells. GAD score–based clustering showed a clear separation of T cell and B cell populations (Fig. 1E). Focusing on B cells, we performed coembedding with a previously published single-cell RNA sequencing (scRNA-seq) dataset (*23*) to guide subclustering and annotation of B cell states. The B cell population could be further clustered into NBC, MBC, GCBC, and PC (Fig. 1F and fig. S1, E and F). Differentially expressed genes identified by scRNA-seq showed concordant differences in GAD scores, including canonical markers of cell identity (Fig. 1G).

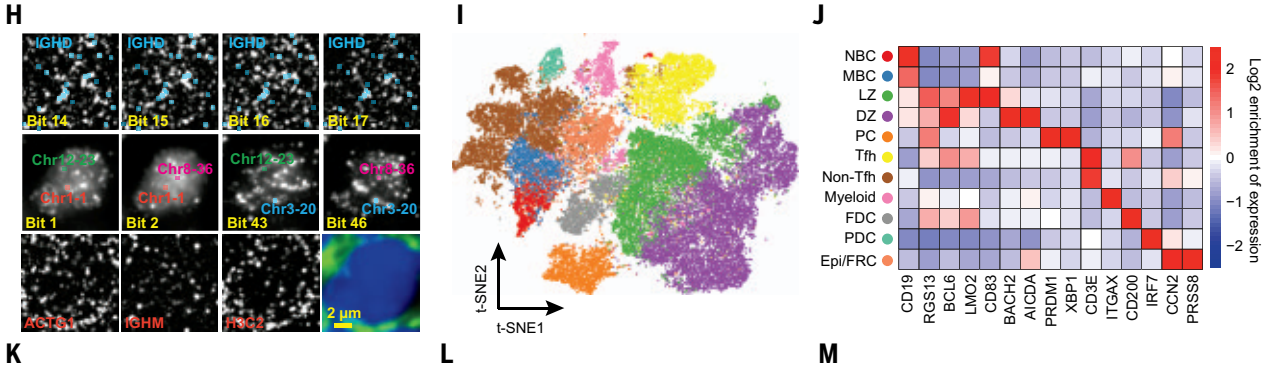
In the MINA analysis, we designed a genome-wide chromatin tracing library targeting a total of 1151 genomic loci, including (previously defined) (*8*) 49 SHM-susceptible (hot) TADs, 97 SHM-resistant (cold) TADs, and 1005 filler regions that spanned the human genome with near uniform genomic intervals (fig. S2, A and B). To integrate genome folding measurements with cell type identification and cell cycle status, we designed an RNA MERFISH library targeting 447 RNA species, including selected tonsil cell type markers, cell cycle markers, and immunity-related genes. Nine RNA species, which are either too short or overly abundant, were included in a separate single-molecule FISH (smFISH) library designed for sequential readout. Human tonsil tissue was fresh-frozen and cryosectioned into 8-μm-thick slices for experiments. We applied MINA (integrative highly multiplexed DNA, RNA, and protein imaging) to tonsil sections, combining genome-wide chromatin tracing, RNA MERFISH, sequential smFISH, and protein staining on the

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA. [2]Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA. [3]Department of Cell Biology, Yale University School of Medicine, New Haven, CT, USA. [4]Yale Combined Program in the Biological and Biomedical Sciences, Yale University, New Haven, CT, USA. [5]Molecular Cell Biology, Genetics and Development Program, Yale University, New Haven, CT, USA. [6]Biochemistry, Quantitative Biology, Biophysics, and Structural Biology Program, Yale University, New Haven, CT, USA. [7]M.D.-Ph.D. Program, Yale University, New Haven, CT, USA. [8]Yale Center for RNA Science and Medicine, Yale University School of Medicine, New Haven, CT, USA. [9]Yale Liver Center, Yale University School of Medicine, New Haven, CT, USA. [10]Yale Cancer Center, Yale University School of Medicine, New Haven, CT, USA. ***Corresponding author. Email: david.schatz@yale.edu (D.G.S.); siyuan.wang@yale.edu (S.W.)** †These authors contributed equally to this work.


A
Tonsillectomy surgery
Single-cell dissociation
Tissue section
N = 4 reps from 2 donors for sequencing
N = 5 reps from 3 donors for imaging
Sequencing-based analysis by Droplet Hi-C
Image-based analysis by MINA
Germinal center light zone
Cell boundary (by WGA)
Nucleus (by DAPI)
CD35 staining
456 genes in MERFISH/smFISH
1,151 loci in genome-wide chromatin tracing
10-kb resolution fine-scale chromatin tracing (in cultured cells)


B
All chr
Raw counts
150
0
C
chr2
Raw counts
10
0
D
chr2:63,950,000-65,550,000
Imputed contact number
0.63
0.001
1.6 Mb


E
UMAP2
UMAP1
B cells
T cells
F
UMAP2
UMAP1
GCBC
NBC
MBC
PC
G
Gene expression
GAD score
T cells
NBC
MBC
GCBC
PC
Z-score
250 top differentially expressed genes


H
IGHD
Bit 14
Bit 15
Bit 16
Bit 17
Bit 1
Bit 2
Bit 43
Bit 46
ACTB1
IGHM
HSC2
2 µm
I
t-SNE2
t-SNE1
J
NBC
MBC
LZ
DZ
PC
Tfh
Non-Tfh
Myeloid
FDC
PDC
Epi/FRC
CD19
RGS13
BCL6
LMO2
CD83
BACH2
AICDA
PRDM1
XBP1
CD3E
ITGAX
CD200
IRF7
CCN2
PRSS8
Log2 enrichment of expression

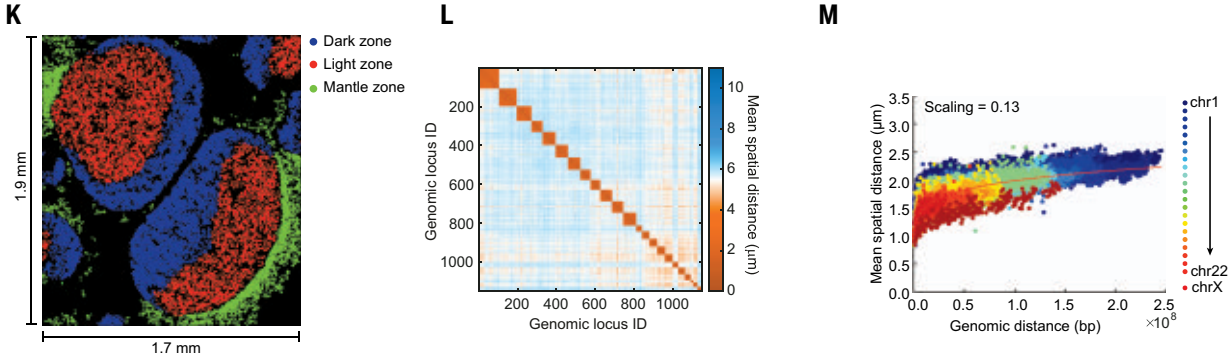

K
Dark zone
Light zone
Mantle zone
1.9 mm
1.7 mm
L
Genomic locus ID
Genomic locus ID
Mean spatial distance (µm)
M
Scaling = 0.13
Mean spatial distance (µm)
Genomic distance (bp)
chr1
chr22
chrX

Fig. 1. A single-cell 3D genome atlas of human tonsil. (**A**) Schematic illustration of the experimental workflow. (**B** to **D**) Droplet Hi-C–derived interaction map at whole-genome (B), single-chromosome (C), and chromatin domain (D) levels. (**E** to **F**) Uniform manifold approximation and projection (UMAP) visualization of Droplet Hi-C–based cell type clusters (E) and B cell subclusters (F). $n = 5066$ cells in (E) and 3707 cells in (F) from one representative biological replicate. Data points represent individual cells. (**G**) Heatmap of RNA fragments per kilobase per million mapped reads and mean scGAD scores of differentially expressed genes in each cell type. (**H**) Examples of raw MERFISH foci (first row), chromatin tracing foci (second row), smFISH foci, 4′,6-diamidino-2-phenylindole (DAPI), and WGA staining (third row) in human tonsil. Examples of MERFISH and chromatin tracing decoding results were color-labeled. Scale bars, 2 μm. (**I**) Single-cell clusters identified from MERFISH data displayed with t-SNE; $n = 36{,}039$ cells from one representative biological replicate. Data points represent individual cells. (**J**) Log_2 enrichment of marker genes in each identified cell types measured by MERFISH. (**K**) The spatial map of the identified dark zone, light zone, and mantle zone. (**L**) Matrix of mean interloci distances between all genomic loci in tonsil cells. Genomic locus IDs are ordered by chromosomes (1 to 22, followed by X) and by genomic coordinates. (**M**) Relation between mean spatial distance and the genomic distance of each pair of loci from each chromosome. The red line represents the power-law fitting curve. Data points represent pairs of genomic loci.

same sample (Fig. 1A). CD35, a follicular dendritic cell marker, was stained by coimmunofluorescence for GC LZ labeling. Wheat germ agglutinin (WGA) staining was performed for cell membrane labeling for cell segmentation. We obtained clear and decodable FISH signals in tonsil tissue (Fig. 1H). Decoded RNA MERFISH copy numbers correlated with RNA-seq results at the bulk level ($R = 0.61$, fig. S2C), and high correlation values among different RNA MERFISH datasets indicated biological reproducibility (fig. S2, D and E). Segmented single cells contained on average 201 copies of RNA molecules (total count of RNA molecules of all probed RNA species) per cell (fig. S2F) and 80 detected genes (probed RNA species) per cell (fig. S2G). By performing single-cell RNA profile clustering and annotation, we identified all major cell types in human tonsil, including NBC, MBC, LZ centrocytes, DZ centroblasts, PC, $CD4^+$ and $BCL6^+$ T follicular helper (Tfh) cells, $BCL6^-$ non–T follicular helper (non-Tfh) cells, myeloid cells, follicular dendritic cells (FDC), plasmacytoid dendritic cells (PDC) and a combined group of epithelial cells and fibroblastic reticular cells (Epi/FRC) (Fig. 1I), with known marker genes (*23*–*30*) expressed in the corresponding cell types (Fig. 1J and fig. S2H).

Projecting cell identities onto the tissue image revealed the GC structure—clearly distinguishing DZ, LZ, and mantle zone cells (Fig. 1K and fig. S3, A to C)—with the corresponding spatial expression patterns of marker genes (fig. S3D). The pattern of LZs and DZs aligned with the CD35 coimmunofluorescence staining pattern observed in the same tissue section (fig. S3E). The cell cycle markers included in our RNA MERFISH library allowed us to further define the cell cycle stages of individual cells, and three distinct clusters corresponding to G1, G2/M, and S phases were identified. Projecting cell cycle stages onto the t-distributed stochastic neighbor embedding (t-SNE) of cell type clusters revealed varying proportions of cell cycle stages among different cell types (fig. S3F). The DZ B cell cluster contained 55 and 26% of cells in the G2/M and S stages, respectively (fig. S3G), which is consistent with values reported in literature (*31*) and the highly proliferative nature of B cells in the DZ (*32*).

After the decoding of genome-wide chromatin tracing data, we reconstructed chromatin traces in single cells. To validate the quality of the genome-wide chromatin tracing, we calculated the average spatial distance between each pair of targeted genomic loci and generated an interloci spatial distance matrix (Fig. 1L). Genomic loci on the same chromosome exhibited shorter distances between one another compared with loci on different chromosomes, which is consistent with the existence of chromosome territories (*33*). Smaller chromosomes with shorter genomic lengths exhibited overall smaller spatial sizes (Fig. 1M). The intrachromosomal spatial distance between genomic loci scaled with their genomic distance with a power law scaling factor of 0.13 (Fig. 1M), which is consistent with previous observations in human cell cultures and mammalian tissues (*18*, *20*). Among five different genome-wide chromatin tracing datasets performed on tonsils from three individuals, the high correlation coefficients of the mean intrachromosomal distances between replicates demonstrated high biological and technical reproducibility (fig. S3, H and I).

To systematically characterize the 3D genome organization in tonsil in a cell type–specific fashion, we analyzed the 3D genome conformations in the different cell types using our Droplet Hi-C and MINA datasets. With Droplet Hi-C, we derived cell type–specific A/B compartment organization (fig. S4A) in which compartment A is often associated with active chromatin and compartment B with inactive chromatin (*34*). We identified 3874 genomic bins with differential A/B compartment scores among the five tonsil cell types at 100-kb resolution (Fig. 2A), with the remaining ~80% of genomic regions exhibiting a conserved compartment state across cell types (Fig. 2B). To link compartment dynamics to gene regulation, we focused on the NBC-to-GCBC, GCBC-to-MBC, and GCBC-to-PC transitions along the B cell developmental trajectory (fig. S1A). For each pairwise comparison, genes up-regulated at the RNA level in one cell type relative to the other exhibited higher delta A/B compartment scores, indicating a coordinated shift toward a more transcriptionally permissive A compartment (Fig. 2C and fig. S4, B and C). Correlation analysis of both compartment scores and GAD scores across all five cell types recapitulated the known developmental relationships (*35*) (Fig. 2, D and E). Among the five cell types, differential gene expression better correlated with GAD scores than with A/B compartment scores (fig. S4D), which is consistent with previous observations that A/B compartment changes are moderately associated with gene expression changes (*36*).

We examined compartment and GAD scores of select genes known to be involved in B cell state transitions (Fig. 2F). In the NBC-to-GCBC transition, *BCL6* (a vital transcription factor for GCBC and Tfh cell development) (*37*, *38*), *AICDA* (encoding AID), and SHM are induced and cells enter a highly proliferative state, reflected by the up-regulation of cell cycle–related genes such as *MKI67*. *AICDA, BCL6*, and *MKI67* are expected to be up-regulated in the NBC-to-GCBC transition and down-regulated upon GC exit. Consistent with this, both GAD scores and A/B compartment scores for the *BCL6*, *AICDA*, and *MKI67* gene regions increased in the NBC-to-GCBC transition and decreased in the GCBC-to-MBC and GCBC-to-PC transitions (Fig. 2F). At the *AICDA* locus, these GAD score and A/B compartment score changes were accompanied by the same trend of chromatin accessibility changes measured by single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) from a previous study (*23*) (fig. S4E).

As GC-specific gene expression features progressively shut down in the GCBC-to-MBC and GCBC-to-PC transitions, the GCBC-to-PC transition is accompanied by robust up-regulation of PC-associated effector genes including *PRDM1* and *JCHAIN*. *PRDM1*, encoding the transcription factor BLIMP-1, plays critical roles in silencing GC-specific transcriptomic features and supporting PC differentiation and up-regulation of the secretory machinery (*39*). *JCHAIN* facilitates the assembly of polymeric Ig's in secretory PC (*40*). The GAD scores for *PRDM1* and *JCHAIN* again faithfully captured the expected directional changes across all transitions (Fig. 2F). Two other markers, *CD27* and *CD38*, both of which are known to be highly expressed in PC (*41*), likewise showed the highest GAD scores in PC (Fig. 2F). In comparison, compartment score and chromatin accessibility changes are often less concordant with the expected expression changes of these marker genes (Fig. 2F and fig. S4E). To further investigate the gene-level regulation, we examined chromatin loops at the *PRDM1* locus. In PC, *PRDM1* displayed markedly elevated enhancer-promoter looping (Fig. 2G), with

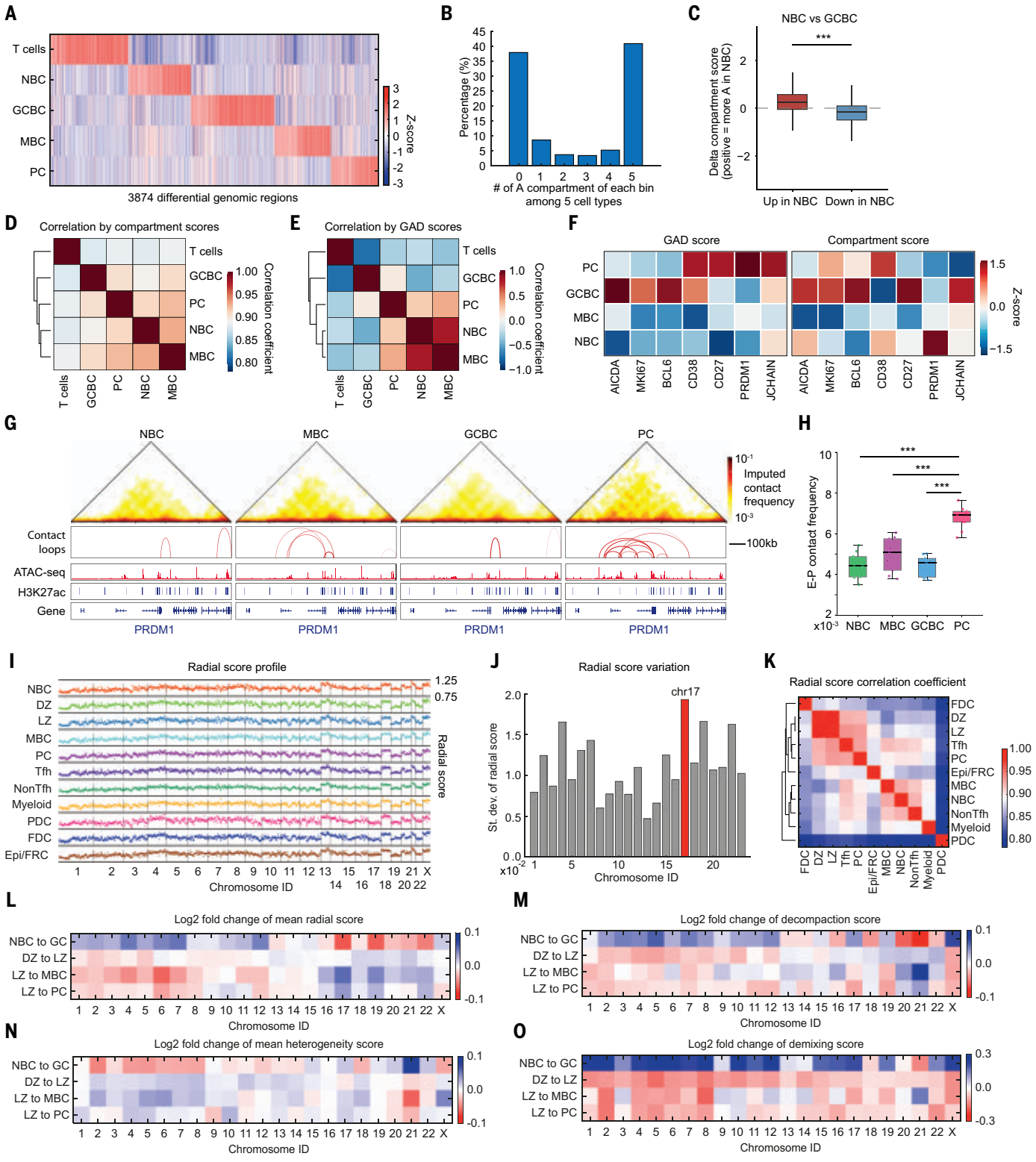


Fig. 2. Three-dimensional genome organization in tonsil. (**A**) Heatmap showing the *z* scores of compartment scores of differential compartments among cell types. (**B**) Percentages of regions with different levels of compartment switch among five cell types. Along the *x* axis, "0" indicates regions that are always in compartment B among the 5 cell types, and "5" indicates regions that are always in compartment A among the 5 cell types. (**C**) Box plots of delta A/B compartment scores for genes differentially expressed in NBC versus GCBC. Up-regulated genes tend to exhibit positive delta compartment scores, indicating enhanced A compartment identity. Boxes show the interquartile ranges (IQR), central lines represent the medians, and whiskers denote the maximum and minimum values up to 1.5 × IQR away from the box. (**D**) Correlation of compartment scores between each pair of cell types with hierarchical clustering. (**E**) Correlation of GAD scores between each pair of cell types with hierarchical clustering. (**F**) Heatmap of GAD scores and A/B compartment scores for selected B cell state transition genes across B cell states. (**G**) Imputed Hi-C contact maps at the *PRDM1* locus (chromosome 6) in PC, GCBC, NBC, and MBC. Enhancers were nominated by colocalization of PC ATAC-seq and Ramos B cell H3K27ac ChIP-seq signals. (**H**) Quantification of

enhancer-promoter contact frequency at the *PRDM1* locus. Each dot represents an interaction between two 10-kb bins corresponding to the *PRDM1* promoter and an enhancer. Boxes show the IQR; central lines represent the medians; and whiskers denote the maximum and minimum values up to 1.5 × IQR away from the box. (**I**) Radial score profile of each chromosome in each cell type. (**J**) Standard deviation of the mean radial score values of each chromosome across all cell types. (**K**) Correlation of radial scores of all targeted foci between each pair of cell types with hierarchical clustering. (**L**) Log_2 fold change of mean radial score along B cell developmental trajectory. (**M** to **O**) Log_2 fold changes of chromosome territory decompaction score (M), mean value of chromatin conformation heterogeneity score (N), and chromatin demixing score (O) of each chromosome along B cell developmental trajectory. Data merged from four biological replicates are shown in (A) to (H) ($n = 12,308$ cells). Data merged from five biological replicates are shown in (I) to (O) ($n = 121,373$ cells). Significance in (C) and (H) was defined as $***P < 0.001$. *P* values were calculated by two-sided Wilcoxon rank-sum tests.

the enhancers nominated by colocalization of PC-specific pseudobulk scATAC-seq signal (*23*) and Ramos B cell H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) (Fig. 2G). Enhancer-promoter contact frequency was substantially higher in PC than in all other cell types (Fig. 2H, $P < 0.001$). Overall, these findings suggested that whereas A/B compartment organization established the broad permissive chromatin framework during B cell differentiation, the precise regulation of lineage-determining genes was governed more locally, mediated by GAD dynamics and enhancer-promoter contacts at individual loci.

On the basis of the finely resolved cell type annotations obtained from RNA MERFISH, we quantified 3D genome conformations across 11 distinct cell types using a set of metrics derived from genome-wide chromatin tracing (fig. S5). We analyzed the radial positioning of all target loci in the cell nucleus, with a higher radial score indicating more association with the nuclear periphery (Materials and methods). Overall, the differences in radial scores between different chromatin regions and different chromosomes were greater than the difference of the same chromatin regions between different cell types (Fig. 2I and fig. S6A). For example, we observed that gene-dense chromosome 19 tended to be located more toward the nuclear interior, whereas the similarly sized but gene-poor chromosome 18 was positioned more toward the periphery of the nucleus, which is consistent with previous reports (*42*, *43*). Previous studies also reported that human chromosomes 16, 17, and 22 tend to be in the nuclear interior, whereas chromosomes 4 and 13 tend to be located at the nuclear periphery in a lymphoblastoid cell line (*44*), diploid lymphoblasts, and primary fibroblasts (*42*), which was recapitulated in our tonsil data (Fig. 2I and fig. S6A), indicating that these characteristics are likely conserved across human cell types. Radial score profiles are generally negatively correlated with A/B compartment score profiles (fig. S6B), which is consistent with the greater tendency of compartment B to associate with the nuclear lamina as compared with compartment A (*20*, *45*).

In addition to cell type–invariant features, radial score profiles also show cell type–specific differences (Fig. 2I and fig. S6A). Chromosome 17 showed the greatest mean radial score variations among different cell types (Fig. 2J) and a correlation analysis of radial scores among different cell types showed clustering of LZ B, DZ B, Tfh, and PC (Fig. 2K). Radial score clustering of LZ B, DZ B, and Tfh cells—GC-resident cell types that share BCL6 as a lineage-determining transcription factor (*37*, *38*)—suggests developmentally related similarities in genome architecture. Along with the B cell developmental trajectory, individual chromosomes exhibited distinct changes in their radial positioning (Fig. 2L). For example, chromosomes 17, 19, and 22 showed a further decrease in radial score during the transition from NBC to GCBC, followed by reversion toward higher radial scores as the cells progress to the MBC and PC states (Fig. 2L).

In addition, we analyzed other 3D genome features, including chromosome territory compaction (measured by the mean intrachromosomal interloci spatial distances) (figs. S6C and S7A), chromatin folding conformational heterogeneity (measured by the coefficient of variation (COV) of intrachromosomal interloci spatial distances) (figs. S6D and S7B), and demixing score (which reflects the degree of chromatin intermingling over long ranges) (fig. S6E). These features varied across cell types and exhibited distinct dynamic changes along the B cell developmental trajectory on different chromosomes (Fig. 2, M to O). For example, from NBC to GCBC, chromosome 2 exhibited an overall increase in intrachromosomal spatial distances, whereas chromosome 20 showed a decrease (Fig. 2M). Closer inspection of the log_2 fold change matrix of chromosome 2 revealed increased long-range interloci distances accompanied by decreased short-range interloci distances (fig. S7C). By contrast, chromosome 20 displayed more homogeneous changes across the distance matrix (fig. S7D). The COV matrix also showed distinct patterns of changes on different chromosomes (fig. S7, E and F).

In sum, tonsil cell types showed both cell type–specific and cell type–invariant 3D genome features that are partly associated with gene expression regulation.

SHM susceptibility is associated with 3D genome architectures

We previously showed that TADs delineate susceptibility to SHM through a genome-wide reporter assay in active chromatin regions in the human Burkitt lymphoma cell line Ramos, allowing the identification of a subset of TADs as particularly susceptible (hot) or resistant (cold) to SHM (*8*). We hypothesized that hot and cold TADs form distinct clusters or occupy specialized positions in the B cell nucleus, which could either prime them for or protect them from SHM. To test this, using our tonsil cell MINA data, we separately analyzed the spatial clustering of hot TADs and cold TADs, as annotated in our previous study (*8*), and filler loci, which represent regions that are not well-defined for SHM activity. We calculated a normalized spatial distance between each pair of TADs in each category to remove the genomic distance contribution to spatial distance (Materials and methods). We observed that cold TADs exhibited lower normalized distances between each other compared with hot TADs and filler loci in DZ and LZ B cells (Fig. 3, A to H), indicating a tendency of cold TADs to cluster. This characteristic was conserved across all major cell types examined, albeit with varying degrees of magnitude (Fig. 3, A to L, and figs. S8, A to T, and S9, A to L). We reasoned that the higher clustering tendency of cold TADs might be associated with a preferential distribution in the nuclear interior. Indeed, further analyses of radial scores of the hot, cold, and filler loci revealed that cold TADs had lower radial scores than hot TADs and filler loci in all major cell populations (Fig. 3, M to O, and figs. S8, U to Y, and S9, M to O). These observations were recapitulated in Ramos and in the diffuse large B cell lymphoma cell line TMD8 (fig. S9, P to Y).

Given that the SHM-associated 3D genome organization identified here is conserved in all cell types rather than occurring only in DZ B cells, we propose a model in which a preexisting large-scale 3D organization of TADs (particularly their nuclear radial positioning and clustering) contributes to their SHM susceptibility. Cold and hot TADs exhibited different dynamics in these 3D genome features along the B cell developmental trajectory (fig. S10A). During the transition from NBC to GCBC, cold TADs showed an overall decrease in radial scores as well as reduced distances to other cold TADs, whereas hot TADs showed no systematic changes in their radial scores and distances to other hot TADs (fig. S10, B and C). These changes of cold TADs in GCBC might further promote their resistance to SHM targeting.

At the compartment level, both cold and hot TADs were strongly enriched in the A compartment (compartment scores >0), as expected given that both were derived from active regions of the genome (*8*) and did not exhibit consistent differences between their compartment

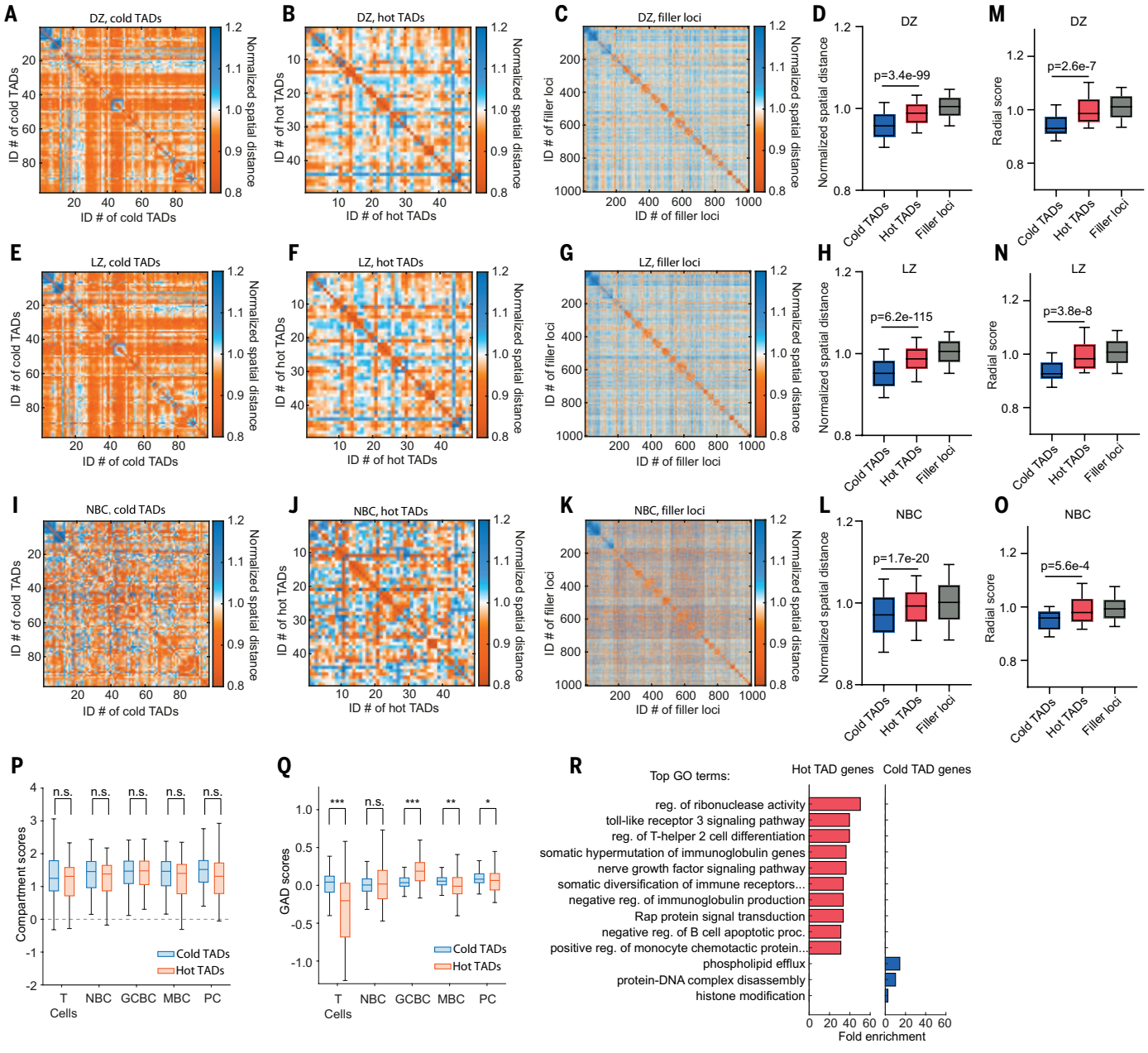


Fig. 3. SHM susceptibility is associated with 3D genome organization. (**A** to **C**) Matrix of normalized interloci distances between cold TADs, hot TADs and filler loci in DZ B cells; $n = 26{,}039$ cells. (**D**) Box plot of normalized distances in (A) to (C). (**E** to **G**) Matrix of normalized interloci distances between cold TADs, hot TADs, and filler loci in LZ B cells; $n = 24{,}136$ cells. (**H**) Box plot of normalized distances in (E) to (G). (**I** to **K**) Matrix of normalized interloci distances between cold TADs, hot TADs, and filler loci in naïve B cells; $n = 11{,}183$ cells. (**L**) Box plot of normalized distances in (I) to (K). (**M** to **Q**) Box plot of radial scores of cold TADs, hot TADs, and filler loci in DZ B cells (M), LZ B cells (N), and naïve B cells (O). In (D), (H), and (L) to (O), boxes show the IQR, central lines represent the medians, and whiskers denote the 10th and 90th percentiles. (P) Box plot of compartment scores of cold and hot TADs in different groups of cells. (Q) Box plot of GAD scores of genes in cold and hot TADs in different groups of cells. In (P) to (Q), boxes show the IQR, central lines represent the medians, and whiskers denote the maximum and minimum values up to 1.5 × IQR away from the box. (**R**) Top GO terms of genes located in hot TADs (left) and cold TADs (right). *P* values in all box plots were calculated by two-sided Wilcoxon rank-sum tests. Levels of significance in (P) and (Q) were defined as $P > 0.05$ (n.s.), $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Data from five biological replicates were merged.

scores (Fig. 3P). In GCBC, hot TADs exhibited higher GAD scores than cold TADs, whereas the opposite trend was observed in T cells (Fig. 3Q). Gene ontology (GO) analysis showed that genes in hot TADs are strongly associated with the GC reaction and B cell development, including "somatic hypermutation of immunoglobulin genes," whereas few specific GO terms were enriched among cold TAD genes (Fig. 3R). These observations indicated that genes in hot TADs gain more chromatin interactions specifically in SHM-active cells, which is consistent with our previous finding that the cohesin loader NIPBL is strongly enriched in hot versus cold TADs (*8*) and with previous studies linking SHM susceptibility to highly active, interconnected chromatin (*12*, *13*). We therefore set out to test the hypothesis that cohesin, a major driver of 3D-chromatin organization, plays an important function in SHM.

Degradation of RAD21 disrupts SHM

To allow for rapid inactivation of cohesin function, we engineered a degradation tag (dTAG) at the C terminus of core cohesin component RAD21 (Fig. 4A) in our previously described rapid activation of SHM (RASH)–1 cell line system (*46*). This Ramos-based cell line supports rapid, efficient SHM at the variable regions of the Ig heavy chain gene (*IGH-V*) and Ig lambda light chain gene (*IGL-V*) and at a green fluorescent protein (GFP)–based SHM reporter (*GFP reporter*) integrated in the *IGH* locus, upon doxycycline (dox) induction of AID7.3, a hyperactive form of AID (*46*, *47*). The degron tag did not substantially perturb RAD21 protein levels (fig. S11A). RAD21 chromatin binding, as assessed by Cleavage Under Target & Release Using Nuclease (CUT&RUN) using both RAD21-specific and hemagglutinin (HA) epitope (a component of the degron tag)–specific antibodies, was strongly enriched in hot versus cold TADs, as was the active enhancer mark H3K27ac (Fig. 4B), strengthening the association of cohesin with SHM susceptibility. Treatment with degrader compound dTAGV-1 resulted in rapid, efficient RAD21 degradation as measured by Western blot and flow cytometry (Fig. 4A and fig. S11B). Degradation of RAD21 for 20 hours did not decrease AID protein expression compared with cells treated with the dTAGV-1 negative control compound (NEG) (Fig. 4C). EdU (5-ethynyl-2'-deoxyuridine) incorporation in the cell cycle remained robust for at least 12 hours after RAD21 degradation (fig. S11C), while cell viability was unaffected for at least 16 hours (fig. S11D).

To assess the functional consequences of RAD21 loss on SHM, RAD21 was degraded for 20 hours with AID induced for the last 14 hours, and mutation accumulation was measured by high-throughput sequencing (Mut-Seq). RAD21 degradation reduced mutation accumulation at *GFP reporter*, *IGH-V*, and *IGL-V* nearly to the background levels defined by the no-dox (no AID) controls (Fig. 4D). Time course experiments, with dox added 2 hours before addition of NEG or dTAGV-1, demonstrated a steady increase in mutations in control (NEG-treated) cells after 6 hours (8 hours of AID induction) that were almost completely abrogated in cells lacking RAD21 (Fig. 4E). Meta-analysis of RAD21 CUT&RUN data confirmed efficient, genome-wide depletion of RAD21 from chromatin after 6 hours of degradation (Fig. 4F). We conclude that loss of RAD21 severely compromises SHM.

To determine whether RAD21 also supports "off-target" SHM of non-Ig loci, we created a sensitized system in which the DNA repair factor uracil-DNA glycosylase (UNG) was knocked out in RAD21 degron RASH-1 cells (fig. S11E). Loss of UNG allows a larger fraction of AID-mediated deamination events to be detected as mutations (*48*). In the absence of UNG, mutations were detected above background levels at AID target loci *BCL6* and *POU2AF1* in NEG-treated cells but not in dTAGV-1–treated cells (fig. S11, F and G). The low mutation frequency observed is attributable to the short AID induction period (to ensure cell viability) combined with the inherent inefficiency of SHM at non-Ig loci. These results showed that loss of RAD21 strongly compromises SHM at both Ig and non-Ig loci.

RAD21 regulates chromatin architecture, transcription, and AID recruitment associated with SHM

Loss of RAD21 and cohesin function could compromise SHM by multiple mechanisms, including effects on chromatin architecture, transcriptional dynamics, and access of AID to its targets. To determine how acute loss of RAD21 affects chromatin architecture at *IGH* and *IGL*, we applied chromatin tracing (Fig. 4, G and H) and chromosome conformation capture (3C)–high-throughput genome-wide translocation sequencing (3C-HTGTS) (Fig. 4, I and J)—a method that incorporates a linear amplification step for high-resolution, quantitative assessment of looping interactions from a bait site (*49*)—to RAD21-degron RASH-1 cells. In the chromatin tracing analysis of *IGH*, many 3D chromatin features, particularly in the region encompassed by the two *IGH* super-enhancers (*3′RR1* and *3′RR2*), exhibited a gradual decline in intensity, remaining readily detectable 12 hours after loss of RAD21 (Fig. 4G and fig. S12, A and B). Interactions between *IGH-V* and the super-enhancers declined noticeably during this time period (Fig. 4G and fig. S12B, yellow box). Consistent with this time course, 3C-HTGTS revealed focal interactions of the *IGH-V* bait site with *3′RR1* and *3′RR2* that declined by ~40% by 12 hours and precipitously 24 hours after loss of RAD21 (Fig. 4I and fig. S12D). Thus, looping interactions involving the *IGH* super-enhancers and *V* region declined but were not lost in the absence of RAD21 during the 6-to-18-hour time period when SHM was abrogated. Many other interactions in the traced region were rapidly lost. For example, a "stripe" of interactions that emanated from the vicinity of the cluster of CTCF binding elements (CBEs) flanking 3′RR2 (*CBE-2*) was no longer detectable 2 hours after RAD21 degradation (Fig. 4G and fig. S12A, magenta box), suggesting that cohesin function was required for its maintenance.

At *IGL*, chromatin tracing revealed a similar loss of interactions in the region involving the *IGL* super-enhancer (*IGL-E*) and *IGL-V* after RAD21 degradation (Fig. 4H and fig. S12C, yellow box). Similarly, 3C-HTGTS showed reduced interactions involving the bait site at *IGL-E* and other portions of the locus, including a ~50% decrease with *IGL-V*, at 2 and 6 hours after loss of RAD21, although because of the relatively short distance (~30 kb) between *IGL-E* and *IGL-V*, a focal interaction was more difficult to appreciate (Fig. 4J and fig. S12E). We concluded that the 3D chromatin architecture of *IGH* and *IGL* was sensitive to the loss of RAD21 and that a complete loss of enhancer-promoter interactions was not the sole explanation for the rapid cessation of SHM after RAD21 degradation.

To determine how loss of RAD21 affects transcription, we profiled RNA polymerase II (Pol II) binding and nascent transcription after RAD21 degradation. After 6 hours of RAD21 degradation, we observed modest but reproducible decreases in levels of RPB1 (total Pol II), Ser5-Pol II, and Ser2-Pol II (Pol II phosphorylated at serine 5 and serine 2, respectively, of its C-terminal domain repeats) genome-wide and at AID targets (Fig. 5, A to C, and fig. S13, A and B). These decreases became much more pronounced at the 18-hour time point (fig. S13, C to G), likely reflecting a progressive destabilization of chromatin architecture, long-range enhancer-promoter interactions, and Pol II transcriptional dynamics (*50*–*52*). TT-TimeLapse-seq (*53*) with a 5-min pulse of 4-thiouridine revealed that transcriptional activity was largely preserved 2 hours after loss of RAD21 (Fig. 5D) and declined to variable extents, but typically less than 2-fold, at 6 hours (Fig. 5, D and E), with little or some further decrease seen at 12 hours (Fig. 5, B to E, and fig. S13, A and B). Substantial transcriptional activity was retained at *IGH* and *IGL* over the 12-hour time course with *IGH* better preserved than *IGL* (Fig. 5E).

These results demonstrated that acute RAD21 degradation does not immediately abrogate transcription—which is consistent with previous studies (*52*, *54*)—and that loss of transcription elongation is not the primary cause of the abrogation of SHM seen after RAD21 degradation. Using an optimized AID CUT&RUN protocol (*47*), we found that loss of RAD21 for 6 hours reduced AID binding modestly (typically twofold or less), whereas after 18 hours, AID binding was severely reduced, sometimes nearly to background levels defined by the no-dox control; this was the case at *IGH* and *IGL* (Fig. 5, F to J) and non-Ig AID binding loci (Fig. 5, J to L, and fig. S14, A to D). We conclude that RAD21 is broadly required for efficient AID binding at Ig and non-Ig loci and that cohesin likely exerts its essential function in SHM by supporting the 3D architectural and transcriptional landscape required for sustained AID recruitment and activity.

"Curl" structure identified at the *IGH* 3′RR super-enhancers

One possible explanation for the substantial resistance of 3D chromatin architecture at portions of the *IGH* locus to acute RAD21 loss is the presence of two large super-enhancers in *IGH*, which might provide structural and functional stability. Consistent with this idea, the *IGH* chromatin tracing spatial distance matrix exhibited a distinctive "curl"

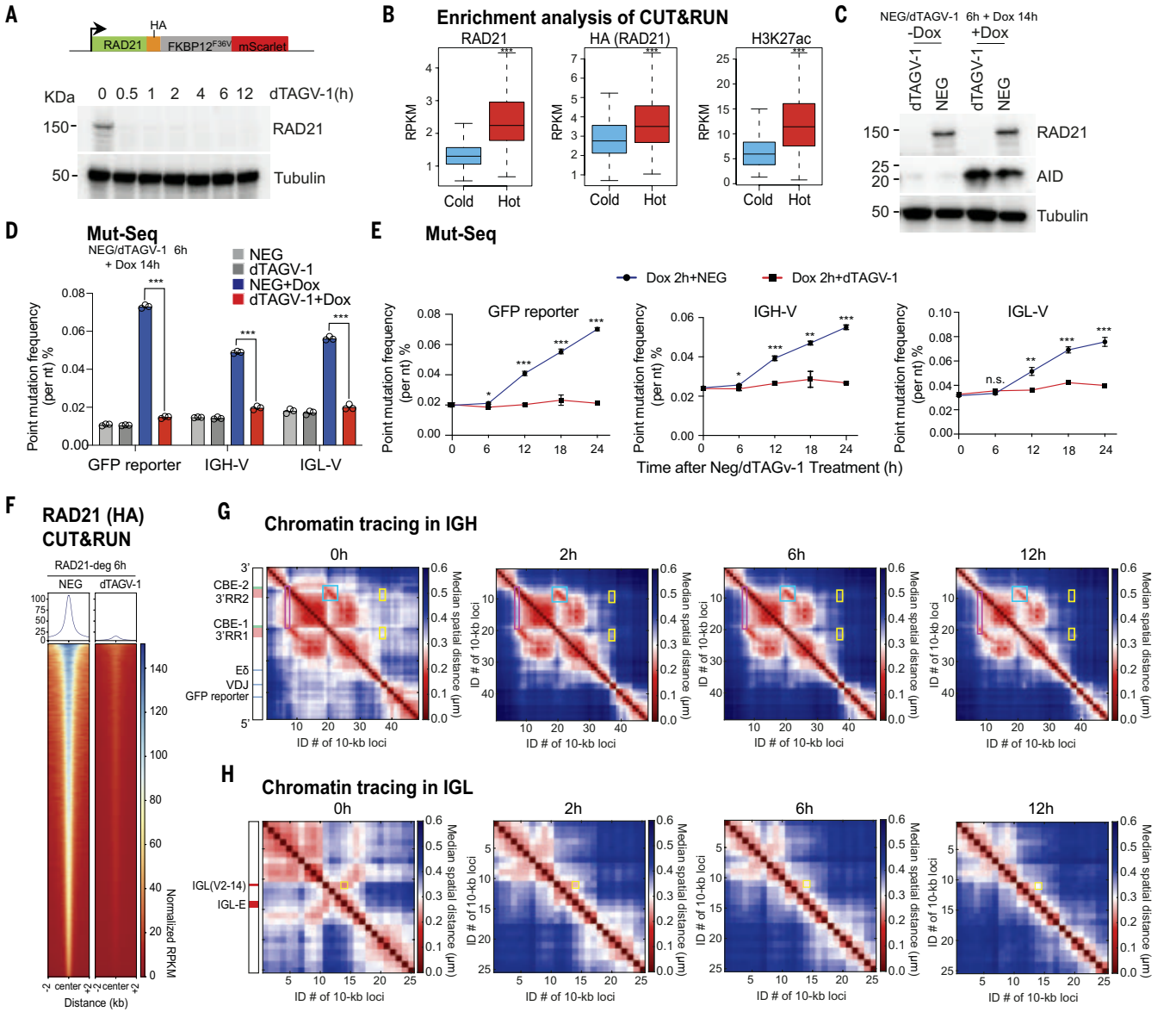

A
HA
RAD21
FKBP12F36V
mScarlet
KDa
0 0.5 1 2 4 6 12 dTAGV-1(h)
150
RAD21
50
Tubulin
B
Enrichment analysis of CUT&RUN
RAD21
HA (RAD21)
H3K27ac
RPKM
Cold
Hot
C
NEG/dTAGV-1 6h + Dox 14h
-Dox
+Dox
dTAGV-1
NEG
150
RAD21
25
20
AID
50
Tubulin
D
Mut-Seq
NEG/dTAGV-1 6h + Dox 14h
NEG
dTAGV-1
NEG+Dox
dTAGV-1+Dox
Point mutation frequency (per nt) %
GFP reporter
IGH-V
IGL-V
E
Mut-Seq
Dox 2h+NEG
Dox 2h+dTAGV-1
GFP reporter
IGH-V
IGL-V
n.s.
Time after Neg/dTAGv-1 Treatment (h)
F
RAD21 (HA) CUT&RUN
RAD21-deg 6h
NEG
dTAGV-1
Normalized RPKM
Distance (kb)
G
Chromatin tracing in IGH
0h
2h
6h
12h
3'
CBE-2
3'RR2
CBE-1
3'RR1
Eδ
VDJ
GFP reporter
5'
ID # of 10-kb loci
Median spatial distance (μm)
H
Chromatin tracing in IGL
IGL(V2-14)
IGL-E
I
3C-HTGTS in IGH
CBE-2
3'RR2
CBE-1
3'RR1
Eδ
IGH
GFP reporter
Bait Site
50 kb
H3K4me3
H3K27ac
HA-RAD21 6h
NEG
dTAGV-1
R1
R2
6h
12h
18h
24h
J
3C-HTGTS in IGL
IGL(V2-14)
IGL-E
Bait Site
50 kb
2h
6h

Fig. 4. RAD21 is essential for SHM. (**A**) Time course of degradation of RAD21 in RAD21-degron RASH-1C cells upon dTAGV-1 treatment. (**B**) Box plot showing enrichment of RAD21, HA-RAD21, and H3K27ac in hot and cold TADs. Boxes show the IQR, central lines represent the medians, and whiskers denote the full range of the distribution. (**C**) Western blot showing protein levels of AID and RAD21 in RAD21-degron RASH-1C cells with NEG or dTAGV-1 treatment for 6 hours followed by 14 hours doxycycline (dox) treatment (together with NEG or dTAGV-1). (**D**) Point mutation frequencies in *HTS7* region of *GFP reporter*, *IGH-V*, and *IGL-V* from RAD21-degron RASH-1C cells treated with NEG or dTAGV-1 for 6 hours followed by 14 hours dox treatment (together with NEG or dTAGV-1). Each dot represents an individual replicate. Bars indicate the mean values. (**E**) Point mutation frequencies in *HTS7* region of *GFP reporter*, *IGH-V*, and *IGL-V* from RAD21-degron RASH-1C cells treated with dox for 2 hours followed by NEG or dTAGV-1 treatment for 6, 12, 18, or 24 hours (together with dox). Mean ± SD from three biological replicates are shown. (**F**) HA-RAD21 CUT&RUN profiles generated from RAD21-degron cells treated with dox for 2 hours followed by NEG or dTAGV-1 treatment for 6 hours (together with dox). Heatmaps display signals across genomic regions spanning ±2 kb from the centers of RAD21 peaks. Two biological replicates were merged for the plots. (**G** and **H**) Chromatin tracing results in *IGH* (G) and *IGL* (H). The interactions between 3′RRs (cyan box), stripe stemming from CBE-2 (magenta box), and *IGH-V* interaction with 3′RRs (yellow boxes) are highlighted in (G). The interaction between *IGL-V* and *IGL-E* is highlighted (yellow box) in (H). Data from two biological replicates were merged. $n = 7256$, 11,850, 12,006, and 4977 cells for 0, 2, 6, and 12 hours dTAGV-1treatment. (**I**) Representative 3C-HTGTS interaction profiles at the *IGH* locus in RAD21-degron RASH-1C cells treated with NEG or dTAGV-1 for 6, 12, 18, or 24 hours. The bait site is indicated by a green arrow. Two independent datasets are shown from libraries normalized to 1 million reads. See data S7 for total junction reads. (**J**) Representative 3C-HTGTS interaction profiles at the *IGL* locus in RAD21-degron RASH-1C cells treated with NEG or dTAGV-1 for 2 and 6 hours. The bait site is indicated by a green arrow. Two independent datasets are shown from libraries normalized to 1 million reads. See data S7 for total junction reads. *P* values in (B), (D), and (E) were calculated by two-sided Student's *t* tests. Levels of significance were defined as $P > 0.05$ (n.s.), $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

feature, involving bins 7 to 10 (spanning *3′RR2*) and bins 19 to 22 (spanning *3′RR1*), that manifested as a line of decreased distances running parallel to the diagonal (Fig. 4G, cyan box, and fig. S12A). This feature suggested a preferential parallel alignment of *3′RR1* and *3′RR2* (fig. S15A), an arrangement that could be observed in individual chromatin traces (fig. S15B). A meta-analysis quantitating the angles between the trajectories of the two interacting regions (fig. S15C; 0° angle = parallel; 180° angle = antiparallel) revealed a strong enrichment of small angles, whereas control pairs from the same traces did not show any directional preference (fig. S15D).

3′RR1 and *3′RR2* arose from a primate-specific genome duplication event (*55*) and have substantial sequence similarity. To rule out the possibility that the curl feature arises from probe cross-hybridization between the two regions, we designed a dedicated probe library targeting these eight bins using more stringent probe design criteria and a competitive hybridization strategy (Materials and methods). The curl structure was recapitulated with the new probe library in the median spatial distance matrix, in individual traces, and through analysis of trajectory angles (fig. S15, E to G). RAD21 degradation led to a gradual decline in signal for the curl feature (fig. S15E), further arguing that this structure is not caused by probe cross-hybridization.

Hi-C contact matrices from Ramos and lymphoblastoid cell line GM12878 have signals consistent with a close parallel juxtaposition of *3′RR1* and *3′RR2* (fig. S16, A and B), providing evidence for the curl feature with an independent methodology. Parallel alignment of *3′RR1* and *3′RR2* predicts that a 3C-HTGTS bait positioned adjacent to one 3′RR region should detect a signal gradient of defined directionality across the other 3′RR region. This prediction was tested using a 3C-HTGTS bait primer adjacent to bin 19 of *3′RR1*, revealing a gradient of signal that, as predicted, declined from bin 7 to bin 10 across *3′RR2* (fig. S16C). Among bins 7 to 10, bin 7 has the longest genomic distance to the bait site, thus the declining trend of interactions cannot be explained by regions closer on the genomic map tending to make more contacts. We conclude that *3′RR1* and *3′RR2* have a propensity to align in a distinctive parallel configuration proximal to one another. While the functional relevance of this *IGH* super-enhancer curl structure and the mechanisms underlying its formation and maintenance remain to be determined, close alignment of *3′RR1* and *3′RR2* together with recently reported three-way interactions between them and *IGH-V* (*9*) could contribute to the substantial resistance of *IGH* 3D architectural structure to loss of RAD21 (Fig. 4, G and I).

Discussion

In this work, we generated a single-cell 3D genome atlas of human tonsil by integrating Droplet Hi-C with an image-based 3D genome and spatial transcriptome toolkit for clinical tissue. Using the atlas, we characterized cell type–specific and cell type–invariant 3D genome features as well as their progression during B cell differentiation. SHM mistargeting is strongly associated with B cell lymphomagenesis (*2*–*7*); however, it remains unclear how mistargeting occurs and why certain locations in the genome are selected. We found that regions of the genome that are located in the interior of the nucleus are more protected from SHM. AID predominantly (>90%) localizes in the cytoplasm and actively shuttles between the cytoplasm and nucleus through nuclear pores (*16*), and a more interior nuclear location places cold TADs farther from nuclear pores. Among TADs in compartment A, those with a more interior positioning in the nucleus might be better protected from SHM activity. This observation is particularly notable given the dynamics of AID subcellular localization during the cell cycle. Nuclear envelope reformation at the beginning of G1 traps substantial levels of AID in the nucleus, which is then rapidly (in 40 to 60 min) exported to the cytoplasm (*56*). Notably, the preponderance of AID-mediated deamination events detected at an *IGH* switch region occurred during this nuclear pore–mediated AID export process (*56*), arguing that a greater distance from nuclear pores is likely to be protective during the genome's time window of greatest vulnerability. We found that the radial positioning of hot and cold TADs in the nucleus is preset in a largely cell type–independent fashion before cells undergo SHM and that the cold TAD positioning is further enhanced in SHM-active germinal center B cells. SHM susceptibility has also been linked to convergent transcription, divergent transcription, super-enhancers, and high levels of looping interactions (*8*, *12*, *13*, *57*). Furthermore, super-enhancers, which are enriched in hot TADs (*8*), can localize to and interact physically with nuclear pores (*58*–*60*). Radial positioning and these other 3D genome and transcriptional features likely collectively contribute to SHM susceptibility and resistance by controlling local AID concentration and the transcriptional environment for AID function. As a result, genes in hot TADs encoding important regulators of GC B cell development and physiology, such as *BCL6*, *PAX5*, and *MYC*, are more prone to the action of AID, which can lead to genome instability, translocations, and the development of B cell lymphoma (*61*, *62*). Because TADs delineate susceptibility to SHM (*8*), the genes that locate in the same TAD are also prone to mutation, such as *LPP* (in *BCL6* TAD) and *ZCCHC7* and *GRHPR* (in *PAX5* TAD), which have also been reported to be associated with B cell lymphoma (*63*–*66*).

Chromatin loop extrusion plays a key role in targeting and orchestrating class switch recombination (CSR), another AID-dependent reaction (*67*, *68*), and cohesin components RAD21 and NIPBL (*8*) are enriched in hot versus cold TADs. A role for cohesin and loop extrusion in SHM had not previously been established. Our findings demonstrate that RAD21 is required for SHM and that its loss interferes with looping architecture, transcription, and AID binding at *IGH-V* and *IGL-V* SHM targets. The kinetics of these perturbations after RAD21 loss

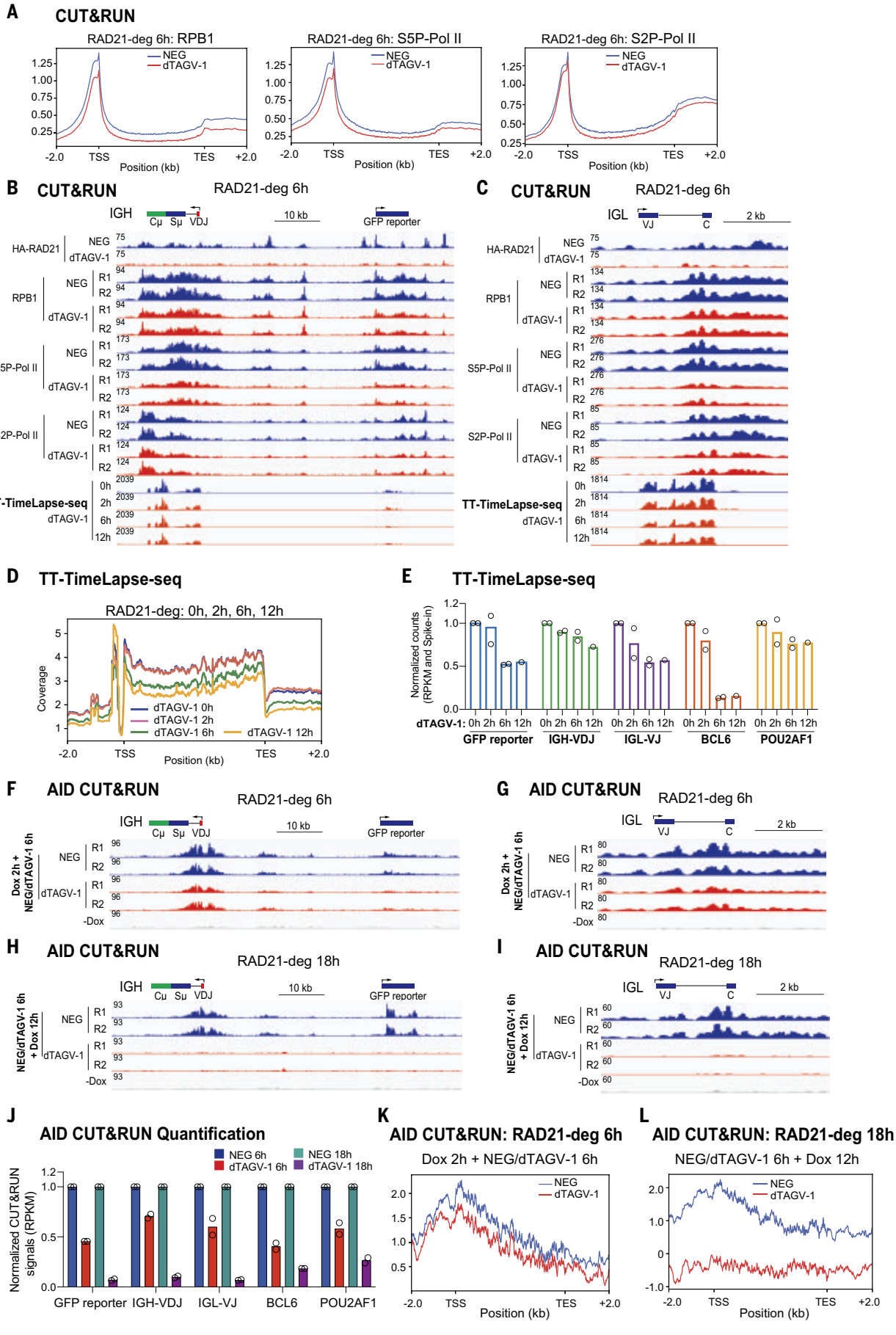

A
CUT&RUN
RAD21-deg 6h: RPB1
RAD21-deg 6h: S5P-Pol II
RAD21-deg 6h: S2P-Pol II
NEG
dTAGV-1
Position (kb)
B
CUT&RUN
RAD21-deg 6h
IGH
Cμ Sμ VDJ
10 kb
GFP reporter
HA-RAD21
RPB1
S5P-Pol II
S2P-Pol II
TT-TimeLapse-seq
C
IGL
VJ
C
2 kb
D
TT-TimeLapse-seq
RAD21-deg: 0h, 2h, 6h, 12h
Coverage
dTAGV-1 0h
dTAGV-1 2h
dTAGV-1 6h
dTAGV-1 12h
E
TT-TimeLapse-seq
Normalized counts (RPKM and Spike-in)
dTAGV-1: 0h 2h 6h 12h
GFP reporter
IGH-VDJ
IGL-VJ
BCL6
POU2AF1
F
AID CUT&RUN
RAD21-deg 6h
Dox 2h + NEG/dTAGV-1 6h
-Dox
G
AID CUT&RUN
RAD21-deg 6h
H
AID CUT&RUN
RAD21-deg 18h
NEG/dTAGV-1 6h + Dox 12h
I
AID CUT&RUN
RAD21-deg 18h
J
AID CUT&RUN Quantification
NEG 6h
dTAGV-1 6h
NEG 18h
dTAGV-1 18h
Normalized CUT&RUN signals (RPKM)
K
AID CUT&RUN: RAD21-deg 6h
Dox 2h + NEG/dTAGV-1 6h
L
AID CUT&RUN: RAD21-deg 18h
NEG/dTAGV-1 6h + Dox 12h

Fig. 5. RAD21 degradation affects transcription and AID recruitment associated with SHM (**A**) CUT&RUN metagene profiles of RPB1, S5P-Pol II, and S2P-Pol II in RAD21-degron cells treated with dox for 2 hours followed by NEG or dTAGV-1 treatment for 6 hours (together with dox). Two biological replicates were merged for the profiles. TSS, transcription start site. TES, transcription end site. (**B** and **C**) Genome browser tracks showing HA-RAD21, RPB1, S5P-Pol II, S2P-Pol II CUT&RUN, and TT-TimeLapse-seq signals in *IGH-VDJ*, *GFP reporter* regions (B), and *IGL-VJ* region (C). For TT-TimeLapse-seq, tracks at 0, 2, and 6 hours after dTAGV-1 treatment represent merged data from two biological replicates, whereas the 12-hour time point is shown from a single biological replicate. (**D**) Metagene profile of TT-TimeLapse-seq in RAD21-degron RASH-1C cells treated with dox for 2 hours followed by 0, 2, 6, and 12 hours dTAGV-1 treatment (together with dox). Profiles at 0, 2, and 6 hours represent merged data from two biological replicates, whereas the 12-hour time point is shown from a single biological replicate. (**E**) Normalized counts of indicated genes by TT-TimeLapse-seq. Each dot represents an individual replicate. Bars indicate the mean values. (**F** and **G**) Genome browser tracks showing AID CUT&RUN at the *IGH* (F) and *IGL* (G) loci for RAD21-degron RASH-1C cells treated with dox for 2 hours followed by NEG or dTAGV-1 treatment for 6 hours (together with Dox). Cells treated with NEG but without dox served as the negative control. (**H** and **I**) Genome browser tracks showing AID CUT&RUN at the *IGH* (H) and *IGL* (I) loci for RAD21-degron RASH-1C cells treated with NEG or dTAGV-1 for 6 hours followed by dox treatment for 12 hours (together with NEG or dTAGV-1). Cells treated with NEG but without dox served as the negative control. (**J**) AID CUT&RUN signal quantified over indicated genomic regions in RAD21-degron RASH-1C cells treated with NEG or dTAGV-1. Signal was calculated as reads per kilobase per million mapped reads (RPKM) within defined gene regions. Each dot represents an individual replicate, and bars indicate mean values. (**K**) CUT&RUN metagene profiles of AID in RAD21-degron cells treated with dox for 2 hours followed by NEG or dTAGV-1 treatment for 6 hours (together with dox). Two biological replicates were merged for the profiles. (**L**) CUT&RUN metagene profiles of AID in RAD21-degron cells treated with NEG or dTAGV-1 for 6 hours followed by 12 hours dox treatment (together with NEG or dTAGV-1). Two biological replicates were merged for the profiles.

suggest a multifaceted role for RAD21 in SHM. Although SHM is abolished as early as it can be detected, 6 to 12 hours after RAD21 degradation, looping interactions and transcription remain substantial, particularly for *IGH-V*, for 12 to 18 hours. AID binding is decreased less than twofold at 6 hours and is greatly diminished at Ig and non-Ig regions by 18 hours. Cohesin is therefore required to maintain long-term AID binding to its SHM targets, in keeping with a prior study that implicated a physical association of AID with cohesin in CSR (*67*). Together, our findings argue that a cohesin-dependent chromatin architecture establishes a permissive framework for SHM and that the role of RAD21 in SHM is not merely to facilitate promoter-enhancer interactions, with loss of SHM occurring before loss of promoter-enhancer looping and before abrogation of transcription. Cohesin functions by sustaining the transcriptional environment and AID targeting, as well as the chromatin architecture, including but not limited to promoter-enhancer looping, which collectively contribute to SHM. Additional mechanistic roles for cohesin in SHM can be envisioned given recent findings implicating factors that influence Pol II elongation and stalling in the transcription unit in SHM (*47*, *69*–*71*) and identifying functional interactions between cohesin and Pol II looping activity (*51*). Cohesin might also have undiscovered functions that contribute to SHM.

Materials and methods

Probe design and synthesis

Template probe design: In genome-wide chromatin tracing, we selected our targets from three categories: (i) hot TADs and (ii) cold TADs identified in our previous publication (*8*); (iii) space filler loci to span each chromosome. Finally, 49 hot TADs, 97 cold TADs, and 1005 filler TADs were targeted in our library. These 1151 targets were separated into two sublibraries. Targets in each sublibrary were assigned a unique 50-bit Hamming Distance 2 (HD2) binary code with Hamming Weight 2. For each target, 400 primary probes were designed to cover the central 100-kb region or the entire target if it was shorter than 100 kb, using the following criteria: (i) probe length of 40 nucleotides (nt); (ii) target sequence melting temperatures ranging from 76° to 100°C; (iii) melting temperatures of potential secondary structures below 76°C; (iv) GC content between 20 and 90%; and (v) no consecutive repeats of four C's or G's or six A's or T's. The targeting sequences were further blasted against (i) repetitive sequence database to exclude strong binding to repetitive genomic regions; (ii) nonrepetitive genome to make sure unique binding in the genome; (iii) unspliced transcriptome to be compatible with RNA MERFISH experiments by avoiding transcript binding. Finally, the template probes were assembled with six regions (from 5′ to 3′): (i) a 20-nt forward priming region; (ii) a 20-nt adapter binding region; (iii) a 40-nt genome targeting region; (iv) a second 20-nt adapter binding region; (v) a third 20-nt adapter binding region; and (vi) a 20-nt reverse priming region. Each template probe is 140 nt in length. Template probes were ordered as oligo pools from Twist Bioscience. The sequences of the ordered probes are listed in data S1.

In RNA MERFISH probe design, we selected our targets from four categories: (i) cell type markers selected from cell types' differentially expressed genes by analyzing a published scRNA-seq data (*28*); (ii) cell type markers further identified with NS-Forest (*72*); (iii) cell cycle markers identified in a previous study (*73*) and (iv) hand-picked immune related genes. Finally, 456 genes were selected: 447 of them were targeted in MERFISH library and 9 of them (too short or abundant in cell) were targeted in smFISH library by sequential readout. 30-nt targeting sequences of MERFISH and smFISH probes were designed with the following criteria: (i) target sequence melting temperatures ranging from 70° to 100°C; (ii) melting temperatures of potential secondary structures below 76°C; (iii) GC content between 30 and 90%; and (iv) no consecutive repeats of four C's or G's or six A's or T's. Each MERFISH target was assigned a unique 32-bit Hamming Distance 4 (HD4) binary code with Hamming Weight 4. Finally, the template probe was assembled and ordered the same way as genome-wide chromatin tracing. Each template probe is 130 nt in length. The sequences of the ordered probes are listed in data S2.

In the fine scale chromatin tracing library design, we designed two libraries targeting the IGH and IGL loci. The IGH library covered 48 consecutive 10-kb segments spanning chr14:105,496,000 to 105,976,000 (modified hg38 for Ramos). The IGL library covered 25 consecutive 10-kb segments spanning chr22:22,657,531 to 22,907,531 (modified hg38 for Ramos). For each 10-kb segment, we designed 200 probes to ensure robust signal detection and spatial resolution. 30-nt targeting sequences were designed with following criteria: (i) target sequence melting temperatures ranging from 66° to 100°C; (ii) melting temperatures of potential secondary structures below 76°C; (iii) GC content between 30% and 90%; and (iv) no consecutive repeats of six A's, T's, C's or G's. A unique adapter binding sequence was assigned to the probes targeting each segment. Finally, the template probe was assembled and ordered the same way as above (with the same adapter binding sequence repeated three times in each template probe). Each template probe is 130-nt in length. The sequences of the ordered probes are listed in data S3.

High stringent probe design for bins 7 to 10 (spanning *3′RR2*) and bins 19 to 22 (spanning *3′RR1*) of IGH tracing library: First, we identified all potential 30-nt targeting sequences based on the criteria described above. Specifically, each of the eight 10-kb bins was scanned using a 1-nt sliding window, and all candidate probes meeting the design criteria were retained. Candidate probes were then evaluated for specificity using BLAST with the "-task blastn-short" setting against the entire IGH tracing region. Probe selection was performed in two rounds. In the first round, probes with no off-target alignments

longer than 23 nt were classified as high-specificity probes and directly retained. In the second round, we selected probes with no off-target alignments longer than 27 nt. For each such probe, we required a competing probe that perfectly matched the corresponding homologous sequence in the other *3′RR* region to be retained together. This competitive hybridization strategy has been previously demonstrated in SNP FISH (*74*). We further required all selected probes overlap by no more than 10 nt. Finally, we had more than 100 probes designed for each 10-kb bin (data S3).

Primary probe synthesis: To amplify the oligo pools, we followed a previously established experimental workflow (*20*, *75*). Briefly, this process included limited-cycle PCR, in vitro transcription, reverse transcription, alkaline hydrolysis, and oligo purification. The limited-cycle PCR primers and reverse transcription primers were obtained from IDT, with detailed sequences provided in data S4.

Human material

De-identified fresh human tonsil tissues were obtained from the Yale Pathology Tissue Service (YPTS)-Tissue Distribution and Analysis (TDA) on the basis of Yale Human Investigation Committee protocol no. 0304025173, which allows retrieval of tonsil tissues from surgical pathology that was consented or has been approved for use with waiver of consent. Patient age and sex are provided in data S5.

MINA experiments in human tonsil

MINA procedure with genome-wide chromatin tracing: Collected tonsil tissue was promptly embedded in O.C.T. compound, and rapidly frozen using dry ice. Frozen tissue blocks were stored at −80°C until use. To prepare MINA samples, 40-mm coverslips (Bioptechs; 40-1313-0319) were first pretreated with a 1:1 mixture of 34 to 37% (v/v) hydrochloric acid (HCl, Sigma-Aldrich; HX0607-1) and methanol (Sigma-Aldrich; 179337-4L-PB), followed by poly-D-lysine solution (Millipore, A-005-C) treatment at room temperature (RT, 22°C). Next, tissue blocks were equilibrated in a cryostat (Leica CM3050S) at −20°C for 20 min and then sectioned into 8-μm slices at −20°C. Tissue slices were attached onto pretreated coverslips followed by 4% PFA fixation (Electron Microscopy Sciences; 15710) for 30 min at RT. Next, the samples were washed with Dulbecco's Phosphate Buffered Saline (DPBS, Sigma-Aldrich; D8537-500ML) three times for 3 min each time. Samples were permeabilized with 70% ethanol at 4°C for one overnight. Next day, the samples were first washed with DPBS for two times for 3 min each time, and then further permeabilized with 0.5% Triton X-100 solution (Sigma-Aldrich; 93443) in DPBS supplemented with 0.1% (v/v) murine RNase inhibitor (MRI) (New England Biolabs, M0314L) for 20 min. After permeabilization, the samples were briefly washed with DPBS and incubated with blocking buffer [1% w/v bovine serum albumin (Sigma-Aldrich, A9647), 0.3% Triton X-100, 0.1% (v/v) MRI in DPBS] for 30 min at RT. Next, samples were incubated with 1:100 (v/v) diluted CD35 antibody (Invitrogen, A51014; RRID: AB_2633298) in blocking buffer with 1% (v/v) MRI for 1 hour (h) at RT. Next, samples were washed three times with 0.1% (v/v) Tween-20 in DPBS, 5 min each. CD35 antibody signal was postfixed with 1% PFA for 5 min at RT, followed with DPBS wash three times, 3 min each. Then, samples were incubated with 0.1 M HCl for 5 min at RT. Next, samples were washed twice with DPBS and once with 2× SSC buffer, prepared by diluting 20× SSC buffer (Invitrogen; 15557-044) with double-distilled water (ddH2O), 3 min each. Samples were then incubated with 2.5 ml prehybridization buffer, composed of 50% formamide (Sigma-Aldrich; F7503-1L), 0.1% (v/v) MRI in 2× SSC, for 30 min at RT. Then, heat denature was performed by incubating in a 90°C water bath in prehybridization buffer for 6 min, followed by a quick rinse with 2× SSC at RT. Next, samples were incubated with 25 μl hybridization buffer, composed of 10% (w/v) dextran sulfate (Sigma-Aldrich, D8906-50G), 0.1 mg/ml yeast tRNA (Invitrogen, AM7119), 50% formamide, 1% (v/v) MRI, 0.04 nM per probe of genome-wide chromatin tracing probe library, 0.3 nM per probe of RNA MERFISH probe library, 1 nM per probe of smFISH probe library, 2 μM per probe of fiducial marker oligos in 2× SSC. The samples were covered with a small piece of parafilm and incubated overnight at 47°C, followed by another overnight incubation at 37°C. Next, samples were washed with 0.1% (v/v) Tween-20 in 2× SSC for two times at 60°C and one time at RT, 15 min each. The samples were rinsed with 2× SSC and ready for imaging. Five biological replicates from three different donors were performed.

Chromatin tracing in B cell line

B cell culture: Human Ramos (RRID: CVCL_0597), TMD8 (RRID: CVCL_A442) and cells derived from Ramos cells were cultured at 37°C with 5% CO_2 in RPMI-1640 medium (Gibco 11875) supplemented with 10% fetal bovine serum (Gemini) and 1% penicillin-streptomycin-glutamine (Gibco 10378016). Human 293T (RRID: CVCL_0063) cells were cultured under the same condition (37°C, 5% CO_2) in DMEM (Gibco 10566) supplemented similarly to the Ramos cells.

Sample preparation: Cultured B cells were fixed with 4% PFA for 10 min at RT in suspension followed with DPBS wash three times, 3 min each. Each group of cells was then labeled with a unique oligo sequence on the cell surface following a published procedure (*76*) described below. Briefly, 25 μl of 500 μM 3′ amine-modified oligo (IDT, listed in data S6) was incubated with 8.2 μl of 10 mM Methyltetrazine-NHS ester (Click Chemistry Tools, 1128-25) and 41.8 μl DMSO in 50 mM sodium borate buffer (Thermo Fisher Scientific, J62902.AP) for 30 min at RT in dark. An additional 8.2 μl of 10 mM Methyltetrazine-NHS ester was then added and the reaction was incubated for 30 min as above, followed by another identical addition and incubation for another 30 min. The reaction was quenched by adding 180 μl of 50 mM sodium borate buffer and 30 μl of 3 M NaCl, followed by the addition of 750 μl ice-cold 100% ethanol. The mixture was precipitated at −80°C overnight. The mixture was centrifuged at 20,000 × *g* for 30 min at 4°C, and the pellet was washed twice with 750 μl of 70% ethanol. The pellet was air-dried and resuspended in 100 μl of cold 10 mM HEPES buffer (Thermo Fisher Scientific, 15630080). 5 million fixed cells were resuspended in 100 μl DPBS and incubated with 4 μl of 1 mM TCO-PEG4-TFP ester (Click Chemistry Tools, 1398-2) for 5 min at RT in the dark. Subsequently, 6 μl of the prepared oligo solution above was added, and the mixture was incubated for 30 min at RT in the dark. Methyltetrazine-PEG4-Amine (Click Chemistry Tools, 1012-100) was then added to 50 μM final concentration. Tris-HCl was added to 10 mM final concentration and the mixture was incubated for 5 min at RT. Cells were then spun down and washed with DPBS three times, 3 min each. Separately labeled cells were pooled, and 500 μl of the cell suspension was seeded onto poly-D-lysine–coated coverslips at a density of 10 million cells/ml in DPBS for 20 min at RT. Excess cells were gently removed, and the attached cells were left on the coverslip for an additional 15 min, followed by DPBS wash for two times, 3 min each. Permeabilization was performed using 0.5% Triton X-100 in DPBS, followed by DPBS wash for three times, 3 min each. Next, the cells were incubated with 0.1 M HCl for 5 min at RT and washed twice with DPBS and once with 2× SSC buffer, with each wash lasting 3 min. Cells were then incubated with 2.5 ml prehybridization buffer, composed of 50% formamide in 2× SSC, for 30 min at RT. Heat denaturation was performed by incubating in a 90°C water bath in prehybridization buffer for 5 min, followed by a quick rinse with 2× SSC at RT. Next, samples were incubated with 25 μl hybridization buffer. For genome-wide chromatin tracing, the buffer was composed of 10% (w/v) dextran sulfate, 0.1 mg/ml yeast tRNA, 50% formamide and 0.04 nM per probe of genome-wide chromatin tracing probe library in 2× SSC and cells were incubated at 47°C overnight. For fine-scale chromatin tracing, the buffer contained 10% (w/v) dextran sulfate, 0.1 mg/ml

yeast tRNA, 50% formamide and 0.4 nM per probe of fine-scale chromatin tracing probe library in 2× SSC and cells were incubated at 37°C overnight. Next, cells were washed with 0.1% (v/v) Tween-20 in 2× SSC as described above. Two biological replicates were collected for each library.

Sequential secondary probe hybridization and imaging

To perform automatic buffer exchange and imaging, we used a Bioptechs FCS2 flow chamber and a computer-controlled, custom-built fluidics and imaging system as described in a previous publication (*75*). For the MINA imaging in human tonsil, CD35 signal was first collected in 647-nm laser channel in oxygen scavenging imaging buffer composed of 50 mM Tris-HCl pH 8.0, 10% (w/v) glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/ml glucose oxidase (Sigma-Aldrich, G2133) and 40 μg/ml catalase (Sigma-Aldrich, C30) in 2× SSC. CD35 signal was then photobleached in 2× SSC. Next, the chromatin tracing and RNA MERFISH targets to be visualized in the first round of readout FISH imaging and fiducial marker targets were hybridized with adapter and readout probes in 3 ml of secondary probe hybridization buffer composed of 20% (v/v) ethylene carbonate (Sigma-Aldrich, E26258) in 2× SSC at RT for 20 min. 7.5 nM Alexa Fluor 488 dye-labeled readout probe was used to label the fiducial marker oligos bound to three repetitive loci in the genome (oligos are listed in data S6). 5 nM of adapter oligos and 7.5 nM of Alexa Fluor 750 dye-labeled readout probes were used for MERFISH imaging. 5 nM of adapter oligos and 7.5 nM of Atto 565 and Atto 647 dye-labeled readout probes were used for chromatin tracing imaging. After secondary probe hybridization, the samples were washed with 2.4 ml of secondary probe hybridization buffer without probes over 7 min, and imaged in the oxygen scavenging imaging buffer. After imaging, the chromatin tracing and MERFISH adapter and readout probes were removed by washing with 5 ml of 65% formamide in 2× SSC for 20 min. Next, the same secondary probe hybridization, washing, imaging and removal procedures were performed iteratively for the following chromatin tracing and MERFISH targets with the corresponding adapter and readout probes supplied at the same concentrations as described above. After all rounds of secondary FISH imaging, 2.4 ml of 500 ng/ml CF-640R dye labeled Wheat Germ Agglutinin (Biotium, 29026) in HBSS buffer (Gibco, 14025092) and 2.4 ml of 1 μg/ml DAPI (Thermo Fisher, 62248) in 2× SSC were applied for 15 min each, washed with 2.4 ml of 2× SSC over 6.5 min, and imaged for cell boundary and nucleus labeling.

For chromatin tracing in cultured B cells, 0.1-μm yellow-green fiducial beads (Invitrogen, F8803) were applied to the samples as fiducial markers. For fine-scale tracing, 5 nM of adapter oligos and 7.5 nM of Alexa Fluor 750 and Atto 647 dye-labeled readout probes with a disulfide bond were used for each round of FISH imaging; the signal was removed by tris(2-carboxyethyl)phosphine (TCEP; Sigma-Aldrich, C4706), which can reduce the disulfide bonds connecting fluorophores to readout probes, as described in previous publications (*77*, *78*). For genome-wide chromatin tracing, adapters and Atto 565 and Atto 647 dye-labeled readout probes at the same concentrations as above were used for each round of FISH imaging. 65% formamide wash was used for signal removal. After chromatin tracing, DAPI was applied and imaged for cell nucleus labeling as above. Finally, cell surface labeling oligos were imaged sequentially with adapters and Atto 565 dye-labeled readout probes at the same concentrations and using the same iterative procedures (with 65% formamide wash for signal removal) as above for each cell group. All the adapter sequences and readout probes are listed in data S6.

Droplet Hi-C in in human tonsil

Tonsil samples were obtained following the same procedure as above. Droplet Hi-C was performed following a published procedure (*21*) and briefly described below. Four biological replicates from two different donors were performed.

Cell fixation: Single cells were dissociated from tonsil samples by smashing the samples with the bottom of a syringe barrel for a few times. Next, single cells were collected and incubated in ACK lysing buffer (Gibco, A1049201) for 5 min at RT and quenched with complete RPMI-1640 medium. Next, cells were fixed with 1% PFA for 20 min at RT and quenched with complete medium. Five million cells were then aliquoted into each tube and stored at −80°C until use.

In situ Hi-C: Cell pellets were lysed in 600 μl ice-cold lysis buffer (10 mM Tris–HCl, pH 8.0 (Thermo Fisher Scientific, 15568025), 10 mM NaCl (Sigma-Aldrich, S5150), 0.2% IGEPAL CA-630 (Sigma-Aldrich, I8896), and 1× protease inhibitor (Roche, 5056489001)) and incubated on ice for 45 min. Nuclei were collected by centrifugation and washed once with lysis buffer. Nuclei were counted, and two million nuclei were aliquoted per reaction. Nuclei were resuspended in 100 μl of 0.5% SDS and incubated at 62°C for 10 min. To quench SDS, 290 μl nuclease-free water and 50 μl of 10% Triton X-100 were added, followed by incubation at 37°C for 15 min at 300 rpm in a thermal mixer (Eppendorf ThermoMixer F1.5). Next, 50 μl of 10× CutSmart buffer (NEB) and restriction enzymes were added, including 100 U DpnII (NEB, R0543L), 130 U MboI (NEB, R0147M), and 15 U NlaIII (NEB, R0125L). Samples were incubated at 37°C for 90 min at 550 rpm. Enzymes were heat-inactivated at 65°C for 20 min and cooled to RT. Nuclei were pelleted by centrifugation and washed with 400 μl ligation buffer (100 μl 10× T4 DNA ligase buffer (NEB, B0202S), 5 μl 20 mg/ml BSA (NEB, B9000S), and 865 μl nuclease-free water). Ligation was performed in 400 μl ligation buffer with 40 μl T4 DNA ligase (NEB, M0202L) at 37°C for 40 min with shaking at 300 rpm. After ligation, single nuclei were purified by FACS sorting with DAPI staining.

Single-cell Hi-C sequencing library construction: Nuclei were resuspended in 1× Nuclei Buffer [diluted from 20× Nuclei Buffer (10x Genomics, PN-2000207)] and submitted to Yale Center for Genome Analysis (YCGA) for library construction. Libraries were constructed following the Chromium Next GEM Single Cell ATAC Reagent Kits v2 User Guide with two modifications: (i) The index PCR elongation time was extended to 1 min; (ii) The double-sided size selection was changed to 1.14× SPRIselect to remove only small fragments. The libraries were sequenced using the Illumina NovaSeq platform [100-base pair (bp) paired-end sequencing].

Plasmid construction

px458 and pFlox_N-Nipbl_Neo_FKBPF36V_mScarlet plasmids were obtained from AddGene (#48138 and #163883). To achieve high-efficiency gene editing, pLsg plasmid was constructed by removing *GFP* and *SpCas9* from px458, resulting in a vector that expresses only a single-guide RNA (sgRNA). sgRNAs were designed using the Broad Institute sgRNA Designer tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and cloned into the pLsg plasmid. sgRNAs' spacer sequences targeting RAD21 and UNG were as follows: sgRAD21, CCTTATATAATATGGAACCT; sgUNG, GCGGCCCGCAACGTGCCCGT. The RAD21-degron donor plasmid was constructed by removing the *Neomycin* and *P2A* sequences from pFlox_N-Nipbl_Neo_FKBPF36V_mScarlet and inserting left and right homology arms flanking the HA-FKBP12^{F36V}-mScarlet sequence.

CRISPR-Cas9 editing in Ramos-derived cell lines

The RAD21-degron cell line was generated from RASH-1C cells which constitutively express Cas9 (*46*). pLsg-sgRAD21 plasmid targeting the RAD21 stop-codon region and the RAD21-degron donor plasmid were delivered into cells via electroporation. Four days later, mScarlet-positive cells were sorted into 96-well plates to obtain single-cell clones. Expanded clones were cultured for 15 days in 96-well plates, analyzed by flow cytometry to detect mScarlet expression, and subsequently validated by Sanger sequencing. RAD21-degron UNG knockout

cells were generated by electroporating the pLsg-sgUNG plasmid into RAD21-degron cells.

High-throughput mutation sequencing

Mutation analysis was performed using a protocol modified from the Illumina "16S Metagenomic Sequencing Library Preparation" protocol. Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, 69504) according to the manufacturer's instructions. Genomic regions were amplified by PCR reactions (primers listed in data S7). For *IGH VDJ*, *HTS7* in *GFP reporter*, and *IGL VJ* regions, PCR reactions were performed using the NEBNext High-Fidelity 2× PCR Master mix (New England Biolabs, M0541L) under the following conditions: an initial denaturation at 98°C for 3 min, followed by 25 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 2 min. For *BCL6* and *POU2AF1* regions, PCR reactions were performed using the Q5U Hot Start High-Fidelity DNA Polymerase (New England Biolabs, M0515L) under the following conditions: an initial denaturation at 98°C for 3 min, followed by 25 cycles of 98°C for 10 s, 60°C for 20 s, and 72°C for 30 s, with a final extension at 72°C for 5 min. The amplified products were sequenced using the Illumina NovaSeq platform (150-bp paired-end sequencing) at YCGA.

CUT&RUN

CUT&RUN was performed following the manufacturer's protocol using the CUTANA ChIC/CUT&RUN Kit (Epicypher, 14-1048). Briefly, RAD21-degron cells were treated with 0.5 μM NEG (Tocris, 6915) or dTAGV-1 (Tocris, 6914) for indicated time followed by 200 ng/ml dox (Sigma-Aldrich, D3447) treatment. Approximately 0.5 million RAD21-degron cells were collected for nuclei extraction. The isolated nuclei were bound to activated concanavalin A-coated magnetic beads and incubated overnight at 4°C on a nutator with HA antibody (Santa Cruz, sc-7392), RAD21 antibody (Abcam, ab992), Rpb1 CTD antibody (Cell Signaling Technology, 2629), S2P-RNAPII antibody (Cell Signaling Technology, 13499), S5P-RNAPII antibody (Cell Signaling Technology, 13523), H3K27ac antibody (Abcam, ab4729) or AID antibody (Invitrogen, 39-2500; RRID: AB_2533403). After two washes with cell permeabilization buffer, 2.5 μl of pAG/MNase was added to the resuspended beads in 50 μl of cell permeabilization buffer and incubated for 10 min at RT. Following another two washes with cell permeabilization buffer, 1 μl of 100 mM CaCl2 was added, and the samples were incubated at 4°C for 2 hours on a nutator. The reaction was stopped by adding 33 μl of stop buffer, followed by incubation at 37°C for 10 min. DNA was purified using SPRIselect reagent (beads). Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645S) and sequenced on an Illumina NovaSeq platform (100-bp paired-end sequencing) with an average depth of 20 million reads per sample.

TT-TimeLapse-seq

TT-TimeLapse-seq was performed as previously described (*53*). Approximately 25 million RAD21-degron cells were labeled with 1 mM 4-thiouridine (s4U) (Fisher Scientific, 13957-31-8) for 5 min. Cells were collected and washed by PBS and lysed by Trizol (Invitrogen, 15596026). Total RNA was extracted and treated with TURBO DNase (Thermo Fisher Scientific, AM2238) to remove genomic DNA. Genomic DNA–depleted Drosophila S2 spike-in RNA (5-min s4U labeled) was added at 5% of total RNA. RNA was fragmented by mixing with 2× RNA fragmentation buffer (150 mM Tris-HCl, pH 7.4, 225 mM KCl, 9 mM $MgCl_2$) and incubating at 94°C for 3.5 min. Fragmentation was stopped by adding EDTA to a final concentration of 50 mM followed by incubation on ice for 2 min. RNA was purified using the RNeasy Mini Kit (Qiagen, 74104).

Fragmented RNA was biotinylated using 10% MTSEA-biotin-XX (VWR, 89139-636) prepared in DMF. Biotinylated RNA was purified and captured using 10 μl Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen, 65001). s4U-labeled RNA was eluted by reducing the disulfide bond between biotin and 4-thiouridine using 25 μl elution buffer (100 mM DTT, 1 mM EDTA, 100 mM NaCl, 0.05% Tween-20), leaving biotin bound to the streptavidin beads.

TimeLapse chemistry was performed as previously described (*53*) to induce U-to-C conversion. Briefly, 20 μl of eluted RNA was treated with TimeLapse reaction mixture containing 0.835 μl 3 M sodium acetate (pH 5.2), 0.2 μl 500 mM EDTA, 1.365 μl nuclease-free water, and 1.3 μl trifluoroethylamine. Freshly prepared sodium periodate (1.3 μl, 200 mM) was added, and samples were incubated at 45°C for 1 hour. Chemically treated RNA was purified and subsequently incubated in reducing buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA) at 37°C for 30 min, followed by a second purification step.

Final RNA was used for library preparation with the Takara Pico Mammalian RNA-seq V3 kit (Takara, 634938). Libraries were sequenced on an Illumina NovaSeq X platform to an average depth of 30 million reads per sample.

3C-HTGTS

3C-HTGTS was performed as previously described (*79*). Briefly, 3 million cells were crosslinked with 2% (v/v) formaldehyde for 10 min at RT and quenched with glycine at a final concentration of 125 mM. Cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100, and protease inhibitors (Roche, 11836153001). Nuclei were digested overnight at 37°C with 350 units of NlaIII (NEB, R0125), followed by ligation under dilute conditions at 16°C overnight. Crosslinks were reversed, and samples were treated with Proteinase K (Roche, 03115852001) and RNase A (Invitrogen, 8003089) prior to DNA precipitation.

The resulting 3C libraries were sonicated using a Diagenode Bioruptor sonicator. LAM-HTGTS libraries were subsequently prepared and analyzed as previously described (*49*).

ChIP-seq

H3K4me3 and H3K27ac ChIP-seq were performed as described before (*8*) with minor modifications. RASH-1 cells were fixed in 1% formaldehyde for 10 min at RT and collected for nuclei extraction. Isolated nuclei were digested with Micrococcal Nuclease (Cell Signaling Technology, 10011) to obtain optimally sized chromatin fragments. Digestion was stopped by addition of 0.5 M EDTA. Nuclear lysates were subjected to immunoprecipitation with 5 μg of H3K27ac antibody (Abcam, ab4729) or H3K4me3 antibody (Abcam, ab8580) overnight. Magnetic beads (Cell signaling Technology, ChIP-Grade Protein G Magnetic Beads #9006) were then incubated with the antibody-bound nuclear lysates for 2 hours at 4°C. Beads were washed sequentially with ice-cold RIPA, RIPA500, LiCl, and TE buffers for 5 min each. DNA was eluted using elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65°C for 15 min. After reversal of cross-links, DNA was purified by phenol-chloroform-isoamyl alcohol extraction. Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit and sequenced on an Illumina NovaSeq platform (100-bp paired-end sequencing), with an average depth of 20 million reads per sample.

Imaging data analysis

Genome-wide chromatin tracing analysis: A total of 1151 loci were imaged using 560-nm and 647-nm laser channels over 50 imaging rounds. To decode each locus, the color shift between the 560-nm and 647-nm laser channels was first corrected using TetraSpeck bead images (Invitrogen, T7279). Sample drift during sequential hybridization and imaging was corrected using fiducial marker images. Cell segmentation was then performed using the watershed algorithm, with the WGA signal (in tonsil tissue) or DAPI signal (in B cell lines). 3D positions (x, y, z coordinates) of DNA foci were fitted with a 3D radial center algorithm (*80*) in each round of imaging and assigned to

each single cell. Foci were decoded with a previously described algorithm (*81*). Briefly, all valid spot pairs corresponding to a valid barcode (with Hamming Weight 2) were identified if the two fitted DNA spots were within a spatial distance threshold of 500 nm. For different spot pairs sharing the same spot, the highest-quality pairs were retained based on signal intensity and spatial distance. Next, we attempted to link the selected spot pairs into chromatin traces. Through an iterative process, the quality of each foci pair was further assessed based on their intensity, the spatial distance between the two spots, and the distance of each spot pair to the nearest corresponding chromosome territory centroid updated in each iteration. The iteration continues until either more than 99% of spot pairs remain unchanged between iterations or more than 97% remain unchanged after 10 iterations. Finally, the selected spot pairs were retained and linked into chromatin traces. Additionally, any missing loci were recovered using the unlinked spot pairs if they were within 6 pixels of the chromosome territory periphery.

Fine-scale chromatin tracing analysis: First, sample drift during sequential hybridization and imaging was corrected using fiducial marker images. Next, cell segmentation was performed with CellPose (*82*) package using DAPI signals. DNA foci were fitted, with approximately two foci per nucleus in each round of imaging. The fitted foci were further linked into traces based on their spatial clustering, using a distance threshold of 0.8 μm between a new focus and the average position of all previous foci in the trace. Finally, we attempted to refit the 3D positions of any missing loci within the chromatin trace region in the corresponding rounds of imaging using 3D Gaussian fitting.

RNA MERFISH analysis: We perform MERFISH decoding with a previously reported pixel based decoding strategy (*78*). In brief, sample drift during sequential hybridization and imaging was corrected by 2D fitting of fiducial markers. Next, for each *z*-stack, background removed image from each round was concatenated vertically into a 3D matrix. A "pixel vector" was derived from the 3D matrix along the third dimension for each pixel. The pixel vectors were normalized to become unit vectors and compared with unit vectors of all available MERFISH codes. The closest code for each pixel was identified within a distance threshold of 0.65. The connected pixels with the same code were merged as one focus. Next, three parameters were calculated for each focus: focus area, unit vector distance to the matched code, and the magnitude of the pixel vector before normalization. All the three parameters from 20 randomly selected FOVs built up a 3D parameter space. The parameter space was divided into small bins, and the error rate for each bin was calculated. Foci within bins where the error rate was below a defined threshold were selected. An adaptative thresholding strategy was performed here so that the final error rate was below 5%. Finally, all MERFISH foci within the selected parameter bins were decoded.

RNA MERFISH cell clustering: Cell clustering and annotation were performed using gene expression enrichment-based approaches in combination with manual annotation. After decoding, single-cell clustering and cell type annotation were performed using Seurat (v5) (*83*). Raw counts were first normalized with SCTransform, followed by PCA, UMAP, and t-SNE dimensionality reduction, neighborhood detection (FindNeighbors), and unsupervised clustering (FindClusters) to get initial cell groups. Marker gene sets for GCBC, NBC/MBC, T-cell subsets (Tfh and non-Tfh), PC, myeloid cells, PDC, FDC, and epithelial cells were obtained from a published tonsil scRNA-seq dataset (*23*) and used as input reference files. The resulting RNA MERFISH clusters were initially annotated based on enrichment scores calculated using the AUCell method implemented in irGSEA (*84*), followed by validation through checking the expression of individual canonical marker genes. Next, a second round of finer annotation was performed. GCBC, NBC/MBC, and T cells were each independently subclustered by integrating marker gene expression, spatial coordinates, and colocalization with CD35 immunofluorescence staining. Within the GC, LZ and DZ B cells were validated by colocalization with CD35 immunofluorescence and differential *LMO2* expression (*23*), with higher *LMO2* marking LZ cells. Within the NBC/MBC compartment, NBCs were preferentially localized to the follicular mantle zone encircling the GC, while MBCs were enriched in the outer follicular and subepithelial regions (*85*), and the two were further distinguished by higher *CD83* expression in NBC. Tfh cells were distinguished from non-Tfh $CD4^+$ T cells by their spatial enrichment within the CD35-defined GC region and elevated *BCL6* expression (*38*).

MINA phenotype analysis

Data from the five replicates were merged prior to downstream phenotypical analysis.

Radial score: For each cell, we first computed the centroid of all detected genomic loci. The radial score for each locus was then defined as its Euclidean distance to the centroid, normalized by the mean distance of all loci in that cell to the centroid.

Demixing score: The demixing score was defined as the standard deviation of all mean interloci distances across a chromosome in a given cell state divided by the average value of mean interloci distances across the entire chromosome.

Normalization of interloci distance for spatial clustering analyses of hot TADs, cold TADs and filler loci: To normalize intrachromosome distances and remove genomic distance contribution to spatial distance, we first performed a power-law fitting with the observed mean spatial distances and the corresponding genomic distances between target loci in each chromosome. Next, we computed the expected spatial distance for each pair of loci using the fitted function and normalized the observed spatial distance to the expected distance. To normalize interchromosome distances, we first calculated the average spatial distances between all loci for each chromosome pair. The interchromosome distances were then normalized by the corresponding average interchromosome distance of the given chromosome pair.

Droplet Hi-C analysis

Processing of tonsil Droplet Hi-C data: The upstream analysis was performed following a published pipeline (*21*) with minor modifications. After demultiplexing using 10x Genomics cellranger-atac mkfastq (v2.2.0) with the GRCh38 2024-A reference genome, raw sequence data were obtained as FASTQ files (R1, R2 and R3). The R1 and R3 reads were aligned against the 10x Genomics ATAC v2 barcode whitelist using Bowtie v1.3.1 (*86*). The resulting SAM files were converted back to FASTQ format with barcode information using HiCTools. Then, the barcode-annotated paired-end reads were trimmed by Trim Galore v0.6.7 to remove the adaptor and low-quality reads, followed by genome alignment using BWA-MEM2 and the format conversion from SAM to BAM using SAMtools v1.21. Next, the contact pair parsing, sorting and deduplication were performed by using Pairtools v1.1.3 (*87*) to generate the pairs file with barcode information. High-quality single nuclei were selected based on the total counts in each single cell using an automated elbow-point detection algorithm (PHCrankPair function).

ScGAD score calculation, coembedding with published scRNA-seq, and clustering: Per-cell .pairs files were extracted from the deduplicated bulk pairs file by splitting based on the barcode information, followed by the format conversion to .cool and BED-format files at 5-kb

resolution with hg38 gene annotations for the single-cell GAD score calculation (*22*). After obtaining the GAD score × cell matrix, principal components analysis (PCA), UMAP embedding, *k*-nearest neighbor graph with symmetric edge completion and Louvain community detection were performed. The resulting clustering was manually reviewed and merged into T and B cell groups based on the UMAP distribution and well-known marker genes. The scGAD scores from the four B cell groups (NBC, MBC, GCBC, and PC) were integrated with published tonsil scRNA-seq data (*23*) to achieve finer clustering of B cell groups in Droplet Hi-C data. Highly variable genes (HVGs) were selected by ranking genes by variance across the Hi-C dataset and the top 2000 genes were selected. Well-known B cell subtype marker genes were added to HVG list to ensure that cell type–defining features were retained during the coembedding. The top 30 Seurat PCA coordinates were selected from the published datasets for the RNA PCA reference space for the projection of Hi-C cells. A joint UMAP was computed on the Harmony-corrected embeddings of Droplet Hi-C and scRNA-seq data. Then cell type label transfer from scRNA-seq to Hi-C was performed using *k*-nearest neighbor algorithm. The cell type information was used for the following analyses.

A/B compartment analysis: The four replicates were merged prior to downstream analysis. Based on the clustering information, single cells were categorized into five different groups, including T cells, GCBC, NBC, MBC, and PC. We then performed A/B compartment analysis for each group following a previously established procedure (*34*). The merged bulk Hi-C contact matrices of individual groups were loaded from .cool files for the A/B compartment analysis at 100-kb resolution. After filtering out genomic bins with extremely high or low coverages, Vanilla Coverage (VC) normalization was performed, followed by the fitting of a genome-wide power-law distance decay model. The observed/expected (O/E) matrices were generated by dividing VC-normalized matrices by the corresponding power-law predicted value. Pearson correlation matrices were computed from the O/E matrices and PCA analysis was performed per chromosome. The coefficients of the first eigenvector from the PCA analysis (PC1) were used as compartment scores. The orientation of the eigenvector was adjusted so that the compartment scores correlated positively with CpG density across the corresponding 100-kb genomic bins, thereby associating positive scores with the more active A compartment. For the pairwise comparisons of A/B compartment values among different groups, dcHiC v2.1 (*88*) was used with PC1 bedGraph files as input.

Chromatin loop calling: Group-level chromatin loop calling was carried out using the scHiCluster package (*89*) at 10-kb resolution, employing a modified SnapHiC workflow (*90*) and leveraging cell type annotation information. Briefly, single-cell imputed contact matrices were generated from single-cell raw contact matrices using hicluster embed-concatcell-chr and hicluster embed-mergechr. Then, a background model was computed using hicluster loop-bkg-cell with the default parameters. Per-group pseudo-bulk loop calling was performed by first aggregating per-cell imputed contact matrices chromosome by chromosome using hicluster loop-sumcell-chr with the group information based on the cluster assignments mentioned above. Finally, per-chromosome aggregated matrices were merged across the genome using hicluster loop-mergechr to generate genome-wide loop files in BEDPE format for the five groups.

Statistical analysis

Data were analyzed statistically and plotted using GraphPad Prism. Single comparisons were conducted using a two-sided Student's *t* test or Wilcoxon test as indicated. Statistical significance was defined as $P < 0.05$ (*), $P < 0.01$ (**), $P < 0.001$ (***), and $P < 0.0001$ (****).

Custom cell line reference genome generation

To enable accurate alignment and interpretation of sequencing data in RASH-1 cells, we generated a custom reference genome based on hg38 that incorporates modified *IGH* and *IGL* locus configurations to reflect their rearrangement by V(D)J recombination, the AID7.3 expression cassette (inserted at the *AAVS1* locus), and the lentiviral *GFP* reporter vector that is inserted ~38 kb upstream of *IGH-V* (*46*). These modifications reflect the altered genome architecture of the RASH-1 cell system, thereby minimizing mapping ambiguity and improving detection of *IGH*- and *IGL*-associated sequence reads. All read alignment and downstream analyses were performed using this custom reference genome (hereafter referred to as the RASH-1 genome).

Mutation-sequencing analysis

Sequence reads were processed using a custom pipeline. Paired-end reads were merged using the fastq-join tool, requiring a minimum 10-bp overlap with a mismatch rate of ≤8%. Merged reads were aligned to their respective reference sequences using the Burrows-Wheeler Aligner (BWA) with default parameters. The Python utility Pysamstats was used to calculate position-wise statistics against the reference sequences based on the aligned BAM files. Statistics were generated for reads with a mapping quality of at least 60 and a minimum base quality of 30. For each sample, the aligned BAM file was converted to TSV format using the Java-based utility sam2tsv. An in-house AWK script was used to generate per-read summary files for each observed variation. Only nucleotide bases with a Phred quality score of ≥30 were considered for point mutations. Finally, WRC sites (AID hotspots) were identified from the position-based statistics files generated by Pysamstats using R script.

3C-HTGTS analysis

3C-HTGTS sequencing data were processed using published HTGTS analysis pipeline (*91*). Briefly, raw paired-end sequencing reads were aligned to the RASH-1 genome and filtered using standard pipeline parameters. To eliminate local background signal arising from self-ligation events, junctions proximal to the bait region were removed prior to downstream analysis. Following artifact removal, interaction profiles were normalized by per-million scaling to enable quantitative comparison across samples. Normalization was performed using a custom Perl script adapted from publicly available 3C-HTGTS utilities (https://github.com/Yyx2626/HTGTS_related/tree/main/3CHTGTS_related), which scales interaction counts based on the total number of valid junctions per sample.

CUT&RUN analysis

Paired-end CUT&RUN sequencing reads were aligned to the RASH-1 genome using Bowtie2 with "local" and "very-sensitive" alignment parameters. Only properly paired, uniquely mapped reads were retained for downstream analyses. Biological replicates were merged prior to signal generation. Genome-wide signal tracks were generated using deepTools (v3.5.1) (*92*) as RPKM-normalized bigWig files following subtraction of the corresponding IgG control. These normalized bigWig files were used for most of the metagene profiles and aggregate signal plots. Region-specific signal quantification and genome-wide heatmap generation were performed using bamCoverage. For AID CUT&RUN, raw counts were obtained for all genes for both dox and no-dox treated samples. Differential analysis was performed using DESeq2 (*93*) to identify statistically significant peaks, defined as those with a $\log_2$ fold change >2 and $P <= 0.01$.

TT-TimeLapse-seq analysis

Nascent RNA synthesis was measured in RAD21-degron RASH-1 cells using TT-TimeLapse-seq with Drosophila spike-in RNA. Raw sequencing reads were processed using a published Snakemake-based pipeline

(*94*). Briefly: Reads were aligned to a combined reference genome consisting of RASH-1 genome and *Drosophila* genome (dm6). Only uniquely mapped reads with high mapping quality were retained for downstream analysis. TimeLapse-specific T-to-C conversions were identified from the aligned BAM files after applying stringent filters to exclude low-quality bases and read-end artifacts. Nascent RNA synthesis was quantified as the number of unique T-to-C conversions per read. Genome-wide coverage tracks were generated from the filtered BAM files and normalized to RPKM. Spike-in scaling factors were computed using DESeq2 on the Drosophila reads. Final spike-in–normalized, RPKM-scaled bigWig files were produced for downstream analyses, including metaplot generation.

Correlation coefficients

Pearson correlation coefficients were used to quantify correlations between datasets and between different measurements, such as compartment score and radial score.

REFERENCES AND NOTES

1. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007). doi: 10.1146/annurev.biochem.76.061705.090740; pmid: 17328676
2. V. H. Odegard, D. G. Schatz, Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **6**, 573–583 (2006). doi: 10.1038/nri1896; pmid: 16868548
3. M. Liu, D. G. Schatz, Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* **30**, 173–181 (2009). doi: 10.1016/j.it.2009.01.007; pmid: 19303358
4. A. Nussenzweig, M. C. Nussenzweig, Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010). doi: 10.1016/j.cell.2010.03.016; pmid: 20371343
5. K. Basso, R. Dalla-Favera, Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* **15**, 172–184 (2015). doi: 10.1038/nri3814; pmid: 25712152
6. R. Casellas *et al.*, Mutations, kataegis and translocations in B cells: Understanding AID promiscuous activity. *Nat. Rev. Immunol.* **16**, 164–176 (2016). doi: 10.1038/nri.2016.2; pmid: 26898111
7. L. Pasqualucci, R. Dalla-Favera, Genetics of diffuse large B-cell lymphoma. *Blood* **131**, 2307–2319 (2018). doi: 10.1182/blood-2017-11-764332; pmid: 29666115
8. F. Senigl *et al.*, Topologically Associated Domains Delineate Susceptibility to Somatic Hypermutation. *Cell Rep.* **29**, 3902–3915.e8 (2019). doi: 10.1016/j.celrep.2019.11.039; pmid: 31851922
9. U. E. Schoeberl *et al.*, Regulation of somatic hypermutation by higher-order chromatin structure. *Mol. Cell* **85**, 2701–2717.e9 (2025). doi: 10.1016/j.molcel.2025.06.003; pmid: 40580953
10. E. Pinaud *et al.*, The IgH locus 3′ regulatory region: Pulling the strings from behind. *Adv. Immunol.* **110**, 27–70 (2011). doi: 10.1016/B978-0-12-387663-8.00002-8; pmid: 21762815
11. J. M. Buerstedde, J. Alinikula, H. Arakawa, J. J. McDonald, D. G. Schatz, Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLOS Biol.* **12**, e1001831 (2014). doi: 10.1371/journal.pbio.1001831; pmid: 24691034
12. F. L. Meng *et al.*, Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159**, 1538–1548 (2014). doi: 10.1016/j.cell.2014.11.014; pmid: 25483776
13. J. Qian *et al.*, B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**, 1524–1537 (2014). doi: 10.1016/j.cell.2014.11.013; pmid: 25483777
14. S. Le Noir *et al.*, The IgH locus 3′ cis-regulatory super-enhancer co-opts AID for allelic transvection. *Oncotarget* **8**, 12929–12940 (2017). doi: 10.18632/oncotarget.14585; pmid: 28088785
15. J. Sun, G. Rothschild, E. Pefanis, U. Basu, Transcriptional stalling in B-lymphocytes: A mechanism for antibody diversification and maintenance of genomic integrity. *Transcription* **4**, 127–135 (2013). doi: 10.4161/trns.24556; pmid: 23584095
16. S. P. Methot, J. M. Di Noia, Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv. Immunol.* **133**, 37–87 (2017). doi: 10.1016/bs.ai.2016.11.002; pmid: 28215280
17. G. D. Victora, M. C. Nussenzweig, Germinal Centers. *Annu. Rev. Immunol.* **40**, 413–442 (2022). doi: 10.1146/annurev-immunol-120419-022408; pmid: 35113731
18. S. Wang *et al.*, Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016). doi: 10.1126/science.aaf8084; pmid: 27445307
19. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi: 10.1126/science.aaa6090; pmid: 25858977
20. M. Liu *et al.*, Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020). doi: 10.1038/s41467-020-16732-5; pmid: 32518300
21. L. Chang *et al.*, Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat. Biotechnol.* **43**, 1694–1707 (2025). doi: 10.1038/s41587-024-02447-1; pmid: 39424717
22. S. Shen, Y. Zheng, S. Keleş, scGAD: Single-cell gene associating domain scores for exploratory analysis of scHi-C data. *Bioinformatics* **38**, 3642–3644 (2022). doi: 10.1093/bioinformatics/btac372; pmid: 35652733
23. R. Massoni-Badosa *et al.*, An atlas of cells in the human tonsil. *Immunity* **57**, 379–399.e18 (2024). doi: 10.1016/j.immuni.2024.01.006; pmid: 38301653
24. B. A. Heesters *et al.*, Characterization of human FDCs reveals regulation of T cells and antigen presentation to B cells. *J. Exp. Med.* **218**, e20210790 (2021). doi: 10.1084/jem.20210790; pmid: 34424268
25. J. Tellier *et al.*, Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. *Nat. Immunol.* **17**, 323–330 (2016). doi: 10.1038/ni.3348; pmid: 26779600
26. S. Ning, J. S. Pagano, G. N. Barber, IRF7: Activation, regulation, modification and function. *Genes Immun.* **12**, 399–414 (2011). doi: 10.1038/gene.2011.21; pmid: 21490621
27. Q. Hu, T. Xu, W. Zhang, C. Huang, Bach2 regulates B cell survival to maintain germinal centers and promote B cell memory. *Biochem. Biophys. Res. Commun.* **618**, 86–92 (2022). doi: 10.1016/j.bbrc.2022.06.009; pmid: 35716600
28. H. W. King *et al.*, Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci. *Sci. Immunol.* **6**, eabh3768 (2021). doi: 10.1126/sciimmunol.abh3768; pmid: 34623901
29. V. G. Martinez *et al.*, Fibroblastic Reticular Cells Control Conduit Matrix Deposition during Lymph Node Expansion. *Cell Rep.* **29**, 2810–2822.e5 (2019). doi: 10.1016/j.celrep.2019.10.103; pmid: 31775047
30. D. E. Peters *et al.*, The membrane-anchored serine protease prostasin (CAP1/PRSS8) supports epidermal development and postnatal homeostasis independent of its enzymatic activity. *J. Biol. Chem.* **289**, 14740–14749 (2014). doi: 10.1074/jbc.M113.541318; pmid: 24706745
31. J. R. Lin *et al.*, Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018). doi: 10.7554/eLife.31657; pmid: 29993362
32. N. S. De Silva, U. Klein, Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15**, 137–148 (2015). doi: 10.1038/nri3804; pmid: 25656706
33. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001). doi: 10.1038/35066075; pmid: 11283701
34. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
35. L. Mesin, J. Ersching, G. D. Victora, Germinal Center B Cell Dynamics. *Immunity* **45**, 471–482 (2016). doi: 10.1016/j.immuni.2016.09.001; pmid: 27653600
36. J. R. Dixon *et al.*, Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015). doi: 10.1038/nature14222; pmid: 25693564
37. K. Basso, R. Dalla-Favera, BCL6: Master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv. Immunol.* **105**, 193–210 (2010). doi: 10.1016/S0065-2776(10)05007-8; pmid: 20510734
38. J. Choi, S. Crotty, Bcl6-Mediated Transcriptional Regulation of Follicular Helper T cells (T_{FH}). *Trends Immunol.* **42**, 336–349 (2021). doi: 10.1016/j.it.2021.02.002; pmid: 33663954
39. M. Shapiro-Shelef *et al.*, Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607–620 (2003). doi: 10.1016/S1074-7613(03)00267-X; pmid: 14563324
40. F. E. Johansen, R. Braathen, P. Brandtzaeg, Role of J chain in secretory immunoglobulin formation. *Scand. J. Immunol.* **52**, 240–248 (2000). doi: 10.1046/j.1365-3083.2000.00790.x; pmid: 10972899
41. S. C. Eisenbarth *et al.*, A roadmap for defining "extrafollicular" B cell responses. *Immunity* **58**, 2627–2645 (2025). doi: 10.1016/j.immuni.2025.08.007; pmid: 40987283
42. S. Boyle *et al.*, The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**, 211–219 (2001). doi: 10.1093/hmg/10.3.211; pmid: 11159939
43. G. Girelli *et al.*, GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* **38**, 1184–1193 (2020). doi: 10.1038/s41587-020-0519-y; pmid: 32451505
44. L. Tan, D. Xing, C. H. Chang, H. Li, X. S. Xie, Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018). doi: 10.1126/science.aat5641; pmid: 30166492
45. L. Guelen *et al.*, Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008). doi: 10.1038/nature06947; pmid: 18463634
46. L. Wu *et al.*, HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev.* **36**, 433–450 (2022). doi: 10.1101/gad.349438.122; pmid: 35450882
47. L. Wu *et al.*, Transcription elongation factor ELOF1 is required for efficient somatic hypermutation and class switch recombination. *Mol. Cell* **85**, 1296–1310.e7 (2025). doi: 10.1016/j.molcel.2025.02.007; pmid: 40049160

48. M. Liu *et al.*, Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841–845 (2008). doi: 10.1038/nature06547; pmid: 18273020
49. S. Jain, Z. Ba, Y. Zhang, H. Q. Dai, F. W. Alt, CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. *Cell* **174**, 102–116.e14 (2018). doi: 10.1016/j.cell.2018.04.035; pmid: 29804837
50. N. Q. Liu *et al.*, WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat. Genet.* **53**, 100–109 (2021). doi: 10.1038/s41588-020-00744-4; pmid: 33318687
51. M. Kim *et al.*, Interplay between cohesin and RNA polymerase II in regulating chromatin interactions and gene transcription. *Nat. Struct. Mol. Biol.* **33**, 259–274 (2026). doi: 10.1038/s41594-025-01708-0; pmid: 41530636
52. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017). doi: 10.1016/j.cell.2017.09.026; pmid: 28985562
53. J. A. Schofield, E. E. Duffy, L. Kiefer, M. C. Sullivan, M. D. Simon, TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **15**, 221–225 (2018). doi: 10.1038/nmeth.4582; pmid: 29355846
54. T. S. Hsieh *et al.*, Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022). doi: 10.1038/s41588-022-01223-8; pmid: 36471071
55. P. D'Addabbo, M. Scascitelli, V. Giambra, M. Rocchi, D. Frezza, Position and sequence conservation in Amniota of polymorphic enhancer HS1.2 within the palindrome of IgH 3'Regulatory Region. *BMC Evol. Biol.* **11**, 71 (2011). doi: 10.1186/1471-2148-11-71; pmid: 21406099
56. Q. Wang *et al.*, The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med.* **214**, 49–58 (2017). doi: 10.1084/jem.20161649; pmid: 27998928
57. E. Pefanis *et al.*, Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389–393 (2014). doi: 10.1038/nature13580; pmid: 25119026
58. A. Ibarra, C. Benner, S. Tyagi, J. Cool, M. W. Hetzer, Nucleoporin-mediated regulation of cell identity genes. *Genes Dev.* **30**, 2253–2258 (2016). doi: 10.1101/gad.287417.116; pmid: 27807035
59. P. Pascual-Garcia, M. Capelson, Nuclear pores in genome architecture and enhancer function. *Curr. Opin. Cell Biol.* **58**, 126–133 (2019). doi: 10.1016/j.ceb.2019.04.001; pmid: 31063899
60. M. Hazawa *et al.*, Super-enhancer trapping by the nuclear pore via intrinsically disordered regions of proteins in squamous cell carcinoma cells. *Cell Chem. Biol.* **31**, 792–804.e7 (2024). doi: 10.1016/j.chembiol.2023.10.005; pmid: 37924814
61. L. Pasqualucci *et al.*, Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001). doi: 10.1038/35085588; pmid: 11460166
62. U. Klein, R. Dalla-Favera, Germinal centres: Role in B-cell physiology and malignancy. *Nat. Rev. Immunol.* **8**, 22–33 (2008). doi: 10.1038/nri2217; pmid: 18097447
63. D. E. Tan *et al.*, Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. *Nat. Genet.* **45**, 804–807 (2013). doi: 10.1038/ng.2666; pmid: 23749188
64. J. R. Cerhan, S. L. Slager, Familial predisposition and genetic risk factors for lymphoma. *Blood* **126**, 2265–2273 (2015). doi: 10.1182/blood-2015-04-537498; pmid: 26405224
65. R. J. Leeman-Neill *et al.*, Noncoding mutations cause super-enhancer retargeting resulting in protein synthesis dysregulation during B cell lymphoma progression. *Nat. Genet.* **55**, 2160–2174 (2023). doi: 10.1038/s41588-023-01561-1; pmid: 38049665
66. K. J. Cheung *et al.*, SNP analysis of minimally evolved t(14;18)(q32;q21)-positive follicular lymphomas reveals a common copy-neutral loss of heterozygosity pattern. *Cytogenet. Genome Res.* **136**, 38–43 (2012). doi: 10.1159/000334265; pmid: 22104078
67. A. S. Thomas-Claudepierre *et al.*, The cohesin complex regulates immunoglobulin class switch recombination. *J. Exp. Med.* **210**, 2495–2502 (2013). doi: 10.1084/jem.20130166; pmid: 24145512
68. X. Zhang *et al.*, Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* **575**, 385–389 (2019). doi: 10.1038/s41586-019-1723-0; pmid: 31666703
69. A. Tarsalainen *et al.*, Ig Enhancers Increase RNA Polymerase II Stalling at Somatic Hypermutation Target Sequences. *J. Immunol.* **208**, 143–154 (2022). doi: 10.4049/jimmunol.2100923; pmid: 34862258
70. U. Storb, Why does somatic hypermutation by AID require transcription of its target genes? *Adv. Immunol.* **122**, 253–277 (2014). doi: 10.1016/B978-0-12-800267-4.00007-9; pmid: 24507160
71. P. Dai *et al.*, Transcription-coupled AID deamination damage depends on ELOF1-associated RNA polymerase II. *Mol. Cell* **85**, 1280–1295.e9 (2025). doi: 10.1016/j.molcel.2025.02.006; pmid: 40049162
72. B. Aevermann *et al.*, A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Genome Res.* **31**, 1767–1780 (2021). doi: 10.1101/gr.275569.121; pmid: 34088715
73. I. Tirosh *et al.*, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016). doi: 10.1126/science.aad0501; pmid: 27124452
74. Y. Chen *et al.*, Non-disruptive 3D profiling of combinations of epigenetic marks in single cells. bioRxiv 2025.06.13.659535 [Preprint] (2025). https://doi.org/10.1101/2025.06.13.659535.
75. M. Liu *et al.*, Chromatin tracing and multiplexed imaging of nucleome architectures (MINA) and RNAs in single mammalian cells and tissue. *Nat. Protoc.* **16**, 2667–2697 (2021). doi: 10.1038/s41596-021-00518-0; pmid: 33903756
76. A. Hafner *et al.*, Loop stacking organizes genome folding from TADs to chromosomes. *Mol. Cell* **83**, 1377–1392.e6 (2023). doi: 10.1016/j.molcel.2023.04.008; pmid: 37146570
77. J. H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641–1659.e26 (2020). doi: 10.1016/j.cell.2020.07.032; pmid: 32822575
78. J. R. Moffitt *et al.*, High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11046–11051 (2016). doi: 10.1073/pnas.1612826113; pmid: 27625426
79. R. Dai *et al.*, Three-way contact analysis characterizes the higher order organization of the Tcra locus. *Nucleic Acids Res.* **51**, 8987–9000 (2023). doi: 10.1093/nar/gkad641; pmid: 37534534
80. R. Parthasarathy, Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012). doi: 10.1038/nmeth.2071; pmid: 22688415
81. M. Liu *et al.*, Tracing the evolution of single-cell 3D genomes in Kras-driven cancers. *Nat. Genet.* **57**, 3075–3087 (2025). doi: 10.1038/s41588-025-02297-w; pmid: 40825871
82. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: A generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021). doi: 10.1038/s41592-020-01018-x; pmid: 33318659
83. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024). doi: 10.1038/s41587-023-01767-y; pmid: 37231261
84. C. Fan *et al.*, irGSEA: The integration of single-cell rank-based gene set enrichment analysis. *Brief. Bioinform.* **25**, bbae243 (2024). doi: 10.1093/bib/bbae243; pmid: 38801700
85. L. Montorsi, J. H. Y. Siu, J. Spencer, B cells in human lymphoid structures. *Clin. Exp. Immunol.* **210**, 240–252 (2022). doi: 10.1093/cei/uxac101; pmid: 36370126
86. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
87. N. Abdennur *et al.*, Pairtools: From sequencing data to chromosome contacts. *PLOS Comput. Biol.* **20**, e1012164 (2024). doi: 10.1371/journal.pcbi.1012164; pmid: 38809952
88. A. Chakraborty, J. G. Wang, F. Ay, dcHiC detects differential compartments across multiple Hi-C datasets. *Nat. Commun.* **13**, 6827 (2022). doi: 10.1038/s41467-022-34626-6; pmid: 36369226
89. J. Zhou *et al.*, Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14011–14018 (2019). doi: 10.1073/pnas.1901423116; pmid: 31235599
90. M. Yu *et al.*, SnapHiC: A computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nat. Methods* **18**, 1056–1059 (2021). doi: 10.1038/s41592-021-01231-2; pmid: 34446921
91. J. Hu *et al.*, Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–871 (2016). doi: 10.1038/nprot.2016.043; pmid: 27031497
92. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi: 10.1093/nar/gkw257; pmid: 27079975
93. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
94. I. W. Vock *et al.*, Expanding and improving analyses of nucleotide recoding RNA-seq experiments with the EZbakR suite. *PLOS Comput. Biol.* **21**, e1013179 (2025). doi: 10.1371/journal.pcbi.1013179; pmid: 40609070
95. Y. Cheng *et al.*, Data accompanying manuscript: 3D genome atlas of human tonsil and the role of loop extrusion in B cell somatic hypermutation, version v1, *Zenodo* (2026); https://doi.org/10.5281/zenodo.19959235.

ACKNOWLEDGMENTS

We thank members of the Wang lab and Schatz lab for helpful discussions. We thank J. Craft and I. Kang and their team from Yale Surgical Pathology for providing tonsil samples from tonsillectomy patients with IRB approval at Yale School of Medicine. We thank Yale Pathology Tissue Service-Tissue Distribution and Analysis (YPTS-TDA) for assistance with the collection of tonsil samples. We thank members of F. Alt's laboratory at Harvard Medical School for providing the bait and nested primer sequences in *IGH-VDJ* region and for guidance on 3C-HTGTS methodology. We thank the Yale Center for Genome Analysis (YCGA) and Keck Microarray Shared Resource for Droplet Hi-C library preparation and sequencing. Figures have been drawn in BioRender.com. **Funding:** D.G.S. was partly supported by NIH (U01CA260701 and R01AI127642). S.W. was partly supported by NIH (UH3CA268202, U01CA260701, R01HG011245, DP2GM137414, R01CA292936, R01HG012969, R01HG013503, and P50CA196530-10S1), the Pershing Square Sohn Cancer Research Alliance, the American Federation for Aging Research, and the Hevolution Foundation. This work was supported by NIH 4D Nucleome Consortia grant U01CA260701, awarded to D.G.S. and S.W. **Author contributions:** Conceptualization: D.G.S., S.W.; Funding acquisition: A.H., D.G.S., S.W.;

Investigation: Y.C., J.W., Y.Z., S.J., G.B., D.G.S., S.W.; Methodology: Y.C., J.W., Y.Z., M.L., A.D.Y., D.G.S., S.W.; Supervision: A.H., D.G.S., S.W.; Visualization: Y.C., J.W., Y.Z.; Writing – original draft: Y.C., J.W., Y.Z., D.G.S., S.W.; Writing – review & editing: Y.C., J.W., Y.Z., D.G.S., S.W., with inputs from all authors. **Competing interests:** S.W. is a co-inventor of the MERFISH technology, which is patented by Harvard University. The other authors declare that they have no competing interests. **Data, code, and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. All sequencing data generated for this study have been deposited in the NCBI BioProject database and are available at BioProject ID PRJNA1460889. All analyzed imaging data, sequencing data, data analysis codes generated for this study, sequence maps of newly constructed plasmids, and uncut gel pictures of Western blots are available on Zenodo (*95*). The newly constructed plasmids (pLsg-sgRAD21, pLsg-sgUNG, and RAD21-degron donor plasmid) and the RAD21-degron RASH-1C cell line are available from David G. Schatz under a uniform biological material transfer agreement with Yale University. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adw4243
Figs. S1 to S16; Data S1 to S7; MDAR Reproducibility Checklist
Submitted 1 February 2025; resubmitted 1 March 2026; accepted 20 May 2026
10.1126/science.adw4243

RESEARCH ARTICLE SUMMARY

Dose-dependent sensitivity of human three-dimensional chromatin to a heart disease–linked transcription factor

Zoe L. Grant†, Shuzhen Kuang†, Shu Zhang, Abraham J. Horrillo, Zhe Chen, Kavitha S. Rao, Cemre Celen, Vasumathi Kameswaran, Carine Joubran, Pik Ki Lau, Keyi Dong, Bing Yang, Weronika M. Bartosik, Nathan R. Zemke, Bing Ren, Deepak Srivastava, Irfan S. Kathiriya, Katherine S. Pollard‡*, Benoit G. Bruneau‡*

Full article and list of author affiliations: https://doi.org/10.1126/science.adv5434

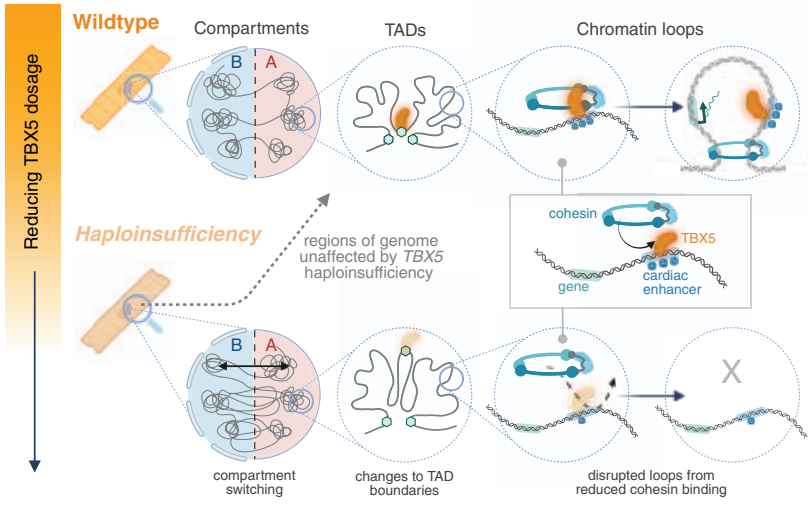
INTRODUCTION: Haploinsufficient mutations of transcription factors (TFs) resulting in reduced gene dosage are an underlying cause of human genetic developmental disorders affecting diverse tissue and cell types. These mutations impact transcriptional regulation and subsequent lineage specification, but exactly how they exert these effects has remained unclear despite decades of study. One key mechanism regulating cell type–specific gene expression is the reorganization of three-dimensional (3D) chromatin during cellular differentiation. In recent years, lineage-specific TFs have emerged as important factors controlling cell type–specific 3D chromatin interactions alongside the ubiquitously expressed loop extrusion complex cohesin and the boundary-associated factor CTCF.

RATIONALE: TBX5 is a cardiac lineage–specific TF. Haploinsufficient mutations in this gene cause congenital heart disease (CHD), and loss of one copy of *TBX5* results in transcriptional dysregulation of both atrial and ventricular cardiomyocytes (CMs) in human patients and stem cell models. We hypothesized that reduced dosage of TBX5 impacts gene regulation through changes in cardiac-specific 3D chromatin organization. Using human induced pluripotent stem cell differentiations into atrial and ventricular CMs, we studied 3D genome dynamics and characterized the role of TBX5 dosage in regulating CM lineage–specific genome organization.

RESULTS: In wild-type cells that carry two copies of the *TBX5* gene, we observed extensive, dynamic reorganization of the 3D genome during atrial and ventricular CM differentiation with clear distinctions between the two CM types, in both bulk populations and single cells. TBX5 was enriched at both topologically associating domain (TAD) boundaries and chromatin loop anchors, 3D genome structures that emerge during cardiac differentiation. Cells harboring a single functional copy of *TBX5* revealed changes in higher-order organization of chromosomal compartments, TADs, and loops. These phenotypes were exacerbated by complete loss of *TBX5*, highlighting the key dosage-sensitive regions of the genome. Genes whose expression is known to depend on TBX5 were associated with chromatin loop changes. Although CTCF was largely unaffected by loss of TBX5, cohesin occupancy was compromised by reduced TBX5 dosage, mainly at TBX5-bound enhancers in chromatin loops, which failed to remodel during CM differentiation.

CONCLUSION: Our findings demonstrate that cell type–specific 3D genome organization is sensitive to the dosage of a lineage-restricted TF and suggest a mechanism by which reduced TF dosage may contribute to CHDs. Future studies will be needed to examine how generalizable this mechanism is to haploinsufficient TFs other than TBX5 and to developmental disorders affecting other cell lineages. □

*Corresponding author. Email: benoit.bruneau@gladstone.ucsf.edu (B.G.B.); katherine.pollard@gladstone.ucsf.edu (K.S.P.) †These authors contributed equally to this work. ‡These authors contributed equally to this work. **Cite this article as Z. L. Grant *et al.*, *Science* 393, eadv5434 (2026). DOI: 10.1126/science.adv5434**



TBX5 dose-dependently organizes cardiac 3D chromatin. Loss of TBX5 during cardiac differentiation disrupts 3D chromatin organization at multiple scales, including a subset of cardiac-specific A/B compartments, TAD boundaries (green hexagon), and chromatin loops. Failed loop formation may be due to reduced binding of cohesin at TBX5-bound enhancer regions. [Figure created with Biorender.com]

RESEARCH ARTICLE

Dose-dependent sensitivity of human three-dimensional chromatin to a heart disease–linked transcription factor

Zoe L. Grant[1]†, Shuzhen Kuang[1]†, Shu Zhang[1,2], Abraham J. Horrillo[1,3], Zhe Chen[1], Kavitha S. Rao[1], Cemre Celen[1], Vasumathi Kameswaran[1], Carine Joubran[1,4], Pik Ki Lau[5,6], Keyi Dong[5,6], Bing Yang[5,6], Weronika M. Bartosik[5,6], Nathan R. Zemke[5,6], Bing Ren[5,6], Deepak Srivastava[1,7,8], Irfan S. Kathiriya[9], Katherine S. Pollard[1,10,11]‡*, Benoit G. Bruneau[1,7,12]‡*

Dosage-sensitive transcription factors (TFs) underlie altered gene regulation in human developmental disorders, and cell type–specific gene regulation is linked to the reorganization of three-dimensional (3D) chromatin during cellular differentiation. In this work, we show dose-dependent regulation of chromatin organization by the congenital heart disease (CHD)–linked, lineage-restricted TF TBX5 in human cardiomyocyte differentiation. Genome organization, including compartments, topologically associated domains, and chromatin loops, was sensitive to reduced *TBX5* dosage in a human model of CHD, with variations in response across individual cells. Cohesin binding was reduced at TBX5-bound enhancer elements in a TBX5 dose-dependent manner, providing a potential mechanism for disrupted loop formation. These results highlight the importance of lineage-restricted TF dosage in cell type–specific 3D chromatin dynamics, suggesting a mechanism for TF-dependent disease.

Our understanding of the genetic foundation of most human developmental disorders primarily stems from dominant mutations in gene regulators, notably transcription factors (TFs) and chromatin-modifying factors (*1*, *2*). Many of these mutations are predicted or shown to result in haploinsufficiency, where loss of a single copy of a gene leads to disease. The TFs involved are well-studied in the context of embryonic development and congenital defects; for example, *PAX6* in aniridia, *SOX9* in campomelic dysplasia, *NOTCH1* in bicuspid aortic valve, and *TBX5*, *NKX2-5,* and *GATA4* in congenital heart disease (CHD) (*2*). The major consequence of TF haploinsufficiency is transcriptional dysregulation. However, how TF haploinsufficiency affects downstream transcriptional networks is not mechanistically understood despite decades of study.

The genome is a three-dimensional (3D) structure organized into hierarchical layers that modulate transcriptional activity through the formation of active (A) and repressive (B) compartments, topologically associating domains (TADs), and chromatin loops (*3*). The formation of 3D, long-range interactions between cis-regulatory elements, such as enhancers and promoters, is especially important for regulating cell type–specific gene expression (*4*, *5*). Indeed, different cell types are characterized by lineage-specific A and B compartments, TADs, and chromatin loops (*6*, *7*). Chromatin loops are held together by CTCF, which is found at loop anchors and TAD boundaries, where it acts to halt cohesin-dependent loop extrusion (*3*, *8*–*10*). Because CTCF and cohesin complex members are expressed in all cell types, other factors must be responsible for cell type–specific 3D genome organization. This includes lineage-constrained TFs as potentially direct regulators of 3D genome organization at one (*11*, *12*) or many loci (*13*–*19*). Given the specificity of TF expression, it is likely that a distinct TF or set of TFs regulates 3D genome organization in each cell type.

The T-box TF TBX5 is a master regulator of cardiac gene expression required for normal heart patterning and morphogenesis (*20*). TBX5 regulates target gene expression in a dose-dependent manner (*20*–*22*). The clinical relevance of this dose dependency is demonstrated by heterozygous loss-of-function mutations in *TBX5*, which lead to Holt-Oram Syndrome (*23*, *24*). These haploinsufficient *TBX5* mutations cause 85% penetrant CHD that are primarily atrial and ventricular septal defects and conduction system defects (*23*, *24*). *TBX5* is also a strong locus for susceptibility to atrial fibrillation (*25*–*27*). Using an induced pluripotent stem cell (iPSC) *TBX5* allelic series consisting of wild-type (WT) and heterozygous or homozygous loss-of-function mutations, we showed that reducing *TBX5* dosage induced widespread changes in gene expression in atrial and ventricular cardiomyocytes (CMs) (*21*, *28*). Whether altered 3D chromatin is part of the mechanism that affects transcriptional dysregulation owing to haploinsufficient CHD-causing *TBX5* mutations and whether TBX5 generally influences how 3D chromatin is established in a lineage-restricted way remain to be determined.

In this work, we present a comprehensive analysis of the impact of *TBX5* dosage on 3D chromatin structure during the differentiation of human iPSC–derived CMs. Our results demonstrate that cell type–specific 3D genome organization is sensitive to the dosage of a lineage-restricted TF and suggest a mechanism by which reduced TF dosage may contribute to CHD.

Dynamic reorganization of the 3D genome accompanies human cardiac differentiation

Analyses of 3D chromatin during CM differentiation have been carried out before (*29*, *30*) but not at the resolution that would enable precise detection of chromatin loops and their anchors. This resolution is particularly important for detecting chromatin interactions that rely on TFs. To this end, we assessed chromatin contacts during atrial CM differentiation by Hi-C 3.0, an iteration of in situ Hi-C that is optimized for high-resolution detection of loops in addition to accurate detection of larger structures, such as TADs and compartments (*31*). We collected two biological replicates from key cardiac differentiation time points, including pluripotency [day 0 (d0)], cardiac mesoderm (d2 to d4), cardiac precursors (d6), and various CM stages (d11, d20, and d45) (Fig. 1A). At d45, ~90% of cells expressed the CM marker cardiac muscle troponin T (cTnT+) as assessed by flow cytometry, indicating highly efficient CM differentiation. We sequenced libraries to an average depth of 1.69 billion raw read pairs and 761 million unique cis interactions per time point, which allowed us to reach a resolution of 5 kilobases (kb) (fig. S1A). Biological replicates were highly reproducible and were pooled for downstream analysis (fig. S1, B and C). To assess correlations between chromatin contacts and gene expression, we also generated bulk RNA sequencing (RNA-seq) libraries at each time point (data S1).

[1]Gladstone Institutes, San Francisco, CA, USA. [2]Bioinformatics Graduate Program, University of California, San Francisco, CA, USA. [3]TETRAD Graduate Program, University of California, San Francisco, CA, USA. [4]Developmental and Stem Cell Biology Graduate Program, University of California, San Francisco, CA, USA. [5]Department of Cellular and Molecular Medicine, University of California, San Diego, School of Medicine, La Jolla, CA, USA. [6]Center for Epigenomics, University of California, San Diego, School of Medicine, La Jolla, CA, USA. [7]Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, USA. [8]Department of Pediatrics and Biochemistry, University of California, San Francisco, CA, USA. [9]Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA, USA. [10]Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, USA. [11]Chan Zuckerberg Biohub, San Francisco, CA, USA. [12]Department of Pediatrics, Cardiovascular Research Institute, Institute for Human Genetics, and the Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA. ***Corresponding author. Email: benoit.bruneau@gladstone.ucsf.edu (B.G.B.); katherine.pollard@gladstone.ucsf.edu (K.S.P.)**
†These authors contributed equally to this work. ‡These authors contributed equally to this work.

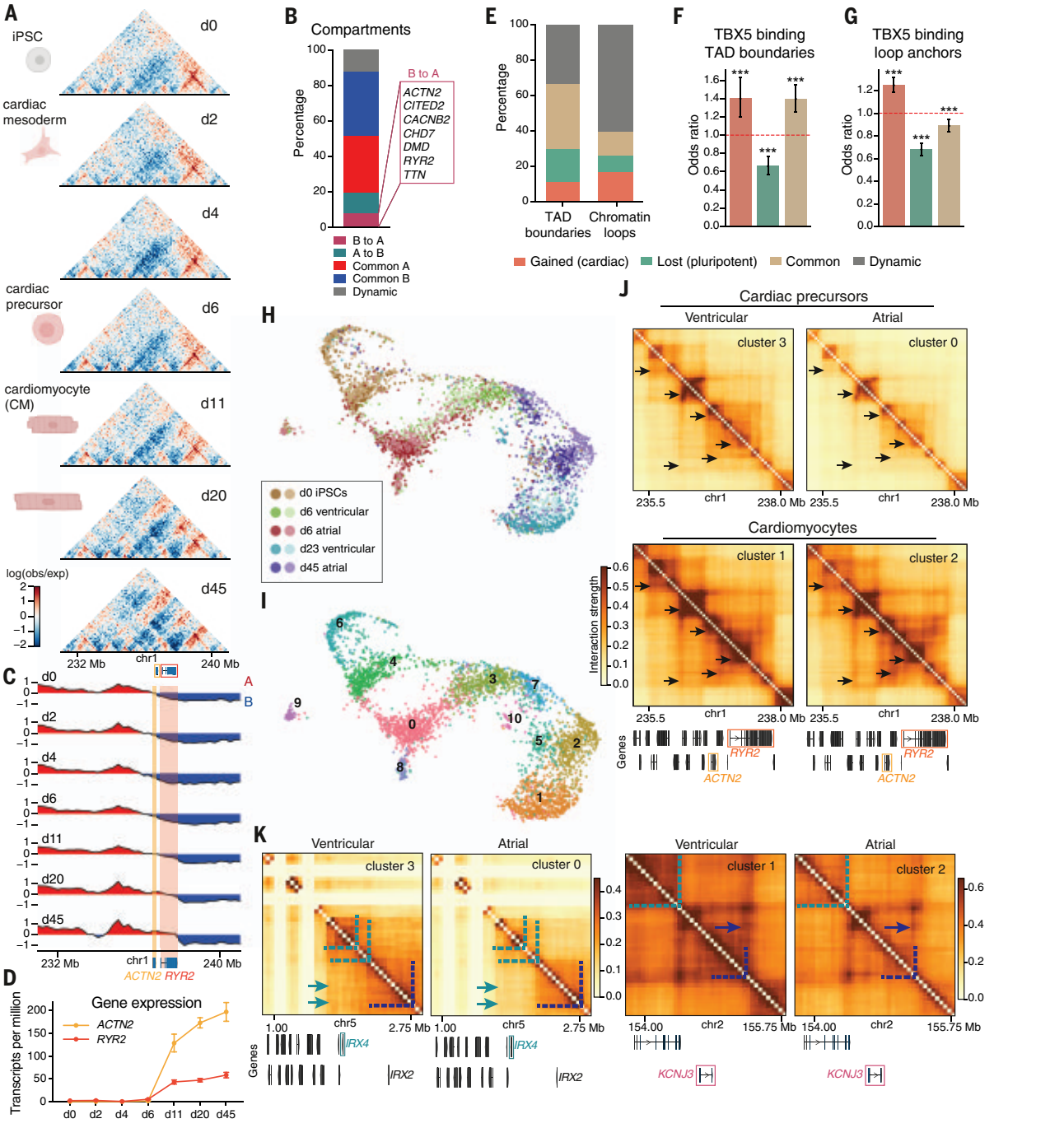


Fig. 1. Dynamic reorganization of human cardiac chromatin. (**A**) Observed over expected contact maps for a 10-Mb region around *ACTN2* and *RYR2* during atrial cardiac differentiation. Red and blue indicate regions with more or fewer contacts than expected, respectively. [Cell differentiation figure created with Biorender.com] (**B**) Percentage of compartments that do or do not switch during differentiation. (**C** and **D**) Principal components scores showing progressive transition of *ACTN2* and *RYR2* loci from B to A compartments during differentiation (C) and corresponding increase in RNA-seq gene expression (mean ± SEM) (D). (**E**) Percentage of TAD boundaries and chromatin loops that are common or change. (**F** and **G**) OR plots of TBX5 binding at gained, lost, and common (F) TAD boundaries or (G) loop anchors. (**H** and **I**) UMAP of atrial and ventricular differentiations, where the features are derived from Fast-Higashi embeddings of Hi-C contacts at 1 Mb resolution. Cells are colored by (H) time point or differentiation, and two shades indicate biological replicates, or (I) Leiden clusters. (**J** and **K**) Higashi-imputed contact matrices of (J) Leiden clusters 0 to 3 around cardiac-enriched genes *ACTN2* and *RYR2* (black arrows indicate regions of increased contact frequency in CM clusters 1 and 2) and (K) ventricular (d23) and atrial (d45) CM-enriched clusters around atrial-enriched gene *KCNJ3* and ventricular-enriched gene *IRX4*. Arrows and dashed lines indicate regions of increased contact frequency in atrial (navy) and ventricular (teal) CMs. Scales in (J) and (K) show the Higashi-imputed contact values, normalized by the range of values in cluster 1. Some gene annotations have been omitted for clarity in (A), (C), (J), and (K). Two merged biological replicates per time point. ORs compare each subset to all others; dotted lines at OR = 1 indicate no enrichment, OR scores > 1 indicate enrichment, and OR scores < 1 indicate depletion; error bars indicate 95% confidence intervals; ***$P < 0.001$.

Consistent with other studies (*29*, *30*), we observed large-scale chromatin reorganization during cardiac differentiation (Fig. 1A). We called A/B compartments for each time point and found that although most of the genome (68.3%) remained in the same compartment throughout differentiation, some regions (19.5%) switched once from A to B or from B to A, and a smaller number (12.2%) switched multiple times (Fig. 1B, fig. S1D, and data S2). Focusing on genes with expression highly correlated with cardiac-specific A compartments (d11 onward, Pearson correlation coefficient $r \geq 0.8$), gene ontology analysis showed terms associated with heart morphogenesis and cardiac function (fig. S1E). This included a number of essential cardiac genes, such as *TTN*, *DMD*, *RYR2*, *ACTN2*, *CITED2*, *CACNB2,* and *CHD7*. We observed that loci containing large structural or contractile protein genes, such as *RYR2/ACTN2* (1147 kb), *TTN* (281 kb), *DMD* (2092 kb), and *CACNB2* (403 kb), were insulated within B compartments at earlier time points and that A compartments expanded to cover these genes as cells differentiated (Fig. 1, B and C; fig. S1F; and data S2). Increased gene expression sometimes preceded these chromatin changes (Fig. 1, A and D, and fig. S1G), suggesting that lineage-specific transcriptional regulation or transcriptional activity itself is a driver of compartment switching.

TADs were more dynamic than compartments, with only 36.6% of TADs shared across all stages of cardiac differentiation (Fig. 1E and fig. S1H). Chromatin loops were even more dynamic, as only 13.5% were shared among all time points (Fig. 1E and data S2). This finding is consistent with observations in other cell types that suggest that chromatin loops form in a highly cell type–specific manner (*7*). We identified a total of 54,381 chromatin loops in CMs, a fifth of which (10,639) were specific to d11 to d45 CMs (Fig. 1E and fig. S1I). As expected, CM-specific chromatin loops had stronger interactions in CMs than in pluripotent cells (fig. S1J).

To understand whether TBX5 is involved in regulating cardiac-specific 3D chromatin, we assessed TBX5 binding during differentiation and correlated its binding profile with changes in 3D chromatin. For this, we engineered a biotin-tagged TBX5 WT iPSC line and assessed TBX5 binding by chromatin immunoprecipitation sequencing (ChIP-seq). We found that TBX5 binding was enriched to similar degrees at both CM-specific TAD boundaries and TAD boundaries common to all time points (Fig. 1F). By contrast, TBX5 binding was more enriched at CM-specific chromatin loop anchors than at loop anchors common to all time points (Fig. 1G). These results suggest that TBX5 may play a role in establishing chromatin loops in a CM-specific manner.

Together, these data highlight that the human genome is dynamically reorganized across all scales, from compartments to TADs to chromatin loops, as pluripotent cells are directed toward the CM cell lineage.

Chamber-specific cardiomyocyte subtypes display distinct 3D chromatin organization

TBX5 is expressed in both atrial and ventricular CMs in human heart development (*32*). Atrial and ventricular CMs have distinct functional and electrophysiological properties (*33*, *34*) and can be distinguished transcriptionally (fig. S2, A and B, and data S1). To determine whether these differences are driven by distinct 3D genome organization, we generated a Hi-C 3.0 time course of ventricular CM differentiation. As with atrial differentiation, we observed dynamic changes in compartments, TADs, and chromatin loops between differentiation stages (fig. S3, A to F, and data S2). Comparing atrial d45 to ventricular d23 CMs, at which point cells had reached similar differentiation maturity in both protocols (fig. S2B), we found that approximately 8.1% of compartments, 33.4% of TADs, and 48.1% of chromatin loops were cell type–specific (fig. S3, G to I). Thus, although 3D chromatin is dynamic during differentiation of pluripotent cells into CMs, the majority of compartments and more than half of the chromatin contacts are conserved between the two CM types that we analyzed. Nevertheless, we found that the expansion of A compartments across large structural protein–encoding genes, such as *TTN* and *RYR2*, began slightly earlier in the ventricular CM differentiations, indicating some differences in timing of identity acquisition at the chromatin level (fig. S3, J to L). Supporting this, ventricular CMs at d6 also appeared to be slightly more advanced transcriptionally than d6 atrial CMs (fig. S2B).

To determine whether the two CM types could be distinguished by their 3D chromatin contacts, we differentiated iPSCs into either atrial or ventricular CMs and profiled chromatin contacts and DNA methylation in individual cells using single-cell methyl–Hi-C sequencing (snm3C-seq) (*35*, *36*). Single-cell resolution allowed us to assess the heterogeneity in 3D chromatin organization within and between CM differentiations. We generated snm3C-seq libraries for two biological replicates at three time points for each differentiation: iPSCs (d0), cardiac precursors (d6 for both atrial and ventricular differentiations), and late CMs (d23 for ventricular and d45 for atrial, which consisted of ~89 and ~81% cTnT+ cells, respectively).

After sequencing, we obtained on average 451,095 chromatin contacts per nucleus for 3304 nuclei passing quality control (fig. S4, A and B). Clustering cells on the basis of their chromatin contacts showed that cells from atrial and ventricular differentiations formed multiple distinct clusters at both cardiac precursor and CM stages (Fig. 1H), thus confirming that atrial and ventricular CMs can be distinguished by genome-wide 3D chromatin organization and have some degree of heterogeneity in their chromatin organization. Clustering by DNA methylation or combining DNA methylation with chromatin contacts revealed temporal differences (fig. S4, C and D). However, all cardiac precursors (d6) clustered together, and differences between atrial and ventricular populations only emerged at later stages of differentiation. This suggests that changes in DNA methylation do not drive atrial versus ventricular fate to the same extent as changes in 3D chromatin. Using Leiden clustering, a graph-based algorithm that optimizes the grouping of connected data points, we identified 11 distinct clusters that clearly separated the different time points and differentiation types from each other (Fig. 1I and fig. S4E). Focusing on the clusters with the largest number of cells, we saw changes in chromatin contacts around cardiac function genes, such as *ACTN2/RYR2* and *TTN* between cardiac precursors and late CMs in both differentiations (Fig. 1J and fig. S5, A and B). In addition, we identified differences in contacts around the atrial-specific gene *KCNJ3* and the ventricular gene *IRX4* when comparing atrial with ventricular cells (Fig. 1K and figs. S5C and S6). Overall, these results show that atrial and ventricular CMs can be distinguished by their 3D genome early during differentiation and that these differences persist at more mature stages.

TBX5 influences A/B compartment switching in a dose-dependent manner

Because we found that TBX5 was enriched at cardiac-specific chromatin loop anchors and to some extent TAD boundaries (Fig. 1, F and G), we asked whether it may organize cardiac chromatin. For this, we performed Hi-C 3.0 in our previously generated cell lines (*21*), representing a *TBX5* allelic series (WT, $TBX5^{+/+}$; heterozygous, $TBX5^{in/+}$; null, $TBX5^{in/del}$) (Fig. 2A). $TBX5^{in/+}$ CMs display sarcomere disarray indicating contractile dysfunction and recapitulate electrophysiological defects observed in human patients in the form of prolonged calcium transients. Both of these phenotypes are exacerbated in $TBX5^{in/del}$ CMs (*21*, *28*).

We generated libraries for Hi-C 3.0 for two biological replicates per *TBX5* allelic series genotype at d20 of atrial CM differentiation. As assessed by flow cytometry, the WT CMs were ~76% cTnT+. The $TBX5^{in/+}$ line consistently produced beating CMs, albeit with reduced cTnT+ populations compared with those in the WT (*28*), and the samples we collected for Hi-C 3.0 contained ~62% cTnT+ cells. $TBX5^{in/del}$ iPSCs rarely differentiate into CMs, and only ~10% of the cells we collected

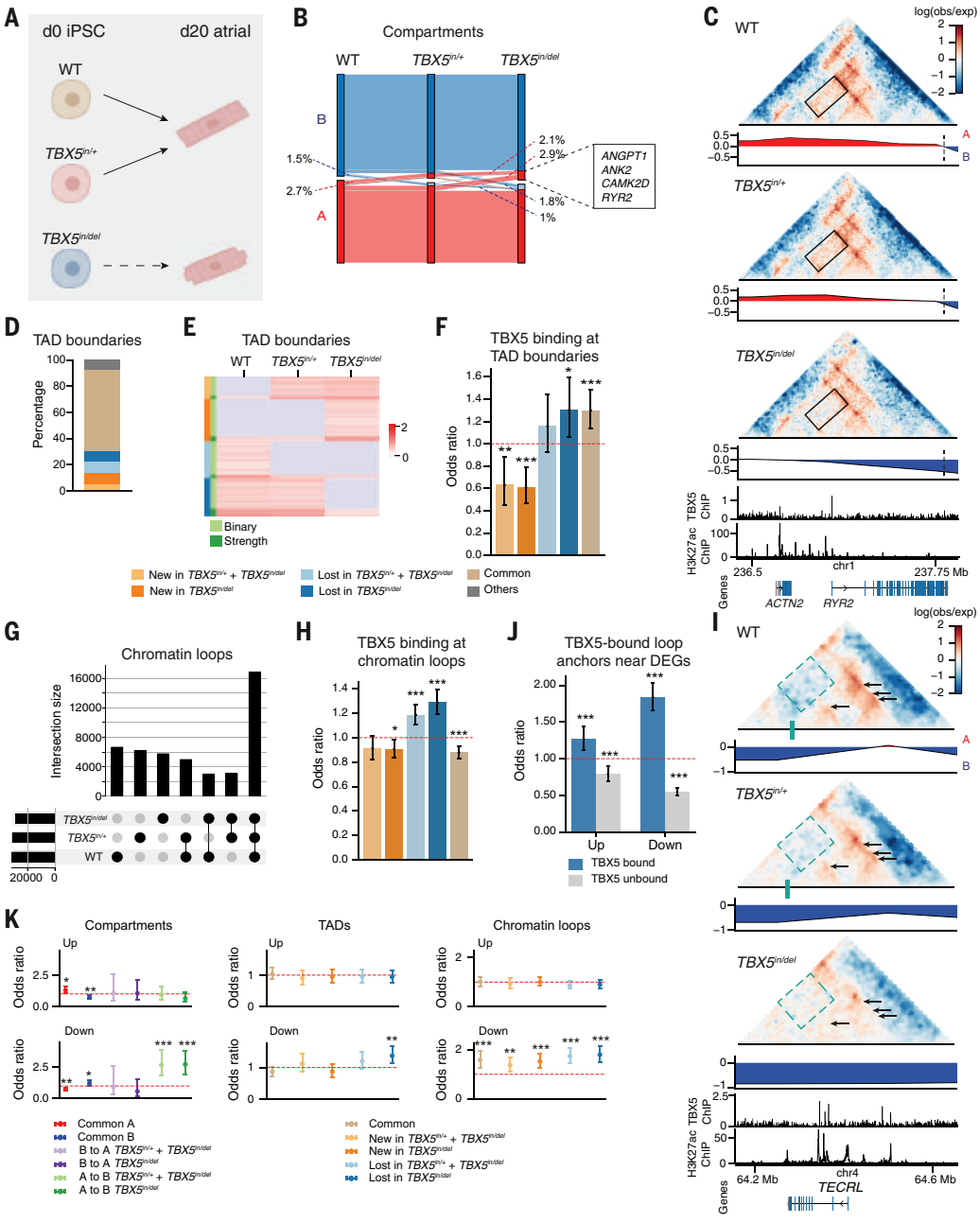


Fig. 2. TBX5 dose-dependently influences 3D chromatin organization. (**A**) *TBX5* allelic series. [Figure created with BioRender.com] (**B**) Sankey plot showing compartment A/B switching. (**C**) Observed over expected contact map with A/B compartment principal components and WT TBX5 and H3K27ac ChIP-seq tracks around *ACTN2* and *RYR2*. The boxed area indicates decreased contact frequency in $TBX5^{in/del}$ cells. The dotted line indicates the WT A/B compartment boundary. (**D**) Percentage of TAD boundaries that are common or genotype specific. (**E**) Heatmap of gained or lost TAD boundaries. Binary and strength indicate methods used for identifying TAD boundary changes. Scale indicates the insulation strength of TAD boundaries. 0, boundaries not defined in a sample. (**F**) OR of TBX5 binding at TAD boundaries. (**G**) Upset plot of consensus chromatin loops between two or more genotypes. (**H**) OR of TBX5 binding at loop anchors. (**I**) Same caption as (C), around *TECRL*. The green bar indicates the TAD boundary in WT and $TBX5^{in/+}$. The boxed area indicates increased contact frequency in $TBX5^{in/del}$ cells. Arrows indicate lost loops in $TBX5^{in/+}$ and/or $TBX5^{in/del}$ cells. (**J** and **K**) OR of the association of up-regulated ($\log_2FC \geq 0.2$, $P_{adj} < 0.05$) or down-regulated ($\log_2FC \leq -0.2$, $P_{adj} < 0.05$) genes (J) within 5 kb of TBX5-bound loop anchors and (K) with common or changed compartments, TADs, and chromatin loops (±5 kb of anchors). Two merged biological replicates per genotype. ORs compare each subset to all others; dotted lines at OR = 1 indicate no enrichment, OR scores > 1 indicate enrichment, and OR scores < 1 indicate depletion; error bars indicate 95% confidence intervals; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

were cTnT+ (fig. S7A). We sequenced Hi-C 3.0 libraries to an average depth of 1.29 billion raw read pairs and 706 million unique cis interactions per genotype (fig. S7B), providing 5-kb resolution. The two $TBX5^{in/+}$ samples differed in their percentage of long cis contacts, which may reflect the phenotypic variation of heterozygous cells (fig. S7, C and D). Nevertheless, $TBX5^{in/+}$ samples were overall more similar to each other than to the WT, and we pooled biological replicates for quantitative analysis (fig. S7E).

We observed clear, dose-dependent compartment changes in $TBX5^{inl+}$ and $TBX5^{in/del}$ cells relative to WT cells (data S3). For example, the

frequency of compartment switches was twice as high in *TBX5*$^{in/del}$ cells (4.9% of A compartments switched to B, and 2.8% of B compartments switched to A) as in *TBX5*$^{in/+}$ cells (2.7 and 1.5%, respectively) (Fig. 2B) and showed dose-dependent changes in compartment strength (fig. S8A). Overall, there was a small increase in the total number of B compartments in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ CMs (fig. S8B). Compartments that switched from A to B in a TBX5 dose-dependent manner included important CM genes, such as *ANGPT1*, *ANK2*, *CAMK2D*, and *SCN10A* (Fig. 2B). In addition, *ACTN2* and *RYR2*, which switched from B to A during differentiation of WT cells, had weaker A compartments in *TBX5*$^{in/+}$ cells and remained in B compartments in *TBX5*$^{in/del}$ cells, indicating that the expansion of some cardiac A compartments depends on TBX5 (Fig. 2C). These findings highlight the importance of TBX5 for maintaining the correct spatial regulation of lineage-defining genes.

TAD boundaries and chromatin loops are dynamically changed with reducing TBX5 dosage

Reduced dosage of TBX5 led to major changes in TAD organization and looping (data S3). We found that in WT, ~38% of all TBX5 binding sites were within 10 kb of a TAD boundary or loop anchor (12 and 35% at TAD boundaries and/or at loop anchors, respectively). Overall, ~40% of TAD boundaries identified across the genotypes were either lost or new in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells compared to WT (Fig. 2, D and E). TAD boundaries that were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells were highly likely to be bound by TBX5 in WT cells (Fig. 2F). TAD boundaries that were common to all three genotypes were also enriched for TBX5 binding, suggesting that only some TAD boundaries depend on TBX5 binding for their formation or maintenance. In contrast, the boundaries of the new *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ TADs tended not to be bound by TBX5 in WT cells (Fig. 2F). Therefore, we focused our analysis on TADs and loops that were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ samples.

To complement our TBX5 ChIP dataset, we generated CTCF ChIP-seq data in each of the three genotypes. Of the WT CTCF binding sites, 11% were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells. Notably, the main impact of reduced TBX5 dosage on CTCF binding was gain rather than loss of binding sites, leading to an increase in total CTCF peaks in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells (fig. S8C). Both new and lost TADs in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells were not associated with changes in CTCF binding (fig. S8D), highlighting that mechanisms other than CTCF binding likely regulate these TAD changes. Of CTCF peaks lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells, only 6% were bound by TBX5. Although this was a small proportion of lost CTCF peaks, they were enriched at TBX5-bound regions compared with other CTCF peaks, suggesting that, in some cases, TBX5 is needed for normal CTCF binding (fig. S8E).

Because TBX5 binding was enriched at CM-specific chromatin loops, we predicted that the major impact of reducing TBX5 would be on chromatin loop formation. Reduced *TBX5* led to both loss of loops and newly formed loops (Fig. 2G), highlighting the dynamic changes associated with *TBX5* loss. As with new TADs, the anchors of new loops were not enriched for TBX5 binding in WT cells (Fig. 2H), indicating that they are regulated by TBX5-independent or downstream mechanisms, which may include partner TF redistribution (*37*). By contrast, TBX5 binding was enriched at the anchors of loops lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells (Fig. 2H), indicating that chromatin loop formation is especially vulnerable to reduced dosage of TBX5.

To determine whether changes in chromatin loops disrupted putative enhancer or promoter interactions, we performed acetylated histone 3 lysine 27 (H3K27ac) ChIP-seq in WT d20 atrial CMs. Loops with anchors involving H3K27ac enhancer regions were lost with heterozygous levels of *TBX5*, whereas loops with anchors at H3K27ac active promoters became sensitized with complete loss of TBX5 (fig. S8F). Of loops lost in both *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ samples, 35.7% involved an active H3K27ac+ enhancer at one loop anchor (fig. S8G). A number of chromatin loops with an anchor at the *TBX5* promoter or gene body were bound by TBX5, either at the anchor or within the looped region, in WT CMs. Some of these loops were lost in *TBX5*$^{in/+}$ CMs, and all of them were lost in the *TBX5*$^{in/del}$ CMs, suggesting that TBX5 regulates its own 3D genome architecture and possibly gene expression (fig. S8H). Only one loop anchor (out of five loops anchored within the *TBX5* locus) spanned exon 3 of *TBX5*, the single guide RNA (sgRNA)–targeted site used to engineer the allelic series (*21*), highlighting a largely locus deletion–independent effect.

Dose-dependent changes in chromatin organization at all scales are exemplified around the TBX5 target gene *TECRL*, whose expression depends on TBX5 dosage (*21*, *28*). In WT CMs, the *TECRL* gene body was within a weak B compartment, with a ~50-kb weak A compartment upstream of the promoter region. The strength of the B compartment gradually increased as TBX5 dosage decreased, whereas the weak A compartment was completely abrogated by the loss of only one copy of *TBX5* (Fig. 2I). The TAD boundary that insulates the *TECRL* TAD in both WT and *TBX5*$^{in/+}$ CM was lost in *TBX5*$^{in/del}$ cells, leading to increased contact frequency between neighboring TADs (boxed area in figure). Many chromatin loops were identified with anchors inside the *TECRL* gene body as well as between the promoter and the upstream euchromatic region in WT CMs. Half of these chromatin loops were lost in *TBX5*$^{in/+}$ CMs, and the majority were lost in *TBX5*$^{in/del}$ cells. We found that TBX5 binding sites were enriched at H3K27ac ChIP-seq peaks within chromatin loops anchored at the *TECRL* promoter, suggesting that the presence of TBX5 is required at these sites for correct 3D genome organization (Fig. 2I).

Reduced transcription of TBX5 target genes is linked to altered 3D chromatin

Given the transcriptional dysregulation associated with reduced TBX5 levels (*21*, *28*) (fig. S9A), we also wanted to understand whether transcriptional changes were linked to changes in 3D chromatin organization and TBX5 binding. Genes down-regulated when *TBX5* was reduced tended to map within 5 kb of loop anchors bound by TBX5 (Fig. 2J). Both new and lost loops were associated with down-regulated genes (Fig. 2K). This suggests that a failure to correctly organize 3D chromatin around these genes disrupts their expression, which may include lost loops to enhancers and new loops to repressor elements. Supporting this, loops lost between enhancers and promoters were preferentially associated with down-regulated genes (fig. S9B). Down-regulated genes were also highly enriched within compartments that underwent an A to B switch and in TADs that were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells (Fig. 2K). These findings are consistent with TBX5 acting as a direct activator of the transcription of these genes.

By contrast, up-regulated genes in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells were not enriched at chromatin structures that changed (Fig. 2K). TBX5-bound loop anchors were enriched within 5 kb of up-regulated genes but not to the same extent as down-regulated genes (Fig. 2J). This suggests that TBX5 binding tends to mediate the activation of target genes rather than their repression. TBX5 has been shown to act as a transcriptional repressor by recruiting the NuRD complex at a number of genes (*38*). It is possible that some up-regulated genes are repressed by TBX5, but others may be due to secondary effects, such as the down-regulation of other genes or a redistribution of TFs that normally cooperate with TBX5, leading to ectopic gene expression, as previously shown (*37*).

Many cardiac loops are preserved but weakened by acute depletion of TBX5 from cardiomyocytes

To determine whether the presence of TBX5 was directly required for loop maintenance in d20 CMs, we used a system in which TBX5 depletion could be induced. We engineered a FKBP^{F36V} degradation tag into the N terminus of TBX5, targeting TBX5 for degradation upon binding its

ligand dTAG^V-1. We treated d20 CMs with dimethyl sulfoxide (DMSO) or dTAG^V-1 for 24 hours before collecting for Hi-C 3.0 library preparation (two biological replicates per treatment) and RNA-seq (three biological replicates per treatment). All CM samples were beating before and after 24 hours of treatment with DMSO or dTAG^V-1 ligand and consisted of ~73% cTnT+ CMs (fig. S10A). TBX5 protein was completely depleted after 24 hours (fig. S10B), which led to a number of transcriptional changes [$|\log_2 FC| > 1$ (FC, fold change), 79 up-regulated and 183 down-regulated genes in 24 hours of dTAG^v-1 compared with 24 hours of DMSO] (fig. S10C). Some of the down-regulated genes were consistent with those found by single-cell RNA sequencing (scRNA-seq) in our *TBX5* allelic series (figs. S9A and S10C).

We sequenced the Hi-C 3.0 libraries to an average depth of 1.44 billion raw read pairs and 674 million unique cis interactions per genotype (fig. S10D), providing 5-kb resolution. Compared with sustained *TBX5* dosage reduction (Fig. 2), acute TBX5 depletion led to fewer A/B compartment switches and fewer TAD boundary changes (fig. S10, E and F, and data S4). By contrast, it caused 31.7% of chromatin loops to change (fig. S10G), and half of the lost loops were also lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ samples ($P < 0.0001$). TBX5 binding was slightly enriched at anchors of lost loops (fig. S10H), indicating that TBX5 binding has some influence on maintaining chromatin loops. Visual inspection of contact frequency maps showed that many of the chromatin contacts appear preserved after acute TBX5 reduction, including at the *TECRL* locus (fig. S10I), which was impacted by sustained depletion (Fig. 2I). However, quantitative analyses indicated that some contacts were weakened, including the *HEY1* locus, which had quantitative changes in loop formation and corresponding down-regulation in mRNA expression in TBX5-depleted cells (fig. S10J). The general preservation but weakened chromatin loops indicates that chromatin loops are stable for at least 24 hours after TBX5 removal and, hence, that the loop changes that we detected in d20 in *TBX5*$^{in/+}$ or *TBX5*$^{in/del}$ cells reflect an earlier role for TBX5 in chromatin contact formation and a later role in maintaining loop strength.

Cardiomyocytes show heterogeneity in chromatin contacts at TBX5 dosage–sensitive loci

Variations in chromatin architecture due to loss of TBX5 may not affect all cells uniformly. We therefore assessed chromatin contacts at the single-cell level using snm3c-seq in WT, *TBX5*$^{in/+}$, and *TBX5*$^{in/del}$ cells at d20 of atrial CM differentiation for two biological replicates per genotype. Samples were ~74% cTnT+ in WT, ~62% cTnT+ in *TBX5*$^{in/+}$ and ~18% cTnT+ in *TBX5*$^{in/del}$ samples (fig. S11A). We obtained on average ~1.5 million reads and 761,900 chromatin contacts per nucleus for 2170 nuclei passing quality control (fig. S11, B and C).

We clustered cells using chromatin contacts with and without integrating methylation data, identifying clear differences between WT and *TBX5*$^{in/+}$ cells compared with *TBX5*$^{in/del}$ cells (Fig. 3A and fig. S11D). Similar to our atrial and ventricular time course data, we found that the separation by methylation was less distinct than separation by chromatin contacts (fig. S11E). Indeed, visualizing methylation genome tracks did not reveal methylation differences (fig. S11F). Focusing on chromatin contact clusters, we performed Leiden clustering and found that most of the WT and *TBX5*$^{in/+}$ cells fell into two separate clusters (clusters 0 and 2). By contrast, *TBX5*$^{in/del}$ samples clustered separately from the WT and *TBX5*$^{in/+}$ samples but also distinctly (clusters 1, 3, and 5). (Fig. 3, A and B, and fig. S11G). This clustering analysis clearly highlighted both inter- and intragenotype heterogeneity, which we could distinguish by examining cluster-specific contact frequency maps (Fig. 3, C and D, and fig. S12A). At the TBX5 dosage–sensitive gene *TECRL*, we found that the two clusters containing a majority of WT and *TBX5*$^{in/+}$ cells showed increased contacts around the gene compared with the clusters containing mostly *TBX5*$^{in/del}$ cells (Fig. 3C, dashed blue lines, and fig. S12B), and *TBX5*$^{in/del}$ cluster 3 cells had a broader interacting domain (Fig. 3C, dashed black lines, and fig. S12B). However, there were also differences in contact frequency between the two WT and *TBX5*$^{in/+}$-enriched clusters (white arrows, Fig. 3C) and between two *TBX5*$^{in/del}$-enriched clusters (black arrows, Fig. 3C). At another locus containing *ANK2* and *CAMK2D*, genes that showed dose-dependent changes in chromatin contacts in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells (Fig. 2B), we saw changes in contacts between WT and *TBX5*$^{in/+}$ clusters and *TBX5*$^{in/del}$ clusters (Fig. 3D, dashed blue lines, and fig. S12C). Again, there were also differences between clusters of the same genotype (blue, white, and black arrows, Fig. 3D), showing that single-cell analysis reveals heterogeneity in 3D organization. That TBX5 dosage sensitivity exhibits a degree of heterogeneity in regulating 3D chromatin organization that is consistent with heterogeneity in gene expression and chromatin accessibility changes (*21*, *28*).

CTCF insulates most TBX5-bound TAD boundaries and loop anchors

Because TBX5 binding was enriched at TAD boundaries that were lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells as well as at TAD boundaries shared across all genotypes (Fig. 2F), this suggests that only some TBX5-bound TAD boundaries are sensitive to loss of TBX5. To explore the reasons behind this differential sensitivity, we assessed co-occupancy by TBX5 and CTCF (fig. S13, A to E). Genome wide, ~26% of TBX5 binding sites overlapped with CTCF binding (fig. S13F). Almost all (95%) of TBX5-bound TAD boundaries were co-occupied by CTCF (fig. S13, A and C). Although TAD boundaries bound by TBX5 alone were a small proportion of all TBX5-bound boundaries, most of them were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells (fig. S13A), indicating that TBX5 is directly required to maintain these TADs. By contrast, CTCF and TBX5 co-occupied boundaries were less likely to be lost with reduced TBX5 dosage compared with TBX5-only boundaries (18%, relative risk 0.26, $P < 0.0001$) (fig. S13A), suggesting that the presence of CTCF may protect TADs from loss of TBX5. In general, we found that lost TADs had lower insulation scores compared with those of TADs common to all genotypes. This was particularly evident for TBX5 only–bound TAD boundaries, which showed lower insulation scores compared with CTCF and TBX5–cobound ones, including at common TAD boundaries (Fig. 4A). Many critical cardiac transcriptional activators, including *HAND2*, *MYOCD*, *NR2F1*, *NR2F2*, and *TBX3*, were inside TADs adjacent to a CTCF and TBX5–cobound lost TAD boundary (Fig. 4B and fig. S13A).

Similar to TAD boundaries, loop anchors bound by TBX5 alone were rare but overrepresented as lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells compared with TBX5 and CTCF cobound anchors (relative risk 1.46, $P < 0.0001$) (Fig. 4C and fig. S13D). Only ~7% of all TBX5-bound loop anchors had TBX5 at both anchors (Fig. 4D). These loops were enriched for genes encoding other cardiac lineage–determining TFs, *GATA4* and *NKX2-5*, as well as coactivator *MYOCD* and key cardiac contractile apparatus genes *TNNT2*, *TNNI3*, *ACTN2*, and *MYL4* (Fig. 4, D and E). Given that these loops are particularly sensitive to TBX5 dosage, failure to normally regulate these genes likely explains why *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells display dose-dependent decreases in CM differentiation efficiency.

Although TBX5 binding was enriched at TAD boundaries and loop anchors, the majority of TADs and chromatin loops that were lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells did not have TBX5 binding sites at their boundaries or anchors (fig. S13G). For the loops, we called these CTCF-only loops. Many of these CTCF-only loops contained TBX5 binding sites inside their looped domain. We analyzed motifs at lost CTCF-only loop anchors and found enrichment for MEF2 and ZNF (C2H2 family) motifs (Fig. 4, F and G). Of the cardiac-expressed MEF2 family proteins, both *MEF2A* and *MEF2C* were expressed by a large proportion of CMs and dysregulated in the absence of TBX5, with *MEF2A* showing dose-dependent down-regulation in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells (fig. S9A). This analysis suggests that 3D chromatin organization is also sensitive to *MEF2A* loss downstream of TBX5. Additionally, *Tbx5* and *Mef2c* genetically interact in mouse heart development (*21*) and

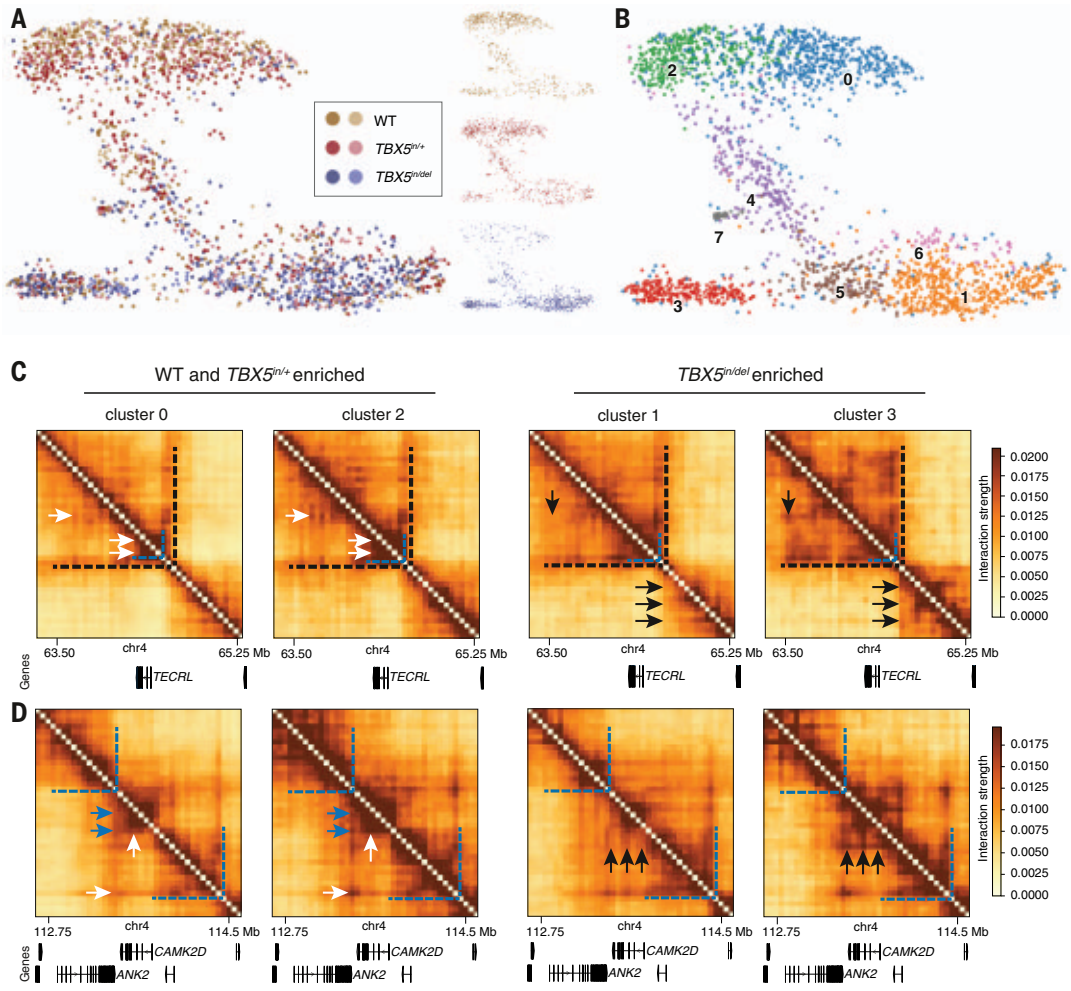


Fig. 3. Chromatin contact heterogeneity associated with reduced dosage of TBX5. (**A**) UMAP of cells from the *TBX5* allelic series, where the features are derived from Fast-Higashi embeddings of Hi-C contacts at 1-Mb resolution. Cells are colored by *TBX5* genotype (and shown separately), with the two shades indicating biological replicates. (**B**) Same UMAP as in (A), but with cells colored by Leiden clusters. (**C** and **D**) Higashi imputed contact matrices of WT and *TBX5*$^{in/+}$-enriched clusters (cluster 0 and 2) and *TBX5*$^{in/del}$-enriched clusters (cluster 1 and 3) around TBX5 dosage–sensitive genes (C) *TECRL*, (D) *ANK2*, and *CAMK2D*. Arrows indicate differences within genotypes (WT and *TBX5*$^{in/+}$ enriched clusters, blue for cluster 0, and white for cluster 2; for *TBX5*$^{in/del}$-enriched clusters, black for cluster 3). Dashed lines indicate differences between WT and *TBX5*$^{in/+}$-enriched clusters (blue) and *TBX5*$^{in/del}$-enriched clusters (black). The scale shows the Higashi imputed contact values, normalized by the range of contacts in cluster 0. Two biological replicates per genotype for snm3c-seq were analyzed and referred to in this figure.

are part of a set of TFs that can reprogram fibroblasts to CMs (*39*). Therefore, MEF2A or MEF2C may be additional cardiac TFs with a putative role at chromatin loop anchors.

Loss of TBX5 disrupts cohesin binding and loop formation

As we noted that a large proportion of TBX5 binding sites were found inside looped regions rather than at loop anchors (Fig. 2I), we predicted that failed TBX5 recruitment to these sites could impact loop formation. Overall, ~57% of cardiac loops contained TBX5 binding sites. TBX5 binding is enriched at active enhancer regions (*37*, *40*) [odds ratio (OR) = 7.6, $P < 0.0001$]. Active enhancers can be regulators of cohesin recruitment that enable cohesin-dependent loop extrusion and loop formation (*41*, *42*), where cohesin loading can be functionally influenced by TFs (*43*). We hypothesized that the loops with CTCF only-bound loop anchors may be impacted by reduced cohesin loading at TBX5-bound enhancers, leading to impaired extrusion of these loops. Supporting this, we found that TBX5 binding was enriched within lost loops with CTCF-only anchors compared with that in other loops (Fig. 5, A and B), especially TBX5 only–bound loops.

To explore the possibility that TBX5 binding within loops enables loop formation together with cohesin, we assessed global genome occupancy of the cohesin complex by RAD21 ChIP-seq in the *TBX5* allelic series. As expected, many RAD21 peaks (90.5%) overlapped CTCF-bound regions and formed highly intense, narrow binding peaks. By contrast, RAD21 peaks not coinciding with CTCF-bound regions tended to be broader with lower intensity, and the majority overlapped with H3K27ac, indicating a general enrichment at active enhancer regions (fig. S14, A and B). Of all TBX5 binding sites, 34% were co-bound by RAD21, and 48% were within 5 kb of a RAD21 peak (fig. S14C). As we predicted, cobound sites were enriched inside chromatin loops rather than at loop anchors and especially at enhancer regions (Fig. 5C).

To determine whether the loss of RAD21 binding could explain changes to chromatin organization from reduced *TBX5* dosage, we examined RAD21 ChIP-seq peaks in the three genotypes. In contrast to CTCF, which largely gained binding sites in the absence of *TBX5*, RAD21 showed a dose-dependent reduction in the number of binding sites (Fig. 5D). The RAD21 binding sites lost in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells

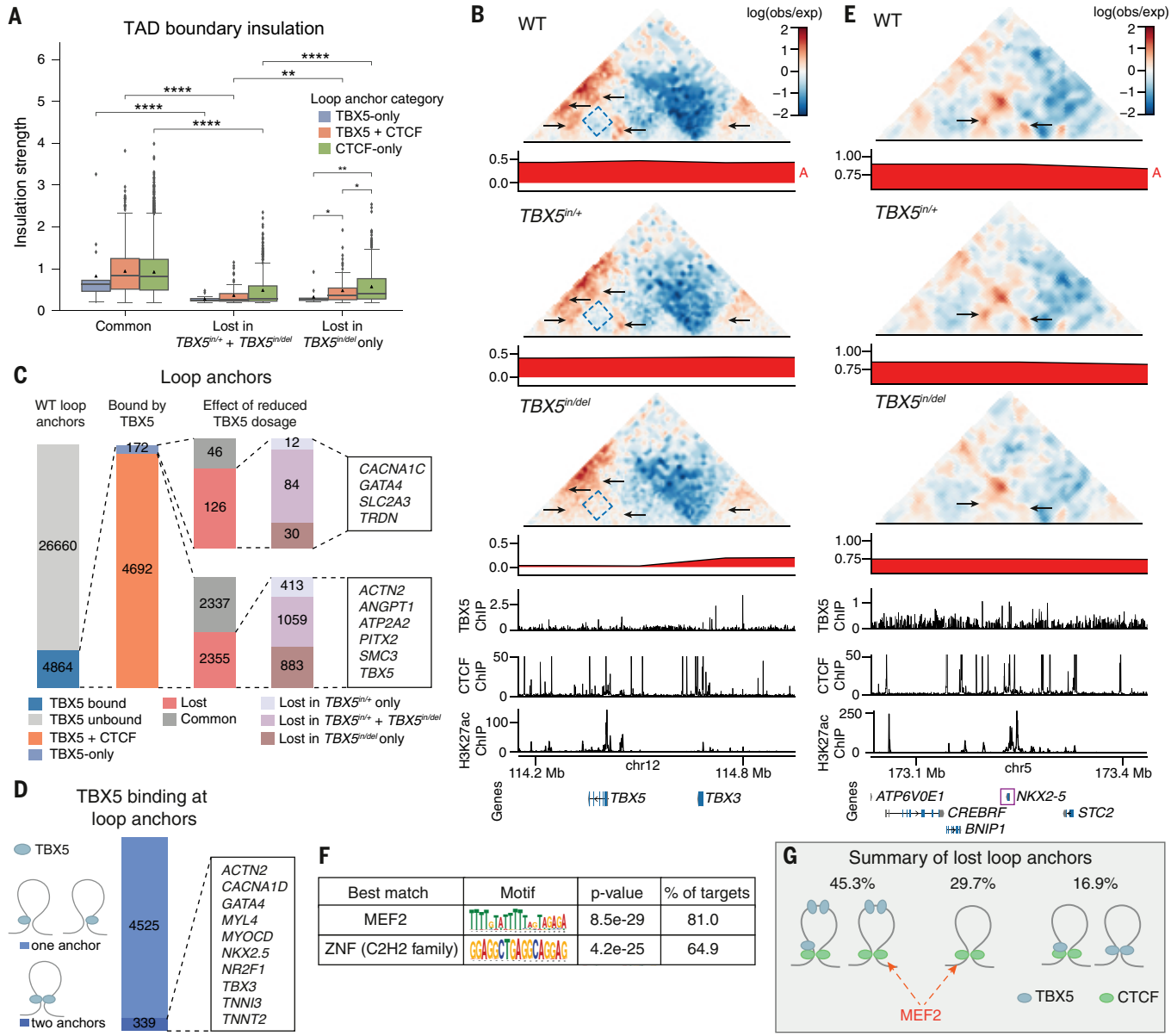


Fig. 4. TBX5 establishes cardiac chromatin contacts through both CTCF-dependent and -independent mechanisms. (**A**) Insulation strength of common and lost TAD boundaries by TBX5 or CTCF binding status. The boxplot represents the interquartile range (IQR) with whiskers set to 1.5 times the IQR, and outliers are shown in points. $^{**}P < 0.01$; $^{****}P < 0.0001$. (**B**) Observed over expected contact maps, principal components scores highlighting A/B compartments, WT TBX5, CTCF, and H3K27ac ChIP around *TBX5* and *TBX3*. Arrows indicate reduced contact frequency in $TBX5^{in/+}$ and/or $TBX5^{in/del}$ cells. The boxed area indicates increased contact frequency in $TBX5^{in/del}$ cells. The scale indicates the log(observed/expected) contact frequency. (**C**) Percentage of TBX5-bound loop anchors segregated into TBX5-only or CTCF and TBX5–co-occupied loops and by sensitivity to loss of TBX5. The gene list indicates a subset of those within 5 kb of lost contacts. (**D**) Proportion of all chromatin loop anchors bound by TBX5 separated into those that are bound at one or both anchors. The numbers in (C) and (D) indicate the number of loops per category. (**E**) Same legend as (B), around *NKX2-5*. Arrows indicate reduced contact frequency in $TBX5^{in/+}$ and/or $TBX5^{in/del}$ cells. (**F**) Motif analysis of CTCF only–bound loop anchors lost in $TBX5^{in/+}$ and/or $TBX5^{in/del}$ cells. All data are Hi-C 3.0 for two merged biological replicates per genotype. (**G**) Summary of chromatin loop anchors lost with reduced TBX5 dosage. The percentage values indicate the proportion of $TBX5^{in/+}$ and $TBX5^{in/del}$ lost loops that are associated with each category.

were enriched inside chromatin loops compared with at loop anchors and were highly enriched at sites normally bound by TBX5 (Fig. 5E). Furthermore, a majority of lost loops in $TBX5^{in/+}$ and $TBX5^{in/del}$ cells (52%) were associated with loss of RAD21 binding (Fig. 5F, $P < 0.001$). The dose-dependent loss of RAD21 binding was evident at cardiac genes, for example, at the *HCN4* locus, which encodes a potassium-sodium transporter crucial for the generation of pacemaker action potentials in the heart (*44*) (Fig. 5G), and cardiac TF *TBX3* (Fig. 5H). Underscoring the transcriptional relevance of correct cohesin binding, we found that genes within 5 kb of anchors of loops associated with lost RAD21 peaks inside the loop were likely to be down-regulated (Fig. 5I).

Our results demonstrate that cohesin binding is sensitive to reduced TBX5 dosage and that loss of cohesin binding likely explains the sensitivity of chromatin contacts to TBX5 dosage. This suggests that TBX5 is required for normal cohesin binding in order to form cardiac-specific chromatin loops. Given our data showing that TBX5 does not

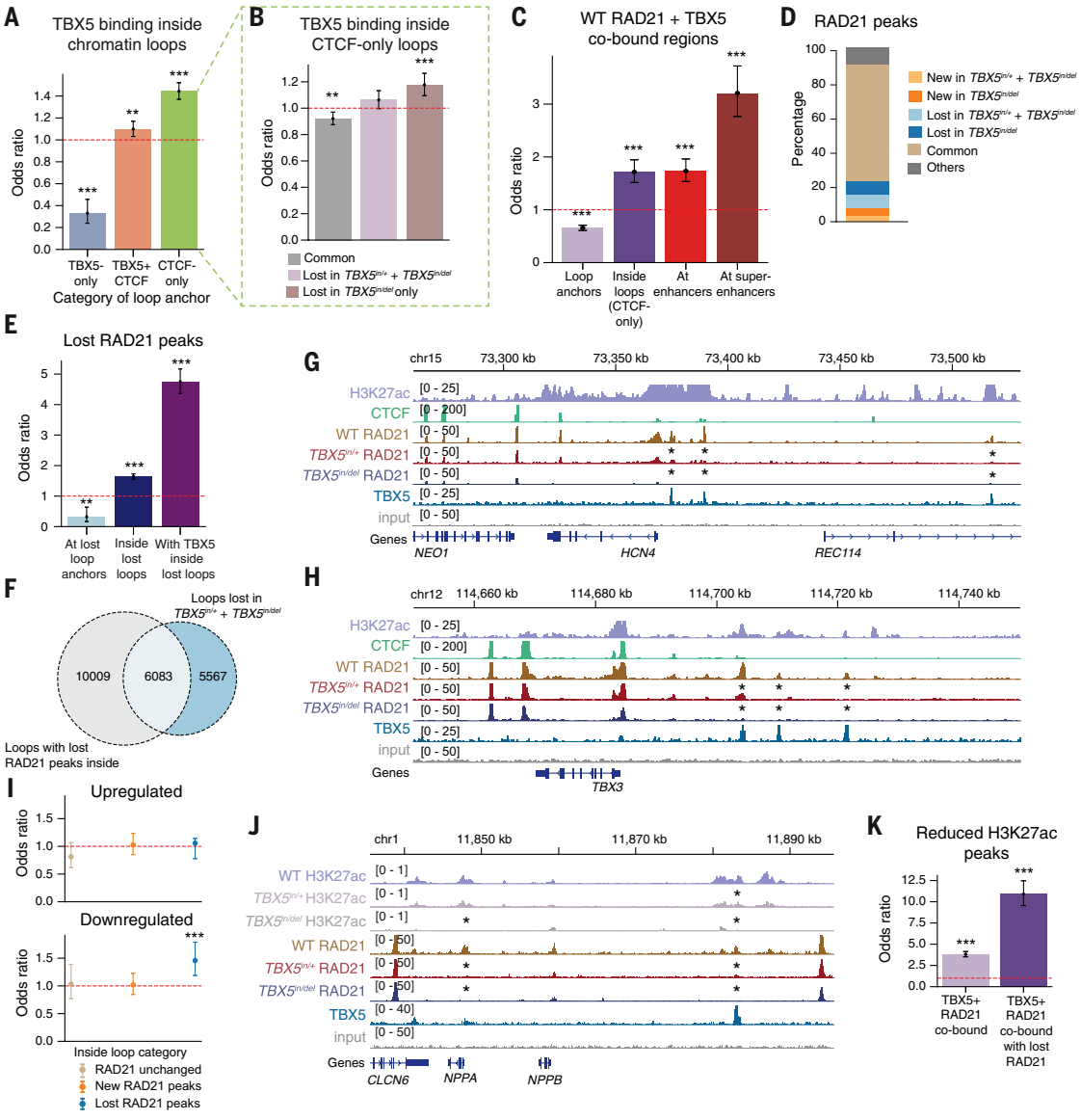


Fig. 5. TBX5 is required for cohesin binding at cardiac enhancers. (**A** to **C**) OR of (A) TBX5 binding within loop anchors bound by TBX5 or CTCF, (B) TBX5 binding within CTCF only–bound loop anchors grouped by sensitivity to reduced TBX5 dosage, and (C) RAD21 peaks cobound with TBX5 at loop anchors, inside CTCF-only loops, or at enhancer regions. Superenhancers are defined as having >2 H3K27ac peaks with no promoters within ±20 kb. (**D**) Percentage of RAD21 peaks unchanged, gained, or lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells. (**E**) OR for lost RAD21 peaks at loops lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells and their association at loop anchors, inside loops, or with TBX5, calculated relative to all other RAD21 peaks. (**F**) Loops lost overlapped with loops associated with lost RAD21 binding in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells. (**G** and **H**) ChIP-seq tracks RAD21 in WT, *TBX5*$^{in/+}$, and *TBX5*$^{in/del}$ around (G) *HCN4* and (H) *TBX3*. WT H3K27ac, CTCF, and TBX5 are shown. Asterisk indicates lost RAD21 peaks cobound by TBX5. (**I**) ORs of *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ up-regulated ($log_2FC \geq 0.2$, $P_{adj} < 0.05$) or down-regulated ($log_2FC \leq -0.2$, $P_{adj} < 0.05$) genes within 5 kb of anchors of loops with unchanged, gained, and lost RAD21 peaks. (**J**) ChIP-seq tracks for H3K27ac and RAD21 in WT, *TBX5*$^{in/+}$, and *TBX5*$^{in/del}$ around *NPPA/B* enhancer. Asterisks indicate reduced H3K27ac and RAD21 peaks cobound by TBX5. (**K**) ORs for H3K27ac peaks reduced in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells and association with RAD21 and TBX5–cobound regions. Two merged biological replicates per genotype. ORs compare each subset to all others; dotted lines at OR = 1 indicate no enrichment, OR scores > 1 indicate enrichment, and OR scores < 1 indicate depletion; error bars indicate 95% confidence intervals; $^{**}P < 0.01$; $^{***}P < 0.001$.

maintain most chromatin loops in CMs (fig. S10), we believe that TBX5 has a role early in cardiac differentiation to enable a permissive chromatin landscape around cardiac enhancers. This allows for their dynamic and ongoing loop formation at subsequent stages of CM differentiation. Supporting this, we found that H3K27ac was reduced, along with RAD21 binding, at well-known cardiac superenhancers for *NPPA/B* (*45*) and *MYH6/7* (*46*) bound by TBX5 and generally at WT RAD21 and TBX5–cobound cardiac enhancer regions in *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells (Fig. 5, J and K, and fig. S14D). This was not due to a global reduction in H3K27ac levels, as 63% of WT peaks were not reduced in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ cells (fig. S14D).

Our data support a model in which TBX5 is required for a subset of cardiac chromatin interactions and is particularly important for establishing chromatin loops. At a small proportion of loop anchors, TBX5 binding is directly required to form cardiac-specific chromatin loops and TADs. For most loops, TBX5 is likely required during loop formation for establishing a permissive chromatin landscape at cardiac-specific enhancers, allowing the cohesin complex to associate

with these regions and contribute to dynamic chromatin loop formation within CMs (fig. S14E).

Discussion

We identified the disease-linked, lineage-specific cardiac TF TBX5 as a dose-dependent regulator of human CM 3D organization. TBX5 is enriched at TAD boundaries and chromatin loops that emerge during CM differentiation. We discovered that diverse features of the 3D genome—chromatin loops, TAD boundaries, and compartments—are sensitive to reduced TBX5 dosage. Of relevance to human disease, *TBX5*-heterozygous CMs had considerable alterations in 3D chromatin folding, which was further exacerbated by the complete loss of *TBX5*. This indicates that chromatin organization is highly sensitive to TBX5 dosage. These alterations in 3D genome organization were correlated with the transcriptional changes that occur with reduced TBX5 dosage, particularly of down-regulated genes. Together, the data suggest that alterations in lineage-specific chromatin organization are a mechanism contributing to haploinsufficient TF transcriptional dysregulation, failed cellular differentiation, and potentially clinical CHD.

We propose that TBX5 regulates cardiac 3D chromatin organization through multiple distinct but functionally overlapping mechanisms. First and most frequently, cardiac chromatin loops are associated with TBX5 binding sites at cardiac enhancers within looped regions where TBX5 is dose-dependently required for normal cohesin complex binding. This is consistent with the observation that CTCF-independent cohesin binding sites tend to overlap with multiple TFs at cis-regulatory regions (*47*), supported by tissue-specific TFs associated with 3D chromatin changes in other cellular contexts (*13*, *43*). These examples, together with our data, imply that TF-guided cohesin binding and/or loading is a major mechanism that regulates lineage-restricted 3D chromatin looping. Indeed, modeling predicts that a distinct set of TFs is required for chromatin folding in different cell types (*48*). Secondly, TBX5 binding at loop anchors suggests that TBX5 is directly required to anchor cardiac-specific loops. Because the majority of anchors are co-occupied by CTCF, we propose that TBX5 may act as a cell type determinant of where loops should be created, with CTCF then structurally insulating them. At the few loops where TBX5 is bound at both anchors, TBX5 may have a direct role in loop formation, possibly through homodimer formation to bring together distant loci (*49*). Lastly, there may be effectors downstream of TBX5 that regulate 3D genome organization.

Our data are consistent with H3K27ac Hi-C combined with ChIP-seq (Hi-ChIP) data from adult mouse atria that show that some H3K27ac-dependent loops are lost in the absence of *TBX5* (*50*). Our data show that H3K27ac is reduced at cardiac enhancers in general, which likely explains the changes by HiChIP. In addition, we found that H3K27ac loss is largely within loops, which might be key to understanding the changes in cohesin binding and chromatin looping within loops. Because acute TBX5 depletion in CMs had a lesser impact on chromatin organization than sustained dosage reduction throughout differentiation, TBX5's role at both cohesin- and CTCF-bound sites likely occurs early during early lineage specification. This is likely through recruitment of chromatin remodeling complexes (*51*, *52*), histone modifying enzymes (*53*), or other cofactors that transform the epigenetic and chromatin landscape at cardiac enhancers, allowing their continued regulation of genes required for CM function.

Many but not all changes in 3D chromatin caused by loss of *TBX5* were associated with dysregulated gene expression, pointing to both structural and transcription-regulating roles for TBX5 in organizing 3D chromatin. Given the dynamic reorganization of the 3D genome that we observed during CM differentiation, it is possible that the transcriptional consequences of lost contacts at some regions would be evident at other stages in differentiation, as shown in datasets for other noncardiac lineage–restricted TFs MYOD and IKAROS (*13*, *17*). Because our allelic series datasets capture a single time point, it is unknown at which stage during differentiation these 3D contacts are established, but this question can be addressed in future experiments.

In disease contexts, known pathogenic alterations in 3D chromatin contacts impact one or a few gene loci, particularly through structural variants that modify the expression of a disease-related gene (*54*). Our results highlight large-scale, disease-relevant mechanisms through which TBX5 haploinsufficiency may lead to dysregulated gene expression in patients with CHD. Mechanisms of haploinsufficiency are poorly understood. One possible mechanism involves changes to chromatin accessibility (*28*, *55*). In this study, we have described an additional important layer: alterations in 3D genome organization. Given the preponderance of haploinsufficient TFs as regulators of developmental processes dysregulated in human syndromes (*2*), these extensive chromatin changes might reflect a more generalizable mechanism of developmental defects caused by TF haploinsufficiency.

Materials and methods

Maintenance of iPS cells and differentiation to cardiomyocytes

Protocols for use of iPSCs were approved by the Human Gamete, Embryo and Stem Cell Research Committee, as well as the Institutional Review Board at UCSF. WTc11 iPSCs (gift from Bruce Conklin, available at NIGMS Human Genetic Cell Repository/Coriell #GM25256) were used in this study. WTc11 iPSCs with genome-edited mutations for *TBX5* (*TBX5*$^{in/+}$ and *TBX5*$^{in/del}$) were previously generated and characterized in both ventricular (*21*) and atrial cardiomyocyte (CM) differentiations (*28*). WTc11 iPSCs used for atrial and ventricular time course Hi-C 3.0 experiments and corresponding RNA-seq were grown on hESC-qualified Matrigel (Corning #354277). Human iPSCs for all other experiments were grown on growth factor-reduced basement membrane Matrigel (Corning #356231) in mTeSR-1 (StemCell Technologies #85850) or mTeSR+ medium (StemCell Technologies #100-0276). iPSCs were routinely passaged using ReLesR (StemCell Technologies #100-0483). For CM differentiations, iPSCs were dissociated using Accutase (StemCell Technologies #07920) and seeded into 6-well or 12-well plates. Cells were grown for three days until 70-90% confluency and induced using Stemdiff Cardiomyocyte Ventricular (StemCell Technologies #05010) or Atrial (StemCell Technologies #100-0215) kits, according to manufacturer's instructions.

TBX5 was induced for depletion from TBX5-FKBP12^{F36V} cells at d20 of atrial differentiation by treatment with 500 nM dTAG^{V}-1 ligand (Tocris #6914) for 24 hours and compared to DMSO vehicle treated control samples.

Generation of TBX5-bio and TBX5-FKBP12^{F36V} cell line

A codon-optimized BioTAP tag (*56*), which can be biotinylated by endogenous biotin ligases, was used to generate an endogenously tagged TBX5 protein for ChIP. An FKBP12^{F36V} with 2x HA (*57*) tag was used to generate a cell line in which TBX5 could be induced for degradation. Each tag was inserted immediately after the start methionine codon in the N terminus of the endogenous TBX5 locus using ribonucleoprotein (RNP) complexes of sgRNA (GCCCTCGTCTGCGTCGGCCA, ordered from Synthego) and SpCas9-NLS 'HiFi' protein (QB3 MacroLab, University of California Berkeley). Donor template sequences for each were provided by using HDR Alt-R gene blocks (IDT). RNPs were prepared by mixing 1 µl of 40 µM Cas9 with 1.2 µl of 100 µM sgRNA and made to a total volume of 10 µl with Lonza P3 transfection buffer (Lonza #197179) and incubated at room temperature for 10 min. 250,000 WTc11 iPSCs were resuspended in 12.2 µl Lonza P3 transfection buffer and nucleofected with 1 µg donor template and 10 µl RNPs using Lonza Nucleofector 4D system in 96-well format using DS-138 settings. After nucleofection, mTeSR1 supplemented with CloneR (StemCell Technologies #05888) was added to cells and incubated for 10 min at 37°C, before transferring cells into a single well of a Matrigel-coated 48-well plate. The following day, the medium was replaced with fresh mTeSR1. The cells were expanded into a 12-well plate. Single cells

were dissociated with Accutase (StemCell Technologies #07920) and were sorted into 96-well plates for colony selection using Namocell Hana single-cell dispenser. Plates were cultured in mTeSR1 supplemented with CloneR for 3 days post sorting to ensure survival and then replaced with mTeSR1. Colonies were screened by polymerase chain reaction (PCR) using primer pairs flanking the whole insertion site (F: 5′-TCCTCAGAGCAGAACCTTGC, R: 5′-CTTACCTGCTGGGTGAAGG). Amplicon from the whole insertion site PCR was amplified and verified by Sanger sequencing. All clones used were determined to have a normal karyotype.

Flow cytometry

During CM harvest for Hi-C 3.0, snm3c-seq or ChIP, about 5% of each sample was kept for flow cytometry analysis of CM differentiation efficiency, as assessed by proportion of cardiac troponin T2+ cells. Cell pellets were resuspended in 4% formaldehyde (from 16% stock concentration, methanol-free, ThermoFisher Scientific #28906) and incubated on nutator for 15 min at room temperature. Samples were centrifuged for 5 min at 200 × g at 4°C. Samples were washed twice in DPSB and stored at 4°C for up to 1 week before preparing for flow cytometry. Cells were permeabilized in FACS buffer (0.5% w/v saponin, 4% FBS in PBS) for 15 min at room temperature. Cells were stained with a mouse monoclonal antibody against cardiac isoform Ab-1 Troponin (ThermoFisher Scientific #MS-295-P, final concentration 5 µg/ml) or the IgG1kappa isotype control (ThermoFisher Scientific #14-4714-82) for 1 hour at room temperature, gently vortexing cells every 20 min. Samples were washed with FACS buffer and the stained with donkey anti-rabbit IgG Alexa 488 (ThermoFisher Scientific #A21206, 1:200 dilution, final concentration 5 µg/ml) for 1 hour at room temperature, gently vortexing cells every 20 min. Cells were washed with FACS buffer and filtered into tubes with 35-micron mesh caps (Corning #352235). 10 mg/ml DAPI (Invitrogen #D1306) was added to final concentration of 1 µg/ml. At least 10 000 cells per sample were analyzed using the BD LSRFortessa X-20 (BD Bioscience) and final results were processed using FlowJo (FlowJo, LLC).

Hi-C 3.0 library preparation and sequencing

Cells were grown in 3 wells of a 6-well dish per biological replicate, with all 3 wells pooled together after dissociation for the crosslinking step (approximately 10 million cells). iPSCs were dissociated in Accutase (StemCell Technologies #07920) for 5 min. Differentiated cells were dissociated in 0.25% Trypsin-EDTA (5 min for d2, d4 or d6 cells; 10 min for d11, d20, d23, d45 CMs; Gibco #25200056), followed by quenching in 15% FBS in DPBS. Cells were lifted off the dish and resuspended to single cells by repeatedly using P1000 pipette, then transferred to a 15 ml tube and centrifuged at 800 rpm for 5 min at room temperature. Pellets were washed once in 5 ml HBSS before resuspending the entire cell pellet in exactly 10 ml HBSS. 16% methanol-free formaldehyde (ThermoFisher Scientific #28908) was added to a final concentration of 1% and samples incubated on nutator at room temperature for 10 min. Samples were quenched with 2.5 M glycine to a final concentration of 0.125 M and incubated on nutator at room temperature for 5 min, then 15 min at 4°C and followed by centrifugation at 1000 × g for 5 min at room temperature. Samples were washed once in DPBS. Pellets were resuspended in 1 ml of 3 mM DSG in PBS, transferred to 1.5 ml tubes, and incubated on nutator for 40 min. Samples were quenched with 2.5 M glycine to a final concentration of 0.125 M and incubated on nutator at room temperature for 5 min. Samples were centrifuged at 2000 × g for 5 min at RT. Cell pellets were resuspended in 5 mg/ml BSA in DPBS, to prevent cells sticking to tube after DSG crosslinking, and centrifuged at 2000 × g for 5 min at 4°C. Supernatant was carefully removed and cell pellets were snap frozen in liquid nitrogen and stored at -80°C.

Two biological replicates per time point were processed for Hi-C 3.0 and library preparation (with the exception of ventricular d2 and d4 where n = 1 was used). Samples were processed for Hi-C using Arima-HiC+ kit (Arima Genomics). >1 µg per sample was used as determined by Arima's Estimating Input Amount Protocol using Qubit Fluorometer and dsDNA HS Assay Kit (ThermoFisher Scientific #Q32854). Samples were processed as per Arima's user guide for mammalian cell lines with one exception: DNA was digested using 50 U DpnII (New England Biolabs #R0543M) and 50 U DdeI (New England Biolabs #R0175L) restriction enzymes in 1x NEB Buffer 3.1 for 60 min at 37°C. All samples passed Arima-QC1 Quality Control. Sequencing libraries were prepared as per Arima's user guide for Library Prep Using KAPA Hyper Prep Kit (Roche #7962347001). Between 1.5-4 µg of DNA was fragmented in 130 µl microtube (Covaris #520045) using a Covaris S2 sonicator for 55s at 10% duty cycle, intensity 4, 200 cycles per burst in a water bath maintained at 4°C. An average fragment size of 400 bp was verified using Agilent 2100 Bioanalyzer. Between 400-1000 ng of size-selected DNA was used for biotin enrichment, end repair, dA-tailing and adapter ligation. The number of PCR cycles was determined following Arima-QC2 Quality Control using KAPA Library Quantification Kit (Roche KK4824) and libraries were amplified to at least 10 nM. All samples passed Arima-QC2 Quality Control. DNA cleanup steps were done using Ampure XP beads (Beckman Coulter #A63881). Final libraries were quantified using Qubit Fluorometer and dsDNA HS Assay Kit and assessed on Agilent BioAnalyzer. Pooled, equimolar libraries were sequenced using NovaSeq technology (Illumina). All atrial time course samples were sequenced in one batch, all TBX5 allelic series were sequenced in one batch, DMSO and dTAG^{V}-1 samples were sequenced in one batch, and ventricular time course samples were sequenced in two batches (one biological replicate per batch) using 150 bp paired-end reads.

Hi-C 3.0 data processing

Hi-C 3.0 data were processed using 4D Nucleome (4DN) Hi-C data Processing Pipeline (v0.2.7) (https://data.4dnucleome.org/resources/data-analysis/hi_c-processing-pipeline) with the hg38 reference genome, which generated .mcool file with iterative correction and eigenvector decomposition (ICE) normalization and .hic file with Knight-Ruiz (KR) normalization. For the analyses that do not explicitly compare replicates, data from biological replicates for each timepoint or genotype were merged together to generate a combined Hi-C matrix for high resolution after checking their reproducibility using HiCRep (v0.2.6) (*58*). Compartments and TAD boundaries were called using cooltools (v0.5.4) from .mcool files (*59*) at the resolution of 250 kb and 5 kb, respectively. HiCCUPS (Juicer tools v1.8.9) (*60*) and Peakachu (v2.3) (*61*) were used to detect chromatin loops at both 5 kb and 10 kb resolution from .mcool files and .hic files, respectively. The loops called by each tool at the two resolutions were first merged and then the resulting loops were further combined to get a union set of loops by removing the ones with both anchors within 20 kb of another one.

A/B compartment analyses

Compartments for each group of samples (atrial or ventricular time course, *TBX5* allelic series, or DMSO and dTAG^{V}-1 samples) were classified as common A, common B, B to A, A to B, or dynamic. Compartments that consistently remained in the A or B state across all time points, genotypes or treatments were classified as common A or common B, respectively. For the time series data, compartments that switched from B to A at a specific time point and remained in the A compartment for all subsequent time points were classified as B to A ("cardiac specific"). Conversely, those that switched from A to B at a given time point and remained in B were classified as A to B. For the atrial time course data, to evaluate whether the compartment change is ahead of gene expression change or the other way, the earliest time point for compartment change was identified by finding the time point at which the compartment switched to a different status and kept the status for all subsequent time points.

For the *TBX5* allelic series, compartments that were in the B compartment in WT samples and shifted to A in $TBX5^{in/+}$ and $TBX5^{in/del}$ samples, or those in B in both WT and $TBX5^{in/+}$ but switched to A in $TBX5^{in/del}$, were classified as B to A. Similarly, compartments that changed from A to B across genotypes as above were classified as A to B. For DMSO and $dTAG^{v}$-1, compartments shifting from B in DMSO to A in $dTAG^{v}$-1 were classified as B to A, while those shifting from A in DMSO to B in $dTAG^{v}$-1 were classified as A to B. All other compartments exhibiting changes across time points or genotypes, but not fitting these categories, were classified as dynamic.

Differential compartment analyses for the *TBX5* allelic series was performed using dcHiC (v2.0) (*62*), to identify compartments with statistically significant changes in their compartment scores, regardless of changes between the A and B states. First, dcHiC quantile normalized the compartment scores, which were identified using cooltools, across samples. Then, a distance score based on Mahalanobis distance was calculated for each compartment per genotype based on these normalized compartment scores, representing the degree of compartment score change for one compartment compared to all compartments within the *TBX5* allelic series. As the squared Mahalanobis distance follows a chi-square distribution, these distances were converted into p-values from that distribution by dcHiC. These p-values were then adjusted for multiple testing with Independent Hypothesis Weighting, and an FDR threshold of 0.01 was used to determine significant differential compartments. Finally, significant compartments were clustered with k-means.

TAD boundary analyses

The .bigwig files generated by cooltools during TAD boundary calling were converted to .wig format using bigWigToWig (v377). Regions lacking insulation scores were identified and extracted as gap regions for each sample. For each group of samples, these TAD boundaries and gap regions were combined by merging those within 10 kb of each other using bedtools (v2.31.0) merge, respectively. The presence of each TAD boundary in a sample was identified using bedtools intersect. TAD boundaries were filtered out from further analyses if the TAD regions formed by a boundary and its adjacent boundaries overlapped with gap regions by more than 20% in more than 60% of the samples containing that boundary. Genes associated with a TAD boundary were defined as the ones located within the TADs formed by the TAD boundary and its adjacent boundaries.

For the time course data, the TAD boundaries were categorized into four groups: common, gained, lost, and dynamic. TAD boundaries that either appeared (binary) or showed higher insulation strength ($\log_2$FC >0.6) at a specific time point and remained consistent in all subsequent time points were classified as gained ("cardiac specific"). Conversely, boundaries that disappeared (binary) or had lower insulation strength ($\log_2$FC < -0.6) from a certain time point onwards were classified as lost. TAD boundaries that were present across all time points but did not meet the criteria for gained or lost were categorized as common. All other boundaries that exhibited changes but did not fit into these categories were classified as dynamic. For the *TBX5* allelic series data, TAD boundaries were similarly classified as (1) common, (2) gained in $TBX5^{in/+}$ and $TBX5^{in/del}$, (3) gained in $TBX5^{in/del}$, (4) lost in $TBX5^{in/+}$ and $TBX5^{in/del}$, (5) lost in $TBX5^{in/del}$, and (6) dynamic, based on their binary presence or changes in insulation strength ($\log_2$FC > 0.6 or $\log_2$FC < -0.6). Following the same criteria, the TAD boundaries for DMSO and $dTAG^{v}$-1 samples were classified as common, gained (in $dTAG^{v}$-1) or lost (in $dTAG^{v}$-1).

Chromatin loop analyses

Similar to the TAD boundaries, a union set of chromatin loops for each group was generated by removing the ones with both anchors within 20 kb of another one. The presence of a chromatin loop in each sample was identified using pgltools (v2.2.0) intersect. For the time series data, the chromatin loops were categorized as common, gained ("cardiac specific"), lost, and dynamic others. For the *TBX5* allelic series, the chromatin loops for the TBX5 allelic series were classified as (i) common, (ii) gained in $TBX5^{in/+}$ and $TBX5^{in/del}$, (iii) gained in $TBX5^{in/del}$, (iv) lost in $TBX5^{in/+}$ and $TBX5^{in/del}$, (v) lost in $TBX5^{in/del}$, and (vi) dynamic others, based on their binary presence. For strict analyses of the DMSO and $dTAG^{v}$-1 samples, loops detected in DMSO but absent in WT (from the TBX5 allelic series), as well as loops present in WT and $dTAG^{v}$-1 but absent in DMSO, were excluded. After filtering, loops were classified as common (present in both DMSO and $dTAG^{v}$-1), gained (present in $dTAG^{v}$-1 only), or lost (present in DMSO only).

Aggregate peak analysis (APA) for the gained or lost loops in cardiac cells was performed using coolpup.py and visualized using plotpup.py.

snm3C-seq library preparation and sequencing

Cells were grown in one well of a 6-well dish per biological replicate. iPSCs were dissociated in Accutase (StemCell Technologies #07920) for 5 min. Differentiated cells were dissociated in 0.25% Trypsin-EDTA (5 min for d2, d4 or d6 cells; 10 min for d11, d20, d23, d45 CMs; Gibco #25200056), followed by quenching in 15% FBS in DPBS. Cells were lifted off the dish and resuspended to single cells by repeatedly using P1000 pipette and then transferred to a 15 ml tube and centrifuged at 300 × g for 5 min at room temperature. Pellets were washed once in 1 ml DPBS and counted using Countess II Automated Cell Counter with optimized setting. 2 million cells were aliquoted into a 1.5 ml tube and centrifuged at 300 × g for 5 min at room temperature. Cell pellets were resuspended in 1 ml DBPS, and 16% methanol-free formaldehyde (ThermoFisher Scientific #28908) was added to final concentration of 2% and incubated on nutator at room temperature for 5 min. Samples were quenched with 2.5 M glycine to a final concentration of 0.2 M and incubated on nutator at room temperature for 5 min, followed by centrifugation at 1000 × g for 5 min at 4°C. Samples were washed once in ice-cold DPBS and cell pellets were snap frozen in liquid nitrogen and stored at -80°C.

Two replicates per time point were processed for snm3C-seq and library preparation. Cells were subjected to in situ 3C procedure following the steps in the Arima-3C BETA Kit (Arima Genomics), with some modifications. Cells were lysed for 15 min on ice in 20 µl of Lysis Buffer. 24 µl of Conditioning Solution was added, and cells were transferred to a 0.2 mL PCR tube, and mixed gently by pipetting. Tubes were incubated at 62°C for 30 min in a thermal cycler, with the lid temperature set to 85°C. 20 µl of Stop Solution 2 was added, and mixed gently by pipetting. Tubes were incubated at 37°C for 15 min in a thermal cycler with a lid temperature set to 85°C. 28 µl of the restriction enzyme digestion mix (9.8 µl water, 9.2 µl Buffer H, 4.5 µl Enzyme H1, 4.5 µl Enzyme H2) was added and gently pipette mixed. Tubes were incubated in a thermal cycler for 37°C for 60 min, 65°C for 20 min, then 25°C for 10 min. 82 µl of ligation mix was added (70 µl Buffer C, 12 µl Enzyme C) and gently mixed by pipetting, then incubated at 25°C for 15 min. Nuclei were transferred to a tube containing 850 µl 1% BSA in DPBS on ice and filtered through a 30 µm Celltrics strainer to a new tube on ice. Nuclei were centrifuged at 1000 × g for 10 min at 4°C (in bucket rotor). Supernatant was removed and pellets were resuspended each sample in 800 µl of 1% BSA in DPBS. Nuclei were stained with DRAQ7 at 1:200 and mixed gently with a pipette and incubated on ice for 5 min. One tube of Proteinase K (Zymo #D3001-2-D) was dissolved in 1 mL of Proteinase K Resuspension Buffer (Zymo #D3001-2-B) then added to a mix of 15 mL M-Digestion Buffer 2x (Zymo #D5021-9) and 14 mL water to make pK Digestion Buffer. 2 µl of pK Digestion Buffer were added to each well of an Eppendorf twin.tec PCR Plate 384 LoBind, skirted, 45 µl, PCR clean (Eppendorf #0030129547). Single DRAQ7 positive nuclei were sorted into individual wells of the 384-well plate using a Sony SH800 cell sorter. One 384-well plate was used per biological replicate. Plates were sealed then incubated in a thermal cycler at 50°C for 20 min with a heated lid at 85°C for nuclei digestion.

Library preparation and Illumina sequencing for snm3C-seq

The snm3C-seq bisulfite conversion procedure and library preparation protocol was carried out as previously described (*63*, *64*). This procedure was automated using the Beckman Coulter Biomek i7 liquid handlers for all reactions in 384 and 96-well plates. The snm3C-seq libraries were shallow sequenced on the Illumina NovaSeq 2000 to 2-10 M reads to check quality. Successful libraries were deeply sequenced on the NovaSeq6000 or NovaSeqX Plus instruments using 150 bp paired-end reads.

Mapping and analysis of snm3C-seq

Raw sequence FASTQ files were mapped using YAP (Yet Another Pipeline) software (cemba-data v1.6.9), as previously described (*64*). Briefly, FASTQ files were demultiplexed by cell barcodes and read quality assessment (cutadapt, v.2.10). Two-pass mapping was performed with bismark (v0.20, with bowtie2 v2.3) for alignment to hg38 reference genome assembly. BAM file processing and QC was performed using samtools (v1.9) and Picard (v3.0.0). Chromatin contacts were called and methylome profiles were generated using ALLCools software (v1.0.23).

Single-cell methylation analyses were performed with the ALLCools suite (*63*). Cells were first filtered based on their mapping rates (>50) and methylation fraction at mCG (>0.5), mCH (<0.2), and mCCC (<0.03) sites. For the atrial and ventricular time course datasets, 3304 nuclei (from total 3840 nuclei) passed quality control filters. For *TBX5* allelic series datasets, 2170 nuclei (from total 2304 nuclei) passed quality control filters. To cluster single cells, the fractions of mCH and mCG methylation in 100 kb bins across the genome were used. These bins were filtered by coverage (>50 and <3000), and sites in ENCODE blacklist regions were not considered. The top 20000 most variable regions with a support vector regression (SVR) were selected. Next, Principal Component Analysis was performed on these top regions for dimensionality reduction. The top principal components were visualized with Uniform Manifold Approximation and Projection (UMAP), and cells were colored according to their time point, cell type, and/or genotype.

Single-cell Hi-C analyses were performed with Fast-Higashi (v0.1.1) (*65*). In brief, the sparse single-cell Hi-C maps per cell were imputed with a partial random walk with restart algorithm for 1-Mb bins across each chromosome. Next, the imputed scHi-C tensors are decomposed to produce cell embeddings and chromosome-specific bin weights, transformation, and meta-interaction matrices. 256-dimension embeddings were obtained to represent each cell and used for downstream visualization and clustering. Leiden clustering was performed for cells using the single-cell methylation and single-cell Hi-C principal components and embeddings. The Leiden clusters were then used to pseudo bulk the scHiC contacts. With cooler (v0.10.2) (*66*), the contacts from cells in each cluster were merged at 50 kb resolution, and the resulting bulk matrices were balanced with ICE normalization. Finally, to integrate the single-cell methylation and single-cell HiC clusters, the respective principal components and embeddings were concatenated.

To impute the single-cell Hi-C maps for visualization, we used Higashi (v.0.1.0a0) (*67*). Maps for each cell were imputed at 50 kb resolution. For pseudobulking, maps were aggregated per Leiden cluster and normalized by the number of cells per cluster.

Bulk RNA-seq library preparation

For the atrial differentiation, samples from WT cells were collected at d0 (iPSCs), d2, d4, d6, d11, d20, and d45 for three biological replicates. For the ventricular differentiation, samples from WT cells were collected at d2, d4, d6, d10, d12, d15, d30 for three biological replicates. Cells were grown in 1 well of a 6- or 12-well dish per biological replicate. iPSCs were dissociated in Accutase (StemCell Technologies #07920) for 5 min. Differentiated cells were dissociated in 0.25% Trypsin-EDTA (5 min for d2, d4, or d6 cells; 10 min for time points past d11; Gibco #25200056), followed by quenching in 15% FBS in DPBS. Cells were lifted off the dish and resuspended to single cells by repeatedly using P1000 pipette and then transferred to a 1.5 ml tube and centrifuged at 300 × g for 5 min at room temperature. Pellets were snap frozen in dry ice. Total RNA was isolated using Qiagen RNeasy micro kit (Qiagen #74004) with QIAshredder (Qiagen #79656) and eluted in 25 µl water. RNA concentration was determined using Nanodrop and quality was checked using Agilent Bioanalyzer.

Atrial CM RNA-seq libraries were generated using Illumina Stranded Total RNA Prep with Ribo-Zero Plus kit as per manufacturer's instructions using 150 ng of input RNA. Ventricular CM RNA-seq libraries were generated using NuGEN Ovation Ultralow System V2 kit as per manufacturer's instructions. Library concentration was quantified using Qubit Fluorometer and dsDNA HS Assay Kit and quality assessed on Agilent Bioanalyzer. Libraries were pooled and sequenced on a NovaSeq X (Illumina) using 50 bp paired-end reads for atrial differentiation samples or on a NextSeq500 using 75 single-end reads for ventricular differentiation samples.

For DMSO and dTAG^{V}-1 TBX5 depletion samples, 150 ng total RNA was sent to Novogene for library generation, sequencing and data analysis. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesized using random hexamer primers followed by the second strand cDNA synthesis. The library was ready after end repair, A-tailing, adapter ligation, size selection, amplification, and purification. The library was checked with Qubit and real-time PCR for quantification and Agilent BioAnalyzer for size distribution detection. Libraries were sequenced on a NovaSeq X (Illumina).

Bulk RNA-seq analysis

For atrial and ventricular time course data, the fastq files were analyzed using nf-core Nextflow RNA-seq pipeline (v3.12.0; DOI: 10.5281/zenodo.1400710) (*68*) aligning to hg38 reference genome and using STAR for alignment (v2.6.1d) (*69*) and Salmon quantification (v1.10.1) (*70*). Differential gene expression analyses were performed for each pair of timepoints using DESeq2 (v1.46.0) (*71*). The time point that a gene shows increased ($\log_2$FC≥1, P_{adj}<0.05) or decreased ($\log_2$FC≤-1, P_{adj}<0.05) expression compared to previous time points and remained consistent in all subsequent time points was considered as the earliest point for gene expression changes.

For DMSO and dTAG^{V}-1 TBX5 depletion samples, data analysis was performed by Novogene. Briefly, quality control of fastq files was performed using fastp (*72*), aligned to hg38 reference genome using HISAT2 (v2.2.1) (*73*) and quantification of gene expression using featureCounts (v2.0.6) (*74*). Differential gene expression analysis was performed using DESeq2 (v1.42.0) (*71*).

scRNA-seq

scRNA-seq was performed in the *TBX5* allelic series at d20 atrial CM differentiation for two biological replicates and available from GSE285169 (*28*). From differential gene expression, genes that showed increased ($\log_2$FC>0.2, P_{adj}<0.05) or decreased ($\log_2$FC <-0.2, P_{adj}<0.05) expression in both $TBX5^{in/+}$ and $TBX5^{in/del}$ or $TBX5^{in/del}$ alone were identified to investigate the association of gene expression with 3D chromatin changes.

Western blotting

~2 million CMs were snap frozen in liquid nitrogen after trypsin digestion and making a single cell suspension. For protein extraction, cell pellets were thawed on ice with RIPA buffer (Sigma-Aldrich #R0278-50ML) + 1x protease inhibitor (ThermoFisher Scientific #78438) for 15 min, then placed on a rotator in 4C for 15 min. The lysed cells were centrifuged at full speed at 4°C for 20 min, and the supernatant was collected. BCA was performed using Pierce BCA protein assay kit (Thermo Scientific #23227) to quantify the protein concentration. 20 µg of protein was collected from each sample, mixed with 4x Laemmli

Sample Buffer (Bio-Rad #1610747) containing 10% 2-mercaptoethanol, and boiled at 95°C for 10 min. The samples were loaded on SDS-PAGE gel (Thermo Fisher Scientific #NW04122BOX) and ran at 200 V for 40 min. The gel was transferred onto a PVDF membrane at 100 V for 2 hours in 4°C. The PVDF membrane was blocked in 5% milk in TBST (Boston BioProduct # LBB-180X) for 1 hour at room temperature, washed with TBST for 3 × 5 min, then incubated with primary antibody in 5% milk in TBST at 4°C overnight. Primary antibodies used were rabbit anti-TBX5 (Sigma-Aldrich #HPA008786, 1:400) and rabbit anti-PCNA (Abcam #ab92552, 1:2000). The next day, the primary antibody was washed off with TBST for 3x 5 min and the membrane was incubated with secondary antibody (anti-rabbit IgG HRP conjugated, Cell Signaling Technology #7074S, 1:3000) at room temperature for 1 hour, followed by 3x 5 min washes in TBST. The blot was developed with a mixture of Amersham ECL Prime Western Blotting Detection Reagent (Cytiva #45-002-401) and Amersham ECL Western Blotting Detection Reagent (Cytiva #RPN2109). The blot was then imaged using the Bio-Rad ChemiDoc MP imaging system.

ChIP-seq library preparation

For TBX5 ChIP, WT TBX5-bio CMs were collected at d11 (n = 2), d20 (n = 6), and d45 (n = 4) during atrial differentiation. For CTCF ChIP-seq, WT, *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ CMs were collected at d20 of atrial differentiation (n = 2 per genotype). CMs were grown in one 6-well plate per replicate, with all 6 wells pooled together after dissociation for the crosslinking step (approximately 20 million cells). Each plate was one biological replicate. CMs were dissociated in 0.25% Trypsin-EDTA (Gibco #25200056) for 10 min, followed by quenching in 15% FBS in DPBS. CMs were lifted off the dish and resuspended to single cells by repeatedly using P1000 pipette and then transferred to a 15 ml tube and centrifuged at 800 rpm for 5 min at room temperature. Pellets were washed once in 10 ml DPBS plus 1x protease inhibitors (DBPS+PI) before resuspending the entire cell pellet in exactly 10 ml DPBS+PI. 16% methanol-free formaldehyde (ThermoFisher Scientific #28908) was added to a final concentration of 1% and samples incubated on nutator at room temperature for 10 min. Samples were quenched with 2.5 M glycine to a final concentration of 0.125 M and incubated on nutator at room temperature for 5 min, followed by centrifugation at 500 × g for 5 min at 4°C. Samples were washed once in ice cold DPBS+PI, before being transferred to 2 ml protein lo-bind rubes, centrifuged at 2500 × g at 4°C for 5 min, and cell pellets were snap frozen in liquid nitrogen and stored at -80°C.

TBX5-bio cross-linked cells were thawed on ice and resuspended in 1 ml of ice-cold cell lysis buffer (25 mM Tris-HCl, pH 7.4, 85 mM KCl, 0.25% Triton X-100 and 1x Protease inhibitors in ddH_2O) and incubated at 4°C rotating for 30 min. Samples were centrifuged at 2500 × g for 5 min at 4°C and resuspended in 1 ml nuclear lysis buffer (0.5% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, and 1x protease inhibitors in ddH_2O). Samples were transferred to 1 ml millitube (Covaris #520130) and sheared using a Covaris S2 sonicator for 15 min at 5% duty cycle, intensity 5, 200 cycles per burst in a water bath maintained at 4°C. After shearing, 5 µl of each sample was kept for reverse crosslinking and the remaining sonicated sample was transferred to a new protein lo-bind tube and snap frozen in liquid nitrogen. The 5 µl sample was made up to 100 µl and incubated for 30 min at 37°C with 1 µl RNase A, followed addition of 12 µl 5 M NaCl and reverse crosslinking with 1 µl proteinase K at 65°C overnight. The following day the reverse crosslinked samples were purified using Qiagen MinElute PCR Purification kit (Qiagen #28006) and DNA quantified using Qubit Fluorometer and dsDNA HS Assay Kit (ThermoFisher Scientific #Q32854). Samples were run on an Agilent 2100 Bioanalyzer to verify that chromatin was sheared to fragments between 100-500 bp in size. Based on these measurements, the amount of DNA in each lysed sample could be extrapolated and equal amounts of sheared chromatin could be used per ChIP. Sheared samples were thawed quickly at 37°C and the equivalent of 55 µg of chromatin was transferred to a 5 ml protein lo-bind tube for ChIP and made up to 1 ml in nuclear lysis buffer. 50 µl (5%) of each sample was collected as input and stored at 4°C until reverse crosslinking step. ChIP samples were then diluted 1:5 with ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 0.01% SDS, 1.1% Triton X-100, 167 mM NaCl, and 1x protease inhibitors in ddH_2O). 50 µl of pre-washed Dynabeads MyOne Streptavidin T1 beads (ThermoFisher Scientific #65601) were added to each sample and incubated overnight with rotation at 4°C. The following day, beads were collected using a magnetic stand and transferred to 2 ml protein lo-bind tubes and washed with 1 ml twice of the following wash buffers: wash buffer 1 (2% SDS in ddH_2O), wash buffer 2 (0.1% sodium deoxycholate, 1% Triton X-100, 1 mM EDTA, 50 mM HEPES pH 7.5, 500 mM NaCl in ddH_2O), wash buffer 3 (250 mM LiCl, 0.5% IGEPAL CA-630 [Sigma #I8896], 0.5% sodium deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.0 in ddH_2O) and TE buffer (10 mM Tris-HCl pH 7.5 and 1 mM EDTA ddH_2O). After adding each wash buffer, beads were vortexed for 15 s, centrifuged briefly, and collected against a magnet. DNA was eluted with 150 µl elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0 in ddH_2O) shaking at 1000 rpm at 65°C overnight. The following day, beads were collected on a magnetic stand and supernatant was transferred to a new DNA lo-bind tube. Both ChIP and input samples were incubated at 37°C for 30 min with 1.5 µl RNase A, followed by addition of 18 µl 5 M NaCl and reverse crosslinking with 1.5 µl proteinase K at 65°C for 6 hours. DNA was purified using Qiagen MinElute PCR Purification kit (Qiagen #28006).

CTCF, RAD21, and H3K27ac ChIP-seq was performed generally as above with changes detailed here. The following buffers were used: cell lysis buffer (20 mM Tris-HCl pH 8.0, 85 mM KCl, 0.5% IGEPAL CA-630, and 1x protease inhibitors in ddH_2O), nuclear lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, and 1x protease inhibitors in ddH_2O), wash buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100 in ddH_2O), wash buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100 in ddH_2O), wash buffer 3 (250 mM LiCl, 1% IGEPAL CA-630 [Sigma #I8896], 1% sodium deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.0 in ddH_2O), and elution buffer (1% SDS, 100 mM $NaHCO_3$ in ddH_2O). 70 µg of chromatin was used per sample. After diluting in dilution buffer, each ChIP sample was incubated rotating at 4°C overnight with antibody (CTCF: combination of both Active Motif #61311 [discontinued] and Abcam #ab128873 used in each sample; RAD21: Abcam #ab992, H3K27ac: Active Motif #39133) and the following day incubated with 50 µl prewashed Magna ChIP protein A/G beads (Sigma #16-663) for 4 hours at 4°C. Bead washes were performed by resuspending beads in 1 ml buffer using P1000 pipette, spinning briefly and collecting on magnet. Following the bead washes, DNA was eluted in 50 µl shaking at 37°C, transferring to new tube and repeating bead elution to pool eluates to a total volume of 100 µl. Both ChIP and input samples were incubated at 37°C for 30 min with 1 µl RNase A, followed by addition of 12 µl 5 M NaCl and reverse crosslinking with 1 µl proteinase K at 65°C overnight. DNA was purified using ChIP DNA Clean and Concentrator kit (Zymo Research #D5205).

For Illumina sequencing library preparation, the entire ChIP sample or 2 µl of each input was used. End repair was performed in a 100 µl reaction with 15 U T4 DNA polymerase (New England Biolabs #M0203L), 5 U Klenow fragment DNA polymerase (New England Biolabs, #M0210L), 50 U T4 PNK (New England Biolabs #M0201L), 400 µM dNTP (Promega #U1511), and 1x T4 DNA ligase buffer with 10 mM ATP (New England Biolabs #B0202S) in nuclease-free water for 30 min at room temperature, followed by 1.6x bead:sample AmpureXP purification. Entire eluate was used for A-tailing in a 50 µl reaction with 1 mM dATP (New England Biolabs #N0440S), 15 U Klenow 3′>5′ exonuclease (New England Biolabs #M0212L), and 1x NEB buffer 2 in nuclease-free water for 30 min at 37°C, followed by 1.6x bead:sample AMPureXP purification. Entire eluate was used for adapter ligation in a 50 µl reaction with 6000 U

T4 DNA ligase (New England Biolabs #M0202L), 20 nM annealed and uniquely indexed adapters, and 1x T4 DNA ligase buffer with 10 mM ATP (New England Biolabs #B0202S) in nuclease-free water for 2 hours at room temperature, followed by 1x bead:sample AMPureXP purification. Adapters were prepared by annealing following HPLC purified oligos: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T and 5′Phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG where * represents a phosphothiorate bond and NNNNNN is a Truseq index sequence.The entire eluate was used for PCR amplification in a 50 µl reaction with 1x NEB Next High-Fidelity 2x PCR master mix (New England Biolabs #M0541L), 10 µM primers (F: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCC-TACACGA, R: 5′-CAAGCAGAAGACGGCATACGAGAT) in nuclease-free water, using thermocycler settings 98°C for 30 s; 16 cycles of 98°C for 10 s, 58°C for 40 s and 72°C for 30 s; 72°C for 5 min and followed by 0.8x bead:sample AMPureXP purification. DNA was quantified using Qubit Fluorometer and dsDNA HS Assay Kit and quality assessed on Agilent Bioanalyzer. TBX5-bio, RAD21 and H3K27ac ChIP-seq libraries were sequenced on a NovaSeq X (Illumina) using 50 bp paired-end reads. CTCF ChIP-seq libraries were sequenced on a NextSeq2000 (Illumina) using 50 bp paired-end reads.

ChIP-seq analysis

Fastq files were aligned to the hg38 reference genome using bowtie2 (v2.5.4) (*75*). Reads were filtered to include those with a mapq score of 30 or greater and duplicate reads were removed using SAMtools (v1.2) (*76*). Blacklisted regions were removed using bedtools (v2.31.1) (*77*). ChIP-seq peaks were called on each individual replicate relative to input using macs2 (v2.2.9.1) and the narrowpeaks parameter for CTCF and TBX5 samples and broadpeaks parameter for H3K27ac and RAD21 samples (*78*). CTCF, RAD21 and WT H3K27ac ChIP bigwig files for visualization were generated using deepTools2 bamCoverage (v3.5.6) (*79*). H3K27ac ChIP-seq data for WT, *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ samples were analyzed using the nextflow chip-seq pipeline (v2.0.0) with macs2 broadpeaks parameter (*68*). These data are shown in Fig. 5 – bigwig ChIP tracks and differential peak analysis.

To define CM TBX5 binding sites, the union set of peaks detected in d11, d20, and d45 samples was used. TBX5 ChIP bigwigs for visualization were generated as $\log_2$FC over input using deepTools2 bigwigCompare (v3.5.6) (*79*).

To determine the changes in ChIP peaks between samples, CTCF peaks from replicates of each genotype were first merged, and then merged again across genotypes using bedtools (v2.31.1) merge. For RAD21 and H3K27ac, consensus peaks in at least two replicates were calculated for each genotype using bedtools intersect and then merging peaks for each genotype using bedtools merge (*77*). Lost CTCF and RAD21 peaks were defined as peaks present in the merged genotype dataset but absent in the *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ for the corresponding individual peaks files (merged peaks for CTCF; consensus peaks for RAD21 and H3K27ac). Reduced H3K27ac peaks were determined using bedtools multicov to count the number of reads under the merged peaks, followed by differential count analysis using DESeq2 (v1.44.0) (*71*). Peaks were considered reduced if $\log_2\text{FC} \leq -0.6$ and $P_{adj} < 0.05$ (data S5).

Using H3K27ac ChIP-seq data, super enhancers were defined as more than 2 H3K27ac peaks with no promoters within ±20 kb. Other enhancers were defined as a single H3K27ac peak with no promoters within ±20 kb. For H3K27ac peaks with reduced signal, ORs were computed to evaluate their enrichment at TBX5–RAD21 or TBX5–lost RAD21 co-bound regions compared with all other H3K27ac peaks.

Annotation of ChIP-seq and RNA-seq data at TAD boundaries and loop anchors and inside loops

TBX5 and CTCF occupancy at TAD boundaries were identified by intersecting TAD boundaries (±5 kb) with their respective ChIP peaks. ORs were calculated to evaluate whether TBX5 binding were preferentially located at specific subsets of TAD boundaries (common, gained, or lost) relative to all other TAD boundaries. The relative risk analysis of TBX5 and CTCF co-bound lost TADs was calculated by the percentage of TBX5+CTCF co-bound TAD boundaries that are lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ (17.83%) divided by the percentage of TBX5-only bound TAD boundaries that are lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ (68.63%). Similar to the analysis of TBX5 binding, we calculated ORs to test whether up- or down-regulated genes were more likely to be associated with a given subset of TAD boundaries compared with all other TAD boundaries.

The occupancy of TBX5, CTCF, RAD21, and H3K27ac peaks, as well as promoters (defined as the 500 bp regions upstream of transcription start sites), at loop anchors was assessed by their overlap within ±5 kb of anchor regions. Loop anchors annotated with both promoters and H3K27ac peaks were defined as active promoters, whereas anchors annotated only with H3K27ac peaks were defined as active enhancers. TBX5, H3K27ac, RAD21, and subsets of RAD21 peaks (common, gained, or lost) were annotated within loops by assessing their overlap with loop domains, which were defined as the interval from 5 kb downstream of anchor 1 to 5 kb upstream of anchor 2. ORs were calculated to evaluate whether TBX5 binding was preferentially located at anchors of a subset of loops (common, gained, or lost) relative to all other loops. The relative risk analysis of TBX5 and CTCF co-bound sensitive loops was calculated by the percentage of TBX5-only bound loop anchors that are lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ (73.26%) divided by the percentage of TBX5+CTCF co-bound loop anchors that are lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$ (50.19%). ORs were used to evaluate whether TBX5 binding was enriched within loops whose anchors were bound by TBX5-only, TBX5+CTCF, or CTCF-only relative to all other loops. Additionally, ORs were calculated to assess the enrichment of TBX5 binding inside lost CTCF-only loops compared to all remaining CTCF-only loops. Enrichment of RAD21 peaks co-bound by TBX5 at loop anchors, within loops or at enhancer regions was evaluated by calculating ORs relative to all other RAD21 peaks in WT. Likewise, enrichment of lost RAD21 peaks at lost loop anchors, within lost loops or at TBX5 binding sites inside lost loops were assessed using ORs relative to all other RAD21 peaks in WT. We calculated ORs to test whether up- or down-regulated genes were preferentially located at anchors (±5 kb) of specific loop subsets compared to all other loops.

Motif enrichment analysis

To explore other factors that could regulate the loops sensitive to TBX5 loss, we applied MEME (v5.5.2) with classical mode (*80*) to identify the motifs enriched in anchors of CTCF-only bound loops that were lost in *TBX5*$^{in/+}$ and/or *TBX5*$^{in/del}$. The discovered motifs were further compared with the known vertebrate TF PWMs in the JASPAR 2022 database using Tomtom (v5.5.2) (*80*, *81*).

REFERENCES AND NOTES

1. S. N. Nees, W. K. Chung, The genetics of isolated congenital heart disease. *Am. J. Med. Genet. C. Semin. Med. Genet.* **184**, 97–106 (2020). doi: 10.1002/ajmg.c.31763; pmid: 31876989
2. J. G. Seidman, C. Seidman, Transcription factor haploinsufficiency: When half a loaf is not enough. *J. Clin. Invest.* **109**, 451–455 (2002). doi: 10.1172/JCI0215043; pmid: 11854316
3. J. Dekker, L. A. Mirny, The chromosome folding problem and how cells solve it. *Cell* **187**, 6424–6450 (2024). doi: 10.1016/j.cell.2024.10.026; pmid: 39547207
4. J. M. Dowen *et al.*, Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014). doi: 10.1016/j.cell.2014.09.030; pmid: 25303531
5. S. Schoenfelder, P. Fraser, Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019). doi: 10.1038/s41576-019-0128-0; pmid: 31086298
6. F. Grubert *et al.*, Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737–743 (2020). doi: 10.1038/s41586-020-2151-x; pmid: 32728247

7. R. Stadhouders, G. J. Filion, T. Graf, Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019). doi: 10.1038/s41586-019-1182-7; pmid: 31092938
8. E. P. Nora *et al.*, Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.* **11**, 5612 (2020). doi: 10.1038/s41467-020-19283-x; pmid: 33154377
9. E. P. Nora *et al.*, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22 (2017). doi: 10.1016/j.cell.2017.05.004; pmid: 28525758
10. I. F. Davidson *et al.*, CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023). doi: 10.1038/s41586-023-05961-5; pmid: 37076620
11. R. Drissen *et al.*, The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* **18**, 2485–2490 (2004). doi: 10.1101/gad.317004; pmid: 15489291
12. C. R. Vakoc *et al.*, Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**, 453–462 (2005). doi: 10.1016/j.molcel.2004.12.028; pmid: 15694345
13. Y. Hu *et al.*, Lineage-specific 3D genome organization is assembled at multiple scales by IKAROS. *Cell* **186**, 5269–5289.e22 (2023). doi: 10.1016/j.cell.2023.10.023; pmid: 37995656
14. T. M. Johanson *et al.*, Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat. Immunol.* **19**, 1257–1264 (2018). doi: 10.1038/s41590-018-0234-8; pmid: 30323344
15. Z. Liu, D. S. Lee, Y. Liang, Y. Zheng, J. R. Dixon, Foxp3 orchestrates reorganization of chromatin architecture to establish regulatory T cell identity. *Nat. Commun.* **14**, 6943 (2023). doi: 10.1038/s41467-023-42647-y; pmid: 37932264
16. Q. Shan *et al.*, Tcf1 and Lef1 provide constant supervision to mature $CD8^+$ T cell identity and function by organizing genomic architecture. *Nat. Commun.* **12**, 5863 (2021). doi: 10.1038/s41467-021-26159-1; pmid: 34615872
17. R. Wang *et al.*, MyoD is a 3D genome structure organizer for muscle cell identity. *Nat. Commun.* **13**, 205 (2022). doi: 10.1038/s41467-021-27865-6; pmid: 35017543
18. W. Wang *et al.*, TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nat. Immunol.* **23**, 1052–1062 (2022). doi: 10.1038/s41590-022-01232-z; pmid: 35726060
19. A. Magli *et al.*, Pax3 cooperates with Ldb1 to direct local chromosome architecture during myogenic lineage specification. *Nat. Commun.* **10**, 2316 (2019). doi: 10.1038/s41467-019-10318-6; pmid: 31127120
20. B. G. Bruneau *et al.*, A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709–721 (2001). doi: 10.1016/S0092-8674(01)00493-7; pmid: 11572777
21. I. S. Kathiriya *et al.*, Modeling Human TBX5 Haploinsufficiency Predicts Regulatory Networks for Congenital Heart Disease. *Dev. Cell* **56**, 292–309.e9 (2021). doi: 10.1016/j.devcel.2020.11.020; pmid: 33321106
22. A. D. Mori *et al.*, Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev. Biol.* **297**, 566–586 (2006). doi: 10.1016/j.ydbio.2006.05.023; pmid: 16870172
23. C. T. Basson *et al.*, Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* **15**, 30–35 (1997). doi: 10.1038/ng0197-30; pmid: 8988165
24. Q. Y. Li *et al.*, Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat. Genet.* **15**, 21–29 (1997). doi: 10.1038/ng0197-21; pmid: 8988164
25. H. Holm *et al.*, Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117–122 (2010). doi: 10.1038/ng.511; pmid: 20062063
26. A. Pfeufer *et al.*, Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153–159 (2010). doi: 10.1038/ng.517; pmid: 20062060
27. A. V. Postma *et al.*, A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circ. Res.* **102**, 1433–1442 (2008). doi: 10.1161/CIRCRESAHA.107.168294; pmid: 18451335
28. I. S. Kathiriya *et al.*, Reduced TBX5 dosage undermines developmental control of atrial cardiomyocyte identity in a model of human atrial disease. bioRxiv 2025.08.16.669546 [Preprint] (2025); https://doi.org/10.1101/2025.08.16.669546.
29. A. Bertero *et al.*, Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nat. Commun.* **10**, 1538 (2019). doi: 10.1038/s41467-019-09483-5; pmid: 30948719
30. Y. Zhang *et al.*, Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019). doi: 10.1038/s41588-019-0479-7; pmid: 31427791
31. B. Akgol Oksuz *et al.*, Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **18**, 1046–1055 (2021). doi: 10.1038/s41592-021-01248-7; pmid: 34480151
32. M. Asp *et al.*, A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647–1660.e19 (2019). doi: 10.1016/j.cell.2019.11.025; pmid: 31835037
33. L. Cyganek *et al.*, Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight* **3**, e99941 (2018). doi: 10.1172/jci.insight.99941; pmid: 29925689
34. A. P. Walden, K. M. Dibb, A. W. Trafford, Differences in intracellular calcium homeostasis between atrial and ventricular myocytes. *J. Mol. Cell. Cardiol.* **46**, 463–473 (2009). doi: 10.1016/j.yjmcc.2008.11.003; pmid: 19059414
35. D. S. Lee *et al.*, Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019). doi: 10.1038/s41592-019-0547-z; pmid: 31501549
36. G. Li *et al.*, Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019). doi: 10.1038/s41592-019-0502-z; pmid: 31384045
37. L. Luna-Zurita *et al.*, Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell* **164**, 999–1014 (2016). doi: 10.1016/j.cell.2016.01.004; pmid: 26875865
38. L. Waldron *et al.*, The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev. Cell* **36**, 262–275 (2016). doi: 10.1016/j.devcel.2016.01.009; pmid: 26859351
39. M. Ieda *et al.*, Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010). doi: 10.1016/j.cell.2010.07.002; pmid: 20691899
40. Y. S. Ang *et al.*, Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**, 1734–1749.e22, 22 (2016). doi: 10.1016/j.cell.2016.11.033; pmid: 27984724
41. N. J. Rinzema *et al.*, Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat. Struct. Mol. Biol.* **29**, 563–574 (2022). doi: 10.1038/s41594-022-00787-7; pmid: 35710842
42. E. S. M. Vos *et al.*, Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Mol. Cell* **81**, 3082–3095.e6 (2021). doi: 10.1016/j.molcel.2021.06.008; pmid: 34197738
43. K. Bansal *et al.*, Aire regulates chromatin looping by evicting CTCF from domain boundaries and favoring accumulation of cohesin on superenhancers. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2110991118 (2021). doi: 10.1073/pnas.2110991118; pmid: 34518235
44. J. Stieber *et al.*, The hyperpolarization-activated channel HCN4 is required for the generation of pacemaker action potentials in the embryonic heart. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15235–15240 (2003). doi: 10.1073/pnas.2434235100; pmid: 14657344
45. J. C. K. Man *et al.*, Genetic Dissection of a Super Enhancer Controlling the *Nppa-Nppb* Cluster in the Heart. *Circ. Res.* **128**, 115–129 (2021). doi: 10.1161/CIRCRESAHA.120.317045; pmid: 33107387
46. A. M. Gacita *et al.*, Genetic Variation in Enhancers Modifies Cardiomyopathy Gene Expression and Progression. *Circulation* **143**, 1302–1316 (2021). doi: 10.1161/CIRCULATIONAHA.120.050432; pmid: 33478249
47. A. J. Faure *et al.*, Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* **22**, 2163–2175 (2012). doi: 10.1101/gr.136507.111; pmid: 22780989
48. 4D Nucleome Consortium, An integrated view of the structure and function of the human 4D nucleome. bioRxiv 2024.09.17.613111 [Preprint] (2024); https://doi.org/10.1101/2024.09.17.613111.
49. L. Pradhan *et al.*, Intermolecular Interactions of Cardiac Transcription Factors NKX2.5 and TBX5. *Biochemistry* **55**, 1702–1710 (2016). doi: 10.1021/acs.biochem.6b00171; pmid: 26926761
50. M. E. Sweat *et al.*, Tbx5 maintains atrial identity in post-natal cardiomyocytes by regulating an atrial-specific enhancer network. *Nat. Cardiovasc. Res.* **2**, 881–898 (2023). doi: 10.1038/s44161-023-00334-7; pmid: 38344303
51. H. Lickert *et al.*, Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107–112 (2004). doi: 10.1038/nature03071; pmid: 15525990
52. J. K. Takeuchi *et al.*, Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat. Commun.* **2**, 187 (2011). doi: 10.1038/ncomms1187; pmid: 21304516
53. S. A. Miller, A. C. Huang, M. M. Miazgowicz, M. M. Brassil, A. S. Weinmann, Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. *Genes Dev.* **22**, 2980–2993 (2008). doi: 10.1101/gad.1689708; pmid: 18981476
54. D. M. Ibrahim, S. Mundlos, Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Curr. Opin. Cell Biol.* **64**, 1–9 (2020). doi: 10.1016/j.ceb.2020.01.003; pmid: 32036200
55. S. Naqvi *et al.*, Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat. Genet.* **55**, 841–851 (2023). doi: 10.1038/s41588-023-01366-2; pmid: 37024583
56. A. A. Alekseyenko, A. A. Gorchakov, P. V. Kharchenko, M. I. Kuroda, Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2488–2493 (2014). doi: 10.1073/pnas.1400648111; pmid: 24550272

57. B. Nabet *et al.*, The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018). doi: 10.1038/s41589-018-0021-8; pmid: 29581585
58. T. Yang *et al.*, HiCRep: Assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017). doi: 10.1101/gr.220640.117; pmid: 28855260
59. N. Abdennur *et al.*, Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLOS Comput. Biol.* **20**, e1012067 (2024). doi: 10.1371/journal.pcbi.1012067; pmid: 38709825
60. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
61. T. J. Salameh *et al.*, A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat. Commun.* **11**, 3428 (2020). doi: 10.1038/s41467-020-17239-9; pmid: 32647330
62. A. Chakraborty, J. G. Wang, F. Ay, dcHiC detects differential compartments across multiple Hi-C datasets. *Nat. Commun.* **13**, 6827 (2022). doi: 10.1038/s41467-022-34626-6; pmid: 36369226
63. H. Liu *et al.*, DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021). doi: 10.1038/s41586-020-03182-8; pmid: 34616061
64. N. R. Zemke *et al.*, Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390–402 (2023). doi: 10.1038/s41586-023-06819-6; pmid: 38092918
65. R. Zhang, T. Zhou, J. Ma, Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. *Cell Syst.* **13**, 798–807.e6 (2022). doi: 10.1016/j.cels.2022.09.004; pmid: 36265466
66. N. Abdennur, L. A. Mirny, Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020). doi: 10.1093/bioinformatics/btz540; pmid: 31290943
67. R. Zhang, T. Zhou, J. Ma, Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022). doi: 10.1038/s41587-021-01034-y; pmid: 34635838
68. P. A. Ewels *et al.*, The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020). doi: 10.1038/s41587-020-0439-x; pmid: 32055031
69. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
70. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017). doi: 10.1038/nmeth.4197; pmid: 28263959
71. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
72. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018). doi: 10.1093/bioinformatics/bty560; pmid: 30423086
73. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019). doi: 10.1038/s41587-019-0201-4; pmid: 31375807
74. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi: 10.1093/bioinformatics/btt656; pmid: 24227677
75. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
76. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
77. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278
78. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
79. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016). doi: 10.1093/nar/gkw257; pmid: 27079975
80. T. L. Bailey *et al.*, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37** (Web Server), W202–W208 (2009). doi: 10.1093/nar/gkp335; pmid: 19458158
81. J. A. Castro-Mondragon *et al.*, JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50** (D1), D165–D173 (2022). doi: 10.1093/nar/gkab1113; pmid: 34850907

ACKNOWLEDGMENTS

We thank members of the Gladstone Stem Cell Core, the Flow Cytometry Core, and the UCSF Center for Advanced Technologies for their expert assistance. We thank B. Conklin for the WTc11 iPSC line, M. Bernardi of the Gladstone Genomics Core for assistance with bulk RNA-seq library preparation, T. Sukonnik for generating ventricular CMs for RNA-seq libraries, R. Thomas of the Gladstone Bioinformatics Core for valuable feedback on bioinformatics analysis, F. Chanut for editorial assistance, T. Tolpa for design assistance, and members of the Bruneau and Pollard labs for discussions and comments. **Funding:** NIH 4D Nucleome Project (NHLBI U01 HL157989 to B.G.B., K.S.P., and I.S.K.; UM1HG011585 to B.R.); NHLBI R01 HL155906 (B.G.B. and I.S.K.); California Institute for Regenerative Medicine (Z.L.G.); Additional Ventures Innovation Award (B.G.B., K.S.P., D.S.); Additional Ventures Catalyst to Independence Award (Z.L.G.); The Gladstone Institutes (B.G.B., K.S.P., D.S.); The Roddenberry Foundation (B.G.B., D.S.); The Younger Family Fund (B.G.B., D.S.); UCSF Pediatric Heart Center Catalyst Award (I.S.K.); UCSF Anesthesia Research Support (I.S.K.); The Saving Tiny Hearts Society (I.S.K.); The National Science Foundation Graduate Research Fellowship (S.Z.). The Gladstone Flow Cytometry Core is funded by NIH S10 RR028962 and the James B. Pendleton Charitable Trust, and DARPA. Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. **Author contributions:** B.G.B., K.S.P., I.S.K., and Z.L.G. conceived the study. Z.L.G. generated atrial and ventricular Hi-C 3.0 datasets, bulk atrial CM RNA-seq, and TBX5 and RAD21 ChIP-seq datasets and all snm3C-seq samples. Z.L.G. and A.J.H. generated *TBX5* allelic series Hi-C 3.0 datasets. S.K. processed and analyzed all Hi-C 3.0 data, except dcHiC, which was analyzed by S.Z. Z.C. generated and analyzed TBX5-dTAG RNA-seq datasets. Z.C. generated H3K27ac ChIP-seq datasets. Z.L.G. generated CTCF ChIP-seq datasets with assistance and optimization from C.J. Z.L.G. and Z.C. processed ChIP-seq and bulk RNA-seq data, which were further analyzed by S.K. S.Z. analyzed snm3C-seq datasets. K.S.R. analyzed scRNA-seq datasets for *TBX5* allelic series with supervision and input from I.S.K. C.C. generated the TBX5-dTAG line with supervision and input from D.S. V.K. generated and analyzed bulk ventricular CM RNA-seq data. P.K.L., K.D., B.Y., and W.M.B. generated and processed snm3C-seq data with supervision and input from N.R.Z. and B.R. Z.L.G., S.K., S.Z., I.S.K., K.S.P., and B.G.B. interpreted the data. Z.L.G., S.K., and S.Z. prepared figures. Z.L.G., S.K., S.Z., and N.R.Z. wrote the original draft manuscript. Z.L.G., S.K., S.Z., I.S.K., K.S.P., and B.G.B. edited the manuscript with contributions from all coauthors. All authors approved the manuscript. **Competing interests:** B.G.B., K.S.P., and D.S. hold stock in Tenaya Therapeutics. C.C. and V.K. are current employees of Genentech and shareholders of Roche. B.R. is a cofounder of Epigenome Technologies and has equity in Arima Genomics. All other authors declare that they have no competing interests. **Data, code, and materials availability:** Cell lines are available under a material transfer agreement by contacting the corresponding authors. All original data are available on NIH GEO DataSets: Hi-C 3.0 (GSE285274), acute TBX5 depletion Hi-C 3.0 (GSE307018), bulk atrial and ventricular CM RNA-seq (GSE285277), acute TBX5 depletion bulk RNA-seq (GSE307021), TBX5 and CTCF ChIP-seq (GSE285268), RAD21 and H3K27ac ChIP-seq (GSE307019), and snm3c-seq (GSE285419). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv5434
Figs. S1 to S14; Reproducibility Checklist; Data S1 to S5
Submitted 23 December 2024; resubmitted 4 September 2025; accepted 6 November 2025
10.1126/science.adv5434

RESEARCH ARTICLE SUMMARY

Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure

Yang Xie†, Luca Tucciarone†, Elie N. Farah†, Lei Chang†, *et al.*

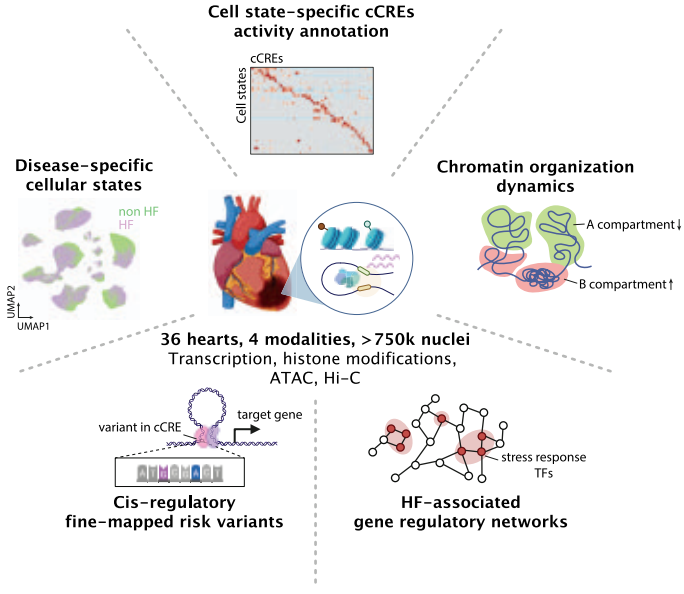
Full article and list of author affiliations: https://doi.org/10.1126/science.ady6893

INTRODUCTION: Heart failure represents the end stage of diverse cardiovascular pathologies, yet the molecular mechanisms that transform adaptive responses into maladaptive remodeling remain to be defined. Large-scale genetic studies have uncovered numerous disease-associated variants, the majority of which reside in noncoding regions of the genome, implicating their gene-regulatory roles as central drivers of pathogenesis. However, the heart comprises highly heterogeneous cell types, each with distinct transcriptional programs and stress responses. This cellular complexity poses a challenge for conventional bulk genomic assays to pinpoint the specific regulatory events that occur within individual cell types during disease progression.

RATIONALE: Gene regulation is controlled across multiple hierarchical layers, including chromatin architecture, DNA accessibility, and histone modifications. To determine how these regulatory layers are disrupted in heart failure, we applied a suite of complementary single-cell genomic approaches to all cardiac chambers of human hearts, including joint profiling of transcription and chromatin accessibility (10x Multiome), paired measurement of transcription and histone modifications (Droplet Paired-Tag), and single-cell chromatin conformation mapping (Droplet Hi-C). Through integrative analyses, we revealed cell type–specific disruptions to the gene-regulatory landscape in heart failure, identified rewiring of core transcriptional programs, and linked disease-associated risk variants to specific alterations across cardiac cell populations.

RESULTS: We constructed a multimodal cell atlas of 36 human hearts spanning atrial and ventricular chambers, cataloging major cardiac cell types and their associated disease states. Across all cell types, heart failure was accompanied by widespread remodeling of chromatin states at gene-regulatory elements, which was coupled to coordinated changes in transcription factor activity and target gene expression. Cardiomyocytes and fibroblasts in particular exhibited extensive rewiring of gene-regulatory programs, with trajectory analysis revealing intermediate cellular states bridging healthy and diseased conditions. These transitional states were marked by activation of stress-responsive transcription factors, implicating them as critical stages in disease progression.

Single-cell chromatin conformation analysis further revealed extensive and cell type–specific reorganization of the genome architecture in heart failure. In cardiomyocytes, disease was associated with disruption of chromatin compartments and gain in long-range chromatin interactions. These structural changes were closely linked to altered gene regulation, indicating that heart failure involves coordinated dysregulation at both the level of *cis*-regulatory elements and higher-order genome organization. Finally, integration with genetic association data further revealed that genetic risk variants for heart failure are enriched within cell type–specific active regulatory elements, especially in cardiomyocytes, and preferentially exert their effects through long-range chromatin interactions.



Single-cell multiomic analysis of human heart failure. Integrated single-cell multimodal profiling revealed multilayered gene regulation in failing human hearts. Transcription defines disease-associated states, and chromatin profiling annotates regulatory elements and uncovers disrupted genome organization. Together, these data expose disease-linked regulatory networks, connecting risk variants to target genes to provide a blueprint for therapeutic discovery. [Figure created with BioRender.com]

CONCLUSION: This study provides a cell type–resolved atlas of regulatory landscape remodeling in human heart failure, integrating transcriptional, chromatin state, and chromatin organization data across distinct cardiac cell types. Our findings demonstrate that heart failure pathogenesis involves the coordinated disruption of cell type–specific gene-regulatory networks encompassing altered transcription factor activity, dynamic changes in *cis*-regulatory element usage, and reorganization of genome structures. By linking heart failure–associated genetic variants to cell type–specific regulatory circuitry and putative target genes, these findings establish a mechanistic framework for decoding the molecular basis of developing more precise therapeutic strategies for heart failure. □

Corresponding authors: Neil C. Chi (nchi@health.ucsd.edu); Bing Ren (biren@health.ucsd.edu); Kyle J. Gaulton (kgaulton@health.ucsd.edu); Stavros G. Drakos (stavros.drakos@hsc.utah.edu); Allen Wang (a5wang@health.ucsd.edu) †These authors contributed equally to this work. Cite this article as Y. Xie *et al.*, *Science* 393, eady6893 (2026). DOI: 10.1126/science. ady6893

4D NUCLEOME

RESEARCH ARTICLE

Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure

Yang Xie[1,2]†, Luca Tucciarone[3]†, Elie N. Farah[4]†, Lei Chang[1,2]†, Qian Yang[5], Thirupura S. Shankar[6], Weston Elison[3,7], Shaina Tran[4], Jovina Djulamsah[5], Audrey Lie[1], Timothy Loe[1], Alyssa R. Holman[4], Sierra Corban[3], Justin Buchanan[5], Sainath Mamde[1,8], Haowen Zhou[4], Ruth M. Elgamal[3,7], Eleni Tseliou[6,9], Vincent Huang[6,9], Zhaoning Wang[1,10], Jeffrey Huey-Chuan Chiu[1,7], Rebecca Melton[3,7], Emily Griffin[3], Qingquan Zhang[4], Jacinta Lucero[5], Sutip Navankasattusas[6], Daofeng Li[11], Chanrung Seng[11], Eugin Destici[4], Craig H. Selzman[6,12], Agnieszka D'Antonio-Chronowska[5], Ting Wang[11,13], Allen Wang[5]*, Stavros G. Drakos[6,9]*, Kyle J. Gaulton[3,14]*, Bing Ren[1,2,5,10,14,15,16]*, Neil C. Chi[4,14,17]*

Heart failure is a leading cause of morbidity and mortality, yet gene-regulatory mechanisms driving cell type–specific pathologic responses remain undefined. Here, we present the cell type–resolved transcriptomes, chromatin accessibility, histone modifications, and chromatin organization of 13 nonfailing and 23 failing human hearts across all cardiac chambers. Integrative analyses revealed dynamic changes in cell type composition, gene-regulatory programs, and chromatin organization, particularly in cardiomyocytes and fibroblasts. Mapping cell type–specific enhancer-gene interactions from these analyses enabled the illumination of likely causal genetic contributors to heart failure from genetic association data. Together, these findings provide multimodal gene-regulatory maps of the human heart in health and disease, offering a framework for designing precise, cell type–targeted therapies for treating heart failure.

Heart failure (HF) remains a major global cause of morbidity and mortality (*1*). Its etiologies are multifactorial, encompassing acquired cardiovascular conditions such as ischemia, diabetes, hypertension, infections, and toxins, as well as genetic causes including monogenic dilated, hypertrophic, and restrictive cardiomyopathies (*2*–*4*). Recent genome-wide association studies (GWASs) have uncovered numerous HF-associated risk loci, underscoring the contribution of polygenic variation to HF susceptibility and ventricular traits (*5*–*8*). Although a minority of fine-mapped variants localize to coding regions, most reside within noncoding regions (>85%) and remain functionally uncharacterized (*5*–*8*). Delineating how these noncoding variants affect specific cardiac cell types will offer opportunities for developing therapeutic or disease prevention strategies.

The heart comprises diverse cardiac cell types that coordinately act to sustain cardiac function and circulation (*9*). However, how these cell types preserve heart performance and become perturbed in ischemic (ICM) and nonischemic (NICM) HF remains incompletely understood. Single-cell transcriptomic studies have begun to identify specific cell types and their interactions in HF (*10*, *11*); however, the regulatory elements and networks driving pathologic outcomes remain unresolved. Because most noncoding disease variants likely disrupt transcription factor (TF) binding and gene expression, elucidating gene-regulatory programs governing cell type–specific responses to HF is paramount to understanding its pathogenesis and developing targeted therapies. Investigating the chromatin landscape, including epigenomic and structural regulation, provides a framework to uncover changes in genome organization underlying cell type–specific cardiac stress responses (*12*) and to identify HF-associated noncoding variants that alter cell type–specific gene regulation. Recent advancements in single-cell technologies now enable such analyses across distinct cell types within human tissues (*13*, *14*).

Toward this end, we performed single-cell multiomic assays to investigate the transcription, chromatin accessibility, histone modification, and three-dimensional (3D) chromatin conformation of individual cells from major cardiac chambers of 36 adult non-HF and HF (ICM and NICM) human hearts. Integrative analyses revealed key cell type–specific changes in gene-regulatory and transcriptional programs during HF, and together with GWAS data, identified disease-enriched gene-regulatory networks underlying HF and cardiovascular risk.

Results

Cell type–specific transcriptomic, epigenomic, and 3D genome maps of human hearts

To investigate cell type–specific transcriptomic and epigenomic landscapes in non-HF and HF hearts, we collected 36 hearts, including both atrial and ventricular chambers, from 23 patients with HF and 13 non-HF patients for single-cell multimodal profiling (Fig. 1A and table S1). The HF cohort comprised 13 ICM and 10 NICM cardiomyopathy patients, with markedly reduced left ventricular ejection fraction (LVEF) (20.1 ± 7.3%) compared with non-HF controls (66.8 ± 9.4%) (table S1). Whole-genome sequencing (WGS) identified likely pathogenic variants in the *LDLR* (two ICM samples) and *FLNC* (one NICM sample) genes (tables S2 to S4). The cohort included 23 males and 13 females, with mean ages of 57 ± 11.3 (HF) and 47.6 ± 15.9 (non-HF), respectively (Fig. 1B and tables S1 and S2). Clinical metadata including genetically identified sex were incorporated as covariates in downstream analyses (see the Materials and methods).

To increase throughput and to minimize batch effects, we implemented a pooled sampling strategy (see the Materials and methods and fig. S1A). Tissues from 30 hearts (10 non-HF and 20 HF, including 10 ICM and 10 NICM) were combined into four chamber-specific pools for multimodal profiling using single-cell multiome [assay for transposase-accessible chromatin (ATAC) + RNA], Droplet Paired-Tag (histone modification + RNA), and Droplet Hi-C (single-cell Hi-C) (Fig. 1, B to D; fig. S1, B to J; and Materials and methods) (*13*, *14*). Cells were assigned to subjects using genotype information (table S2). Within pooled libraries,

[1]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [2]New York Genome Center, New York, NY, USA. [3]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA. [4]Department of Medicine, Division of Cardiology, University of California San Diego, La Jolla, CA, USA. [5]Center for Epigenomics, Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [6]Nora Eccles Harrison Cardiovascular Research and Training Institute (CVRTI), University of Utah, Salt Lake City, UT, USA. [7]Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA. [8]Bioengineering Graduate Program, University of California San Diego, La Jolla, CA, USA. [9]Division of Cardiovascular Medicine, University of Utah, Salt Lake City, UT, USA. [10]Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY, USA. [11]Department of Genetics, The Edison Family Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St Louis, MO, USA. [12]Division of Cardiothoracic Surgery, University of Utah, Salt Lake City, UT, USA. [13]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA. [14]Institute for Genomic Medicine, School of Medicine, University of California San Diego, La Jolla, CA, USA. [15]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA. [16]Department of Systems Biology, Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, New York, NY, USA. [17]Institute of Engineering in Medicine, University of California, San Diego, La Jolla, CA, USA. ***Corresponding author. Email: nchi@health.ucsd.edu (N.C.C.); biren@health.ucsd.edu (B.R.); kgaulton@health.ucsd.edu (K.J.G.); stavros.drakos@hsc.utah.edu (S.G.D.); a5wang@health.ucsd.edu (A.W.)** †These authors contributed equally to this work.

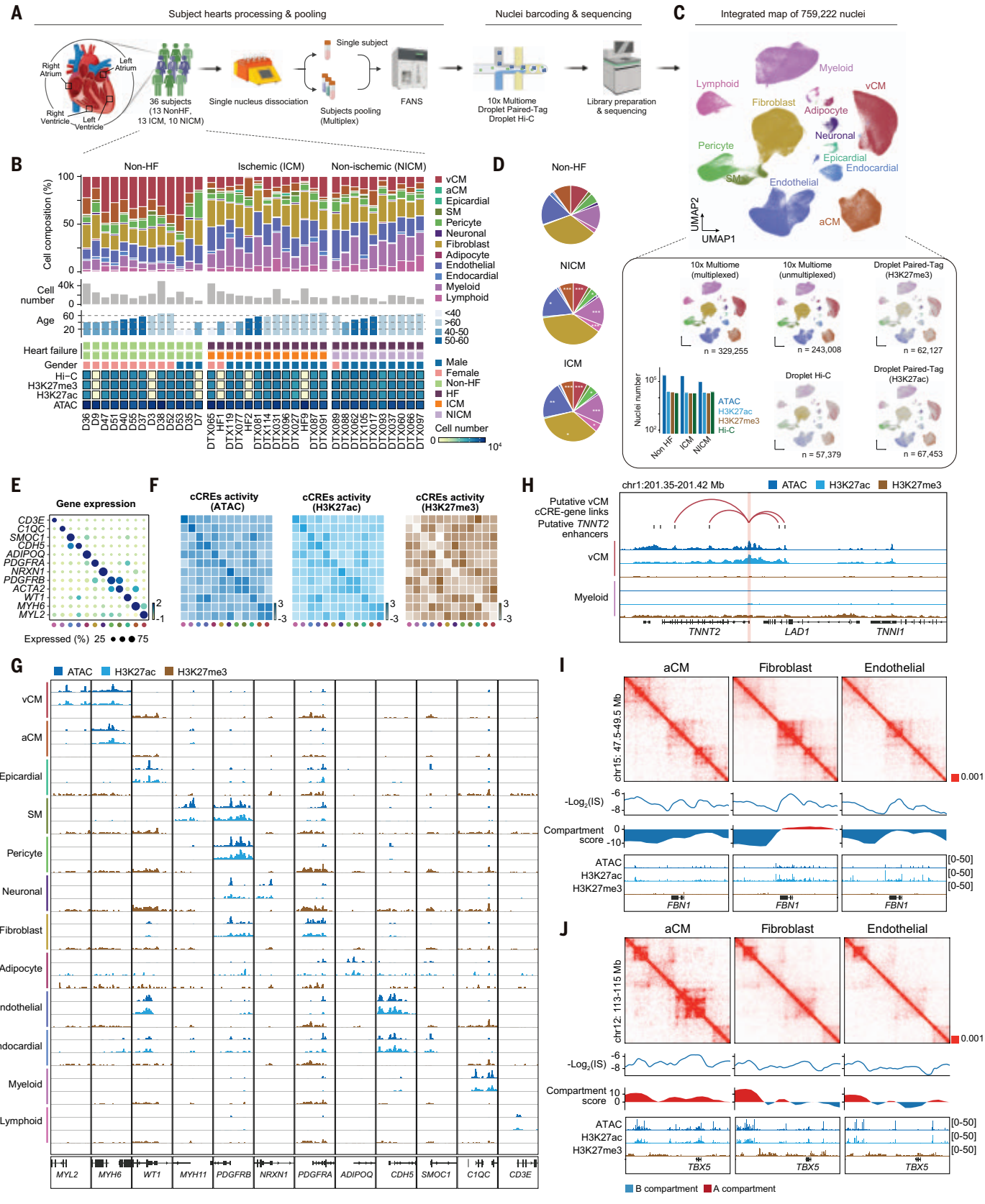


Fig. 1. Single-cell multiomic, histone modification, and Hi-C analyses uncover cell type–specific transcriptome, epigenome, and 3D genome maps of adult human hearts. (**A**) Schematic outlines of the workflow of single-cell 10x Multiome, Droplet Paired-Tag, and Droplet Hi-C studies of cardiac chambers from HF and control (non-HF) subjects. HF patients include those with ICM and NICM cardiomyopathy. (**B**) Stacked bar charts showing the cardiac cell type proportions and total nuclei number for each subject (top). Clinical metadata (subject age, sex, and disease condition), along with the number of nuclei profiled for each single-cell sequencing modality is shown below.

(**C**) Top, uniform manifold approximation and projections (UMAPs) showing the clustering of 759,222 nuclei into major cardiac cell types based on transcriptome or gene activity features from all single-cell sequencing modalities. Bottom, UMAPs showing the nuclei contribution from single-cell sequencing modalities to each major cardiac cell type cluster. Bottom left includes a barchart describing the number of nuclei captured per modality across disease conditions. The *x* axis is displayed on a $\log_{10}$ scale. (**D**) Pie charts displaying the percentage of cardiac cell types in non-HF and HF samples. (**E**) Expression of selected marker genes shown in dot plot across the 12 major cardiac cell types. (**F**) Heatmaps displaying the scaled, averaged epigenetic signals (ATAC, H3K27ac, and H3K27me3) for the selected marker genes in (E). (**G**) Genome browser tracks revealing the aggregated epigenetic profiles for all major cardiac cell types at selected marker gene loci. (**H**) Genome browser track showing an example of cCRE-gene links for *TNNT2* in vCMs (top). Putative *TNNT2* enhancers and epigenomic landscape for vCM versus myeloid lineage are displayed (bottom). (**I** and **J**) Representative examples of cell type–specific Hi-C contact maps for *FBN1* (I) and *TBX5* (J) loci at 10-kb resolution. The -log2 insulation score (IS), compartment score, and genome browser tracks for chromatin accessibility and histone modifications are shown below.

nuclei contributions were consistent across modalities despite modest variation in yield (figs. S1, B to D, and S2, A to F). To assess potential pooling bias, we generated unpooled single-cell multiome data from 10 individual hearts, which confirmed that pooling preserved data quality, cell representation, and doublet rates while increasing recovery (fig. S2, G to M, and table S5). Consistent with these findings, pseudobulk correlations showed minimal bias in gene expression and accessibility (fig. S2, A, D, and K).

Clustering of 329,255 nuclei from pooled multiome data identified 12 major cardiac cell types based on known markers (*10*, *15*, *16*), including atrial cardiomyocytes (aCMs), ventricular cardiomyocytes (vCMs), fibroblasts (FBs), endothelial cells, myeloid cells, pericytes, lymphoid cells, endocardial cells, smooth muscle (SM) cells, neuronal cells, epicardial cells, and adipocytes (figs. S3 and S4, A and B). Further subclustering resolved 65 subpopulations capturing both lineage diversity (e.g., endothelial subtypes) and disease-associated states (e.g., activated FBs) as described previously (figs. S3 and S4, C and D) (*10*, *17*). Cell proportions varied across chambers and disease states (Fig. 1D and fig. S4E), with HF hearts showing increased myeloid cells and FBs and reduced cardiomyocytes (*10*, *18*) (Fig. 1D). Chamber-specific effects included loss of atrial and ventricular cardiomyocytes, as well as prominent FB expansion in the LV (fig. S4F). To enhance coverage, 243,008 nuclei generated using a nonpooled single-cell multiome were mapped onto the pooled dataset.

We next integrated Droplet Paired-Tag (67,453 H3K27ac and 62,127 H3K27me3 nuclei) and Droplet Hi-C (57,379 nuclei) data with multiome data using gene expression or activity scores to generate a unified multimodal map (Fig. 1C and figs. S5 and S6A). This combined dataset comprised 759,222 nuclei spanning all major cell types (Fig. 1C, fig. S6B, and table S6). Integration showed balanced sex representation while preserving disease-specific structure (fig. S6, C to G) and consistent annotations across modalities (fig. S6H). Gene expression correlations recapitulated known cellular hierarchies (fig. S6, I to K). Consistent with expression patterns, chromatin accessibility and H3K27ac signals were enriched at the transcription start sites of cell type–specific genes (*MYL2*, *MYH6*, and *PDGFRA*), whereas H3K27me3 marked these regions in other lineages (Fig. 1, E to G, and fig. S7A). To map the regulatory elements directing these genes, we predicted putative enhancers for genes in each cell type using the Activity-by-Contact (ABC) model (*19*) and integrating Droplet Hi-C and H3K27ac Droplet Paired-Tag data. Enhancer activity represented by H3K27ac and chromatin accessibility correlated with gene expression, whereas H3K27me3 was depleted at these enhancers in nonexpressing cell types (Fig. 1H and fig. S7, B and C).

Finally, Droplet Hi-C revealed cell type–specific chromatin architecture changes aligned with transcriptional and epigenomic programs. Specifically, domain boundaries (fig. S7, D and E) and compartments (fig. S7, F to H) varied by cell type and correlated with gene expression. For example, *FBN1* resided in active A compartments with high H3K27ac in FBs but shifted to B compartments in cardiomyocytes and endothelial cells (Fig. 1I). Conversely, *TBX5* was localized within aCM-specific domain–like structures with active chromatin features, which were repressed in other lineages (Fig. 1J). Altogether, this single-cell multimodal dataset enables systematic determination of the gene-regulatory programs underlying cell type–specific responses during HF.

Integrative exploration of single-cell epigenomic studies identifies cell type–specific gene-regulatory programs of adult human hearts

Integrating multiple single-cell epigenomic profiles enabled both functional annotation of candidate *cis*-regulatory elements (cCREs) and classification of gene-regulatory programs across cardiac cell types and conditions. We performed iterative peak calling on aggregated chromatin accessibility profiles from 10x Multiome data to identify cCREs, and then applied non-negative matrix factorization (nNMF) to group cCREs into cell type–associated modules (Fig. 2A and table S7) (*20*). This analysis identified 285,873 cCREs covering 2.82% of the genome, expanding upon ENCODE SCREEN (*21*) and heart snATAC-seq annotations (*22*) and adding 49,848 (17.4%) previously uncharacterized elements (fig. S8A).

Clustering of cCREs identified 11 modules, including 10 that were cell type–enriched and one (M0) enriched for promoter elements with constant chromatin accessibility (Fig. 2A and fig. S8B). To characterize these elements, we used ChromHMM on pseudobulk profiles to define chromatin states by integrating chromatin accessibility, H3K27ac, and H3K27me3 (Fig. 2B and table S8) (*23*). A five-state model best captured major chromatin patterns without overfitting (fig. S8, C and D, and supplementary text) (*24*), classifying regions as active, open, weak, repressive, or unclassified based on epigenetic signals and genomic features overlap (Fig. 2B; fig. S8, E to H; and table S8). Despite using fewer histone marks, this model recapitulated previous ENCODE chromatin annotations on adult human hearts (fig. S8G) (*25*). Genome-wide annotation revealed that 3.75 and 3.02% of the genome was classified as open and active, respectively (fig. S8F). When evaluating cCREs, the percentage of open, active, and repressive states increased to 14.90, 15.24, and 11.89% (Fig. 2C). These states were largely cell type–dependent, reflecting lineage-specific regulation, with many cCREs active or open in their corresponding modules (except for M0) but repressed elsewhere (Fig. 2D). Notably, 75.1% of repressive cCREs were active or open in at least one other cell type (fig. S8I) and were linked to lineage-specific functions (fig. S8J), supporting coordinated regulation across cell types.

Building on this framework, we used likelihood ratio testing and specificity scoring [empirical false discovery rate (FDR) < 0.05; see the Materials and methods] to identify 20,267 cell type–specific cCREs, which grouped into 10 classes corresponding to the major cardiac cell types (table S9). Two classes were shared between vCMs and aCMs and between pericytes and SM cells, consistent with their lineage similarities (*9*) (fig. S9A). Most cell type–specific cCREs were active (~86.3%), with lower activity in epicardial cells and adipocytes likely reflecting limited coverage (fig. S9B). Motif enrichment analysis identified 333 known TF motifs with cell type–specific enrichment patterns (fig. S9C and table S10), including MEF2A and MEIS2 in cardiomyocytes, SOX4 in endocardial cells, and SPIB in myeloid lineages (fig. S9C). Gene set enrichment analysis (GSEA) (*26*, *27*) linked these cCREs to cell type–specific functions, such as cardiac muscle contraction in cardiomyocytes and T cell receptor signaling pathway in lymphoid cells (fig. S9D and table S11). Together, these analyses show that cell type–specific cCREs likely function as enhancers that regulate essential genes in distinct cardiac lineages.

To connect distal cCREs to target genes, we applied the ABC model and identified 66,317 distal cCRE-promoter pairs (fig. S9, E to H, and

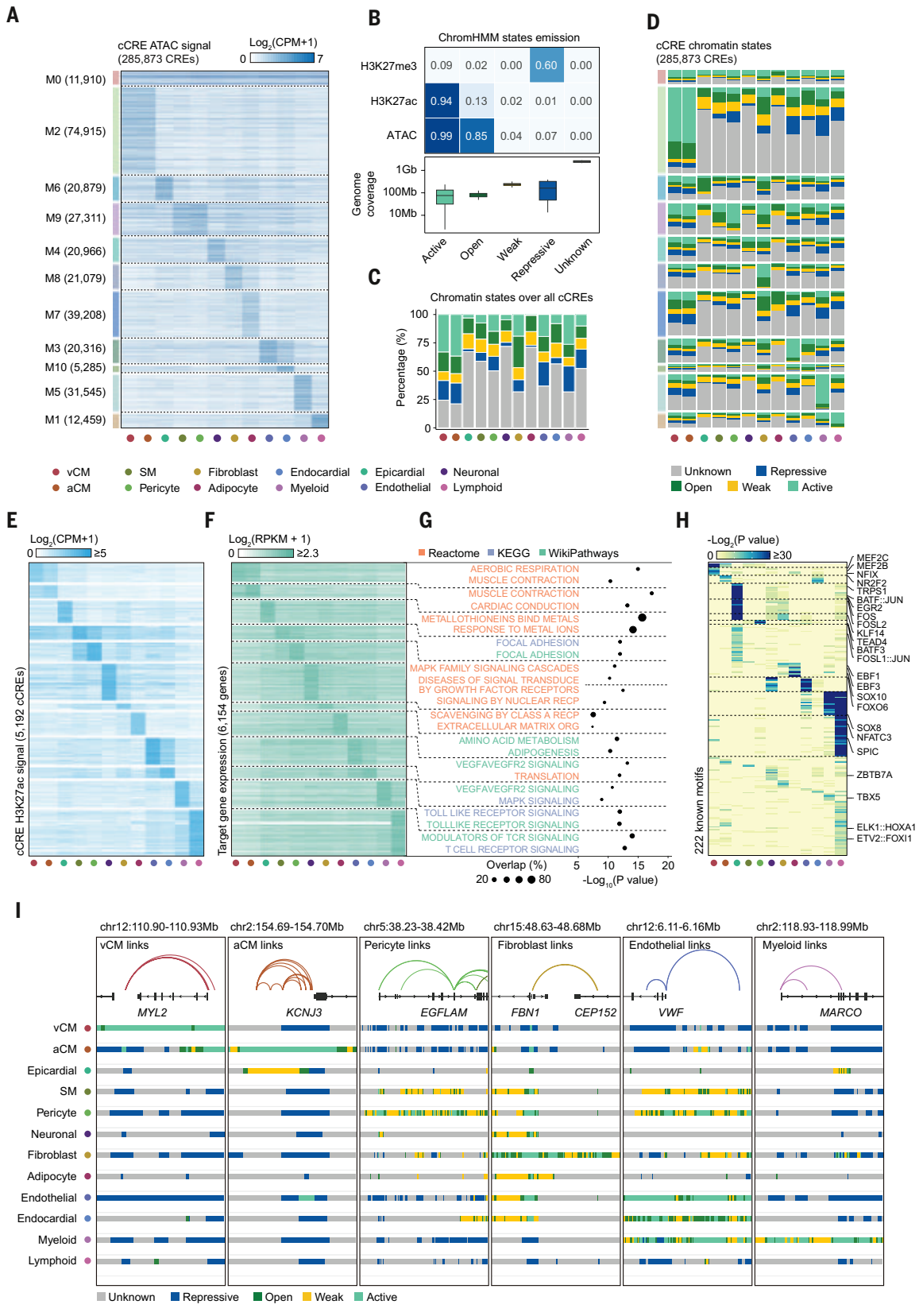


Fig. 2. Comprehensive epigenomic profiling of human cardiac cell types reveals their gene regulation. (**A**) Heatmap showing chromatin accessibility signal of cCREs in 11 nNMF modules (M0 to M10) across 12 major cell types. (**B**) Heatmap showing the emission probability of three epigenetic marks across five chromatin states as identified by ChromHMM (top). Boxplot shows the genomic coverage of ChromHMM-annotated chromatin states (bottom). The *y* axis represents $\log_{10}$-scaled genomic coverage. (**C**) Stacked bar charts showing the percentage of cCREs annotated in each chromatin state across all major cell types. (**D**) Stacked bar charts showing the proportion of chromatin states over cCREs in each nNMF module described in (B) across all major cell types. (**E**) Heatmap showing the H3K27ac signal on cCREs in cell type–specific distal cCRE-gene links.

(**F**) Heatmap showing the gene expression level of target genes associated with cCREs from (E) in cell type–specific distal cCRE-gene links. (**G**) Dot plot showing top enriched GO terms from GSEA analysis of cell type–specific genes targeted by distal cCREs in each cell type. *P* values were calculated using a permutation test in the GSEA tool. (**H**) 222 HOMER known TF motifs enriched in cCREs from cell type–specific distal cCRE-gene links. Example motifs in different cell types are shown. (**I**) Representative genome browser tracks for ChromHMM chromatin states across all major cell types on selected cell type–specific distal cCRE-gene links, including *MYL2* (vCM specific), *KCNJ3* (aCM specific), *EGFLAM* (pericyte specific), *FBN1* (FB specific), *VWF* (endothelial cell specific), and *MARCO* (myeloid cell specific).

table S12), including 11,938 cell type–specific pairs (Fig. 2, E and F). These links were enriched for chromatin contacts (fig. S9I) and active chromatin states (fig. S9B) in the corresponding cell types, and their target genes displayed concordant expression. GSEA of target genes revealed pathways consistent with cell type–specific functions (Fig. 2G and table S13), and motif analysis identified 222 TF motifs associated with these regulatory interactions (Fig. 2H and table S14). Examples include vCM-specific *MYL2* linked to a downstream cCRE, aCM-specific *KCNJ3* linked to multiple upstream cCREs, and endothelial *VWF* linked to an upstream cCRE (Fig. 2I). Altogether, these findings highlight how cell type–specific epigenetic landscapes orchestrate precise gene regulation across cardiac cell types.

Global gene-regulatory and chromatin structural changes of cardiac cell types during HF

Multimodal single-cell data integration revealed cell type–specific regulatory and chromatin architecture changes underlying HF. DESeq2 analysis identified 10,400 differentially expressed genes and 52,738 differentially accessible cCREs (FDR < 0.05) between HF and non-HF hearts (*28*). Differential cCREs were mostly distal (84.1%) and were enriched in CMs (4373 in aCMs and 28,557 in vCMs), FBs (18,292), and myeloid cells (4121), correlating with their transcriptional changes (Fig. 3A; fig. S10, A to D; and tables S15 and S16). Up-regulated cCREs primarily transitioned from open to active states, whereas down-regulated elements showed reduced active and open states (Fig. 3B). Motif analysis identified TF motifs enriched in both up- and down-regulated cCREs (fig. S10E). Although many motifs were cell type specific, several were shared across cell types, including stress-response AP-1 motifs (JUN, FOS, and ATF), suggesting global stress-response activation across cardiac cell types during HF. Conversely, motifs such as MEF2A and ESRRB were enriched among HF-down-regulated cCREs in vCMs but not aCMs, indicating more extensive gene-regulatory remodeling in vCMs during HF (fig. S10E). To compare molecular features between ICM and NICM cardiomyopathies, we performed differential analyses between these conditions, which displayed similar cellular composition and molecular changes, as described previously (*29*) (figs. S4E and S10, F to H), but selective differences in angiogenic and inflammatory pathways (fig. S10, I to M). Epigenetic remodeling was pronounced in the cardiomyocyte, FB, endothelial, pericyte, and myeloid populations (Fig. 3A). H3K27ac changes largely paralleled chromatin accessibility; however, in vCMs, a substantial fraction of cCREs gained H3K27ac without increased accessibility (52.3%, 1630 of 3116; "H3K27ac-exclusive" cCREs) (fig. S11A). Chromatin state analysis indicated that these elements preferentially transitioned to active rather than open states in HF, suggesting H3K27ac-driven enhancer activation independent of chromatin opening (fig. S11B). These elements were associated with cardiac muscle genes, including those implicated in HF (fig. S11C) (*30*–*32*), and may represent primed enhancers facilitating rapid transcriptional response in HF. Concurrently, 37,924 genomic regions exhibited reduced H3K27me3 across cell types (Fig. 3A and table S17), indicating global derepression and disruption of PRC2-mediated gene silencing during HF (*33*, *34*).

To link these distal cCREs to target genes, we applied the ABC model and identified 66,317 cCRE-promoter pairs, including disease-associated connections in cardiomyocytes (*NKX2-5*, *ADRB2*, and *CKM*), FBs (*NPPC* and *FAP*), and myeloid lineages (*NLRP3* and *CD163*) (table S18), which showed concordant changes in gene expression and chromatin contacts (Fig. 3, C and D). For example, *PM20D2* up-regulation in HF cardiomyocytes was associated with enhanced promoter activity through activation of a distal enhancer (Fig. 3E). Gene ontology (GO) analysis revealed HF–up-regulated links enriched for muscle adaptation in vCMs and extracellular matrix organization in FBs, whereas HF–down-regulated links were associated with collagen metabolism in FBs or heart contraction in vCMs (fig. S11D). In summary, these findings reveal how distal cCREs regulate target gene activities in a cell type–specific manner in HF.

To identify TFs regulating cell type–specific responses during HF, we constructed gene-regulatory networks (GRNs) by integrating single-cell transcriptomic and chromatin accessibility data. Disease-associated GRNs were classified based on differential TF-targeting cCRE activity between disease and control conditions (fig. S12A and tables S19 and S20) (*35*), identifying 41 cell type–specific and 23 shared TFs (Fig. 3F). As examples, BACH2 and RUNX2 were selectively activated in cardiomyocytes and FBs, respectively, whereas HEY2 was down-regulated in vCMs during HF. Despite transcriptional similarities, aCMs and vCMs furthermore displayed distinct GRNs and TF-targeted cCREs (fig. S12, B to E). For instance, >70% of BACH2-targeted cCREs displayed differing accessibilities (fig. S12C). Across disease etiologies in vCMs, NICM and ICM shared core TF motifs but particularly diverged in stress-related regulators (fig. S12, F and G), with BACH2 and FOS more activated in NICM, corresponding to increased expression of target genes such as *NPPB* (fig. S12H). Furthermore, we identified "bidirectional" TFs with opposing activities across cell types in the GRN analysis, indicating cell type– and condition-specific TF activity and regulatory mechanisms (fig. S12I). For example, MEF2A repressed targets in vCMs but activated them in FBs during HF (fig. S12, J to L). Additional analysis of vCM GRNs revealed dynamic shifts in transcriptional regulation between non-HF and HF hearts (Fig. 3G). GO analysis of HF-associated regulons identified shared pathways among target genes, such as wound healing, guanosine triphosphatase (GTPase) regulation, and kinase activation, yet their targeted cCREs showed distinct activities across conditions (Fig. 3H). Although certain TFs, such as FOS, exhibited elevated GRN activity in multiple cell types, its expression increased significantly [adjusted *P* value (P_{adj}) < 0.05] only in endothelial cells (Fig. 3F). Despite its broad GRN activity, FOS targeted distinct cCREs and genes in a cell type–specific manner (Fig. 3, I and J, and fig. S12M). These target genes were linked to both stress responses and lineage-specific functions (fig. S12N). Furthermore, chromatin state mapping indicated that vCM-specific FOS-targeted cCREs were predominantly open or active in vCMs but were more variable in myeloid cells (Fig. 3J). These vCM FOS-associated cCREs became increasingly accessible and active during HF, suggesting that they are primed under non-HF conditions and activated upon cardiac stress (Fig. 3J). Altogether, these findings uncover key transcriptional programs in HF across many major cardiac cell types including CMs, thus highlighting the distinct gene-regulatory changes underlying HF progression.

Finally, chromatin interaction profiles revealed not only cCRE-gene links but also the global organization pattern of nuclei. Further Droplet Hi-C analyses identified 5221 genomic bins undergoing compartment switching across cell types (Fig. 3K), coinciding with coordinated changes in cCRE activity and gene expression (Fig. 3L and fig. S13A). For example, *PRELID2* down-regulation in cardiomyocytes was associated with A-to-B compartment switching, whereas *DIAPH3* up-regulation

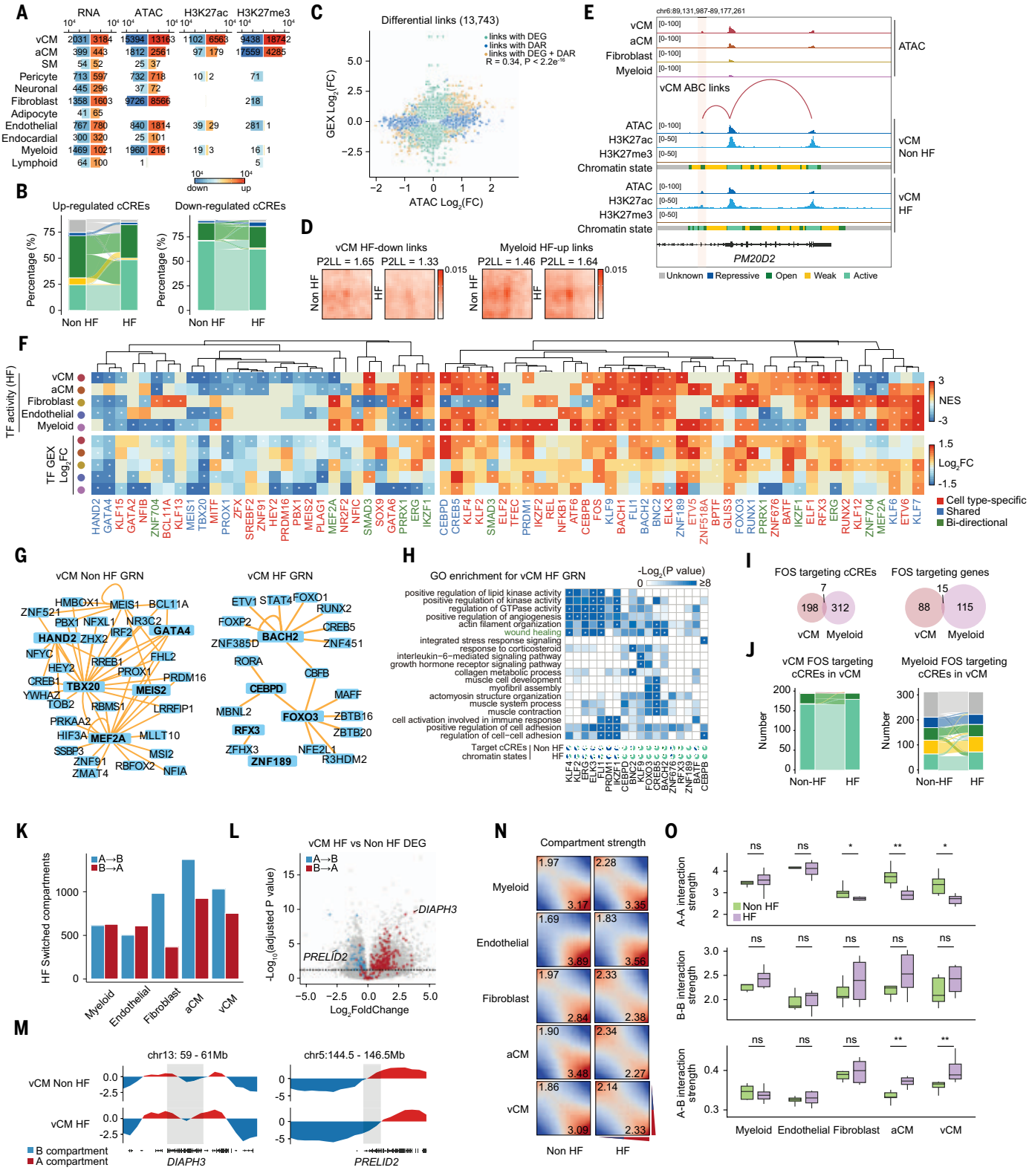


Fig. 3. Cell type–specific epigenetic changes direct gene-regulatory responses during HF. (**A**) Bar plot showing the number of genomic regions and genes identified from differential analysis between HF and non-HF conditions across cell types. *x* axes are displayed on a $\log_{10}$ scale. (**B**) Sankey plots showing the chromatin state dynamics for differentially accessible cCREs in HF and non-HF hearts. Averaged percentage of chromatin states across cell types is shown. (**C**) Scatter plots comparing the log2FCs of gene expression (GEX) and chromatin accessibility (ATAC) for genes and cCREs from putative ABC links between HF and non-HF conditions. Links are colored based on whether the associated gene and/or cCRE is classified as differentially expressed (DEG) or differentially accessible (DAR). (**D**) Normalized aggregate peak analysis (APA) scores of Droplet Hi-C signals in vCM and myeloid cells for vCM down-regulated and myeloid up-regulated ABC links. APA scores are represented as P2LL (peak-to-lower-left) values. (**E**) Genome browser tracks showing an example of a differential ABC link targeting *PM20D2* in vCM. Pseudobulk chromatin accessibility tracks across selected cardiac cell types (top), as

well as differential epigenetic signals and chromatin states in HF versus non-HF vCMs (bottom), are shown. (**F**) Heatmap showing the enrichment of HF- or non-HF–associated TF-regulated GRNs calculated from fGSEA across selected cell types (top). Only GRNs with an absolute NES >2 and with consistent differential TF expression results are shown. Asterisk indicates significant enrichment ($P_{adj} < 0.05$), with *P* values estimated based on an adaptive multilevel split Monte-Carlo scheme in the *fgsea* package. The log2FCs of each TF corresponding to the respective GRN are also shown (bottom). Asterisk indicates significant enrichment ($P_{adj} < 0.05$), calculated using Wald test in the *DESeq2* package and corrected for multiple testing using Benjamini-Hochberg FDR correction. (**G**) Network plots showing representative vCM GRNs in non-HF versus HF conditions. (**H**) Heatmap showing enrichment of GO terms for HF-associated TF GRN-targeted genes in vCMs (top). Pie charts below show the chromatin state percentage for HF-associated GRNs-targeting cCREs under non-HF and HF conditions (bottom). Asterisk indicates significant enrichment ($P_{adj} < 0.05$). *P* values were calculated using hypergeometric tests in enrichGO and corrected for multiple testing using Benjamini-Hochberg FDR correction. (**I**) Venn diagrams showing the overlap of FOS target genes and cCREs in vCMs versus myeloid. (**J**) Sankey plot showing the dynamic changes of vCM chromatin states of FOS-targeted chromatin regions in HF versus non-HF conditions for vCMs and myeloid cells. (**K**) Bar chart showing the number of switched compartments in HF versus non-HF for selected major cell types. (**L**) Volcano plot showing the log2FCs and P_{adj} value of differentially expressed genes in vCMs between non-HF and HF conditions. Genes that overlap with switched compartments are colored in blue (A to B) or red (B to A). Example genes shown in (M) are labeled. *P* values were calculated and corrected as described in (F). (**M**) Representative genome browser track of compartment scores in HF versus non-HF vCMs for *DIAPH3* and *PRELID2*. Compartments that shifted are shaded in gray. (**N**) Saddle plots showing compartmentalization strength comparison in selected major cell types between non-HF and HF conditions. Numbers indicate A-A (bottom right) and B-B (top left) compartment interaction strengths. (**O**) Boxplots comparing intracompartment (A-A or B-B) and intercompartment (A-B) interaction strengths in non-HF versus HF cell types in young (≤60 years) subjects.

corresponded to B-to-A transitions (Fig. 3, L and M, and fig. S13, B to D). Notably, cardiomyocytes and FBs exhibited increased A-to-B switching and reduced A-A interactions, indicating loss of active chromatin organization (Fig. 3, K and N). The reduced A-A interactions were associated with decreased insulation, suggesting rewiring of enhancer-promoter interactions (fig. S13E). Additionally, cardiomyocytes displayed increased cross-compartment interactions (Fig. 3O and fig. S13F), reflecting their disrupted nuclear architecture in HF. Specifically in vCMs, these changes involved gains in long-range interactions within inactive B compartments (fig. S13, G to I) and loss of short-range interactions in active compartments (Fig. 3O and fig. S13, I and J). Overall, these results demonstrate coordinated transcriptional, epigenetic, and chromatin structural remodeling across cardiac cell types during HF and have been made available at https://epigenome.wustl.edu/HeartEpigenome/.

Cardiomyocytes display a broad spectrum of cellular states in control and diseased hearts

Consistent with cell types affected during HF, CMs displayed increased heterogeneity among cell types between nondiseased and diseased hearts (Figs. 4 and 5 and figs. S14 to S18). Specifically, we identified seven vCM and five aCM subpopulations that expressed vCM-specific *MYL2* and aCM-specific *NR2F2* (*9*, *36*) within corresponding cardiac chambers (Fig. 4 and fig. S14). All CM subpopulations were detected in both non-HF and HF hearts, but their relative abundances varied with disease status (Fig. 4B and figs. S14, A and B, S19, and S20) and corresponded with previously reported CM subpopulations (fig. S3) (*10*, *17*).

Focusing on the ventricles, the primary cardiac chambers affected during HF, we identified three vCM subpopulations primarily from non-HF ventricles (vCM1 to vCM3), two mainly from HF ventricles (vCM6 and vCM7), and two containing vCMs from both non-HF and HF vCMs (vCM4 and vCM5) (Fig. 4, A and B, and fig. S19, A to C). These vCM subpopulations displayed distinct spatial distributions and regional enrichment across heart tissues, varying with disease conditions (fig. S20). Among non-HF vCM subpopulations (vCM1 to vCM3), which corresponded to those previously described (fig. S3) (*10*, *17*), many genes and chromatin regions were commonly up-regulated and accessible, including *MYL2* and *TBX20* genes, but those associated with HF-related vCM states (vCM6 and vCM7) were down-regulated (Fig. 4, B and C, and tables S21 and S22). Of the highly disease-enriched vCM subpopulations (vCM6 and vCM7), vCM7 expressed HF-related genes (*NPPA* and *NPPB*) and inflammatory-related genes (*TGFB2* and *IL1R1*) and were related to the stressed vCMs observed in NICM cardiomyopathy hearts (fig. S3) (*10*). Conversely, vCM6 HF-associated subpopulations were distinct from vCM7 and stressed vCMs and displayed up-regulated genes related to DNA repair and metabolism, including *FOXO1*, *FOXO3*, and *LINC02388*, which has been suggested to promote blood vessel formation in response to hypoxia (*37*) (Fig. 4, B and D, and table S22). By contrast, vCM4 and vCM5 displayed genes and accessible chromatin regions that were differentially up-regulated in either nondisease or disease-enriched vCMs but at lower levels (Fig. 4B and table S22), supporting that they may represent intermediate or transition vCM states during HF. Stress-response, wound-healing, and inflammatory-related genes were up-regulated in vCM5 (Fig. 4B and tables S22 and S23), suggesting that inflammatory and wound-healing pathways regulating activated FBs (*38*, *39*) may also initiate adaptive stress and disease responses in vCMs. Consistent with this notion, vCM5 highly expressed *NR4A1* and *NR4A3*, transcriptional regulators that modulate cardiomyocyte hypertrophy under stress and control key cardiac neurohormonal pathways, including the renin-angiotensin-aldosterone system and β-adrenergic sympathetic signaling (*40*). Supporting these findings, cell label transfer analysis revealed that vCM4 and vCM5 corresponded to previously described vCM states with similar inflammatory and wound-healing gene profiles (*10*, *17*) (fig. S3). Illuminating the disease relationship of these vCM subpopulations, scVelo RNA velocity analysis revealed disease state trajectories initiating from nondisease vCMs (vCM1 to vCM3), progressing through intermediate disease states (vCM4 and vCM5), and culminating in disease-associated states (vCM6 and vCM7) (Fig. 4A and table S24).

SCENIC+ (*35*) was applied to the integrated single-cell multiome data for all cardiac cell subpopulations to predict their changes in gene-regulatory programs including TF and chromatin accessibility (Figs. 4C and 5C, figs. S14 to S18, and table S20). Building upon our cell type–specific HF GRN analyses (Fig. 3F and fig. S12A), we investigated how gene-regulatory programs of identified vCM states may be regulated across heart conditions (Fig. 4C). In particular, several known CM-specific TFs, including GATA4, HEY2, MEF2A, MEF2C, and TBX20, were expressed and displayed high target gene accessibility in non-HF vCM states but were down-regulated across intermediate and disease-related vCM states, suggesting a loss of vCM identity during cardiac disease and stress. Conversely, transcriptional regulators that progressively increased their expression from non-HF to HF-enriched vCM states were identified, including FOXO3 (vCM6) and CEBPB (vCM7), a TF implicated in adaptive CM hypertrophic and injury responses (*41*, *42*) (Fig. 4C). Corresponding to the up-regulation of stress-response and wound-healing genes, vCM4 and vCM5 highly expressed JUN, FOS, ATF, EGR, and SMAD related TFs, supporting that these transient stress-response programs may initiate a cascade of responses leading to chronic disease vCM states (Fig. 4C). As such, vCM5 exhibited large complex GRNs with TFs that displayed high centrality scores indicating their elevated interconnectedness with other TFs within and across vCM subpopulation GRNs spanning both nondiseased and diseased states (Fig. 4E). Consistent with these findings, differential chromatin marks across vCM states revealed loss of H3K27me3 from non-HF to HF vCM states, affecting not only stress-response and wound-healing genes but also noncardiac developmental genes (fig. S19, D to H).

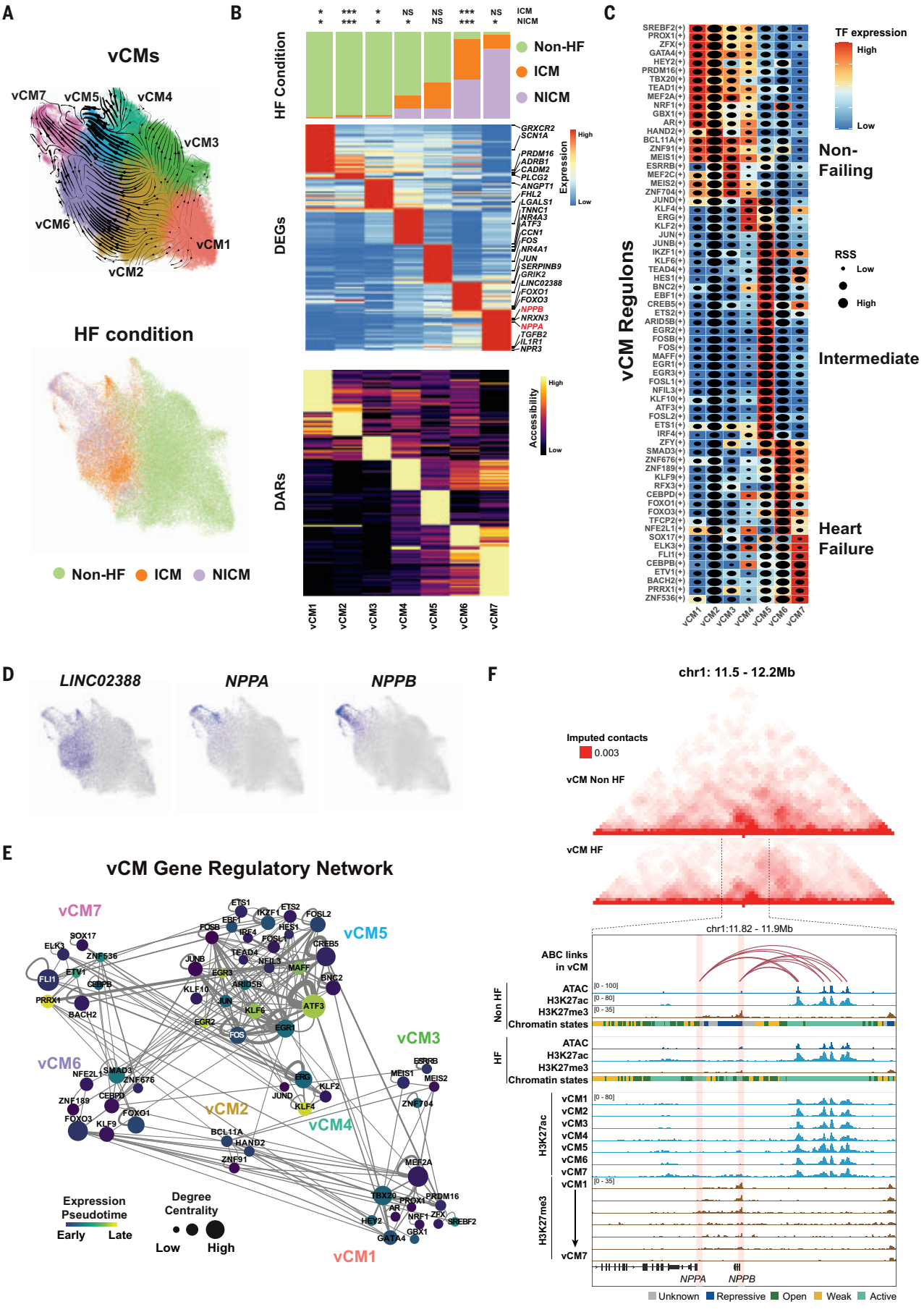

A
vCMs
vCM7
vCM5
vCM4
vCM3
vCM6
vCM1
vCM2
HF condition
Non-HF
ICM
NICM
B
HF Condition
Non-HF
ICM
NICM
DEGs
Expression
High
Low
GRXCR2
SCN1A
PRDM16
ADRB1
CADM2
PLCG2
ANGPT1
FHL2
LGALS1
TNNC1
NR4A3
ATF3
CCN1
FOS
NR4A1
JUN
SERPINB9
GRIK2
LINC02388
FOXO1
FOXO3
NPPB
NRXN3
NPPA
TGFB2
IL1R1
NPR3
DARs
Accessibility
High
Low
vCM1
vCM2
vCM3
vCM4
vCM5
vCM6
vCM7
C
vCM Regulons
TF expression
High
Low
RSS
Low
High
Non-Failing
Intermediate
Heart Failure
SREBF2(+)
PROX1(+)
ZFX(+)
GATA4(+)
HEY2(+)
PRDM16(+)
TBX20(+)
TEAD1(+)
MEF2A(+)
NRF1(+)
GBX1(+)
AR(+)
HAND2(+)
BCL11A(+)
ZNF91(+)
MEIS1(+)
ESRRB(+)
MEF2C(+)
MEIS2(+)
ZNF704(+)
JUND(+)
KLF4(+)
ERG(+)
KLF2(+)
JUN(+)
JUNB(+)
IKZF1(+)
KLF6(+)
TEAD4(+)
HES1(+)
BNC2(+)
EBF1(+)
CREB5(+)
ETS2(+)
ARID5B(+)
EGR2(+)
FOSB(+)
FOS(+)
MAFF(+)
EGR1(+)
EGR3(+)
FOSL1(+)
NFIL3(+)
KLF10(+)
ATF3(+)
FOSL2(+)
ETS1(+)
IRF4(+)
ZFY(+)
SMAD3(+)
ZNF676(+)
ZNF189(+)
KLF9(+)
RFX3(+)
CEBPD(+)
FOXO1(+)
FOXO3(+)
TFCP2(+)
NFE2L1(+)
SOX17(+)
ELK3(+)
FLI1(+)
CEBPB(+)
ETV1(+)
BACH2(+)
PRRX1(+)
ZNF536(+)
vCM1
vCM2
vCM3
vCM4
vCM5
vCM6
vCM7
D
LINC02388
NPPA
NPPB
E
vCM Gene Regulatory Network
vCM7
vCM5
vCM3
vCM6
vCM2
vCM4
vCM1
Expression Pseudotime
Early
Late
Degree Centrality
Low
High
F
chr1: 11.5 - 12.2Mb
Imputed contacts
0.003
vCM Non HF
vCM HF
chr1:11.82 - 11.9Mb
ABC links in vCM
Non HF
HF
ATAC
H3K27ac
H3K27me3
Chromatin states
[0 - 100]
[0 - 80]
[0 - 35]
H3K27ac
H3K27me3
vCM1
vCM2
vCM3
vCM4
vCM5
vCM6
vCM7
NPPA
NPPB
Unknown
Repressive
Open
Weak
Active

Fig. 4. GRN analysis of ventricular cardiomyocytes uncovers genetic programs controlling their cellular states. (**A**) UMAPs showing vCM cell subpopulations and their RNA velocity–derived trajectories (top) and the contribution of HF and non-HF vCMs to each subpopulation (bottom). (**B**) Each vCM cell subpopulation exhibits distinct cellular contributions (top, bar graph) from each heart condition as well as gene expression (middle, heatmap) and chromatin accessibility (bottom, heatmap). *$P_{adj} < 0.1$, **$P_{adj} < 0.05$, ***$P_{adj} < 0.01$; NS, not significant. (**C**) SCENIC+ TF GRNs for each vCM cell subpopulation shown in heatmap. The box is colored by TF expression, and the dot size shows the RSS. (**D**) UMAPs showing the gene expression of marker genes for distinct vCM disease states. (**E**) Analyzing the TF network within the vCM GRN inferred the interactions between distinct GRNs across non-HF and HF vCM subpopulations. Nodes are colored by pseudotime of expression. Node size was determined by their degree of centrality in the network, and edges represent TF-TF connections. (**F**) Hi-C contact maps (top) and representative genome browser tracks (bottom) at *NPPA* and *NPPB* locus showing disease- and subpopulation-specific pseudobulk epigenetic signals, chromatin states, and ABC links associated with *NPPA* and *NPPB*.

Integrating vCM-specific gene-regulatory programs with corresponding epigenetic signals and 3D-chromatin features revealed how pathological vCM gene programs may become activated during HF (Fig. 4F and fig. S19, G to J). Notably, the HF-related genes *NPPA* and *NPPB*, which tandemly reside within the same genomic locus on chromosome 1, were reciprocally regulated in non-HF and HF vCMs (Fig. 4F). Consistent with the low *NPPA* but negligible *NPPB* expression in non-HF vCMs (Fig. 4D and fig. S19G), the *NPPA* locus displayed weak activity, whereas *NPPB* was repressed (Fig. 4F). ABC analyses and vCM-specific Hi-C contact maps revealed a superenhancer ~40 kb upstream of *NPPB* that interacted with the *NPPA* promoter to control its activity, as described previously (*43*) (Fig. 4F). Although the chromatin state for both genes became active in HF vCMs (Fig. 4F), *NPPA* expression remained lower than that of *NPPB* (fig. S19G). This expression pattern corresponded to altered Hi-C contact maps within the *NPPA-NPPB* locus of HF vCMs, revealing reciprocal superenhancer interactions with *NPPA* and *NPPB* (Fig. 4F and fig. S19, I and J). Chromatin profiling across vCM states revealed progressive loss of the repressive mark H3K27me3 at the *NPPB* locus during the transition from non-HF to HF vCMs (Fig. 4F), suggesting that derepression enables *NPPB* activation while contributing to *NPPA* down-regulation. Finally, SCENIC+ analysis identified TFs potentially driving dynamic regulation of the *NPPA-NPPB* locus in diseased vCM states (fig. S19K). Specifically, *CREB5* was predicted to interact with superenhancer CREs to regulate both *NPPA* and *NPPB*, whereas distinct TFs were observed to bind specific CREs of *NPPA* and *NPPB* to control their expression, including *FOXO3* and *CEBPB/BACH2*, respectively (fig. S19K). Overall, these findings support a spectrum of CM subpopulations across ventricular and atrial cardiac chambers (supplementary text and fig. S14) that may represent potential conserved transitions of CMs from non-disease to specific disease states, including intermediate- and end-disease and stress cellular states during HF and cardiac stress.

Gene-regulatory programs activating diseased FBs

Cardiac FBs actively maintain and remodel the heart under a wide range of cardiac conditions. Supporting their dynamic role in cardiac homeostasis and disease, seven cardiac FB states were identified and contributed across all conditions (non-HF, ICM, and NICM). Notably, these FB states correlated with those reported in previous human heart single-cell RNA-sequencing (scRNA-seq) studies and showed region-specific enrichment in diseased heart tissues (Fig. 5 and figs. S3, S20, and S21, A to C). More than 50% of FB2-FB7 subpopulations were from diseased hearts, particularly FB6 and FB7, suggesting their involvement in cardiac stress responses (Fig. 5, A and B, and fig. S21, A to C). Consistent with this notion, diseased-associated FBs differentially expressed genes involved in inflammation and fibrosis, such as transforming growth factor-β signaling, extracellular matrix, wound healing, and stress response genes (Fig. 5B and tables S19 and S20). Furthermore, FB6 and FB7 expressed *POSTN*, *FAP*, and *MEOX1* and *DMD* and *MYOCD*, respectively (*38*) (Fig. 5B), and corresponded to previously reported activated FBs and myo-FBs (fig. S3) (*10*, *17*). Notably, FB7 expressed ion channel genes including *KCNJ3* and *GJC1* (*CX45*), suggesting an electrically active FB, as recently reported (Fig. 5B) (*44*). Similar to CM subpopulations, cardiac FB subpopulations were identified expressing genes observed in both non-HF and HF FB states, suggesting intermediate disease states (Fig. 5B and table S22). Supporting these findings, these FBs (FB4) aligned with those displaying intermediate FB features (*10*, *17*) (figs. S3 and S20). Finally, scVelo RNA velocity uncovered two major cardiac FB trajectories leading toward either an activated FB (FB6) or myo-FB (FB7) state, with intermediate disease-associated FB subpopulations expressing genes related to each FB disease state (Fig. 5, A and B, and table S24), as described previously (*38*).

To illuminate the gene-regulatory programs driving cardiac FB transitions to disease states, we analyzed TF activity and chromatin accessibility across all cardiac FB subpopulations (Fig. 5C). We observed that the Wnt-related TF TCF7L2 and its associated chromatin accessibility were elevated in non-HF FBs but progressively declined along disease trajectories as FBs transitioned from intermediate to disease-associated FB subpopulations (Fig. 5C), as described previously (*45*). Conversely, we discovered TFs that increased along specific disease trajectories including EMT-related RUNX1, RUNX2, and FGF-related ETV1 in FB6-activated FBs and MEF2A/C, TEAD1, and PRDM6 in FB7 myo-FBs (Fig. 5C and fig. S21D). Notably, we discovered that FB4 specifically expressed many stress-response TFs such as FOS-, JUN-, and ATF-related TFs and may be a key transitory state as cardiac FBs become either activated FBs or myo-FBs (Fig. 5C).

Additional GRN and epigenomic analyses revealed how the identified TFs may control the TF network directing FB transition along distinct disease trajectories (Fig. 5D and fig. S21E). FOS- and JUN-related TFs enriched in FB4 were highly connected to many TFs in not only FB4 but also nearly every FB state, as determined by GRN centrality scores, suggesting that FB4 serves as a key intermediate in FB disease progression (Fig. 5, C and D). Accordingly, FB4 exhibited the largest GRN, containing FB4-related TFs that may direct its transition toward either FB6-activated FBs or FB7 myo-FBs (Fig. 5, C and D). Consistent with these findings, TFs linked to FB3, FB4, and FB5 were identified as regulators of RUNX1 and other TFs in FB6-activated FBs and FB7 myo-FBs including MEF2A/TEAD1 (Fig. 5D and fig. S21, F and G). Furthermore, SOX9, a key profibrotic regulator enriched in FB2 subpopulations, was linked to FB4-associated TFs, suggesting that FB2 represents an early prestressed state that precedes FB4, consistent with the FB trajectory predicted by scVelo (Fig. 5, A, C, and D).

Further analysis of FB6 and FB7 GRNs revealed distinct pathological programs regulating activated FBs and myo-FBs, respectively. Within the FB6 GRN, RUNX1 and RUNX2 regulated known activated FB-associated genes including *POSTN*, *MEOX1*, and *FAP* (Fig. 5, E and F) (*38*). Although *RUNX1* is expressed in FB6 (Fig. 5B, fig. S21D, and table S22), its transcriptional expression and activity in FB4 (Fig. 5C and fig. S21D) suggest a predisease role in translating early cardiac stress into broad activation of pathologic FB gene programs, as described previously (*46*, *47*). By contrast, *RUNX2* and *MEOX1*, the expressions of which are more restricted to FB6 (fig. S21D), may regulate gene expression specific to activated FBs as described previously (*38*, *47*). Supporting this notion, RUNX1 was regulated by both FB3-related TFs and FOS, JUN, and other stress-response TFs in FB4, whereas RUNX2, which was down-regulated in FB4, was controlled primarily by FB6-related TFs (fig. S21F). Notably, RUNX1 and RUNX2 were also observed to regulate each other in a possible feedforward mechanism to promote the activation of FB6 FBs (fig. S21F). Conversely, MEF2A and

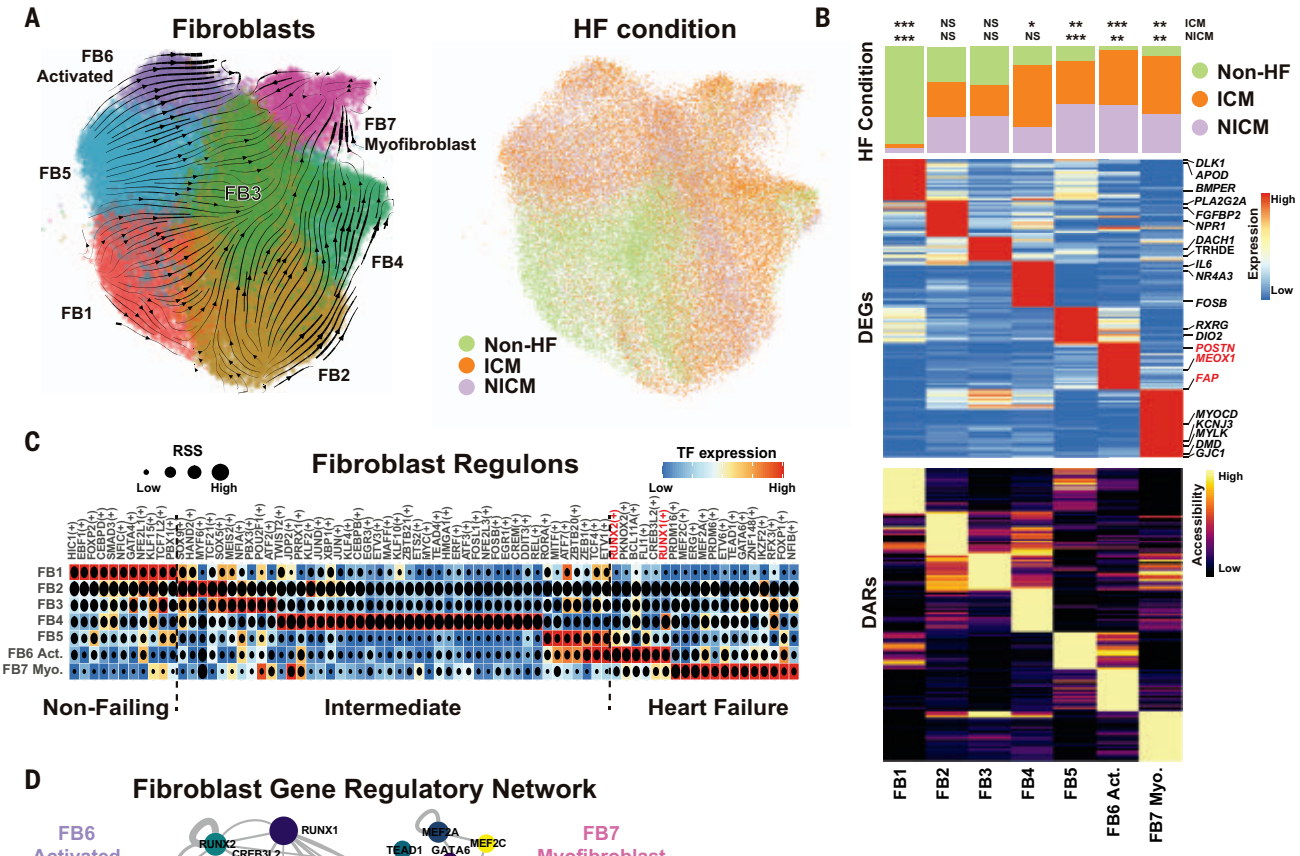

A
Fibroblasts
HF condition
FB6 Activated
FB7 Myofibroblast
FB5
FB3
FB4
FB1
FB2
Non-HF
ICM
NICM
B
HF Condition
ICM
NICM
Non-HF
DEGs
DLK1
APOD
BMPER
PLA2G2A
FGFBP2
NPR1
DACH1
TRHDE
IL6
NR4A3
FOSB
RXRG
DIO2
POSTN
MEOX1
FAP
MYOCD
KCNJ3
MYLK
DMD
GJC1
High
Low
Expression
DARs
Accessibility
FB1
FB2
FB3
FB4
FB5
FB6 Act.
FB7 Myo.
C
RSS
Low
High
Fibroblast Regulons
TF expression
Non-Failing
Intermediate
Heart Failure
D
Fibroblast Gene Regulatory Network
FB6 Activated
GSEA
-Collagen formation
-RUNX2 regulates osteoblast diff.
-ECM organization
-MET promotes cell motility
FB7 Myofibroblast
GSEA
-Smooth muscle contraction
-Regulation of actin cytoskeleton
-Myometrial relaxation and contraction pathways
-Focal adhesion
FB3
FB4
FB5
FB1
FB2
Expression Pseudotime
Early
Late
Degree Centrality
Low
High
E
IL1R1
MEOX1
POSTN
COL1A1
LTBP2
RUNX1
FAP
RUNX2
SERPINE1
F
FB6 Activated Fibroblast
RUNX1-regulated POSTN
RUNX1
RUNX1
FB1
FB2
FB3
FB4
FB5
FB6 Act.
FB7 Myo.
Genes
LINC01048
LINC00547
POSTN
Peaks
Links
Score
High
Low
37450000
37500000
37550000
37600000
chr13 position (bp)

Fig. 5. GRN analysis of cardiac FBs reveals genetic programs directing their cellular states. (**A**) UMAP showing FB cell subpopulations and their trajectories as detected by RNA velocity (left). UMAP labeled by HF condition reveals the contribution of non-HF and HF FB cells to each subpopulation (right). (**B**) Each cardiac FB cell subpopulation displays distinct cellular contributions (top, bar graph) from each heart condition as well as gene expression (middle, heatmap) and chromatin accessibility (bottom, heatmap). $*P_{adj} < 0.1$, $**P_{adj} < 0.05$, $***P_{adj} < 0.01$; NS, not significant. (**C**) Heatmap showing SCENIC+-predicted TF GRNs for each cardiac FB cell subpopulation. The box is colored by TF expression, and the dot size shows the RSS. TFs in red are associated with FB6-activated FBs. (**D**) Investigating the TF network within the cardiac FB GRN identified interactions between distinct GRNs across cardiac FB states. Nodes are colored by pseudotime of expression. Node size was determined by their degree of centrality in the network, and edges represent TF-TF connections. GSEA terms for the FB6-activated FBs and FB7-activaed myofibroblasts are shown. (**E**) Network plot of the RUNX1 and RUNX2 regulons showing their target genes including known gene markers of activated FBs (colored in magenta). (**F**) Representative genome browser tracks at *POSTN* locus showing aggregated chromatin accessibility profiles of each FB subpopulation and coaccessible peaks linked to *POSTN* (below, loops), including two SCENIC+-predicted peaks for the RUNX1 regulon.

TEAD1 may regulate known myo-FB-associated genes such as *MYH9*, *TNS1*, and *YAP1* in FB7 myo-FBs (*48*–*50*) (fig. S21, G and H). These myo-FB–related TFs were discovered to be regulated by MEF2C and NFIB in FB7 (fig. S21I). Thus, these findings reveal a hierarchy of GRNs that direct the transition of cardiac FBs into distinct pathologic FB states during HF.

Genetic effects on heart cell types and risk of HF

To discover how genetic variants may affect distinct cardiac cell types to influence the risk of HF, we integrated our cardiac cell type gene-regulatory programs with GWAS data across cardiovascular traits (table S25). To this end, we used linkage disequilibrium (LD) score regression to test for enrichment of cardiovascular trait-associated variants in cCREs (*51*), and identified significant (FDR < 0.1) enrichment in cell type–specific cCREs grouped by distinct cardiovascular traits (Fig. 6A and table S26). For instance, genetic variants associated with HF, dilated cardiomyopathy (DCM), cardiac arrhythmia, and cardiac trait measurements were strongly enriched in CM-specific cCREs (Fig. 6A), whereas traits related to hypertension, blood pressure, myocardial infarction (MI), and other cardiovascular diseases were associated with cCREs in noncardiomyocyte cell types including FBs and SM cells (Fig. 6A and table S26). Notably, NICM HF-associated (HF and DCM) variants were enriched in cardiomyocyte cCREs, whereas ICM HF (HF and MI) variants were enriched in FB and SM cCREs, suggesting distinct pathologic mechanisms between these two HF conditions (Fig. 6A, fig. S22A, and table S26). To further define these associations, we analyzed chromatin states of cell type cCREs harboring genetic variants. Trait enrichment was largely cell type specific in active cCREs but limited in open states, suggesting that genetic trait associations are primarily driven by active cCRE states (Fig. 6B and table S27).

We then used cardiac cell type cCREs to identify candidate causal genetic variants associated with different cardiovascular traits using fine-mapping data (Fig. 6B, fig. S22A, and table S27). Based on a recent GWAS study identifying numerous DCM risk loci (*7*), we focused on HF due to DCM. We found that 1293 fine-mapped DCM genetic variants representing 77.5% (62/80) of known DCM loci resided within a cardiac cell type cCRE. Notably, genetic variants at 94% (58/62) of these 62 loci overlapped with CM-specific cCREs, indicating strong enrichment in CMs (Fig. 6, C and D, and fig. S22, B to D). Clustering these 62 DCM loci by the cumulative probability of fine-mapped variant-cCRE overlap across cell types revealed predominant CM associations with smaller subsets linked with FBs and other cell types (Fig. 6C and table S28). Integrating ABC links, SCENIC+, and chromatin loop data, we identified putative CM target genes for 33 DCM loci. Among these loci, 22 loci had predicted target genes differing from previous studies (*7*) (table S29), including distal risk-gene connections involving *GATA4* and *MYH7B* (Fig. 6D and table S29).

To assess how HF genetic risk influences GRNs controlling CM states, we overlaid fine-mapped DCM variants onto CM GRN maps. These variants preferentially intersected GRNs regulated by specific TFs, suggesting convergence of pathways altered in cardiomyocytes during HF (Fig. 6, E and F). HF risk variants showed the strongest enrichment in GRNs of specific vCM states, particularly the intermediate vCM5 state regulated by stress-response related TFs (JUN, FOS, and ATF3; Fig. 6, E and F, and table S30). These GRNs were linked to stress-response, inflammatory, and disease pathways, suggesting that risk variants may modulate cardiac stress responses (fig. S22E and table S31). To identify CM subpopulations and states contributing to HF risk, we tested enrichment of genetic risk variants in GRNs active in vCM subpopulations. DCM-associated variants were significantly enriched in vCM5 GRNs (log enrich = 3.7, *fgwas*), with no enrichment detected in other states (all log enrich < 0), underscoring the key role of cardiac stress-responsive networks in HF (table S32).

Finally, we explored how fine-mapped genetic variants may affect cCREs and target genes in cardiac cell types, focusing on cardiomyocyte-specific variants given their strong enrichment for HF risk. We prioritized candidate variants in CM-specific cCREs predicted by ChromBPNet to affect chromatin accessibility (Fig. 6G and table S33), and identified target genes using ABC links, SCENIC+, and chromatin loops. Our analysis confirmed previously reported HF variant gene associations, including rs17099139-*BAG3* (10q26.11) and rs754920-*FLNC* (7q32.1), and delineated the CM cCREs potentially mediating these effects (Fig. 6G and fig. S22, F and G). We also identified additional target genes of HF variants not reported in prior studies (*7*). For example, candidate HF variant rs13277721 (8q24) resides within an active distal cCRE predicted to regulate *TRAPPC9* through long-range chromatin looping in CM Hi-C data (Fig. 6H) and is predicted to reduce chromatin accessibility by disrupting an AP-2α motif (Fig. 6H). Overall, integrating cardiac cell type–regulatory programs with human GWAS data revealed cell types, gene networks, cCREs, and target genes of HF and cardiovascular disease risk that may provide potential therapeutic targets for disease.

Discussion

Our multiomic single-cell sequencing studies elucidate how coordinated transcriptomic and epigenomic landscapes define cell type–specific GRNs and chromatin architecture in non-HF and HF human hearts. Expanding on prior scRNA-seq studies (*9*, *10*, *17*, *18*, *52*, *53*), integration of Droplet Paired-Tag, Droplet Hi-C, and single-cell multiome data revealed additional insights into the gene-regulatory mechanisms across cardiac cell types, enabling cell type–specific investigation of chromatin states and enhancer-gene linkages and identifying dynamic, disease-associated distal enhancers regulating putative target genes through distinct TF combinations. Among the affected cell types, CMs and FBs displayed the most pronounced transcriptional, epigenomic, and chromatin architecture changes during HF, including disrupted chromatin compartmentalization in CMs, supporting previous reports of nuclear remodeling in cardiomyocytes during HF (*54*, *55*). These chromatin architecture changes, similar to those reported in cancer and developmental disorders (*56*, *57*), may reflect HF-driven alterations in nuclear organization and function.

Detailed analysis identified multiple distinct pathologic states consistent with prior HF scRNA-seq studies (*10*, *17*). Network analyses further uncovered key gene-regulatory programs and TFs driving the

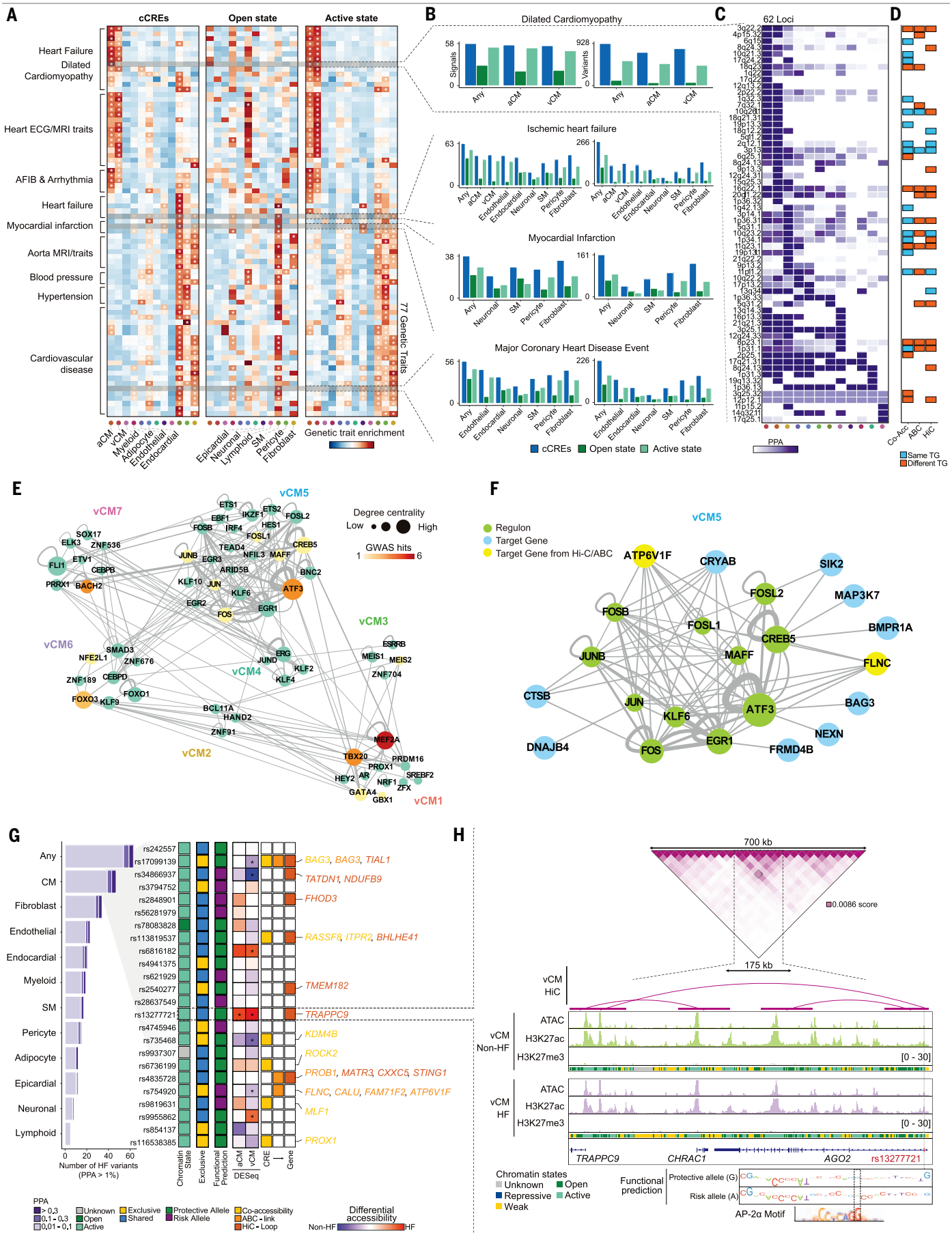

A
cCREs
Open state
Active state
Heart Failure
Dilated Cardiomyopathy
Heart ECG/MRI traits
AFIB & Arrhythmia
Heart failure
Myocardial infarction
Aorta MRI/traits
Blood pressure
Hypertension
Cardiovascular disease
77 Genetic Traits
aCM
vCM
Myeloid
Adipocyte
Endothelial
Endocardial
Epicardial
Neuronal
Lymphoid
SM
Pericyte
Fibroblast
Genetic trait enrichment
B
Dilated Cardiomyopathy
Signals
Variants
Any
aCM
vCM
Ischemic heart failure
Endothelial
Endocardial
Neuronal
SM
Pericyte
Fibroblast
Myocardial Infarction
Major Coronary Heart Disease Event
cCREs
Open state
Active state
C
62 Loci
PPA
D
Co-Acc
ABC
HiC
Same TG
Different TG
E
vCM5
vCM7
vCM3
vCM6
vCM4
vCM2
vCM1
Degree centrality
Low
High
GWAS hits
ETS1
ETS2
IKZF1
FOSL2
EBF1
HES1
FOSB
IRF4
FOSL1
SOX17
ELK3
ZNF536
TEAD4
CREB5
JUNB
EGR3
NFIL3
MAFF
ETV1
FLI1
CEBPB
ARID5B
BNC2
PRRX1
BACH2
KLF10
JUN
KLF6
ATF3
EGR2
FOS
EGR1
ESRRB
MEIS1
MEIS2
SMAD3
NFE2L1
ZNF676
ERG
JUND
KLF2
ZNF704
ZNF189
CEBPD
KLF4
FOXO1
FOXO3
KLF9
BCL11A
HAND2
ZNF91
MEF2A
TBX20
PRDM16
PROX1
HEY2
AR
NRF1
SREBF2
ZFX
GATA4
GBX1
F
Regulon
Target Gene
Target Gene from Hi-C/ABC
ATP6V1F
CRYAB
SIK2
FOSL2
MAP3K7
FOSB
FOSL1
BMPR1A
CREB5
JUNB
MAFF
FLNC
CTSB
JUN
ATF3
BAG3
KLF6
NEXN
DNAJB4
FOS
EGR1
FRMD4B
G
Any
CM
Fibroblast
Endothelial
Endocardial
Myeloid
SM
Pericyte
Adipocyte
Epicardial
Neuronal
Lymphoid
rs242557
rs17099139
rs34866937
rs3794752
rs2848901
rs56281979
rs78083828
rs113819537
rs6816182
rs4941375
rs621929
rs2540277
rs28637549
rs13277721
rs4745946
rs735468
rs9937307
rs6736199
rs4835728
rs754920
rs9819631
rs9955862
rs854137
rs116538385
BAG3, BAG3, TIAL1
TATDN1, NDUFB9
FHOD3
RASSF8, ITPR2, BHLHE41
TMEM182
TRAPPC9
KDM4B
ROCK2
PROB1, MATR3, CXXC5, STING1
FLNC, CALU, FAM71F2, ATP6V1F
MLF1
PROX1
Number of HF variants (PPA > 1%)
Chromatin State
Exclusive
Functional Prediction
DESeq
aCM
vCM
CRE
Gene
PPA
> 0.3
0.1 - 0.3
0.01 - 0.1
Unknown
Open
Active
Exclusive
Shared
Protective Allele
Risk Allele
Co-accessibility
ABC - link
HiC - Loop
Differential accessibility
Non-HF
HF
H
700 kb
0.0086 score
175 kb
vCM HiC
vCM Non-HF
vCM HF
ATAC
H3K27ac
H3K27me3
[0 - 30]
TRAPPC9
CHRAC1
AGO2
rs13277721
Chromatin states
Unknown
Open
Repressive
Active
Weak
Functional prediction
Protective allele (G)
Risk allele (A)
AP-2α Motif

Fig. 6. Fine-mapping of genetic loci associated with cardiovascular traits and HF. (**A**) Heatmap showing the LD score regression (LDSC) enrichment of trait-associated variants in all cCREs as well as cCREs in the open or active chromatin state across distinct cardiac cell types. Colors represent the level of enrichment. (**B**) Bar plots quantifying the number of fine-mapped variants and loci ("signals") overlapping all, open ,and active cCREs in LDSC-enriched cell types. (**C**) Heatmap representing the cumulative posterior probability association (PPA) values of fine-mapped variants associated with DCM across cardiac cell types. The cytogenetic band of the lead variant was used as the locus identifier. (**D**) Heatmap indicating whether the predicted target gene annotations coincide with the prioritized genes from Zheng *et al*. (*7*). SCENIC+-inferred links (Co-Acc), ABC links, or Hi-C loops were used to assign putative target genes. A target gene annotation was considered different if a method did not find a prioritized target gene from Zheng *et al*. (*7*). (**E**) Network plot showing the vCM subpopulation GRN overlap with cCREs harboring fine-mapped variants. Node size represents the centrality score from SCENIC+, and node color intensity indicates the number of variants overlapping GRN-associated cCREs (nodes in teal have no overlapping variants). (**F**) Network plot showing the GRNs and target genes overlapping GWAS variants for the vCM subpopulation containing a high number of GWAS variants (vCM5). (**G**) Stacked bar chart showing the overlap of fine-mapped DCM genetic variants with cCREs across cardiac cell types (left). Colors indicate PPA ranges: 0.01 to 0.1, 0.1 to 0.3, and >0.3. Heatmaps show the overlap between variants and chromatin state classification, exclusivity of variant localization within CM cCREs, ChromBPNet functional predictions, and a log2FC of differential chromatin accessibility between HF and non-HF aCMs and vCMs from DESeq2 (FDR < 0.1, right). Additionally, cCRE-to-gene links are visualized based on coaccessibility, ABC, or Hi-C methods. The colorimetric scheme applies to both boxplots and listed target genes, distinguishing each annotation method. (**H**) Hi-C contact map showing the interactions between a cCRE harboring the HF risk variant rs13277721 and the *TRAPPC9* promoter (top). Genome browser tracks show aggregated chromatin accessibility, histone modifications, and chromatin state heatmap for non-HF and HF vCMs (middle). Motif plot shows ChromBPNet functional predictions comparing the protective allele (G) and risk allele (A), highlighting disruption of an AP-2α motif by the HF risk allele (bottom).

transition of specific cell types from homeostatic to diseased states, including reactivation of fetal cardiac programs (*NPPA* and *NPPB*) that promote pathologic cardiac hypertrophy and remodeling but not reparative cardiomyocyte proliferation characteristic of fetal hearts (*58*), as well as FB activation programs (*RUNX1* and *RUNX2*) that direct FBs from quiescent to activated pathologic states. Although these multimodal datasets have advanced our understanding of gene-regulatory programs across major cardiac cell types during HF, their coverage for rarer populations remains limited due to lower cell numbers sequenced, as noted for other human heart single-cell RNA-seq studies (*9*, *10*, *17*, *53*). Additionally, integrating multiple modalities through shared RNA profiles may also limit the ability to resolve closely related or transitional cell states during cross-modality alignment. However, expanding these single-cell approaches to additional samples and conditions or applying newer multiomic approaches that concurrently profile RNA as well as chromatin interaction, accessibility, and modifications from the same cell (*59*) will refine the human cardiac cell epigenomic atlas and enable deeper interrogation of the gene regulation underlying cardiomyopathies.

Our study, which includes data from both ICM and NICM cardiomyopathy, establishes a gene-regulatory framework with which to guide therapeutic strategies that may reprogram specific cardiac cell types from pathologic to reparative states. Moreover, this cell epigenomic atlas facilitates systematic interpretation of noncoding variants by linking them to gene expression, enhancer activity, and disease susceptibility in a cell type–specific manner. Accordingly, our findings establish a publicly available resource (https://epigenome.wustl.edu/HeartEpigenome/) to help prioritize cardiovascular-disease associated variants (*5*–*8*) for further functional study through large-scale genomic screening assays (*60*–*61*). Overall, our study delineates the cellular heterogeneity and chromatin dynamics driving gene-regulatory programs underlying HF with reduced ejection fraction, but may also offer insights into other HF subtypes with shared mechanisms (*29*).

Materials and methods

Study population

Patients with end-stage cardiomyopathy who underwent heart transplantation at the institutions comprising the Utah Transplantation Affiliated Hospitals Cardiac Transplant Program (University of Utah Health Science Center, Intermountain Medical Center, and the Veterans Administration Salt Lake City Health Care System) were enrolled in this study after obtaining informed consent. Control tissue was acquired from non-HF subjects whose hearts were not allocated for transplantation for noncardiac reasons. Patients with acute forms of HF, defined by symptoms <3 months of duration with no evidence of LV dilation, were excluded. Subjects with hypertrophic or infiltrative cardiomyopathies were also excluded from this study. Clinical metadata was self-reported. The study was approved by the institutional review board of the participating institutions (IRB 00030622).

Chronic ICM cardiomyopathy was defined as a LVEF <40% and any of the following: (i) a history of MI or revascularization; (ii) a history of angina or chest pain and evidence of scarring in noninvasive imaging studies corresponding to previous MI; (iii) the presence of ≥75% stenosis of the left main or proximal left anterior descending artery; or (iv) the presence of ≥75% stenosis of ≥2 epicardial vessels in a patient with unexplained cardiomyopathy. Patients with a LVEF <40% and nonobstructive coronary artery disease without evidence of prior MI or revascularization were considered to have NICM cardiomyopathy, as described previously (*63*).

Myocardial tissue acquisition

Transmural cardiac samples were collected from all each chamber of the heart, which included LV, right ventricle (RV), left atrium (LA), and right atrium (RA), at the University of Utah and snap-frozen in liquid nitrogen at the time of heart transplantation, and stored at −80°C. A total of 13 ICM, 10 NICM cardiomyopathy, and 13 non-HF subject samples were used for this study.

Sample pooling

For each pool, tissues were collected from the same cardiac chamber across multiple subjects and then pooled into the same tube for downstream nuclei extraction, purification, and single-cell sequencing studies. Diseased and nondiseased samples from each chamber were pooled during processing to mitigate potential batch effects. For each sample, ~20 to 30 mg of snap-frozen tissue (tissue samples measuring ~3 mm × 3 mm × 3 mm) was sectioned on dry ice using prechilled scalpels. These sectioned samples were then collected together in a prechilled 5-ml LoBind tube on dry ice. These pooled, sectioned tissue samples were then stored in a −80°C freezer until processed.

Heart nuclei isolation

For each subject, nuclei were isolated from adjacent tissue samples within the same cardiac region or chamber to ensure sample comparability across all assays. Single-nuclei suspensions for single-sample and pooled-sample Multiome studies were prepared from frozen tissues using a gentleMACS Dissociator (Miltenyi Biotec) in magnetic-activated cell sorting buffer containing 5 mM $CaCl_2$ (R040, G-Biosciences), 3 mM Mg-acetate (M2545, Sigma), 10 mM Tris-HCl, pH 8.0 (15568025, Thermo Fisher Scientific), 2 mM EDTA (15575020, Thermo Fisher Scientific), 0.6 mM dithiothreitol (DTT) (D9779, Sigma), 1× Roche cOMPLETE Protease Inhibitor EDTA-Free (11873580001, Sigma), and 1.5 U/μl RNasin Ribonuclease Inhibitor (N2515, Promega). The nuclei suspension was then filtered through 100- and 30-μm CellTrics Filters to remove debris and centrifuged for 5 min at 500*g* at 4°C. The nuclei pellet was resuspended

in iodixanol buffer (50 and 25% iodixanol; OptiPrep Density Gradient Medium, D1556-250ML, Sigma), dilutent buffer (120 mM Tris-HCl, pH 8; 15568025, Thermo Fisher Scientific), 150 mM KCl (AM9640G, Invitrogen), 30 mM $MgCl_2$ (194698, Mp Biomedicals), and molecular biology water (46000-CM, Corning) and centrifuged for 30 min at 4000*g* at 4°C. After resuspension in sorting buffer, nuclei were stained with 2 μM 7-amino-actinomycin D (7-AAD) in sorting buffer [1× phosphate-buffered saline (PBS), 1× protease inhibitor, 1.5 U μl RNasin ribonuclease inhibitor, and 1% bovine serum albumin (BSA)] for 10 min on ice and sorted using an SH800 cell sorter (Sony) for single nuclei. Nuclei were collected in collection buffer (1× PBS, 5× protease inhibitor, 7.5 U/μl RNasin ribonuclease inhibitor, and 5% BSA) and centrifuged for 10 min at 500*g* at 4°C.

For pooled-sample Droplet Paired-Tag, single-nucleus suspensions were also prepared using the gentleMACS Dissociator (Miltenyi Biotec) in MACS buffer as with multiome but with 1 U/μl RNaseOUT and 1 U/μl SUPERaseIn inhibitor. After filtering using 100- and 30-μm CellTrics Filters and centrifugation, Anti-Nucleus MicroBeads (Miltenyi Biotec) were used for debris removal and enrichment of nuclei. Nuclei suspension was centrifuged for 5 min at 500*g* at 4°C. After resuspension in sorting buffer, nuclei were stained with 0.2 mg/ml Hoescht in sorting buffer (1× PBS, 1× protease inhibitor, 1 U/μl RNaseOUT, 1 U/μl SUPERaseIn inhibitor, and 1% BSA) for 10 min on ice and sorted by fluorescence-activated nuclei sorting with an SH800 cell sorter for single nuclei. Nuclei were collected in collection buffer (1× PBS, 5× protease inhibitor, 5 U/μl RNaseOUT, 5 U/μl SUPERaseIn inhibitor, and 5% BSA). For pooled Droplet Hi-C, nuclei were fixed immediately after dissociation using the gentleMACS Dissociator in MACS buffer, as described above.

Genotyping arrays and imputation

Genomic DNA was extracted from 10- to 15-mg sections of frozen tissue obtained from the LV of the 30 subjects. DNA isolation was performed using the Monarch Genomic DNA Purification Kit (NEB, T3010L), with elution in 70 to 100 μl of UltraPure Distilled Water (Invitrogen, 10977-015). When required, DNA concentrations were adjusted to 50 ng/μl using a speed vacuum concentrator.

Genotyping was conducted at the University of California San Diego (UCSD) IGM Genomics Center using the Illumina HumanCoreExome-24v1 array. Genotype calling, cluster positioning, and extraction of single-nucleotide polymorphism (SNP) statistics were performed with GenomeStudio v2.0.5 using the PLINK input report plugin with an hg38 cluster file for reference. The resulting genotypes were converted into PLINK format within GenomeStudio. The germline variants from the genotyping showed a median call rate of 0.99, as well as a median concordance with WGS data of >99% (table S3). For downstream analysis, genotyping data from different arrays were merged using PLINK v1.9. Quality control filtering was applied as follows: variant missing call rate ≤ 0.05, minor allele frequency (MAF) ≥ 0.01, and Hardy-Weinberg equilibrium P value > 1×10^{-5}. Before imputation, variant positions and annotations were standardized using the HRC-1000G-check-bim.pl script (v4.3.0) from the McCarthy Group Tools collection (https://www.chg.ox.ac.uk/~wrayner/tools/).

10x Genomics Epi Multiome ATAC and gene expression assays

The nuclei pellet was resuspended in permeabilization buffer (1 mM DTT, 0.2% IGEPAL-CA630, 1× cOmplete EDTA-free protease inhibitor, 1.5 U/μl RNasin, and 5% BSA in PBS), incubated on ice for 2 min, centrifuged for 5 min at 500*g* at 4°C, and resuspended in 1× nuclei buffer (20× nuclei buffer PN 2000207, 10x Genomics), 1 mM DTT (D9779, Sigma), and molecular biology water (46000-CM, Corning). Suspension of 30,000 permeabilized nuclei were each loaded onto eight lanes and processed using a 10x Genomics Controller following the manufacturer's recommendations. 10x Genomics Epi Multiome ATAC + Gene Expression assays were processed following the manufacturer's recommendations. Generated libraries were sequenced at the Institute for Genomic Medicine at UCSD using the NovaSeq X Plus (Illumina) sequencer to the depth of ~25.6 billion reads for ATAC-seq and ~26.5 billion reads for RNA-seq.

Droplet Paired-Tag data generation

To obtain joint profiles of histone modifications and transcriptomes from single cells in heart samples, Droplet Paired-Tag was performed as described previously (*13*), with minor modifications. In brief, we pooled cardiac chamber tissues from 30 different subjects in equal weights and extracted nuclei using the same protocol as our 10x Genomics single-cell multiome data generation. The nuclei were sorted by fluorescence-activated nuclei sorting using an SH800 cell sorter (Sony) to isolate single nuclei. The isolated nuclei were collected and counted using a cell counter (RWD C100-Pro) with 4′,6-diamidino-2-phenylindole (DAPI) staining. We aliquoted 500,000 nuclei to initiate reactions in individual tubes for each histone modification mark and replicate. Two replicates were included for each tissue pool and histone modification.

First, nuclei were permeabilized with OMNI buffer and incubated with 2 μl of PA-Tn5 (0.4 mg/ml) and 2 μg of antibody in MED#1 buffer overnight at 4°C (*64*). PA-Tn5 and antibodies were then removed by washing with MED#2 buffer. Tagmentation was performed in MED#2 buffer supplemented with 10 mM $MgCl_2$ (Invitrogen, AM9530G) at 550 rpm and 37°C for 60 min in a ThermoMixer (Eppendorf), and the reaction was terminated by adding 2× stop solution. The nuclei were washed in 1× nuclei buffer and counted again. From each tube, 24,000 to 30,000 nuclei were aliquoted and used for a single droplet generation reaction with the Chromium Next GEM Single-cell Multiome kit (10x Genomics, 1000283). Finally, DNA and RNA library amplification was performed according to the Chromium Single-cell ATAC Library kit manual, with the following modifications: the starting amount of preamplification product for DNA library amplification was doubled, 13 amplification cycles were used for the DNA library, and a different SPRI bead size selection were applied for the DNA library (50 μl + 105 μl). For this study, we used antibodies targeting two histone modification marks: H3K27ac (Abcam, ab4729, polyclonal) and H3K27me3 (Abcam, ab192985, recombinant).

Droplet Hi-C data generation

To obtain cell type–specific chromatin architecture, we performed Droplet Hi-C on all heart samples from different subjects, as described previously (*14*). In brief, heart samples from 30 subjects were pooled together, and nuclei were extracted within the same reaction. Nuclei were fixed with paraformaldehyde (Electron Microscopy Sciences, 15714) at a final concentration of 2%, and then quenched with glycine solution at a final concentration of 0.4 M. Nuclei were permeabilized with lysis buffer on ice for 30 min. Then, the nuclei were treated with 0.5% SDS (Promega, V6551), incubated at 62°C for 10 min on a ThermoMixer and then quenched with Triton X-100 (Sigma, 93443) at 550 rpm and 37°C for 15 min. Chromatin in the nuclei was sheared using three restriction enzymes: DpnII (NEB, R0543L), MboI (NEB, R0147M), and NlaIII (NEB, R0125L), at 37°C for 90 min on a ThermoMixer at 550 rpm The three enzymes were then deactivated at 65°C for 20 min, and the mixture was cooled to room temperature. The sheared chromatin was further ligated using T4 DNA ligase (NEB, M0202L). All enzymes were removed from the nuclei, which were then suspended in 1% BSA in 1× PBS with 7-AAD (Invitrogen, A1310) and incubated on ice for 10 min. The nuclei were sorted into collection buffer (5% BSA in 1× PBS) by fluorescence-activated nuclei sorting using an SH800 cell sorter (Sony) to remove all the debris.

The sorted nuclei were collected by centrifugation and washed twice with 1× nuclei buffer. Approximately 25,000 nuclei were aliquoted into each polymerase chain reaction (PCR) tube, and one reaction was performed using the Chromium Next GEM Single-cell ATAC Reagent

Kits v2 (10x Genomics). The tagmentation incubation time was 60 min, and the index PCR elongation time was extended from 20 s to 1 min. Double-sided size selection was adjusted to 1.14x SPRIselect to remove only small fragments. For each cardiac chamber pool, we performed two or three independent reactions.

Single-cell demultiplexing using genotype information

Multiomics, DPT, Droplet Hi-C data were demultiplexed using demuxlet (v2) (*65*) (https://github.com/statgen/popscle). Genotype VCFs for demultiplexing were generated from TOPMed R3 imputed data and filtered for imputation quality ($R^2 > 0.9$) and MAF (>1%) using bcftools. The variants were then subsequently intersected with accessible chromatin peaks identified in the human Cis-element Atlas (CATLAS) and then randomly downsampled to ~1.2 million SNPs. The final panel of variants used for demuxlet consisted of 1,191,480 variants chosen at random, of which 39,044 (3.3%) were directly genotyped and 1,152,436 (96.7%) were imputed. For Droplet Paired Tag data, the filtered VCF and the DNA BAM file (from 10x Multiome or Droplet Paired-Tag, processed with CellRanger) were provided as input to demuxlet. For Droplet Hi-C, cell barcodes were appended to BAM records as "CB" tags to match 10x formatting, and BAM files were sorted before demultiplexing using the same filtered VCF.

Single-cell multiome clustering and annotation

Single-cell multiome samples were processed using Cellranger (v6.0.1) with the reference genome hg38 using the same workflow for both multiplexing and nonmultiplexing libraries. Individual libraries were processed and analyzed in Seurat (v4.3) for quality control (*66*). Cells with >500 expressed genes, >1000 ATAC fragments, ≤5% mitochondrial genes, and ≥2 transcription start site enrichment were kept. Ambient RNA contamination was accounted for by running SoupX (v1.6.2) (*67*) on each sequencing library (for both multiplexing and nonmultiplexing libraries) using its automated algorithm to estimate contamination rates. Gene expression count matrices were adjusted for predicted contamination and rounded to integers [adjustCounts(sc, roundToInt=TRUE)]; these SoupX adjusted counts were used for both clustering and downstream analysis. Then, all libraries were merged; mitochondrial genes were removed; and a first round clustering on the WNN space was performed using 5-kb windows. The data was then batch corrected using harmony (v1.2.0) (*68*) by library, and major cell types and cell compartments were then annotated. We then further cleaned the data by calculating Silhouette scores between any combinations of cellular compartments and removed all cells that did not pass a threshold of −0.5 (1953 cells) and by performing one round of manual cleaning to reach a final number of cells of 329,255 cells. After cleaning, the data were reclustered and reharmonized by library and sex from raw data, and peaks were called by major cell types using Signac (v1.12.0) (*69*) with default parameters. We first limited peak size for all called peaks to 300 bp by centering any peaks larger than 300 bp at their summit and extending coordinates 150 bp in either direction. We then grouped peaks based on overlap to create clusters of peaks using bedops v2.4.41 (*70*). Within each cluster, the peak with the highest read count at its summit was identified as the reference peak for the region. We then generated a list of peaks that did not overlap any of the reference peaks and began the process of clustering and identifying reference peaks again until no peaks remained. Finally, peaks were filtered using quantile thresholding (Q2) with both pileup and variability scores. For the CAREHF dataset, Single-cell Multiomics libraries were processed as described above. Doublets were detected and removed from each library using Scrublet 0.2.3 (*71*) with an automatic threshold, considering an expected doublet rate of 6% for all libraries.

Defining heterogeneity within cell types

For each of the 12 major cell types, further filtering and clustering analyses of the multiome cells were performed using the Seurat package. Nearest neighbor graphs were calculated using the "FindNeighbors" function with 1:50 components for the RNA assay. The generated nearest neighbor graph was then used for graph-based, semiunsupervised clustering (function "FindClusters") and UMAP to project the cells into two dimensions (function "RunUMAP"). Standard criteria were used to call clusters, including a requirement that clusters differentially express several unique genes. Cell identities were assigned to the clusters by cross-referencing their marker genes with known cardiac cell type markers from both human and mouse studies, in addition to in situ hybridization data from the literature. On occasion, a cell cluster would emerge that expressed marker genes representing multiple populations, as well as containing cells with low unique molecular identifier (UMI) and gene counts that escaped the first filtering step. These cells were removed from downstream analyses.

Label transfer with public reference datasets

To assess the accuracy of our cell subtype annotations, we performed semisupervised label transfer using existing human cardiac single-nucleus RNA-seq (snRNA-seq datasets) (*10*, *17*) with single-cell ANnotation using Variational Inference (scANVI) (*72*). For each major cell type, we concatenated our dataset with the corresponding reference cells and first fitted a standard scVI model to learn a shared low-dimensional representation across datasets. The joint cohort identifier, cardiac chamber, and dataset identifier were included as categorical batch covariates, and per-nucleus RNA counts were included as a continuous covariate. We initialized a variational autoencoder (`scvi.model.SCVI`) with three fully connected hidden layers (128 units each), a 50-dimensional latent space, a negative binomial likelihood for gene expression counts, and gene-by-batch dispersion parameters. The model was trained for up to 500 epochs with early stopping.

For semisupervised annotation, we initialized a scANVI model from the pretrained scVI model using "scvi.model.SCANVI.from_scvi_model." The scANVI model was trained for up to 500 epochs with early stopping and "n_samples_per_label = 500" to balance labeled subtypes during training. After training, we computed a scANVI latent representation for all cells and obtained discrete cell-state predictions for every query cell. In addition, scANVI soft assignment probabilities were exported for downstream quality control and thresholding. The final scANVI-derived labels were used for all subsequent comparative analyses.

Preprocessing of Droplet Paired-Tag data

Fastq files from Droplet Paired-Tag were demultiplexed using the "mkfastq" command in cellranger-arc (v2.0.0). After demultiplexing, histone modification and transcriptome modalities were aligned to reference genome (hg38) separately using the "count" command in cellranger-atac (v2.0.0) and cellranger (v6.1.2), respectively. To select high-quality nuclei, we adopted a three-step filtering strategy: First, histone modification data were aggregated at the sample level, and peak calling were performed to identify narrow peaks (for H3K27ac) or broad peaks (for H3K27me3) using MACS2 (v2.1.2) (*73*). Per-cell fragment number and the Fraction of Reads in Peaks (FRiP) were used to filter and retain barcodes with high-quality histone modification profiles. For RNA, barcodes were filtered based on UMI counts per cell. Next, nuclei were selected by pairing the pass-filtered barcodes from both modalities. Finally, based on subject genotype phasing results, only pass-filtered nuclei confidently assigned to a single subject were retained for downstream analysis. Ambient RNA contamination was estimated and removed using SoupX (*67*) for each library separately before clustering.

Genome track visualization

Cell type– or condition-specific alignments for histone modification data were extracted from cellranger bam files using custom scripts. BigWig files were generated with the "bamCoverage" function in

deepTools (v3.5.1) (*74*) using the normalization method "RPGC," with chrY excluded for normalization.

For chromatin accessibility data, valid alignments were extracted, split into single-end reads, and corrected for Tn5 insertions. Reads aligned to the positive strand were shifted +4 bp, and reads aligned to the negative strand were shifted −5 bp. Each alignment was further shifted −100 bp (for the positive strand) or +100 bp (for the negative strand) and then extended to a length of 200 bp to capture the Tn5 insertion points. BigWig files were generated using the "bamCoverage" function with the same normalization method and exclusion criteria as described above for the histone modification data.

Integration of Droplet Paired-Tag data with 10x Multiome

We used the RNA modality from both multiomic assays (10x Multiome and Droplet Paired-Tag) as anchors to perform integration and annotate cell type identities using Seurat (v5.1.0) (*75*). In brief, datasets were first normalized using SCTransform (*76*). The top 3000 integration features shared between the datasets were selected and used to identify integration anchors using canonical correlation analysis with the "FindIntegrationAnchors" function. The joint integrated embedding was calculated using the "IntegrateData" function, followed by principal components analysis (PCA). The first 30 principal components were then used for UMAP visualization and Leiden clustering. To ensure that the joint embedding was not biased by sex bias in the cohorts, we calculated local inverse Simpson's index as described previously (*68*) using sex as the category label for each subclass.

After integration, cell identity for the Droplet Paired-Tag RNA modality was assigned using a custom k-nearest neighbor label transfer procedure analogous to Seurat's TransferData function. At the major cell-type level, we used the 10x Multiome RNA dataset as the reference and for each query cell in the Droplet Paired-Tag dataset, we identified its 25 nearest neighbors in the reference space using the RANN package (v2.6.2; 10.32614/CRAN.package.RANN). Neighbor distances were scaled and standardized and then converted into weights. For each query cell, prediction scores for major cell type labels were obtained by multiplying a binary indicator matrix of reference cell labels by the corresponding neighbor weights. The predicted identity for each query cell was assigned as the label with the highest prediction score.

For subpopulation-level annotation, we applied the same procedure but restricted the neighborhood size to the five nearest reference neighbors. To ensure high-confidence assignments, we retained only query cells with a maximum prediction score ≥0.9 (98.7%) for analyses at the major cell type level and ≥0.8 (65.7%) for analyses at the subtype level; cells below these thresholds were excluded from downstream analyses.

Preprocessing of Droplet Hi-C data

Droplet Hi-C data were processed as described previously (*14*). Fastq files were demultiplexed using the "mkfastq" command in cellranger-atac (v2.0.0). After demultiplexing, the cellular barcode sequences were extracted, aligned to the white list using Bowtie (v1.3.0), and appended to the start of each read name to record the cellular identity (*77*). Potential sequencing adapters were trimmed using Trim-Galore (v0.6.10), and the reads were mapped to the reference genome (hg38) using BWA-MEM (v0.7.17) with the arguments "-SP5M" specified (*78*) (https://doi.org/10.5281/zenodo.7598955). Contact pairs were parsed, sorted, and de-duplicated using Pairtools (v0.3.0) (*79*), with barcode information stored as separate columns in the pairs file.

Before downstream analysis, high-quality nuclei were selected based on the number of per-cell unique contacts in each library and then processed for subject phasing using genotype information. Nuclei confidently assigned to a single subject were retained and split into individual pairs files, with each file containing all contact information from a single nucleus. Imputation was performed using scHiCluster (v1.3.5) (*80*) for individual cells at three different resolutions: 100 kb (for compartment analysis), 25 kb (for domain analysis), and 10 kb (for cell embedding and loop analysis).

Reference mapping of Droplet Hi-C data

We adopted a previously described method to reference map and annotate Droplet Hi-C profiles with gene expression data generated from the same tissue and samples (*81*). In brief, the single-cell gene associating domain (scGAD) score R_{ij} was calculated as the total contact number across the gene body for gene i in cell j at 10-kb resolution. The chromatin interaction profiles can then be represented as a cell-by-gene scGAD score matrix, which has been demonstrated to be correlated with gene activity. This matrix will serve as the input for reference mapping with snRNA-seq data.

To perform reference mapping, the RNA modalities from standard 10x Multiome and Droplet Paired-Tag were integrated and used. The integrated features from the reference dataset were selected and subjected to dimension reduction by PCA. The derived PCA model was then applied to transform the scGAD score matrix with the same features normalized and scaled. Subsequent integration of scGAD score with reference dataset was conducted using canonical correlation analysis. The integrated embedding was then transformed and visualized with the pre-calculated UMAP embedding model from the reference dataset.

To annotate cell identities within using scGAD score, we identified the 15 nearest neighbors in reference RNA data for each Droplet Hi-C cell using PyNNDescent (v0.5.6) (*82*) using Euclidean distance. These distances were scaled and standardized, and neighbors with the shortest distance would be assigned the highest score. The identity of each Droplet Hi-C cell was inferred by nominating the cell type label that garnered the highest standardized score among its top nearest neighbors.

Differential analysis

Differential expression analyses at both major cell type level and subpopulation level were performed using DESeq2 (v1.42.0) on subject-level, aggregated pseudobulk count matrices derived from RNA, ATAC, and histone modification data (*28*). To account for potential confounding factors in the experimental designs, sex and experimental batch were included as covariates in the DESeq2 design formula. For each dataset, features were filtered to retain only those with a minimum of 10 total counts across all samples, ensuring sufficient statistical power and reducing noise from low-abundance features. DESeq2's default normalization and shrinkage methods were applied to control for variance.

For subpopulation-level analyses, the ashr shrinkage method was used to mitigate inflation of $\log_2$-fold change (log2FC) estimates associated with increased data sparsity (*83*). Statistical significance was assessed using the Benjamini-Hochberg FDR procedure to correct for multiple hypothesis testing. At the major cell type level, features with an absolute log2FC >0.58 were considered differentially regulated unless noted. At the subpopulation level, differential features were defined as those with FDR <0.05 and a log2FC >0.58 for genes, or >0 for ATAC-derived cCREs unless noted, to accommodate the increased sparsity inherent to subtype-resolved analyses.

Cell type composition analysis

To test for differences in cell type or cell subtype composition between different conditions, we performed a subject-level proportion analysis. For every cohort, we counted the number of cells assigned to each cell type or cell subtype population and computed the total number of profiled cells for each subject from the 10x Multiome dataset. Cell type counts were then converted to subject-level proportions by dividing the per–cell type counts by the subject's total cell count and scaling these proportions to a common normalizer to account for differences in sequencing depth and sampling effort across subjects. Next, for each

cell type present in a given contrast, we extracted the scaled proportions for diseased (HF or ICM/NICM) and control cohorts, and excluded cell types with fewer than two subjects in either group. We then performed a Wilcoxon rank test to compare subject-level proportions between diseased and control group for each cell type within each contrast. *P* values were adjusted for multiple testing across cell types tested within each contrast using the Benjamini-Hochberg method procedure. We additionally computed the difference in mean proportions as an effect-size measure and annotated significance levels using FDR thresholds of 0.05 (*), 0.005 (**), or 0.0005 (***). All contrast-specific results were used for downstream interpretation and visualization of cell type composition changes.

Predicting enhancer-gene pairs with the ABC model

We used the ABC model (*19*) to infer putative enhancer-gene links for each condition. The ABC model took imputed contact matrices at 10-kb resolution from Hi-C data, along with H3K27ac signal over putative enhancers (in our case, consensus peaks identified in the corresponding cell type from ATAC modalities). Predictions with an ABC score ≥0.02 were considered positive and retained for downstream analysis.

Motif enrichment and GSEA analysis on ABC links

cCREs for each set of cell type–specific cCRE or ABC links were analyzed using HOMER (v4.11.1) with function "findMotifsGenome.pl" (*27*), with peaks from all ABC link candidates serving as the background. Additionally, up to 500 coding target genes from each set of cell type–specific ABC links were analyzed using GSEA (MSigDB). The analysis was conducted using a FDR q-value threshold of <0.05, with gene sets ranging from 10 to 500 genes, and enrichment was evaluated against all pathways within the canonical pathways category. GO analysis of genes located near cell type–specific cCREs was performed using HOMER's annotatePeaks.pl script (*27*) using the "-go" flag.

Annotation of genome using ChromHMM

We used ChromHMM (*23*) to integrate and summarize all three epigenetic modalities (ATAC, H3K27ac, and H3K27me3) into a unified chromatin state annotation. In brief, we first split the cellranger arc generated DNA fragment files by cell type and disease condition and then converted them to BED format. These BED files were binarized on a fixed genomic grid using the ChromHMM "BinarizeBed" function with the "-center" flag enabled. We then trained chromatin state models across all conditions using ChromHMM's "LearnModel" with default parameters.

To determine the optimal number of chromatin states, we followed previously described strategies (*24*). First, we trained a series of candidate models with different state numbers (2 to 8) and applied k-means clustering to the emission probability profiles from all models. For each candidate state number, we quantified cluster separation using the within-cluster sum of squares and selected the smallest number of states at which the separation reached at least 95% of the maximum across all models. As a complementary validation, we used ChromHMM's "CompareModels" function to compare an 8-state "full" model to each simpler model by computing, for each state in the full model, its maximum Pearson correlation with states in the simpler models. We summarized these comparisons by the median correlation across all states in the full model and identified the state number at which this median correlation plateaued. Based on the convergence of these two criteria and the limited number of profiled marks, we adopted a five-state ChromHMM model to balance biological interpretability and model complexity.

To interpret the biological identity of each chromatin state, we applied ChromHMM's "OverlapEnrichment" function using genomic features such as RefSeq genes and transcription start sites. States with strong enrichment at gene promoters and with high ATAC or H3K27ac signal were interpreted as active regulatory states, whereas states with low enrichment of epigenetic marks were considered "weak" or "unknown." Specifically, the "weak" state shows higher enrichment at gene promoters and gene body, indicating its different identity compared with the state "unknown." Final state labels were assigned by manual inspection of emission probabilities and enrichment profiles, guided by prior knowledge of epigenetic mark combinations. We note that, because we did not include additional histone marks commonly used in large chromatin state maps, the state definitions here may not be directly comparable to those from studies based on more extensive mark panels.

Single-cell motif analysis with chromVAR

Motif accessibility z-scores were calculated from snATAC-seq data using chromVAR (v1.22.1) (*84*). The fixed peak sparse count matrix was converted to a SummarizedExperiment object, and GC content bias was estimated using chromVAR's internal method. Human TF motifs were obtained from the JASPAR2022 Bioconductor package and mapped to peaks using motifmatchr (v1.22.0). The motif annotations and SummarizedExperiment were then used as input for chromVAR's computeDeviations function to generate GC bias-corrected motif accessibility z-scores.

Prediction of variants effect on chromatin accessibility signal with ChromBPNet

We trained a ChromBPNet (v0.1.7) model using the default preprocessing parameters outlined in the ChromBPNet tutorial (*85*). To account for Tn5 bias, we trained a separate model using our own 10x Multiome pseudobulk data. Two bias-factorized models were trained on human heart vCMs and aCMs from non-HF individuals.

For training, we used MACS2 narrowPeak regions with a *P* value threshold of <0.05, called on pseudobulk ATAC fragments from all ventricular cardiomyocytes. Chromosomes 1, 3, and 6 were assigned for training, and chromosomes 8 and 20 were reserved for validation. The model was trained using default parameters. ChromBPNet models have two output heads: the profile head, which predicts the shape of the chromatin accessibility profile, and the counts head, which estimates the total read counts within a profile.

For analysis, we used chrombpnet_nobias.h5 models and computed sequence contribution scores for one or both output heads. Additionally, we applied ChromBPNet for variant prediction to assess the impact of genetic variants on chromatin accessibility within our dataset. To plot ChromBPNet results, contribution scores and predicted counts for reference and alternate alleles were generated using the chromBPNet_nobias model. Alternate alleles were introduced into the reference genome using the consensus command from bcftools (v1.10.2) (*86*) to create allele-specific genome sequences. The contribs_bw and pred_bw commands from chromBPNet were then applied separately to the reference and alternate genome versions to compute contribution scores and predicted signal tracks.

Identification of motifs affected by variants with motifbreakR

To identify motifs disrupted by alleles of interest, we ran motifbreakR (v2.8.0) using human TFs from JASPAR2022 and HOCOMOCO_v11 accessed using MotifDb (v1.36.0) (*87*).

GSEA analysis

GSEA (*88*) was performed using fgsea (v1.26.0) on differentially expressed genes identified from DESeq2 results (*89*). For each dataset, pseudobulk differential expression results were processed by first removing ribosomal genes and ranking genes based on their statistical score (stat column), ensuring that nonmapped genes were excluded. The ranked gene lists were sorted in descending order and used as input for fgsea, which was conducted against a curated gene set database, considering pathways containing between 10 and 500 genes. Significantly enriched pathways were identified based on FDR <0.1. To

remove redundancy, a hierarchical pathway reduction approach was applied, generating a final list of nonredundant enriched pathways ranked by normalized enrichment score (NES).

Subpopulation RNA velocity analysis

RNA velocity analysis was conducted using scVelo (v0.2.5) (*90*). Spliced and unspliced read counts were computed with velocyto (v0.17.17) from raw sequencing data using the gex-GRCh38-2020-A reference genome. The spliced and unspliced count matrices were filtered to exclude cells with <20 shared reads, retaining only the top 20,000 variable genes. The data were normalized, and moments were computed using "scvelo.pp.moments," with the top 30 PCs and 50 nearest neighbors. Dynamics were recovered using "scv.tl.recover_dynamics." To refine the dynamics recovery, marker genes representing HF-related gene expression changes were selected. Finally, RNA velocities were calculated using scvelo.tl.velocity, and velocity graphs were constructed with "scvelo.tl.velocity_graph" for latent time inference and visualization.

GRN analysis

To identify genetic programs regulating HF-related changes in cardiac cells, we applied the SCENIC+ (v1.0.1.dev4+ge4bdd9f) GRN tool to each major cell type. The multiome dataset was split by cell type, and SCENIC+ analysis was performed to identify cell subpopulation–specific regulons following the standard pipeline (https://scenicplus.readthedocs.io/en/latest/) (*35*). In brief, SCENIC+ is a three-step workflow that involves identifying candidate enhancers, identifying enriched TF-binding motifs, and linking TFs to candidate enhancers and target genes. For each major cell type, up to 10,000 cells were used for the GRN construction. The cCREs were set to the multiome defined peak set. To identify the top regulon for each subpopulation, regulons were filtered based on their regulon specificity scores (RSSs) and visualized for each subpopulation by degree centrality (Cytoscape) (*91*). GRNs were visualized in Cytoscape for the top regulons by RSS and direct target genes. Pseudotime trajectories per gene were calculated from scVelo results and incorporated as the color scheme in the GRN visualizations.

GWAS enrichments

For 77 genetic traits derived from the Million Veteran Program (MVP) (*92*), CardioVascular Disease Knowledge Portal (CVDKP) (*93*), published studies (*7*, *8*), and FinnGen (*94*), we performed summary statistics munging using the LDSC 1.0 (Linkage Disequilibrium Score Regression LDSC) framework (*51*). Summary statistics were first processed with munge_sumstats.py to ensure standardized formatting and compatibility with LDSC analysis. During this step, SNP alleles were harmonized and filtering was applied based on a MAF threshold of 0.01. Each dataset was aligned to the HapMap3 reference panel (*95*) to retain high-confidence variants. Key parameters included the specification of effect allele (EA), noneffect allele (NEA), *P* value (pvalue), effect size (beta), allele frequency (af_alt), and sample size (TotalSampleSize), with signed sumstats based on effect size directionality. The processed summary statistics were then analyzed using partitioned heritability estimation, leveraging cell type annotations to determine the contribution of specific genomic features to overall trait heritability. First, liftover was applied to convert hg38 peak coordinates to hg19 for each cell type. The resulting hg19 bed files were used to ensure compatibility with reference datasets in subsequent LDSC analyses. Next, annotation files were created for each cell type by mapping the lifted-over peak coordinates to SNPs from the 1000 Genomes Project (Phase 3, European population) (*96*). This was done using make_annot.py, generating one annotation file per chromosome for each cell type. Once annotations were generated, LD scores were computed using ldsc.py with a window size of 1 cM. This step involved thinning the annotation files and calculating LD scores for each SNP while ensuring overlap with baseline LD scores from the 1000 Genomes reference panel. Finally, heritability enrichment analysis was performed using ldsc.py, incorporating cell type annotations alongside baseline LD models. This step was executed for all the processed summary statistics. The analysis estimated the heritability enrichment and statistical significance of each annotation for each trait while controlling for allele frequencies and LD structure.

LDSC heritability enrichment results were aggregated across all cell types and filtered using FDR correction (0.1) to retain significant associations. nNMF (*97*) was applied to decompose the LDSC enrichment matrix into six latent factors (rank = 6), capturing shared heritability signals across cell types and traits. The analysis was run 50 times to ensure stability, generating a basis matrix (W) for cell type contributions and a coefficient matrix (H) for trait associations. Features and traits were then sorted by their dominant factor assignments to enhance interpretability. A heatmap was generated using pheatmap (https://cran.r-project.org/package=pheatmap) to visualize enrichment patterns, with row-wise scaling applied to highlight relative differences. At this step, additional manual sorting was performed to refine biological clustering. Statistical significance was overlaid, and clustering was manually disabled to preserve the factor-based ordering. To assess the contribution of chromatin state to trait heritability, LDSC was rerun using the same parameters but with annotations stratified into open and active chromatin states. This allowed for the evaluation of heritability enrichment specifically within functionally distinct regulatory regions. To ensure comparability across analyses, the same column and row order from the initial nNMF-based visualization was retained, enabling a direct comparison of how different chromatin accessibility states contribute to genetic trait heritability across cell types.

We performed enrichment of subpopulation GRNs using GWAS data for DCM. We retained variants with MAF > 0.05 and calculated Bayes factors for each variant using the approach of Wakefield (*98*). For each cardiomyocyte subpopulation, we obtained the set of cCREs in GRNs with subpopulation specific activity and used liftOver to convert coordinates to hg19. We then tested for enrichment of subtype GRN cCREs for DCM-associated variants using fgwas (v0.3.6) with a window size of 2000 variants.

Fine mapping

Credible sets were either downloaded from FinnGen or generated for a publicly available GWAS for DCM (*7*). Where available, causal variants from multitrait GWAS summary statistics (MTAG) were included. First, 1-Mb windows were determined from reported causal variants identified from single causal variant fine-mapping. Next, approximate Bayes factors were calculated using the Wakefield calculation (corrcoverage v1.2.0), which takes into account variant *P* values and a prior for the SD of the effect size (*98*). Finally, posterior probabilities were calculated using variant z-scores, effect size variances, and the prior for the SD of the effect size. Credible sets with a 95% coverage were produced by determining the minimum number of variants required to reach a posterior probability sum of 0.95.

Fine-mapping integration

For each cell type, genomic intersections between fine-mapped variants and chromatin accessibility peaks were identified using the findOverlaps function from the GenomicRanges package (*99*). Overlapping variants were extracted, and peak information was assigned to each variant overlap event. The total number of overlapping variants, independent signals, and peaks intersecting variants were computed for each cell type across different chromatin states.

Hi-C loops, ABC links, ChromBPNet predictions, and differential accessibility features for each cCRE were integrated using dplyr (*100*) join functions in R. All visualizations were generated using ggplot2 (*101*). The risk and protective allele information was annotated based on the GWAS results.

FORA from the fgsea (v1.26.0) package was used to calculate over-represented canonical pathways in vCM GRNs using all human genes as the background and the following parameters: minSize = 10, maxSize = 500 against all canonical pathways (https://www.gsea-msigdb.org/gsea/msigdb/download_file.jsp?filePath=/msigdb/release/2024.1.Hs/c4.all.v2024.1.Hs.symbols.gmt).

WGS sample processing

Genomic DNA were extracted from 25 to 50 mg of heart tissue using the Monarch Genomic DNA Purification Kit (T3010, NEB). In brief, tissues were incubated in lysis buffer with Proteinase K at 56°C for 2 hours in a Thermomixer with rotation at 1400 rpm. RNase A treatment followed for 5 min at 56°C. The genomic DNA was captured in binding buffer, passed through a column, and eluted with 50 µl of preheated water (65°C). The genomic DNA was quantified using Qubit fluorometry.

For library preparation, 500 ng of genomic DNA from each sample was fragmented by Adaptive Focused Acoustics (E220 Focused Ultrasonicator, Covaris) to produce an average fragment size of 500 bp. Sequencing libraries were generated using the KAPA Hyper Prep Kit (KAPA Biosystems) following the manufacturer's instructions using four cycles of amplification. The quality of the library was assessed using the High Sensitivity D1000 kit on a 4200 TapeStation instrument (Agilent Technologies).

Sequencing was performed using the NovaSeq X Plus Sequencing System (Illumina), generating 150-bp paired-end reads to obtain ~30× coverage.

WGS analysis

We generated WGS data for the cohort to serve as a "ground truth" reference. WGS was attempted for all 30 subjects; however, one sample (DTX60) was excluded due to a technical issue during sequencing library preparation, resulting in 29 samples available for concordance analysis. Raw sequencing reads were assessed using FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to evaluate base quality, GC content, sequence duplication, and adapter contamination. All libraries passed standard quality thresholds and were retained for downstream analysis. Paired-end reads were aligned to the GRCh38 reference genome using BWA-MEM2 v2.2.1 (DOI: 10.1109/IPDPS.2019.0004) with explicit read-group tags (ID, SM, PL, LB, and PU). SAM output files were converted to BAM, coordinate sorted, and indexed with samtools v1.21 (*102*). Alignment quality was verified using samtools coverage and samtools stats. Lane- and library-level BAMs were merged per subject using samtools merge while retaining read-group information, followed by BAM indexing with samtools index. Duplicate reads were marked using GATK MarkDuplicates v46.2.0 (ISBN: 9781491975183; DOI: 10.1101/201178) (*103*), generating duplicate-marked BAMs and associated metrics suitable for downstream recalibration and variant calling. Base quality score recalibration (BQSR) was performed using GATK BaseRecalibrator with standard known-sites resources (Mills_and_1000G_gold_standard.indels, 1000G_phase1.snps.high_confidence, 1000G_omni2.5.hg38, hapmap_3.3, and Homo_sapiens_assembly38.known_indels) obtained from the official GATK resource bundle. Recalibration was applied with GATK ApplyBQSR to generate final subject-level BAMs. Variants were called per subect using GATK HaplotypeCaller in GVCF mode and indexed with GATK IndexFeatureFile. Per-subject GVCFs were combined into chromosome-level databases using GATK GenomicsDBImport and jointly genotyped across all subjects with GATK GenotypeGVCFs using the GRCh38 reference. Resulting per-chromosome VCFs were compressed, indexed with tabix (*102*), and concatenated into a cohort-wide VCF using bcftools v1.21 concat (*86*). Variant quality was refined using GATK VariantRecalibrator and GATK ApplyVQSR using standard hg38 resources (HapMap, Omni, 1000G, Mills, dbSNP). Final recalibrated VCFs were evaluated with bcftools stats and filtered to retain only PASS, biallelic SNPs and INDELs using bcftools view. The filtered, indexed VCF served as the finalized variant set for downstream analyses. Filtered variants were annotated using the Ensembl Variant Effect Predictor (VEP) v115.2 (*104*) in offline mode with the GRCh38 cache and reference FASTA from the Ensembl database. Annotations included predicted molecular consequence, gene context, population frequency, and functional impact using the –everything flag. Output VCFs were compressed and indexed for downstream interpretation. Annotated VCFs were evaluated with bcftools stats and visualized using plot-vcfstats to assess per-subject and cohort-wide variant metrics. Annotation fields were extracted with bcftools +split-vep to generate per-variant and per-gene summary tables, including allele frequencies, functional consequences, and clinical annotations. Genotype matrices were parsed into tabular format using bcftools query and merged with annotation metadata in Python v3.10.14 (pandas v2.3.3) for downstream analyses. Per-library BAM files were then merged by subject using samtools merge, followed by indexing with samtools index. Joint variant calling was then performed with the GATK HaplotypeCaller in GVCF mode to generate per-sample variant call files for downstream joint genotyping.

The variants called from the array (stratified into directly genotyped and imputed) were compared against the WGS variant calls for the same individuals using bcftools stats.

Sex inference from array data

Sex was self-reported but also inferred from the older, pre-VEP VCF using bcftools +guess-ploidy with the hg38 reference.

Variant prioritization and ranking

To identify rare, putatively deleterious variants, we filtered subject-specific variation profiles using a multi-step prioritization strategy. First, we retained only rare variants (gnomAD genome allele frequency < 0.01) and strictly excluded any variants annotated as "benign" or "tolerated" across ClinVar, SIFT, or PolyPhen-2. The remaining variants were required to meet at least one pathogenicity criterion: a "pathogenic" classification in ClinVar, a "deleterious" prediction in SIFT (including low confidence), or a "possibly/probably damaging" prediction in PolyPhen-2. To rank the prioritized candidates, we used a tiered strategy. We first prioritized variants classified as pathogenic in ClinVar. We then ordered the remaining variants based on their predicted functional impact: high-impact mutations (e.g., frameshift, stop-gain) were given precedence, whereas missense variants were ranked using the average of their PolyPhen-2 probabilities and inverted SIFT probabilities (1 − SIFT).

Spatial transcriptomics: Formalin-fixed, paraffin-embedded tissue sectioning and Visium HD

To minimize potential RNAse contamination, dissection surfaces were cleaned with 70% EtoH, sterile water was used for floating sections, and clean new blades were used for microtome sectioning. Formalin-fixed, paraffin-embedded (FFPE) blocks were hydrated for 5 to 10 min on top of an ice-chilled flat metal surface, and sections were cut at 5-µm thickness mounted to Superfrost slides (Fisher Scientific). Slides were dried for 2 hours in the oven at 42°C after sectioning and stored at room temperature with desiccators. All sections were used for experiments within 2 weeks of mounting. To assess the RNA integrity of the tissue, DV200 fragments were assayed. Slices of three to five pieces at 10-µm-thick ribbon slices were saved before mounting. RNA was isolated as recommended in PureLink FFPE Kit (Invitrogen). All samples with DV200 fragments >60% were used for the Visium HD experiments. Deparaffinization and hematoxylin and eosin (H&E) staining conditions were performed as recommended by the 10X Genomics "VisiumHDFFPETissuePrepHandbook_RevA CG000684." Briefly, H&E images were acquired with an Olympus VS200 Slide Scanner at 20× magnification. Immediately after imaging, the slides were de-cross-linked and hybridized overnight with Visium Human Transcriptome

Probes v2 >18,000 genes, >54,000 probe pairs. After permeabilizing and barcoding, all hybridized RNA was transferred to a Visium Slide slide 6.5 mm × 6.5 mm oligo capture area. The ligated and barcoded transcripts were then subjected to amplification and indexing for library construction. The libraries were sequenced between 300 and 500 million bp. Paired-end, dual indexed sequencing conditions were Read1: 43 cycles, i7 Index: 10 cycles, i5 Index: 10 cycles and Read 2S: 50 cycles.

Spatial transcriptomics: Data processing

Visium HD spatial transcriptomic libraries were first processed with spaceranger (10x Genomics, v3.0.0). The subsequent files from nine samples were processed using Seurat. Data were loaded at 2-, 8-, and 16-µm binning, and downstream analyses were performed on the 16-µm bins. For quality control filtering, spots with mitochondrial content <35% and >50 detected features were kept. Standard processing workflow was applied: log normalization, identification of variable features, scaling, PCA, clustering, and UMAP visualization.

Spatially informed tissue domains were additionally identified using BANKSY (v1.1.0) (*105*), followed by PCA, neighbor graph construction, and clustering on the BANKSY assay. Processed samples were merged and corrected for batch effects using Harmony integration before clustering and UMAP embedding in the integrated latent space.

For cell type annotation, the merged Visium HD dataset was converted to AnnData (h5ad) format and annotated using TACCO (v0.4.0.post1) (*106*). Cell type labels were transferred per disease state using the single-cell multiome dataset as a reference. Labels were first transferred on the major population level. Further subpopulation labels were transferred after major population annotation.

REFERENCES AND NOTES

1. M. S. Khan *et al.*, Global epidemiology of heart failure. *Nat. Rev. Cardiol.* **21**, 717–734 (2024). doi: 10.1038/s41569-024-01046-6; pmid: 38926611
2. I. Olivotto, J. E. Udelson, M. Pieroni, C. Rapezzi, Genetic causes of heart failure with preserved ejection fraction: Emerging pharmacological treatments. *Eur. Heart J.* **44**, 656–667 (2023). doi: 10.1093/eurheartj/ehac764; pmid: 36582184
3. S. Heymans, N. K. Lakdawala, C. Tschöpe, K. Klingel, Dilated cardiomyopathy: Causes, mechanisms, and current and future treatment approaches. *Lancet* **402**, 998–1011 (2023). doi: 10.1016/S0140-6736(23)01241-2; pmid: 37716772
4. M. Gigli *et al.*, Pathophysiology of dilated cardiomyopathy: From mechanisms to precision medicine. *Nat. Rev. Cardiol.* **22**, 183–198 (2025). doi: 10.1038/s41569-024-01074-2; pmid: 39394525
5. M. Arvanitis *et al.*, Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. *Nat. Commun.* **11**, 1122 (2020). doi: 10.1038/s41467-020-14843-7; pmid: 32111823
6. S. Garnier *et al.*, Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25.1 and 22q11.23. *Eur. Heart J.* **42**, 2000–2011 (2021). doi: 10.1093/eurheartj/ehab030; pmid: 33677556
7. S. L. Zheng *et al.*, Genome-wide association analysis provides insights into the molecular etiology of dilated cardiomyopathy. *Nat. Genet.* **56**, 2646–2658 (2024). doi: 10.1038/s41588-024-01952-y; pmid: 39572783
8. S. J. Jurgens *et al.*, Genome-wide association study reveals mechanisms underlying dilated cardiomyopathy and myocardial resilience. *Nat. Genet.* **56**, 2636–2645 (2024). doi: 10.1038/s41588-024-01975-5; pmid: 39572784
9. M. Litviňuková *et al.*, Cells of the adult human heart. *Nature* **588**, 466–472 (2020). doi: 10.1038/s41586-020-2797-4; pmid: 32971526
10. D. Reichart *et al.*, Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science* **377**, eabo1984 (2022). doi: 10.1126/science.abo1984; pmid: 35926050
11. A. M. A. Miranda *et al.*, Single-cell transcriptomics for the assessment of cardiac disease. *Nat. Rev. Cardiol.* **20**, 289–308 (2023). doi: 10.1038/s41569-022-00805-7; pmid: 36539452
12. C. Uhler, G. V. Shivashankar, Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat. Rev. Mol. Cell Biol.* **18**, 717–727 (2017). doi: 10.1038/nrm.2017.101; pmid: 29044247
13. Y. Xie *et al.*, Droplet-based single-cell joint profiling of histone modifications and transcriptomes. *Nat. Struct. Mol. Biol.* **30**, 1428–1433 (2023). doi: 10.1038/s41594-023-01060-1; pmid: 37563440
14. L. Chang *et al.*, Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat. Biotechnol.* **43**, 1694–1707 (2024). doi: 10.1038/s41587-024-02447-1; pmid: 39424717
15. R. J. Mills *et al.*, Functional screening in human cardiac organoids reveals a metabolic mechanism for cardiomyocyte cell cycle arrest. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8372–E8381 (2017). doi: 10.1073/pnas.1707316114; pmid: 28916735
16. J. D. Hocker *et al.*, Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci. Adv.* **7**, eabf1444 (2021). doi: 10.1126/sciadv.abf1444; pmid: 33990324
17. C. Kuppe *et al.*, Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766–777 (2022). doi: 10.1038/s41586-022-05060-x; pmid: 35948637
18. M. Chaffin *et al.*, Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* **608**, 174–180 (2022). doi: 10.1038/s41586-022-04817-8; pmid: 35732739
19. C. P. Fulco *et al.*, Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019). doi: 10.1038/s41588-019-0538-0; pmid: 31784727
20. Y. E. Li *et al.*, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021). doi: 10.1038/s41586-021-03604-1; pmid: 34616068
21. J. E. Moore *et al.*, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020). doi: 10.1038/s41586-020-2493-4; pmid: 32728249
22. K. Zhang *et al.*, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021). doi: 10.1016/j.cell.2021.10.024; pmid: 34774128
23. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012). doi: 10.1038/nmeth.1906; pmid: 22373907
24. D. U. Gorkin *et al.*, An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020). doi: 10.1038/s41586-020-2093-3; pmid: 32728240
25. A. Kundaje *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
26. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010). doi: 10.1038/nbt.1630; pmid: 20436461
27. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: 10.1016/j.molcel.2010.05.004; pmid: 20513432
28. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
29. B. Simonson *et al.*, Single-nucleus RNA sequencing in ischemic cardiomyopathy reveals common transcriptional profile underlying end-stage heart failure. *Cell Rep.* **42**, 112086 (2023). doi: 10.1016/j.celrep.2023.112086; pmid: 36790929
30. M. Y. Chan *et al.*, Prioritizing candidates of post-myocardial infarction heart failure using plasma proteomics and single-cell transcriptomics. *Circulation* **142**, 1408–1421 (2020). doi: 10.1161/CIRCULATIONAHA.119.045158; pmid: 32885678
31. Y. H. Cui, C. R. Wu, D. Xu, J. G. Tang, Exploration of neuron heterogeneity in human heart failure with dilated cardiomyopathy through single-cell RNA sequencing analysis. *BMC Cardiovasc. Disord.* **24**, 86 (2024). doi: 10.1186/s12872-024-03739-9; pmid: 38310240
32. Y. Ding, J. Zhu, G. Xu, Q. Cheng, C. Zhu, Single-cell RNA-seq analysis of hearts in patients with fetal tetralogy of Fallot. *Cardiology* **150**, 221–232 (2025). pmid: 39097963
33. A. He *et al.*, Polycomb repressive complex 2 regulates normal development of the mouse heart. *Circ. Res.* **110**, 406–415 (2012). doi: 10.1161/CIRCRESAHA.111.252205; pmid: 22158708
34. C. M. Greco, G. Condorelli, Epigenetic modifications and noncoding RNAs in cardiac hypertrophy and failure. *Nat. Rev. Cardiol.* **12**, 488–497 (2015). doi: 10.1038/nrcardio.2015.71; pmid: 25962978
35. C. Bravo González-Blas *et al.*, SCENIC+: Single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023). doi: 10.1038/s41592-023-01938-4; pmid: 37443338
36. E. N. Farah *et al.*, Spatially organized cellular communities form the developing human heart. *Nature* **627**, 854–864 (2024). doi: 10.1038/s41586-024-07171-z; pmid: 38480880
37. M. André *et al.*, Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels. *Nat. Commun.* **15**, 3352 (2024). doi: 10.1038/s41467-024-47735-1; pmid: 38688933
38. J. M. Amrute *et al.*, Targeting immune-fibroblast cell communication in heart failure. *Nature* **635**, 423–433 (2024). doi: 10.1038/s41586-024-08008-5; pmid: 39443792
39. M. Alexanian *et al.*, Chromatin remodelling drives immune cell-fibroblast communication in heart failure. *Nature* **635**, 434–443 (2024). doi: 10.1038/s41586-024-08085-6; pmid: 39443808
40. L. Medzikovic, C. J. M. de Vries, V. de Waard, NR4A nuclear receptors in cardiac remodeling and neurohormonal regulation. *Trends Cardiovasc. Med.* **29**, 429–437 (2019). doi: 10.1016/j.tcm.2018.11.015; pmid: 30553703

41. P. Boström *et al.*, C/EBPβ controls exercise-induced cardiac growth and protects against pathological cardiac remodeling. *Cell* **143**, 1072–1083 (2010). doi: 10.1016/j.cell.2010.11.036; pmid: 21183071
42. G. N. Huang *et al.*, C/EBP transcription factors mediate epicardial activation during heart development and injury. *Science* **338**, 1599–1603 (2012). doi: 10.1126/science.1229765; pmid: 23160954
43. J. C. K. Man *et al.*, Genetic dissection of a super enhancer controlling the Nppa-Nppb cluster in the heart. *Circ. Res.* **128**, 115–129 (2021). doi: 10.1161/CIRCRESAHA.120.317045; pmid: 33107387
44. Y. Wang *et al.*, Fibroblasts in heart scar tissue directly regulate cardiac excitability and arrhythmogenesis. *Science* **381**, 1480–1487 (2023). doi: 10.1126/science.adh9925; pmid: 37769108
45. O. Contreras, H. Soliman, M. Theret, F. M. V. Rossi, E. Brandan, TGF-β-driven downregulation of the transcription factor TCF7L2 affects Wnt/β-catenin signaling in PDGFRα$^+$ fibroblasts. *J. Cell Sci.* **133**, jcs242297 (2020). doi: 10.1242/jcs.242297; pmid: 32434871
46. J. M. Amrute *et al.*, Defining cardiac functional recovery in end-stage heart failure at single-cell resolution. *Nat. Cardiovasc. Res.* **2**, 399–416 (2023). doi: 10.1038/s44161-023-00260-8; pmid: 37583573
47. Y. Fang *et al.*, RUNX2 promotes fibrosis via an alveolar-to-pathological fibroblast transition. *Nature* **640**, 221–230 (2025). doi: 10.1038/s41586-024-08542-2; pmid: 39910313
48. X. Sun, M. Zhu, X. Chen, X. Jiang, MYH9 inhibition suppresses TGF-β1-stimulated lung fibroblast-to-myofibroblast differentiation. *Front. Pharmacol.* **11**, 573524 (2021). doi: 10.3389/fphar.2020.573524; pmid: 33519439
49. K. Bernau *et al.*, Tensin 1 is essential for myofibroblast differentiation and extracellular matrix formation. *Am. J. Respir. Cell Mol. Biol.* **56**, 465–476 (2017). doi: 10.1165/rcmb.2016-0104OC; pmid: 28005397
50. X. He *et al.*, Myofibroblast YAP/TAZ activation is a key step in organ fibrogenesis. *JCI Insight* **7**, e146243 (2022). doi: 10.1172/jci.insight.146243; pmid: 35191398
51. B. K. Bulik-Sullivan *et al.*, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015). doi: 10.1038/ng.3211; pmid: 25642630
52. M. C. Hill *et al.*, Integrated multi-omic characterization of congenital heart disease. *Nature* **608**, 181–191 (2022). doi: 10.1038/s41586-022-04989-3; pmid: 35732239
53. K. Kanemaru *et al.*, Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023). doi: 10.1038/s41586-023-06311-1; pmid: 37438528
54. J. Wohlschlaeger *et al.*, Hemodynamic support by left ventricular assist devices reduces cardiomyocyte DNA content in the failing human heart. *Circulation* **121**, 989–996 (2010). doi: 10.1161/CIRCULATIONAHA.108.808071; pmid: 20159834
55. V. K. Ninh *et al.*, Spatially clustered type I interferon responses at injury borderzones. *Nature* **633**, 174–181 (2024). doi: 10.1038/s41586-024-07806-1; pmid: 39198639
56. D. G. Lupiáñez *et al.*, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015). doi: 10.1016/j.cell.2015.04.004; pmid: 25959774
57. J. R. Dixon *et al.*, Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* **50**, 1388–1398 (2018). doi: 10.1038/s41588-018-0195-8; pmid: 30202056
58. E. R. Porrello *et al.*, Transient regenerative potential of the neonatal mouse heart. *Science* **331**, 1078–1080 (2011). doi: 10.1126/science.1200708; pmid: 21350179
59. K. Vandereyken, A. Sifrim, B. Thienpont, T. Voet, Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023). doi: 10.1038/s41576-023-00580-2; pmid: 36864178
60. C. P. Fulco *et al.*, Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016). doi: 10.1126/science.aag2445; pmid: 27708057
61. F. M. Chardon *et al.*, Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements. *Nat. Commun.* **15**, 8209 (2024). doi: 10.1038/s41467-024-52490-4; pmid: 39294132
62. V. Agarwal *et al.*, Massively parallel characterization of transcriptional regulatory elements. *Nature* **639**, 411–420 (2025). doi: 10.1038/s41586-024-08430-9; pmid: 39814889
63. J. Wever-Pinzon *et al.*, Impact of ischemic heart failure etiology on cardiac recovery during mechanical unloading. *J. Am. Coll. Cardiol.* **68**, 1741–1752 (2016). doi: 10.1016/j.jacc.2016.07.756; pmid: 27737740
64. M. R. Corces *et al.*, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017). doi: 10.1038/nmeth.4396; pmid: 28846090
65. H. M. Kang *et al.*, Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018). doi: 10.1038/nbt.4042; pmid: 29227470
66. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
67. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020). doi: 10.1093/gigascience/giaa151; pmid: 33367645
68. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: 10.1038/s41592-019-0619-0; pmid: 31740819
69. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021). doi: 10.1038/s41592-021-01282-5; pmid: 34725479
70. S. Neph *et al.*, BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012). doi: 10.1093/bioinformatics/bts277; pmid: 22576172
71. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019). doi: 10.1016/j.cels.2018.11.005; pmid: 30954476
72. C. Xu *et al.*, Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021). doi: 10.15252/msb.20209620; pmid: 33491336
73. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
74. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42** (W1), W187-91 (2014). doi: 10.1093/nar/gku365; pmid: 24799436
75. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024). doi: 10.1038/s41587-023-01767-y; pmid: 37231261
76. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019). doi: 10.1186/s13059-019-1874-1; pmid: 31870423
77. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
78. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
79. N. Abdennur *et al.*, Pairtools: From sequencing data to chromosome contacts. *PLOS Comput. Biol.* **20**, e1012164 (2024). doi: 10.1371/journal.pcbi.1012164; pmid: 38809952
80. J. Zhou *et al.*, Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14011–14018 (2019). doi: 10.1073/pnas.1901423116; pmid: 31235599
81. S. Shen, Y. Zheng, S. Keleş, scGAD: Single-cell gene associating domain scores for exploratory analysis of scHi-C data. *Bioinformatics* **38**, 3642–3644 (2022). doi: 10.1093/bioinformatics/btac372; pmid: 35652733
82. W. Dong, M. Charikar, K. Li, "Efficient K-nearest neighbor graph construction for generic similarity measures" in *Proceedings of the 20th International Conference on World Wide Web, WWW 2011* (ACM, 2011), pp. 577–586; https://dl.acm.org/doi/10.1145/1963405.1963487.
83. M. Stephens, False discovery rates: A new deal. *Biostatistics* **18**, 275–294 (2017). pmid: 27756721
84. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706
85. A. Pampari *et al.*ChromBPNet: bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. bioRxiv 630221 (2025); doi: 10.1101/2024.12.25.630221
86. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi: 10.1093/bioinformatics/btr330; pmid: 21653522
87. S. G. Coetzee, G. A. Coetzee, D. J. Hazelett, R. Motifbreak, motifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015). doi: 10.1093/bioinformatics/btv470; pmid: 26272984
88. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
89. G. Korotkevich *et al.*, Fast gene set enrichment analysis. bioRxiv (2016), p. 060012.
90. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3; pmid: 32747759
91. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi: 10.1101/gr.1239303; pmid: 14597658
92. J. M. Gaziano *et al.*, Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016). doi: 10.1016/j.jclinepi.2015.09.016; pmid: 26441289
93. M. C. Costanzo *et al.*, Cardiovascular Disease Knowledge Portal: A community resource for cardiovascular disease research. *Circ. Genom. Precis. Med.* **16**, e004181 (2023). doi: 10.1161/CIRCGEN.123.004181; pmid: 37814896
94. M. I. Kurki *et al.*, FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023). doi: 10.1038/s41586-022-05473-8; pmid: 36653562

95. D. M. Altshuler *et al.*, Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010). doi: 10.1038/nature09298; pmid: 20811451
96. A. Auton *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
97. R. Gaujoux, C. Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010). doi: 10.1186/1471-2105-11-367; pmid: 20598126
98. J. Wakefield, Bayes factors for genome-wide association studies: Comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009). doi: 10.1002/gepi.20359; pmid: 18642345
99. M. Lawrence *et al.*, Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013). doi: 10.1371/journal.pcbi.1003118; pmid: 23950696
100. H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan, dplyr: A grammar of data manipulation (R package version 1.2.1, 2023); https://dplyr.tidyverse.org.
101. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016); http://link.springer.com/10.1007/978-3-319-24277-4.
102. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
103. A. McKenna *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
104. W. McLaren *et al.*, The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016). doi: 10.1186/s13059-016-0974-4; pmid: 27268795
105. V. Singhal *et al.*, BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024). doi: 10.1038/s41588-024-01664-3; pmid: 38413725
106. S. Mages *et al.*, TACCO unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotechnol.* **41**, 1465–1473 (2023). doi: 10.1038/s41587-023-01657-3; pmid: 36797494
107. Y. Xie, L. Tucciarone, E. N. Farah, L. Chang, Data for: Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure, Zenodo (2026); https://doi.org/10.5281/zenodo.15232789.
108. Y. Xie, Code for: Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure, Zenodo (2026); https://doi.org/10.5281/zenodo.19372224.

ACKNOWLEDGMENTS

We thank the University of Utah patients and their families for providing the valuable myocardial tissue donations used in this study; the Chi, Drakos, Gaulton, and Ren laboratory staff for comments; the University of California San Diego (UCSD) Institute for Genomic Medicine for their help in sequencing data generation; and J. D. Hocker and S. Preissl for comments and help with single-nuclei isolation from heart tissue. **Funding:** This project was supported by the Foundation for the National Institutes of Health (FNIH) (K.J.G., B.R., N.C.C., S.G.D.); the American Heart Association (AHA) Heart Failure Strategically Focused Research Network (grant 16SFRN29020000 to S.G.D.); the National Heart, Lung and Blood Institute (NHLBI) of the NIH (grants R01 HL135121 and 1R01HL166513 to S.G.D. and grant T32HL007444 to E.N.F. and A.R.H.); the US Department of Veterans Affairs (Merit Review Award I01 CX002291 to S.G.D.); the Nora Eccles Treadwell Foundation (S.G.D.); the NIH (grant HG012059 to K.J.G. and grant 1K08HL168315-01 to E.T.); and the Life Sciences Research Foundation (Z.W.). **Author contributions:** Conceptualization: S.G.D., K.J.G., B.R., N.C.C.; Data curation: L.C., T.S.S., S.P., A.L., T.L., Q.Y.; Formal analysis: Y.X., E.N.F., L.T., S.T., H.Z., A.R.H., J.B., D.L., C.S., T.W.; Funding acquisition: S.G.D., K.J.G., B.R., N.C.C.; Investigation: Y.X., E.N.F., L.T., Q.Y., S.T., J.D., J.B., H.Z., E.D., W.E., S.C., R.M.E., S.M., Z.W., J.H.-C.C., R.M., C.M., E.G., J.L.; Methodology: Y.X., L.C., A.W.; Project administration: Y.X., E.N.F., L.T., L.C., T.S.S., A.D.-C.; Resources: T.S.S., E.T., V.H., Q.Z., S.N., C.H.S., S.G.D., N.C.C.; Supervision: Y.X., E.N.F., L.T., A.W., S.G.D., K.J.G., B.R., N.C.C.; Visualization: Y.X., E.N.F., L.T., S.T., H.Z., A.R.H., J.B., D.L., C.S., T.W.; Writing – original draft: Y.X., E.N.F., L.T., K.J.G., B.R., N.C.C.; Writing – review & editing: Y.X., E.N.F., L.T., L.C., A.R.H., E.D., A.W., S.G.D., K.J.G., B.R., N.C.C. **Competing interests:** S.G.D. has received consultancy fees from Abbott. K.J.G. has received consultancy fees from Genentech. K.J.G. has received honoraria from Pfizer, holds stock in Neurocrine Biosciences, and his spouse is employed at Altos Labs. S.G.D. acknowledges research support from Novartis. A.R.H. is an employee of Aspen Neuroscience and holds equity in the company. R.M.E. is an employee and shareholder of Pfizer. B.R. is a shareholder and consultant of Arima Genomics Inc. and cofounder of Epigenome Technologies, Inc. N.C.C. and University of California San Diego are inventors of a filed patent. The remaining authors declare no competing interests. **Data, code, and materials availability:** No new materials were generated in this study. All raw sequence and genotyping data generated in this study have been deposited in dbGaP with accession number phs004650. Processed data including supplementary tables are available on Zenodo (*107*). Processed epigenomic tracks are available to explore on our web portal: https://epigenome.wustl.edu/HeartEpigenome/. Code and analysis notebooks are deposited in Zenodo (*108*) and are also available through GitHub at https://github.com/Xieeeee/FNIH_Heart/. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady6893
Supplementary Text; Figs. S1 to S22

Submitted 1 May 2025; resubmitted 31 December 2025; accepted 12 May 2026

10.1126/science.ady6893

RESEARCH ARTICLE SUMMARY

Single-cell multiomics connects 3D genome and transcriptome alterations in Alzheimer's disease

Yang Zhang, Xinyue Lu, Alexander K. Kunisky, Shahul Alam, Junjie Tang, Ruochi Zhang, Shike Wang, Han Zhang, Jude Baroudi, Walid Ichcho, Deyong Jia, Sahar Ghorbanikalateh, Sahel Ghorbanikalateh, Shihan Wang, David A. Bennett, Hansruedi Mathys*, Zhijun Duan*, Jian Ma*

Full article and list of author affiliations: https://doi.org/10.1126/science.adz1652

INTRODUCTION: Alzheimer's disease (AD) is the most common cause of dementia and is marked by progressive loss of brain function. Many studies have cataloged changes in gene activity across brain cell types; however, the molecular mechanisms underlying these changes remain elusive. Gene activity is controlled not only by DNA sequence and chemical marks on DNA but also by how the genome is folded in three dimensions inside the nucleus. How this three-dimensional (3D) genome organization changes in AD and how such changes relate to cell type–specific gene dysregulation in the human brain remain poorly understood.

RATIONALE: We aimed to determine whether changes in 3D genome folding are linked to the gene expression programs disrupted in AD and whether these links can be detected at single-cell resolution in human brain tissue. To do this, we used GAGE-seq (genome architecture and gene expression by sequencing), a technology that measures, in the same single cell, both gene expression and physical contacts within the genome. We integrated these measurements with chromatin accessibility data from the same donors and with spatial transcriptome maps in intact tissue sections, enabling analyses across molecular, cellular, and tissue scales. We also developed a transformer-based predictive model that integrates DNA sequence with 3D genome features to test when genome structure is necessary to explain AD-related gene expression changes.

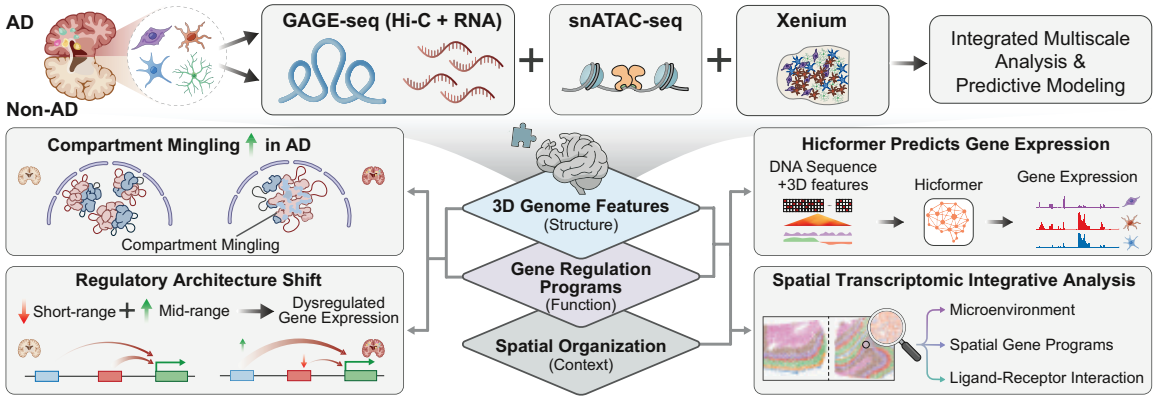
RESULTS: Across major brain cell types, we observed a reproducible shift in genome contact patterns in AD, including reduced short-range interactions and increased longer-range interactions. Although overall compartment patterns were broadly preserved, active and inactive genome regions showed increased mixing, consistent with weakened compartment segregation. These architectural changes were linked to broad, cell type–specific transcriptional remodeling of disease-relevant pathways, and we further related these programs to the current landscape of AD clinical trial targets.

At regulatory elements identified by chromatin accessibility, promoter-proximal interactions weakened, while midrange regulatory interactions became relatively more prominent, particularly at sites associated with chromatin loop organization. In parallel, we observed AD-related changes in key cellular programs, including senescence-related activation in microglia and sex-dependent dysregulation of X-linked genes in females, accompanied by corresponding 3D genome changes at implicated loci.

Our predictive model showed that 3D genome features provide information beyond DNA sequence alone for explaining AD-relevant gene expression, enabling prioritization of distal regulatory elements whose effects are mediated through chromatin contacts. Integrating these molecular features with spatial transcriptomics further placed them in tissue context, revealing altered cell neighborhoods and disrupted spatial coordination of gene programs in AD. Overall, the study connects genome structure, gene regulation, and tissue organization through a unified multimodal analysis.

CONCLUSION: These results provide a multiscale map linking 3D genome remodeling to cell type–specific gene expression changes and spatial tissue organization in AD. The study establishes genome folding as a key regulatory layer associated with AD pathology and provides a framework and resource for mechanistic hypothesis generation, including prioritization of regulatory elements and AD-relevant gene programs for future functional testing and therapeutic exploration. □

*Corresponding author. Email: mathysh@pitt.edu (H.M.); zjduan@uw.edu (Z.D.); jianma@cs.cmu.edu (J.M.) Cite this article as Y. Zhang *et al.*, *Science* **393**, eadz1652 (2026). DOI: 10.1126/science.adz1652



3D genome remodeling in Alzheimer's disease. Single-cell multimodal GAGE-seq maps 3D genome organization and transcriptome in AD. The disease-relevant genome structural shifts track altered gene regulation. Integration with spatial transcriptomics uncovers altered spatial cellular niches reflecting in situ dynamics of the 3D genome in AD. Coupling GAGE-seq with Hicformer, a transformer-based model, further reveals an essential regulatory role of the 3D genome in AD. snATAC-seq, single-nucleus assay for transposase-accessible chromatin using sequencing.

RESEARCH ARTICLE

Single-cell multiomics connects 3D genome and transcriptome alterations in Alzheimer's disease

Yang Zhang[1]†, Xinyue Lu[1]†, Alexander K. Kunisky[2]‡, Shahul Alam[1]‡, Junjie Tang[1]‡, Ruochi Zhang[3]‡, Shike Wang[1], Han Zhang[4], Jude Baroudi[2], Walid Ichcho[5], Deyong Jia[6], Sahar Ghorbanikalateh[2], Sahel Ghorbanikalateh[2], Shihan Wang[2], David A. Bennett[7], Hansruedi Mathys[2]*, Zhijun Duan[5,8,9]*, Jian Ma[1]*

Alzheimer's disease (AD) disrupts brain function through cell type–specific transcriptomic and epigenomic alterations, yet the contribution of three-dimensional (3D) genome organization to AD remains poorly understood. We applied GAGE-seq (genome architecture and gene expression by sequencing) to jointly profile gene expression and 3D chromatin structure in single cells from postmortem brain tissue from AD patients and age-matched individuals without AD, revealing chromatin reorganization linked to cell type–specific dysregulation. Integrations with spatial transcriptomics and chromatin accessibility data uncovered altered niches reflecting genome compartment remodeling and regulatory element reorganization. Hicformer, a deep learning framework, showed that 3D genome features are essential for predicting disease-relevant, cell type–specific gene expression changes. Our results establish higher-order chromatin alterations as a component of AD-associated molecular pathology, providing a multiscale view of transcriptional regulation and 3D genome organization in neurodegeneration.

Alzheimer's disease (AD), the leading cause of dementia, presents a major public health challenge with profound social and economic implications, yet its molecular underpinnings remain incompletely defined (*1*–*3*). Recent advances in single-cell omics technologies have substantially advanced our understanding of the molecular complexity underlying AD pathogenesis, revealing cell type–specific transcriptional and epigenetic alterations associated with disease progression (*4*–*9*). Single-cell RNA sequencing (scRNA-seq) studies comparing postmortem human brain tissues from individuals across the AD clinical spectrum and cognitively unimpaired controls have provided unprecedented resolution in defining AD-related transcriptional perturbations that have informed evolving theories of pathogenesis and progression (*4*, *10*). Extensive single-cell analyses, involving large cohorts such as the Religious Orders Study and the Rush Memory and Aging Project (collectively known as ROSMAP) (*11*), have uncovered distinct transcriptional signatures across neuronal and glial populations linked to AD pathology, cognitive impairment, and resilience, highlighting the crucial role of cellular heterogeneity in AD etiology (*4*, *6*, *10*, *12*). Notably, recent high-resolution scRNA-seq profiling studies identified selectively vulnerable inhibitory neuron subtypes depleted in AD, specific neuronal subsets enriched in individuals maintaining high cognitive function despite substantial pathology, an astrocyte gene expression program linked to cognitive resilience, and glial subpopulations implicated in modulating AD pathology and cognitive decline (*6*, *7*, *10*, *12*). Complementing these insights, single-cell epigenomic assays, such as the single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq), have revealed widespread chromatin accessibility changes in late-stage AD, suggesting that epigenetic dysregulation may play a critical role in neuronal dysfunction and cellular identity loss (*13*).

Despite these advances, the role of three-dimensional (3D) genome architecture, an essential regulatory layer of the epigenome, in shaping cell type–specific transcriptional disruption in AD remains largely unexplored (*14*). The genome organization in 3D space within the cell nucleus critically regulates gene expression and cellular identity through multiscale genome structure features, including A/B compartmentalization, topologically associating domains, and chromatin looping (*15*–*17*). Dysregulation of these higher-order chromatin features can reorganize gene regulatory landscapes and has been implicated in various neurological disorders (*18*–*21*). While recent studies have suggested epigenetic erosion in AD (*13*, *22*), a study comprehensively investigating 3D genome remodeling and transcriptional dysregulation in AD at single-cell and cell type–specific resolution is missing. Furthermore, no study to date has integrated 3D chromatin architecture with the transcriptome, chromatin accessibility, and cellular spatial context from the same samples or conditions. Addressing this gap through multimodal integration is crucial for understanding how 3D genome architecture contributes to AD pathophysiology and for identifying regulatory mechanisms driving selective cellular vulnerability and disease progression (*23*).

We recently developed GAGE-seq (genome architecture and gene expression by sequencing) (*24*), a single-cell multiomic technology that simultaneously profiles transcriptomes and 3D chromatin interactions from the same single cells. In this study, leveraging GAGE-seq, we conducted an integrative and comprehensive single-cell multiomic analysis of human prefrontal cortex (PFC) samples from age-matched individuals with and without AD from ROSMAP (*11*), integrating transcriptomic profiles and 3D genome architecture. We further enriched this approach by incorporating participant-matched snATAC-seq and spatial transcriptome data (Xenium), enabling a unified framework to delineate how alterations in 3D genome structure may influence transcriptional regulation and cellular interactions within the spatially complex architecture of the brain. Our analyses reveal distinct AD-associated alterations in chromatin looping and genome compartmentalization explicitly linked to transcriptional dysregulation in specific neuronal and glial subpopulations, including excitatory neurons, inhibitory neurons, and astrocytes. Furthermore, spatially resolved multiomic integration highlights localized shifts in cell-cell interactions correlated directly with inferred disruptions in 3D genome organization. This study thus provides a high-resolution map linking 3D genome remodeling, gene expression changes, and spatial cellular alterations in AD, considerably advancing our understanding of the complex molecular landscape and regulatory architecture underlying the disease.

Results

High-quality single-cell co-assay of transcriptome and 3D genome

To investigate the relationship between 3D genome architecture and gene expression changes in AD, we applied GAGE-seq (*24*) to postmortem PFC tissue from 20 individuals in ROSMAP (*11*) (Fig. 1A). We selected 10 individuals with clinical and pathological diagnoses of AD (Braak stages 5 and 6; AD group) and 10 individuals without such

[1]Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Department of Neurobiology, University of Pittsburgh, Pittsburgh, PA, USA. [3]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Computer Science Department, University of California, Los Angeles, Los Angeles, CA, USA. [5]Division of Hematology and Oncology, Department of Medicine, University of Washington, Seattle, WA, USA. [6]Department of Urology, University of Washington, Seattle, WA, USA. [7]Rush Alzheimer's Disease Center, Chicago, IL, USA. [8]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA. [9]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. ***Corresponding author. Email: mathysh@pitt.edu (H.M.); zjduan@uw.edu (Z.D.); jianma@cs.cmu.edu (J.M.)**
†These authors contributed equally to this work. ‡These authors contributed equally to this work.

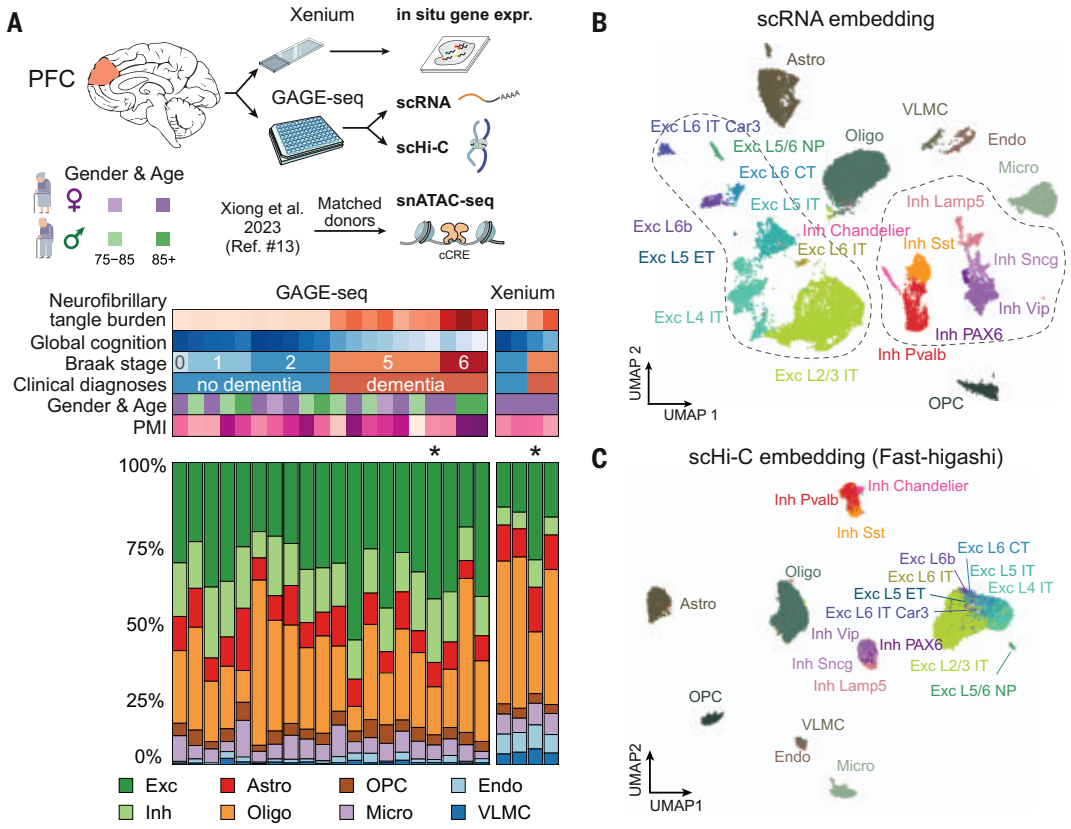


Fig. 1. High-quality joint profiling of transcriptome and 3D genome in AD. (**A**) Overview of the experimental design and sample composition. PFC tissue from 20 individuals (10 AD, 10 non-AD), with matched clinical and pathological annotations, was profiled using GAGE-seq, enabling joint measurement of single-cell RNA expression and single-cell Hi-C from the same nuclei. Matched donor snATAC-seq data from Xiong *et al*. (*13*) were integrated for regulatory element annotation, and additional Xenium spatial transcriptome samples were included for comparison. Heatmaps summarize key clinical and demographic metadata for each donor, including neurofibrillary tangle burden, global cognition, Braak stage, clinical diagnosis, sex, age, and postmortem interval (PMI). Stacked bar plot shows relative abundance of major cell types across individuals. Astro, astrocyte; Endo, endothelial cell; Exc, excitatory; Inh, inhibitory; IT, intratelencephalic; Micro, microglia; Oligo, oligodendrocyte. The asterisk symbol indicates one donor shared between the Xenium and GAGE-seq datasets. (**B**) UMAP visualization of RNA-defined cell types ($n = 23{,}825$ nuclei) showing major neuronal and non-neuronal cell types and subtypes identified from scRNA-seq data. CT, corticothalamic; NP, near-projecting; Pvalb, parvalbumin; Sncg, synuclein gamma; Sst, somatostatin; Vip, vasoactive intestinal peptide. (**C**) UMAP embedding derived solely from 3D chromatin contacts using Fast-Higashi. Cells are colored by cell type identities assigned from the matched scRNA-seq modality [as shown in (B)]. The concordance between the contact-based embedding and RNA-defined cell type identities demonstrates that 3D genome organization alone is sufficient to robustly separate major brain cell types.

diagnoses (Braak stages 0, 1, and 2; non-AD group), matched for age and balanced by sex (see methods for detailed selection criteria). Clinical and pathological annotations provided by ROSMAP confirmed markedly higher neuritic plaque and neurofibrillary tangle burden and lower global cognition scores in the AD group (fig. S1A and methods). Additionally, all 20 donors were part of a recent AD study (*13*) examining the genome-wide chromatin accessibility in the PFC region using snATAC-seq, enabling cross-modality integration.

To complement this multiomic profiling, we conducted spatial gene expression profiling using Xenium with the Prime 5K panel on PFC samples from two AD individuals and two age-matched non-AD individuals from the same cohort. This spatial dataset was later integrated with the GAGE-seq dataset to define spatial gene programs and examine tissue-level dynamics of 3D genome features (see the section "Spatially resolved gene programs and in situ alterations of the 3D genome in AD"). Notably, consistent with intact-section spatial profiling better capturing cerebrovascular niches (*25*, *26*), vasculature-associated cell types such as VLMCs (vascular leptomeningeal cells) appeared proportionally higher in Xenium than in the single-nucleus GAGE-seq dataset.

After stringent quality control of the GAGE-seq data—removing cells with low-coverage RNA and/or low-coverage Hi-C reads and doublets (tables S1 and S2 and methods)—we retained 23,825 high-quality cells. On average, each cell contained 22,704 unique molecular identifiers (UMIs) and 2477 expressed genes, along with 288,931 chromatin contacts (range: 30,088 to 3,034,568). These per-cell measurements reflect the high depth and robust quality of both modalities (scRNA and scHi-C) of GAGE-seq and enables detailed joint analysis of cell type–specific transcriptomic and 3D genome landscapes in these samples.

We annotated cells into eight major cell types and 22 subtypes using known marker genes and prior annotations from the same brain region (*7*), which include nine excitatory and seven inhibitory neuron populations (Fig. 1, A and B, and methods). The overall distribution of cell types is broadly similar across individuals and diagnoses, with some interindividual variability (Fig. 1B and fig. S1, B and C), confirming reproducible recovery of transcriptional identities from GAGE-seq. To evaluate whether 3D chromatin organization alone could resolve cellular identities, we applied Fast-Higashi (*27*) to the scHi-C data from GAGE-seq. The UMAP (uniform manifold approximation and projection) embedding of chromatin contacts recovered major neuronal and glial populations (Fig. 1C), demonstrating that 3D chromatin contact patterns encode broad cell identity and supporting the reliability and modality-specific resolution of GAGE-seq (fig. S1D).

To further validate the cell type annotations, we integrated GAGE-seq transcriptomes with scRNA-seq profiles from matched donors (*10*). The

resulting joint embedding showed strong agreement in clustering and cell type labels (fig. S2, A, B, E, and F), confirming robust cross-modality consistency. Similarly, joint embedding of GAGE-seq transcriptomes and matched-donor snATAC-seq gene activity scores (*13*) also showed high agreement in transcriptional identity and cell type–specific chromatin accessibility (fig. S2, C, D, G, and H). Marker genes consistently exhibited concordant gene expression, ATAC-seq gene activity score, and 3D genome features across cell types (fig. S2I).

Together, these data demonstrate that GAGE-seq yields high-quality, multimodal single-cell joint profiles of transcriptomic and 3D genome features across diverse brain cell types of both AD and non-AD samples. This newly created resource also enables robust integration with orthogonal single-cell and spatial modalities (see later section on integration with Xenium), providing a foundation for dissecting AD-specific gene regulation and chromatin reorganization.

Cell type–specific transcriptional alterations and pathway-level dysregulation in AD

To characterize transcriptome-wide changes in AD, we performed differential gene expression (DEG) analysis between AD and non-AD samples. To validate our findings, we compared GAGE-seq results with scRNA-seq data from Mathys *et al.* (*10*) and observed strong concordance in both direction and magnitude of differential expression across six major cell types [Pearson correlation coefficient (r) ranging from 0.433 to 0.902] (Fig. 2A). Neuronal cells, including excitatory and inhibitory subtypes, exhibited predominantly down-regulated genes, whereas glial cells—astrocytes, microglia, and oligodendrocytes—exhibited markedly fewer DEGs compared with neurons, consistent with previous observations (*10*). These consistent transcriptomic changes further confirmed the robustness of GAGE-seq in capturing AD-related gene expression alterations. The resulting DEG patterns were stable across saturation and leave-one-donor-out analyses (fig. S3, A to C), indicating that the observed transcriptional differences were not driven by sequencing depth or individual donors.

To understand the functional consequences of these changes, we performed pathway enrichment analysis on DEGs using curated gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (*28*), Reactome (*29*), and previously reported AD-related gene sets (*10*, *30*). Excitatory neurons showed enrichment for multiple AD-relevant signatures (Fig. 2B), including up-regulation of mitochondrial function and mRNA metabolic pathways, consistent with a compensatory stress response, coupled with down-regulation of lipid metabolism, mitochondrial oxidative phosphorylation, mitochondrial aminoacyl-tRNA synthetase activity, and synaptic signaling (table S3), indicative of impaired neuronal and metabolic function. The coexistence of up-regulated mitochondrial and mRNA metabolic pathways with global down-regulation in neurons may reflect compensatory transcriptome remodeling under disease-associated stress. Astrocytes showed up-regulation of oxidative stress [FDR (false discovery rate) = 0.0027], lipid storage and programmed cell death (FDR = 0.046), consistent with reactive glial states and down-regulation of angiogenesis (FDR = 1.9×10^{-12}) (Fig. 2B). Notably, Parkinson's disease-related gene sets were up-regulated in the AD group across multiple cell types, suggesting potential molecular convergence between neurodegenerative conditions (*31*, *32*). Gene sets previously reported to be altered in AD, such as BLALOCK_ALZHEIMERS_DISEASE_UP and BLALOCK_ALZHEIMERS_DISEASE_DOWN (*33*), also showed consistent enrichment across multiple cell types.

Beyond disease-specific pathways, canonical biological processes were also dysregulated across multiple cell types. Genes involved in oxidative phosphorylation, glycolysis, and mRNA splicing were broadly up-regulated in AD, whereas several lipid metabolism–related gene sets and pathways (e.g., free fatty acid–regulated insulin secretion) were consistently down-regulated, underscoring the role of metabolic imbalance in AD (*34*). Additionally, pathways related to nervous system function and neurotransmitter metabolism showed consistent down-regulation in AD, indicating a global decline in neuronal function. Among the down-regulated pathways, calcineurin/NFAT (nuclear factor of activated T cells) signaling—a critical regulator of memory, synaptic plasticity, and astrocyte inflammatory reactivity—was notably suppressed across all cell types (Fig. 2B), consistent with prior studies of AD (*35*, *36*).

Glial senescence is increasingly recognized as a key feature of AD (*37*, *38*). Previous snRNA-seq studies of the AD cortex found that microglia exhibit the highest up-regulation of canonical senescence genes (*39*–*41*). To investigate glial senescence at single-cell resolution, we computed a normalized senescence score using previously established canonical senescence pathway (CSP) markers (*42*) (methods). Microglia showed the highest CSP scores among all cell types, with the largest proportion of senescent cells (Fig. 2C and fig. S3, D to F) and significant up-regulation of key senescence regulators *ETS2* (*41*), *CDKN1A* (*43*, *44*), and *CDKN2D* (*41*, *45*) (Fig. 2D and fig. S3G). Other previously reported senescence regulators, such as *ATM*, *CDKN2A*, and *SERPINE1* (*41*, *46*), were also elevated in microglia, whereas previously reported down-regulated genes associated with senescence, such as *PRKCA*, *TLN2*, and *JAK3*, were down-regulated (*41*) (Fig. 2E). Consistent with gene expression changes, we observed elevated gene-body 3D chromatin contact scores (*24*) for up-regulated senescence markers in glial cells compared with neurons (Wilcoxon rank sum test, $P = 0.05$), with the strongest effects in microglia (fig. S3H). Senescent glial cells are also known to acquire a proinflammatory senescence-associated secretory phenotype (SASP) in AD (*40*, *43*, *47*). Astrocytes in GAGE-seq RNA showed the highest SASP scores, calculated using a curated SASP gene panel (fig. S3I and methods). Spatial transcriptomic data largely recapitulated these patterns in situ, with microglia showing the highest CSP scores and senescent cell proportions (fig. S3, D and J). SASP-associated signatures were more broadly distributed across glial types in spatial data (fig. S3, D and K). Notably, VLMCs and endothelial cells exhibited high CSP and SASP scores and elevated proportions of senescent cells across both modalities (fig. S3, D and I to L), especially in spatial data, consistent with emerging evidence of endothelial senescence in AD (*42*) and pointing to a previously unreported senescent state in VLMCs.

We further examined sex as a biological variable and observed sex-dependent differences in both AD-associated transcriptional dysregulation and large-scale 3D genome remodeling. These results highlight a prominent contribution from X-chromosome regulation, with X-chromosome inactivation (XCI) escape genes showing the strongest sex-dependent shifts in expression and X-linked 3D chromatin features (see supplementary text in the supplementary materials and fig. S4). Together, these results reveal widespread transcriptional remodeling in AD across major brain cell types, marked by loss of neuronal identity and functional integrity, alongside differential engagement of senescence programs, particularly in glial populations, as well as sex-dependent changes.

Global chromatin alterations and compartment mingling in AD

To identify chromatin reorganization at different scales, we first examined the A/B compartment score, a metric derived from principal components analysis of Hi-C contact maps that partitions the genome into transcriptionally active (A) and repressive (B) compartments, across all GAGE-seq samples. We observed modest genome-wide changes between AD and non-AD groups, with overall similar checkerboard patterns in the contact map (fig. S5A) and high correlations of compartment scores between the two groups across major cell types (r = 0.914 to 0.989; fig. S5B), consistent with prior bulk Hi-C studies (*14*).

Previous studies reported increased long-range chromatin interactions with aging in human cerebellum samples (*48*). In our data, this was not observed in the non-AD group (fig. S5C), likely because of the advanced age (all >75 years old) and intentionally narrow age range

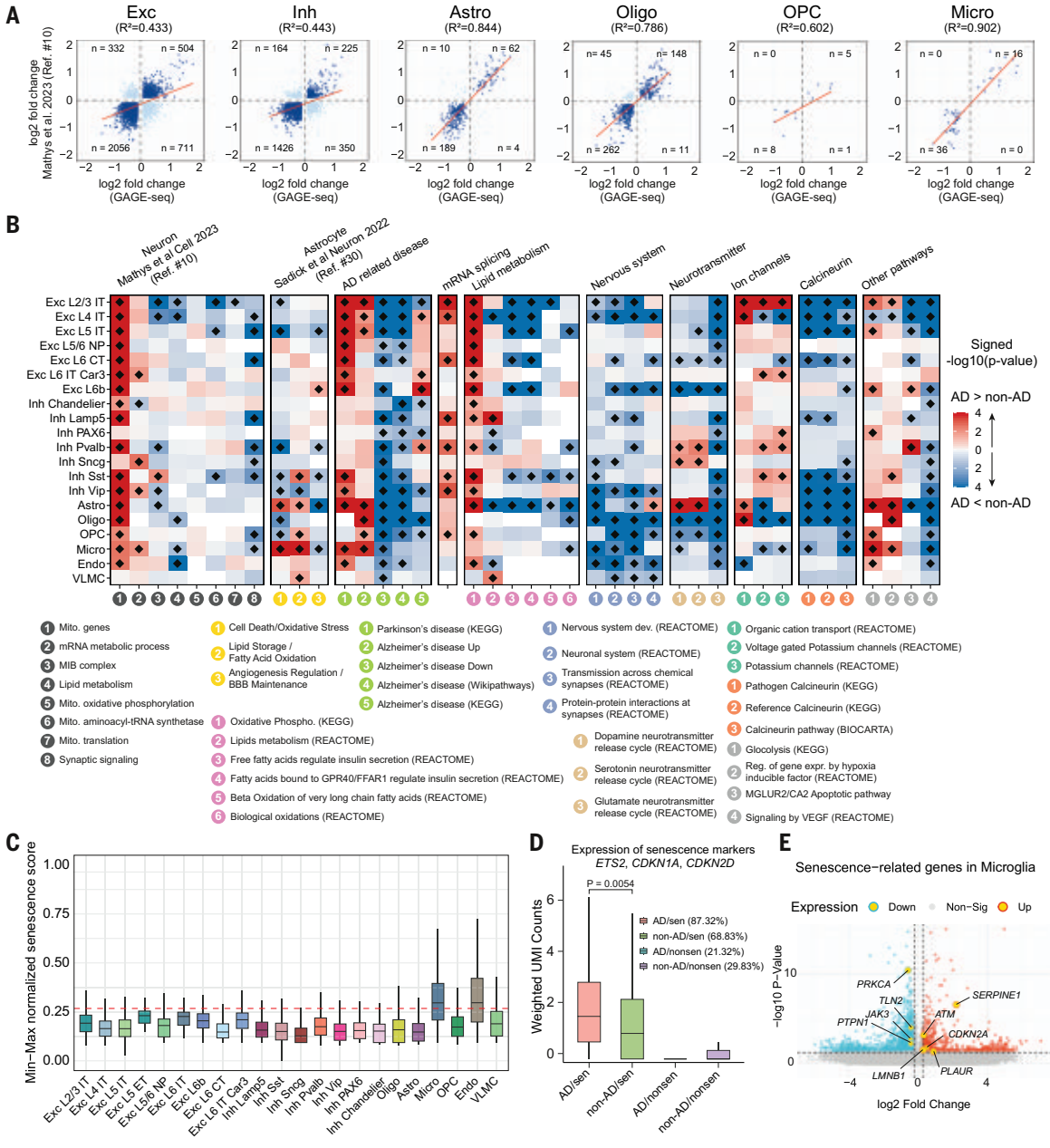


Fig. 2. Transcriptome-wide alterations in AD across brain cell types. (**A**) Cross-study DEG concordance. For each major cell type, we compared AD versus non-AD differential expression $\log_2$ fold changes estimated from our GAGE-seq data (*x* axis) against those reported in Mathys *et al*. (*10*) (*y* axis). Each point is one gene; the red line shows the fitted linear trend, with the corresponding coefficient of determination (R^2) indicated. Numbers in each quadrant report gene counts by sign of fold change in the two studies (concordant versus discordant direction). (**B**) Pathway enrichment analysis of DEGs across 20 annotated cell subtypes using curated AD-related and canonical gene sets. Color scale indicates signed $-\log_{10}$(*P* value); positive values denote enrichment in AD. Black dots indicate adjusted $P < 0.05$. (**C**) Cell type senescence score landscape. Box plot shows per-cell senescence scores (computed from expression of canonical senescence markers) summarized by cell types and min-max normalized across cell types. The dashed line marks the senescence-score cutoff (mean + 2 SD) used to classify senescent cells, highlighting elevated senescence programs in microglia relative to other cell populations. (**D**) Microglial expression of key senescence regulators (*ETS2*, *CDKN1A*, *CDKN2D*). Box plots show aggregated UMI expression per cell, stratified by disease status (AD versus non-AD) and by senescence status [senescent ("sen") versus nonsenescent ("nonsen"); group proportions indicated in the legend]. Expression is significantly higher in senescent microglia from AD donors. Two-sided Wilcoxon rank sum test compares AD versus non-AD within microglia senescent cells. (**E**) Volcano plot of senescence-related gene expression in microglia (AD versus non-AD), showing $\log_2$ fold change versus $-\log_{10}$(*P* value). Notable up-regulated genes include *SERPINE1*, *PLAUR*, and *CDKN2A*; notable down-regulated genes include *PRKCA* and *TLN2*.

of our cohort, which was designed to minimize aging-related confounding effects. However, we observed a consistent and significant increase in the proportion of long-range chromatin interactions in AD compared with non-AD across nearly all cell types examined (Fig. 3A and fig. S5C, right). To further explore this pattern, we empirically binned intrachromosomal distances on a $\log_2$ scale (see methods) and grouped interactions into four categories: short-range (1 to 128 kb), midrange (128 to 4096 kb), long-range (4096 kb to 65.536 Mb), and ultralong (>65.536 Mb). The ultralong group, enriched for interarm and centromere-spanning contacts, was analyzed separately, and combining

both long-range and ultralong interactions did not materially affect the overall trends (fig. S6, A and B). The robustness of the trends was confirmed by power analysis, saturation, leave-one-donor-out, and bootstrapping approaches (fig. S6, C to I). We found that AD samples showed a modest but consistent increase in long-range versus short-range interactions across different chromosomes and cell types (Fig. 3A and fig. S7A). This trend is additionally supported by an increased trans-to-cis ratio of chromatin interactions (Fig. 3A, bottom). Stratification of chromatin interactions into fine-resolution distance groups and contact map visualization further confirmed this observation (Fig. 3, B and C, and fig. S7B).

To explore the mechanistic basis for the observed increase in long-range chromatin interactions, we generated saddle plots of compartment scores to visualize interaction dynamics between A/B compartments. As shown in Fig. 3D, there was a pronounced increase in intrachromosomal mingling (i.e., off-diagonal interactions) between the A and B compartments alongside a modest reduction in A-A and B-B homotypic interactions (i.e., diagonal interactions) (fig. S8A). A similar trend of weakened compartment segregation was observed using interchromosomal interactions (fig. S8B).

To further investigate how these changes in chromatin interaction pattern relate to transcriptional alterations in AD, we plotted the single-cell $\log_2$ ratio of long-range versus short-range interactions against total UMI counts per cell (Fig. 3E and fig. S9, A and B). Across most cell types, particularly neurons, AD cells with elevated long-range interactions exhibited lower UMI counts, suggesting reduced global transcriptional output. Leveraging GAGE-seq's ability to jointly profile chromatin interactions and gene expression in the same cell, we next asked whether A/B compartment shifts correlate with transcriptional changes. Gene sets or individual genes up-regulated in AD generally showed increased compartment scores, suggesting a shift from B to A in single cells (or vice versa) (Fig. 3F and fig. S9C).

We then applied scGHOST (single-cell graph-based Hi-C organization and segmentation toolkit) (*49*) to identify five primary subcompartments (A1, A2, B1, B2, and B3) in each cell, corresponding to distinct chromosome spatial localization in the nucleus (*49*–*51*). Saddle plots of these subcompartments revealed enhanced interactions between active (A1 and A2) and inactive subcompartments (B1, B2, and B3) in AD (fig. S8C), confirming the observations calculated using pseudobulk A/B compartments. On the basis of these annotations, we defined a single-cell A/B compartment mingling score—calculated as the ratio of heterotypic (A-B) to homotypic (A-A and B-B) interactions using interchromosomal interactions alone (see methods)—to quantify chromatin compartment shifts in single cells. AD cells tend to have higher mingling scores and lower UMI counts than non-AD cells (Fig. 3G and fig. S9D). These findings reinforce the existence of a chromatin alteration signature in AD characterized by an increased proportion of long-range interactions and a shift toward greater A/B compartment mingling, accompanied by an overall decrease in transcriptional activity.

We then sought to explore which biological processes might underlie or associate with this AD-specific chromatin compartment shift signature. Although higher mingling scores generally correlated with down-regulated gene expression globally (Fig. 3F and fig. S9D), gene sets involved in biological processes related to oxidative phosphorylation and RNA metabolism exhibited positive correlations with single-cell compartment mingling scores in AD cells (Fig. 3H and fig. S9E). These gene sets were also among the top up-regulated ones. In contrast, housekeeping-like genes (see methods) and neuronal programs—including neurotransmitter signaling, neurodevelopment, and synaptic pathways—showed strong negative correlation with compartment mingling scores (Fig. 3H and fig. S9E). These correlations for both gene sets were stronger in AD than in non-AD cells (Fig. 3I), indicating that AD-associated compartment mingling preferentially associates with up-regulated metabolic and RNA-processing programs while opposing neuronal and housekeeping gene programs.

Taken together, these results suggest that chromatin compartmentalization is globally weakened in AD, with increased long-range interactions and A/B compartment mingling. These chromatin structural changes are associated with decreased transcriptional output and selective dysregulation of disease-relevant pathways (Fig. 3J).

3D genome alterations at cis-regulatory elements in AD

The observed global chromatin reorganization in AD led us to investigate whether candidate cis-regulatory elements (cCREs) associated with chromatin interactions are differentially affected. We integrated GAGE-seq data with snATAC-seq data from the PFC of the same donors (*13*). To stratify chromatin interactions by local regulatory context, we categorized cCREs identified by snATAC-seq into six types according to their proximity to TSSs (transcription start sites) and overlap with bulk CTCF ChIP-seq (chromatin immunoprecipitation sequencing) peaks (see methods).

We then quantified differences in chromatin interactions between AD and non-AD groups, stratified by the presence of specific cCRE types at both anchors of the interaction. Interactions were grouped into three categories: cCRE to cCRE, cCRE to non-cCRE, and non-cCRE to non-cCRE. Notably, chromatin interactions with both anchors at non-cCREs were consistently reduced in AD across both short-range (1 to 128 kb) and midrange (128 to 4096 kb) distances (Fig. 4A and fig. S10A). In contrast, interactions for cCRE to non-cCRE and cCRE to cCRE exhibited more variable patterns. For instance, in glial and other non-neuronal cell types, both short- and midrange cCRE to non-cCRE interactions decreased, whereas in excitatory and inhibitory neurons, midrange interactions increased while short-range interactions declined (Fig. 4A and figs. S10A, S11A, and S12). Among cCRE to cCRE interactions, those marked by CTCF ChIP-seq peaks showed the strongest increase in midrange interaction strength (Fig. 4B; figs. S10B, S11B, and S13A; and methods). These results suggest that whereas short-range chromatin interactions decline globally in AD, midrange cCRE-associated interactions, particularly those involving CTCF, are selectively enhanced. Given CTCF's role in chromatin loop formation, this implicates CTCF-mediated rewiring of regulatory contacts in AD.

To further validate the observed increase in midrange chromatin interactions, we performed loop calling for major cell types (see methods). Aggregated peak analysis (APA) on observed-versus-expected contact maps centered at ±200 kb confirmed a consistent increase in loop signal in AD compared with non-AD (Fig. 4C). Astrocytes, excitatory neurons, and inhibitory neurons showed the most pronounced changes, whereas microglia showed minimal alterations. This trend was consistent regardless of loop size or CTCF presence at anchors and remained robust in leave-one-donor-out analyses (Fig. 4C and fig. S13, B to D). Notably, both the gene expression of loop-extrusion components and the CTCF binding landscape showed little change in AD (fig. S14). Although transcriptional stability does not necessarily imply unchanged protein activity, these results provide no direct evidence for large-scale alterations in loop extrusion, suggesting that the observed compartment mingling may instead reflect weakened compartmental segregation.

To identify transcription factors (TFs) associated with differential chromatin interactions at midrange cCREs, we developed Hi-C ChromVAR, an extension of ChromVAR (*52*) that integrates Hi-C contact frequency with chromatin accessibility to infer TF activities, defined as the bias-corrected enrichment of chromatin interaction strength at genomic bins containing a TF's binding motif within 3D chromatin contexts (see methods). Genomic bins with elevated midrange interactions in AD contained more CREs identified with snATAC-seq data (Fig. 4D) and showed consistently higher motif density (defined as the number of TF motifs annotated per 10-kb bin; see methods) in AD than non-AD (Fig. 4E). We prioritized TFs associated with chromatin alterations in AD by computing Spearman's correlation between TF activity (predicted by Hi-C ChromVAR) and the A/B compartment mingling score across

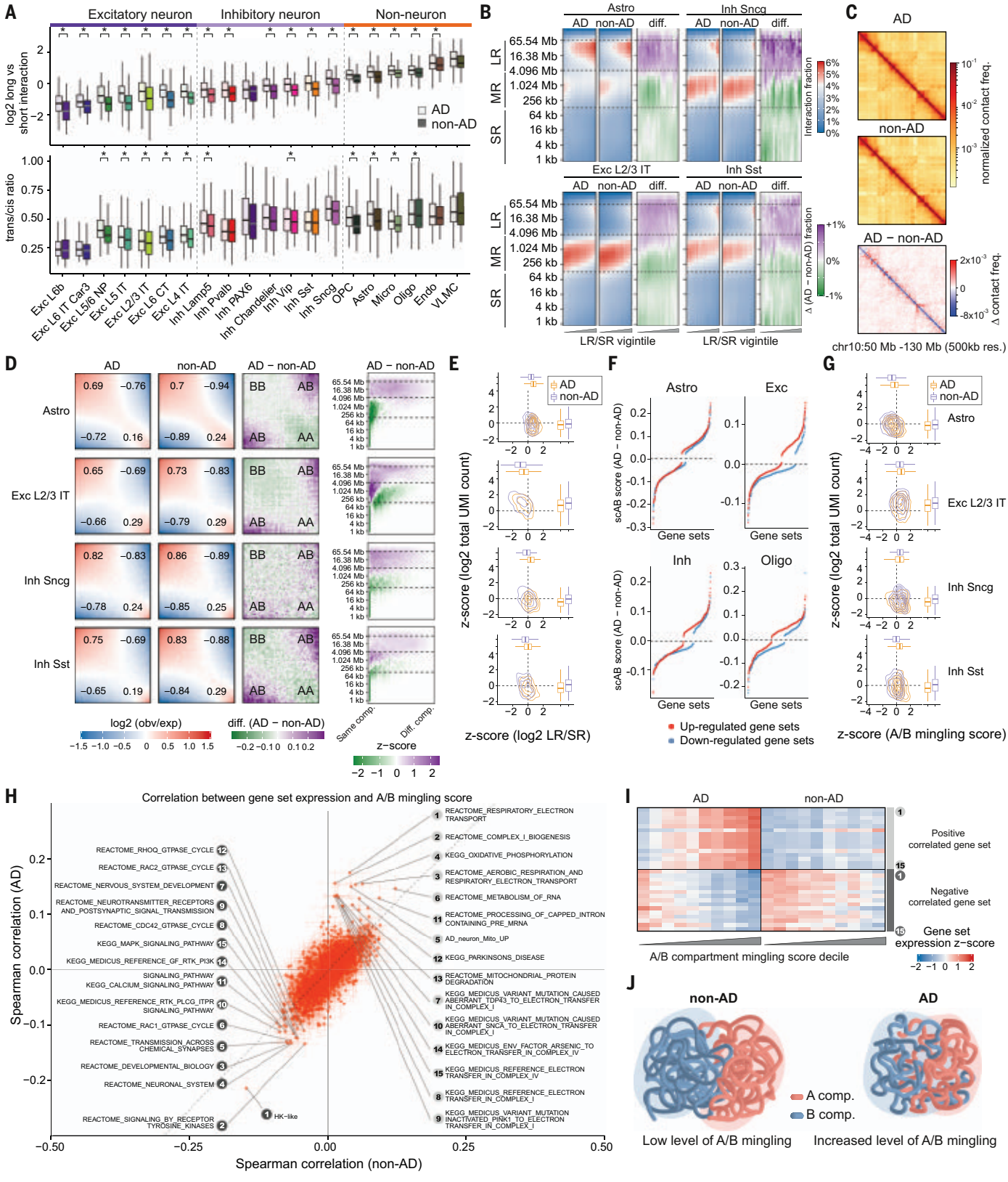


Fig. 3. Global alterations of chromatin interactions and A/B compartment mingling in AD. (**A**) Box plots showing increased $\log_2$ long-range/short-range (LR/SR) interaction ratio (top) and trans/cis contact ratio (bottom) across cell types in AD versus non-AD samples. Each point in the box plot represents a single cell. Two-sided Wilcoxon rank sum test $^*P < 0.05$. (**B**) Heatmaps displaying chromatin interaction distribution in AD, non-AD, and their difference (diff.) across representative cell types. Within each cell type, cells are ordered by increasing LR/SR ratio and grouped into 20 equally sized bins (vigintile). Interaction frequencies are summarized for each bin, with the difference between AD and non-AD shown on the right, highlighting systematic shifts toward longer-range interactions in AD. (**C**) Representative Hi-C contact map (chr10:50 Mb–130 Mb) showing normalized contact frequency in AD and non-AD, and their difference (bottom). (**D**) (Left) Saddle plots showing observed versus expected contact frequencies stratified

by pseudobulk A/B compartment scores (decomposed eigenvector) and divided into 50 bins from B compartment to A compartment. (Right) Heatmap showing *z*-score normalized differences in interaction frequency between AD and non-AD, stratified by compartment score difference (*x* axis) and genomic distance (*y* axis), revealing preferential weakening of short-range interactions and enhanced cross-compartment long-range interactions. (**E**) Contour plots comparing the distribution of LR/SR ratio (*x* axis) versus total UMI count (*y* axis) across AD and non-AD cells. Marginal box plots (top and right) summarize the overall distribution between AD and non-AD cells. (**F**) Single-cell A and B (scAB) compartment score difference between AD and non-AD for up- and down-regulated gene sets. (**G**) As in (E), but with single-cell A/B compartment mingling score on the *x* axis. (**H**) Spearman correlation between gene set expression and single-cell A/B compartment mingling score in AD and non-AD. Each point represents one gene set. Top 15 gene sets with the strongest positive or negative correlation in AD are labeled. (**I**) Heatmap showing gene set expression (*z*-score) across single cells grouped by deciles of compartment mingling score. Gene sets are ordered as in (H), highlighting coordinated transcriptional programs associated with increased compartment mixing in AD. (**J**) Schematic summarizing the global increase in A/B compartment mingling in AD compared with non-AD, illustrating enhanced intercompartment interactions and reduced compartmental segregation.

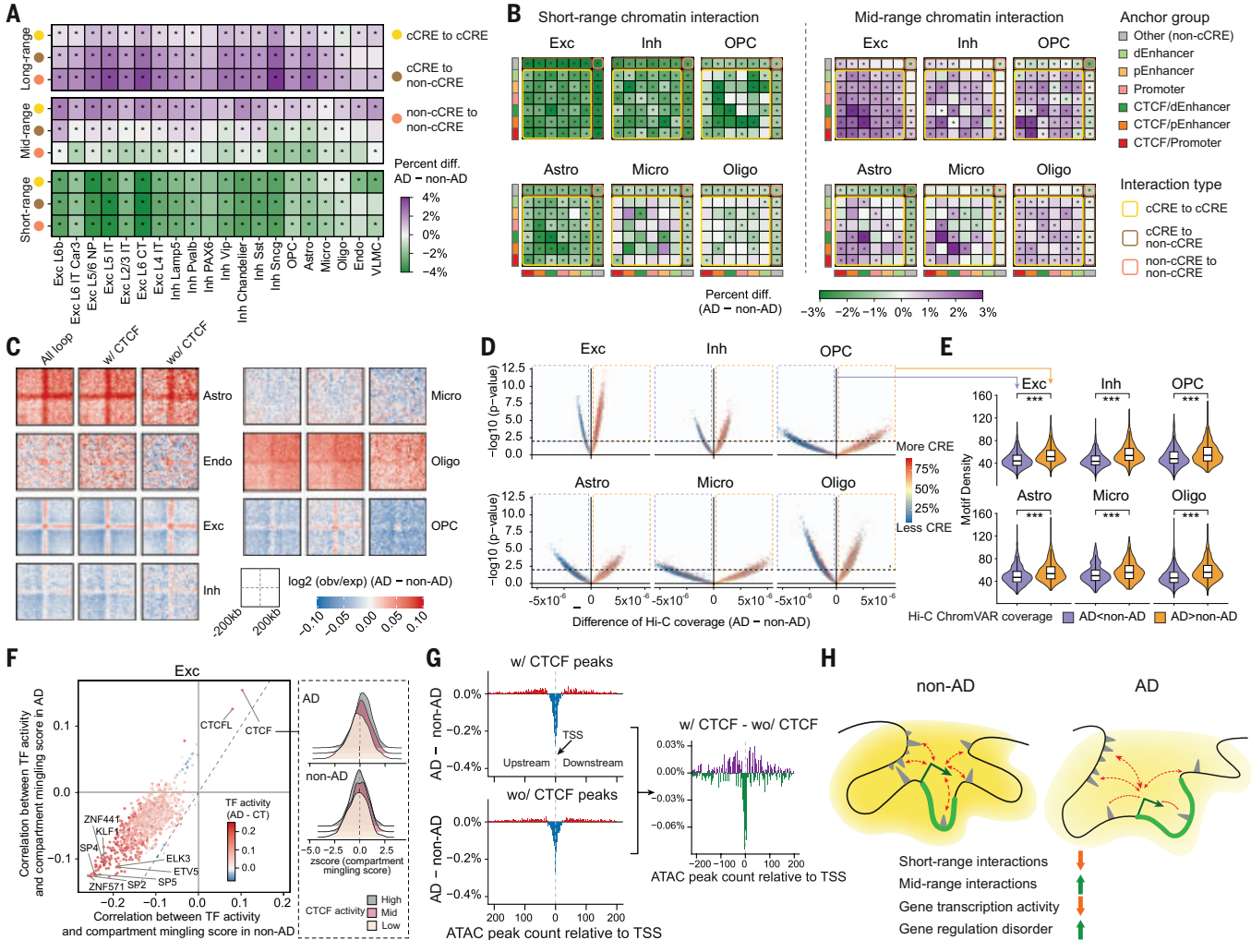


Fig. 4. AD-associated remodeling of cCRE-anchored chromatin interactions. (**A**) Heatmaps showing differences in chromatin interaction frequency between AD and non-AD cells, stratified by the presence of cCRE categories at interaction anchors: non-cCRE to non-cCRE, cCRE to non-cCRE, and cCRE to cCRE. Long-range (top), midrange (middle), and short-range (bottom) interactions are shown separately. Values represent AD minus non-AD differences. Two-sided Wilcoxon rank sum test $*P < 0.05$. (**B**) Heatmaps showing AD versus non-AD differences in chromatin contacts, grouped by finer regulatory elements categories at interaction anchors. Two-sided Wilcoxon rank sum test $*P < 0.05$. (**C**) APA reveals enhanced chromatin loop strength in AD across cCRE-associated interactions. Only midrange chromatin loops were included in the APA analysis. (**D**) Volcano plots showing 10-kb genomic bins with differential midrange interactions between AD and non-AD, highlighting bins enriched for high cCRE density (color scale reflects cCRE count percentiles). (**E**) Violin plots comparing TF motif density in significantly differential bins from (D) between AD and non-AD. Two-sided Wilcoxon rank sum test $***P < 0.001$. (**F**) Scatterplot of TF activity versus compartment mingling score correlations in non-AD (*x* axis) versus AD (*y* axis). Point color indicates change in TF activity; notable TFs, such as CTCF, are labeled. Inset shows distributions of predicted CTCF activity in AD and non-AD. (**G**) (Left) Differences in chromatin interaction frequency from gene TSS to surrounding cCREs (ordered along the *x* axis; within 4096 kb). cCREs are stratified by overlap with CTCF ChIP-seq peaks (top: with CTCF; bottom: without CTCF). (Right) Difference between the two plots on the left. (**H**) Schematic summarizing a model in which AD is characterized by weakened chromatin interactions with local CREs and strengthened midrange interactions, potentially contributing to transcriptional dysregulation and impaired regulatory specificity.

single cells (Fig. 4F, fig. S15A, and methods). Notably, CTCF ranked highest among all TFs in excitatory, inhibitory, and astrocyte cell type, with strong positive correlation with mingling scores. Stratifying cells into quartiles on the basis of predicted CTCF activity revealed that, in AD cells, those with higher CTCF activity also exhibited progressively increased mingling scores, indicating more disordered chromatin and reduced gene expression (Fig. 4F, right panel).

To probe the impact of chromatin reorganization on gene regulation, we analyzed aggregate chromatin interaction decay curves for interactions anchored between gene TSSs and cCREs. In AD, interactions from TSSs to nearby CREs (1st to 20th) decreased, while midrange interactions (20th to 200th) increased (Fig. 4G). This spatial shift in regulatory connectivity, along with weakened interaction separation at topologically associating domain boundaries (fig. S15B), supports a model in which AD cells experience impaired local regulatory interactions and increased mid- to long-range cCRE-gene interactions. These structural alterations likely contribute to widespread transcriptional dysregulation in AD (Fig. 4H), including AD risk genes (fig. S16). These findings highlight a shift in the regulatory architecture of AD cells, marked by weakened local control and aberrant enhancement of midrange chromatin interactions.

Deep learning reveals essential regulatory role of the 3D genome in AD

To investigate how 3D chromatin organization contributes to cell type–specific gene expression in AD, we developed Hicformer, a transformer-based deep learning framework that integrates DNA sequence with multiresolution 3D genome features to predict gene expression. Hicformer operates on pseudo-bulk cell type profiles and captures both local and long-range chromatin interactions (Fig. 5A). Each gene locus (±200 kb around the TSS) is represented by three input modalities: (i) DNA sequence (one-hot-encoded); (ii) cell type–specific 1D Hi-C–derived features, including A/B compartment scores, insulation scores, and gene-body scores; and (iii) 2D local Hi-C contact maps. The model architecture combines a convolutional sequence module, integrated 1D/2D chromatin features, and transformer blocks adapted from Enformer (*53*) (Fig. 5A, fig. S17A, and methods).

Hicformer was trained on 7598 genes (selected for high variability or differential expression) across 26 cell type–condition combinations (13 cell types under two conditions: AD and non-AD). We evaluated model generalization in three held-out scenarios: (i) seen genes in unseen cell types, (ii) unseen genes in seen cell types, and (iii) unseen genes in unseen cell types (methods). In all settings, Hicformer consistently outperformed an Enformer-like sequence-only baseline (Fig. 5B), with performance gains confirmed through ablation studies of Hi-C input features (fig. S17B). Notably, Hicformer maintained strong performance in the most challenging setting (unseen genes in unseen cell types) (Fig. 5B, bottom right), and performance advantages were consistent across all major cell types (fig. S17C).

To understand why 3D chromatin features matter, we analyzed the subset of genes whose expression was much better predicted by Hicformer than by the sequence-only baseline. These genes, especially in AD microglia, were enriched for type I interferon signaling, cytokine production, and regulation of immune response—processes known to be associated with AD (*54*–*56*) (Fig. 5C). We next asked whether AD-associated genes from top pathways strongly correlated with compartment mingling were better predicted by Hicformer. Genes from both positively correlated gene sets (e.g., RNA processing and splicing, protein translation, and amino acid metabolism) and negatively correlated ones (e.g., synaptic signaling, nervous system development) showed improved prediction by Hicformer over the sequence-only baseline (Fig. 5D and fig. S18, A and B). Across cell types, Hicformer also exhibited greater improvement for AD-related DEGs compared with other genes (fig. S17D), indicating that 3D genome features are disproportionately important for predicting AD disease-relevant gene expression. Additionally, we found that genes better predicted by Hicformer were also more enriched for TF binding sites (fig. S17F).

Gradient-based feature attribution analysis revealed that sequence features and insulation scores dominated the prediction in neurons, whereas A/B compartment and gene-body scores contributed more in glia (fig. S17E). In AD cells, model reliance shifted toward 3D chromatin features, with increased attribution from A/B compartment and gene-body scores despite unchanged raw feature values (Fig. 5E and fig. S17, G and H). These findings suggest enhanced regulatory influence of 3D chromatin features on gene regulation in AD.

To evaluate whether 3D genome features capture functionally relevant regulatory changes in AD, we examined *SLC5A11*, a transporter gene recently implicated in oligodendrocyte function and dysregulated in AD (*57*, *58*). Hicformer accurately predicted its up-regulation in AD, concordant with increased gene-body 3D chromatin contact frequency and enhanced local chromatin interactions in the Hi-C maps (Fig. 5F; see additional example in fig. S18C). These results underscore Hicformer's ability to infer gene-level expression shifts mediated by 3D genome remodeling in AD.

We next asked whether Hicformer could prioritize regulatory regions. Using astrocytes—an unseen cell type—as an example, high-activity distal cCREs (i.e., ATAC-seq peak > 2 kb from the TSS; accessibility fell within the upper tertile of peak signal) had broader distributions of *z*-scored gradients, indicating greater regulatory influence on gene expression prediction (Fig. 5G). Notably, the divergence in gradient distributions was more pronounced in AD than non-AD, suggesting that Hicformer effectively prioritizes regulatory elements within AD-associated chromatin landscapes (fig. S17I). We further examined *AQP4*, a key astrocytic water channel involved in glymphatic clearance and neuroinflammation and known to be dysregulated in AD (*59*–*61*). In astrocytes from AD, Hicformer identified a distal cCRE (chr18:26,963,215–26,963,515, ~97.4 kb upstream of the *AQP4* TSS) with high attribution, overlapping open chromatin, a strong CTCF binding site, and a nearby chromatin loop to the *AQP4* promoter region. In silico perturbation of this region reduced predicted *AQP4* expression (Fig. 5H), supporting its potential regulatory importance. Further analysis showed that sequence-only perturbation had little effect, whereas perturbing 3D chromatin features markedly reduced predicted expression (fig. S18D).

We also assessed whether Hicformer captures AD genetic risk by evaluating model gradients and attention at brain cis-eQTLs (cis-expression quantitative trait loci) and AD GWAS (genome-wide association study) loci (supplementary text and figs. S19 and S20). Hicformer preferentially prioritizes variants with stronger fine-mapping or association support and shows gene-specific prioritization at established AD loci, linking 3D-aware regulatory predictions to genetic evidence.

These analyses demonstrate that Hicformer effectively integrates DNA sequence and 3D genome features to predict cell type–specific gene expression. Its interpretability further enables identification of functionally relevant regulatory elements, highlighting the connection between chromatin architecture and transcriptional dysregulation in AD.

Spatially resolved gene programs and in situ alterations of the 3D genome in AD

Although recent spatial transcriptomic studies using Visium and Stereo-seq (spatial enhanced resolution omics-sequencing) have characterized AD-associated gene expression changes and spatial disruptions in brain tissues (*8*, *62*, *63*), the interplay between spatial gene programs, cell-cell communication, and in situ alterations of the 3D genome in AD remains unclear. To address this, we performed spatial transcriptomic profiling on PFC samples from four ROSMAP donors (*11*), two with late-stage AD (AD1 and AD2) and two without dementia (non-AD1 and non-AD2), using the Xenium Prime 5K assay (methods). Donor selection followed the same criteria as in our GAGE-seq analysis, with one donor shared between the Xenium and GAGE-seq datasets (Fig. 1A).

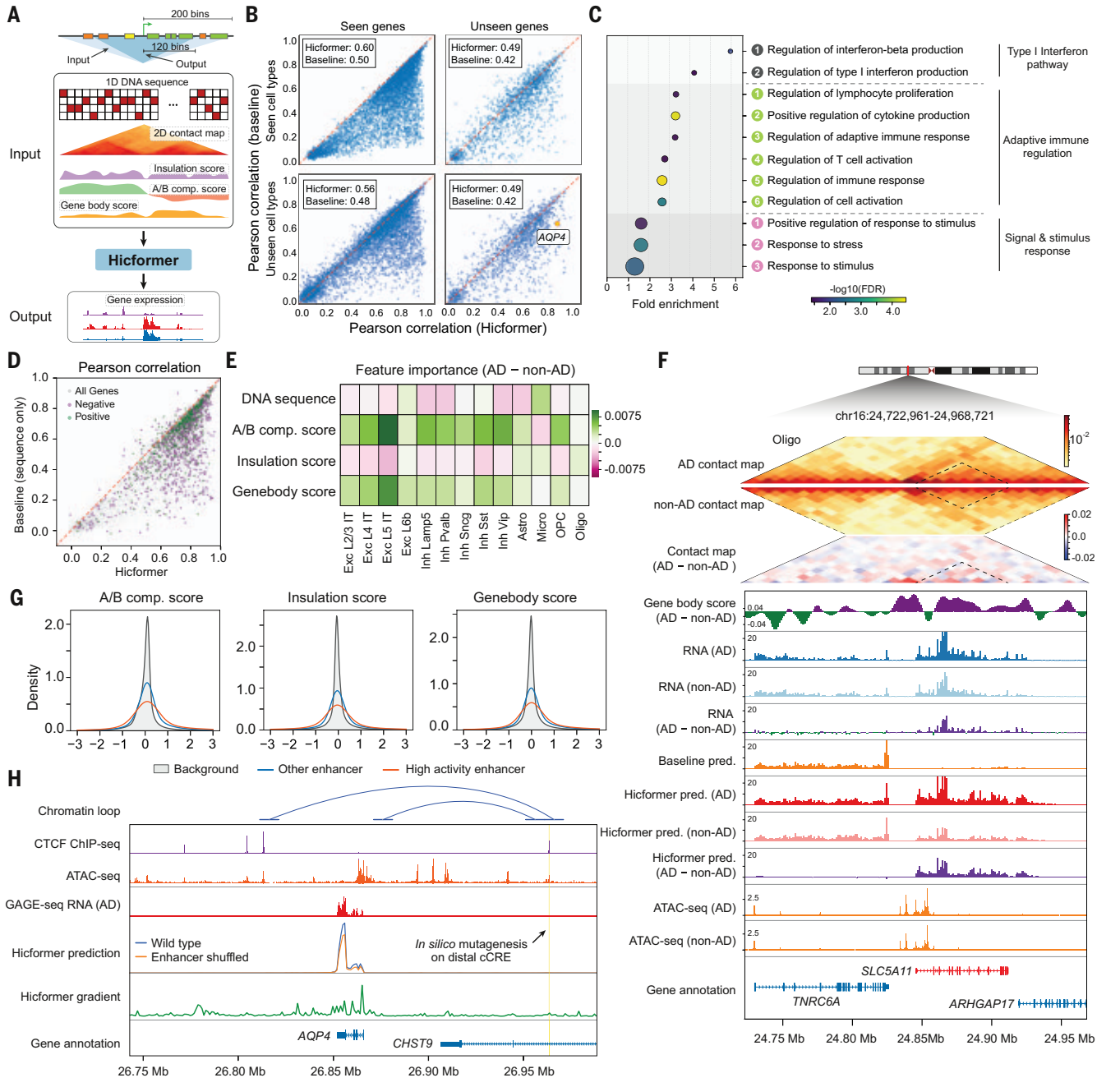


Fig. 5. Hicformer uncovers chromatin-driven transcriptional alterations in AD. (**A**) Schematic of the Hicformer model architecture. Each gene locus (±200 kb around the TSS) is represented by three complementary inputs: one-hot–encoded DNA sequence, 1D Hi-C–derived chromatin scores (A/B compartment, insulation, gene body), and 2D local contact maps. These modalities are jointly integrated to predict cell type–specific gene expression. (**B**) Scatterplots comparing Pearson correlations between observed and predicted gene expression from Hicformer versus a sequence-only baseline across four evaluation settings, stratified by seen versus unseen genes and cell types. Hicformer consistently outperforms the baseline; the AD-relevant gene, *AQP4*, examined in detail in (H), is highlighted. (**C**) Pathway enrichment analysis of genes that are better predicted by Hicformer than by the sequence-only baseline. Enriched pathways are primarily related to interferon signaling, immune activation, and stimulus response. (**D**) Genes from AD-associated pathways linked to altered 3D chromatin structure are better predicted by Hicformer. Highlighted are genes from gene sets with significant positive (green) or negative (purple) correlations with A/B compartment mingling scores. (**E**) Cell type–specific comparison of feature attribution differences between AD and non-AD states. Heatmap shows changes in relative importance of input features (DNA sequence, A/B compartment, insulation score, and gene-body score), revealing increased reliance on 3D chromatin features in AD across multiple cell types. (**F**) Locus-level example illustrating accurate prediction of AD-associated up-regulation of *SLC5A11* in Oligo by Hicformer. Increased predicted expression in AD is accompanied by concordant changes in chromatin accessibility, gene-body contact frequency, and local Hi-C contact intensity. (**G**) Distributions of gradient-based feature attributions in AD astrocytes for A/B compartment score, insulation score, and gene-body score. High-activity distal cCREs exhibit significantly stronger influence on model output compared with other enhancers and background regions (all $P < 0.001$, Levene's test). (**H**) Hicformer prioritizes distal elements through model attribution. Model attribution at the *AQP4* locus identifies a distal cCRE region with high gradient scores, open chromatin, and CTCF binding. In silico perturbation selectively reduces predicted expression, supporting the functional relevance of chromatin architecture.

After quality control, 213,271 cells were retained. Cell types were annotated using canonical marker genes (see methods) and refined through canonical correlation analysis in Seurat with GAGE-seq data (Fig. 6A). Co-embedding showed clear separation of major cell types and subtypes (average Silhouette score: 0.36 for major cell types, 0.29 for subtypes), confirming the transcriptional fidelity of Xenium data. Cortical layers (L1 to L6, white matter) were segmented using layer-specific marker genes from a 10x Visium dataset (*64*) (methods and Fig. 6B).

The proportions of major cell types and subtypes in gray matter were highly consistent across samples (fig. S21, A and B; coefficient of variation: 0.04 to 0.15 for major cell types, 0.04 to 0.35 for subtypes), mirroring results from GAGE-seq (fig. S1). AD-related cell type–specific gene expression changes were also concordant across both platforms (Fig. 6C and fig. S21C; $r = 0.12$ to 0.33 for major cell types and 0.13 to 0.35 for subtypes). Neuronal subtypes generally have a higher number of DEGs and show more down-regulated genes in AD, whereas glial subtypes have fewer DEGs with a more balanced distribution of DEGs in either direction (Fig. 6D), consistent with what we observed in the GAGE-seq data.

To assess spatial microenvironment changes in AD, we quantified neighborhood cell type composition across cortical layers (methods). Spatial shifts in cell type composition were most prominent in deeper cortical layers (L3 to L6), with relatively small changes in superficial layers (L1 and L2) and white matter (fig. S21D). Notably, L4 astrocytes showed reduced spatial proximity to excitatory L4 intratelencephalic (IT) neurons in AD (−63.4% average relative difference in normalized neighborhood cell composition; $P < 1 \times 10^{-15}$, two-sided Wilcoxon rank sum test), suggesting altered neuron-glia coupling or cellular niche disruption in AD in the analyzed individuals (fig. S21E).

To uncover spatially organized gene programs, we applied POPARI (*65*), a new method extending SPICEMIX (*66*), to identify interpretable metagenes and condition-specific spatial affinities (methods). A metagene represents a latent gene program defined as a weighted combination of all measured genes in the spatial transcriptomic sample, with weights normalized to sum to one; genes with higher weights contribute more strongly to the coordinated expression pattern represented by that metagene. POPARI revealed metagenes broadly aligned with cell types (Fig. 6E). For example, metagene *m2* was enriched in microglia, and metagene *m7* was enriched in OPCs (oligodendrocyte precursor cells). A few other metagenes reflected more nuanced cell states, as multiple metagenes may be enriched for a single cell type (e.g., *m5*, *m11*, and *m17* in astrocytes) and some may even associate with multiple cell types (e.g., *m10* across excitatory and inhibitory neurons). Furthermore, several metagenes exhibit cell type–specific AD-related signatures (fig. S22A), organizing AD-associated transcriptional changes into coherent gene programs with distinct spatial distributions across the tissue. Specifically, metagenes *m1*, *m13*, and *m10* were enriched for AD-related DEGs in excitatory L5 and L6 IT neurons; metagenes *m2* and *m16* for microglia; and *m0*, *m4*, *m5*, *m7*, *m19*, and *m11* for oligodendrocytes.

Notably, marker genes strongly correlated with 3D genome compartment mingling were highly enriched in metagene *m15* (fig. S22B). Spatial affinity analysis based on the physical proximity of metagenes in situ revealed several metagene pairs with differential spatial colocalization between AD and non-AD (fig. S22C), including *m0* to *m15*, which showed significantly reduced spatial affinity in AD (Fig. 6, F and G; permutation test $P < 0.05$). Ligand-receptor interactions are known to play key roles in AD progression (*67*, *68*). To explore this further, we extracted ligand-receptor pairs between *m0* and *m15* using CellChat (*69*) annotations and assessed whether cells expressing these genes were spatially colocalized within tissue using bivariate Moran's *I* (methods). These ligand-receptor pairs showed reduced spatial colocalization in AD in both gene expression and A/B compartment scores (Fig. 6, H and I). For example, NCAM1-DPYSL2, a pair involved in cell adhesion (*70*, *71*), neurite elongation and guidance (*72*), and neurodegenerative and neurodevelopmental disorders (*71*, *73*), showed stronger colocalization in non-AD (Fig. 6J). Other pairs, including CNTN2-DPYSL2 and NCAM1-NCAM1, also exhibited greater colocalization in non-AD (fig. S22, D and E) and have been linked to neuronal proliferation, migration, and axon guidance, and thus their disruption may have implications in neurodegenerative disorders (*74*, *75*).

These analyses show that integrating spatial transcriptomics with GAGE-seq enables the identification of spatially coordinated gene programs, shifts in cell-cell communication, and in situ remodeling of 3D genome architecture, providing a framework to study potential chromatin-associated cellular niche reorganization in AD.

Discussion

This study presents a comprehensive single-cell multimodal analysis of the transcriptome and 3D genome organization in the human PFC in AD. Our analyses have uncovered several insights into the 3D genome reorganization associated with AD. In particular, the global shift in chromatin contact patterns and weakened A/B compartmentalization support altered higher-order chromatin organization as an additional layer of regulatory disruption in AD. To further investigate the regulatory impact of 3D genome architecture, we developed Hicformer, a deep learning framework that integrates DNA sequence with both 1D and 2D chromatin features to predict gene expression in a cell type–specific way. Hicformer not only improves predictive accuracy across diverse cell types and conditions but also enables interpretable prioritization of functional regulatory elements. Additionally, integration with spatial transcriptomics places these alterations in tissue context.

Our integrative analyses also provide a translational lens on the regulatory programs associated with 3D genome remodeling in AD. By mapping gene sets linked to compartment mingling onto the current landscape of AD clinical trials and drug targets (figs. S23 and S24 and supplementary text), we find that synaptic and neuronal programs overlap substantially with existing therapeutic focus, whereas metabolic and oxidative-stress programs are comparatively underrepresented. A clear next step is to test whether these chromatin-linked programs can guide patient stratification and nominate repurposing opportunities, including experimental validation of predicted compounds and mechanisms in disease-relevant cellular models.

Several of our key findings were specifically enabled by GAGE-seq. The A/B compartment mingling analysis was only feasible with the joint modalities detected by GAGE-seq. Hicformer further leveraged these matched modalities to predict cell type–specific gene expression. Additionally, GAGE-seq enabled integration with spatial transcriptomics, allowing us to link 3D genome structural features to spatial cell neighborhoods and providing a multiscale view of nuclear and tissue-level organization in AD.

Despite these advances, our study has several limitations. The generation of high-quality single-cell multiomic data for paired RNA and 3D genome profiles from GAGE-seq remains resource intensive, which currently limits sample throughput, although our cohort already reveals reproducible, cell type–specific patterns. Although we provide solid evidence of altered cCRE-associated contacts, a full reconstruction of single-cell regulatory element networks remains an open challenge. Furthermore, the data only represent a static snapshot and correlative associations rather than causality. Future time-resolved studies will be essential to track the progressive remodeling of 3D genome organization during disease progression and to determine whether chromatin compartment shifts precede, coincide with, or follow classical pathological hallmarks. Finally, functional perturbation experiments that manipulate compartment states or chromatin loops are needed to validate the causal role of specific chromatin changes. Nonetheless, the resources and approaches we present here provide a foundational platform for hypothesis generation, mechanistic inference,

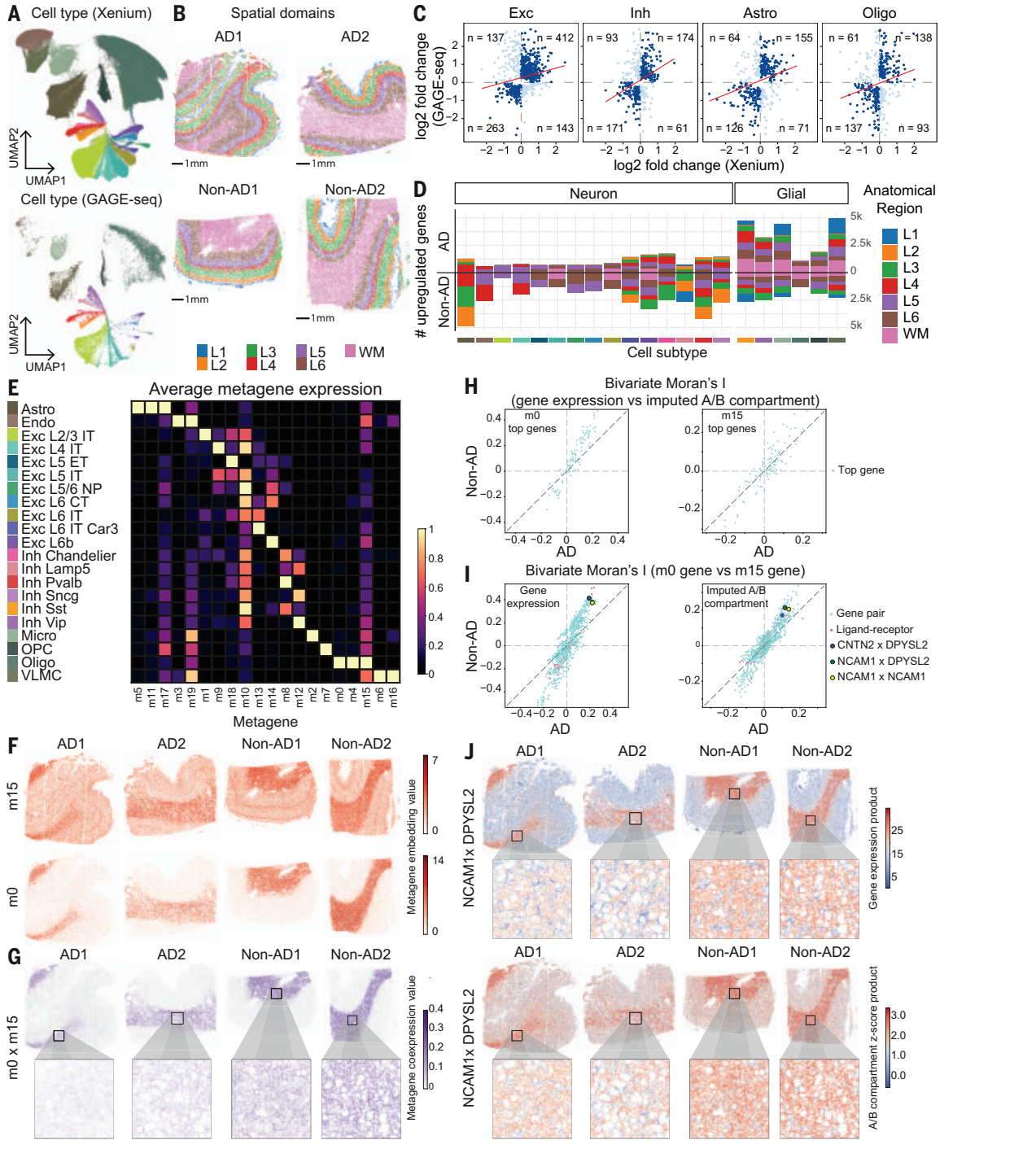


Fig. 6. Spatial transcriptome and 3D genome integrative analysis. (**A**) UMAP of Xenium and GAGE-seq cell types confirms consistent annotation across modalities. The cell type color legend is shown in (E). (**B**) In situ spatial domain identification of cortical layers (L1 to L6, white matter) across AD and non-AD samples (labeled AD1, AD2, non-AD1, and non-AD2) based on layer-specific marker genes. Scale bars: 1 mm. (**C**) Comparisons of differential gene expression (log_2 fold change between AD and non-AD) between Xenium and GAGE-seq across representative major cell types. (**D**) Stacked bar chart showing the number of DEGs (AD versus non-AD) in each cell subtype. The cell type color legend is shown in (E). (**E**) Average metagene expression (POPARI) across annotated cell types. Most metagenes map specifically to distinct cell types, while a few others capture shared programs. (**F**) In situ expression of metagenes *m15* (top row) and *m0* (bottom row). Cells are colored by raw POPARI metagene embedding value. (**G**) In situ coexpression between *m0* and *m15* metagenes across AD and non-AD samples. Spatial edges are colored by the edgewise *m0-m15* coexpression score (see methods for details). The inset highlights a representative 1 mm by 1 mm region of white matter. (**H**) Bivariate Moran's *I* comparing spatial colocalization of gene expression and imputed A/B compartment scores in non-AD versus AD, for genes associated with *m0* and *m15*. (**I**) Bivariate Moran's *I* comparing spatial colocalization of *m0-m15* gene pairs in non-AD versus AD for both gene expression and A/B compartment. (**J**) In situ spatial localization of NCAM1-DPYSL2, a key ligand-gene pair, shows reduced colocalization in AD for both expression and A/B compartment scores. Spatial edges are colored by the corresponding edgewise coexpression score.

and predictive modeling. Integration with additional modalities, such as histone modifications and proteomics, could further enrich our understanding of the regulatory landscape in AD. Spatial transcriptomic extensions, as demonstrated through integration with Xenium data in this work, are particularly valuable for linking chromatin states to tissue architecture and local cell-cell interactions.

This work establishes a high-resolution, multimodal map of transcriptional and 3D genome alterations in AD. Through joint analysis of single-cell transcriptome and chromatin architecture, we identify disease-specific regulatory changes, structural hallmarks, and predictive features of gene dysregulation. These findings offer both insights and a broadly applicable framework for understanding genome organization in neurodegenerative disease.

Materials and methods are available in the supplementary materials.

REFERENCES AND NOTES

1. A. D. Korczyn, L. T. Grinberg, Is Alzheimer disease a disease? *Nat. Rev. Neurol.* **20**, 245–251 (2024). doi: 10.1038/s41582-024-00940-4; pmid: 38424454
2. D. S. Knopman *et al.*, Alzheimer disease. *Nat. Rev. Dis. Primers* **7**, 33 (2021). doi: 10.1038/s41572-021-00269-y; pmid: 33986301
3. P. Scheltens *et al.*, Alzheimer's disease. *Lancet* **388**, 505–517 (2016). doi: 10.1016/S0140-6736(15)01124-1; pmid: 26921134
4. H. Mathys *et al.*, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019). doi: 10.1038/s41586-019-1195-2; pmid: 31042697
5. N. Sun *et al.*, Single-nucleus multiregion transcriptomic analysis of brain vasculature in Alzheimer's disease. *Nat. Neurosci.* **26**, 970–982 (2023). doi: 10.1038/s41593-023-01334-3; pmid: 37264161
6. H. Mathys *et al.*, Single-cell multiregion dissection of Alzheimer's disease. *Nature* **632**, 858–868 (2024). doi: 10.1038/s41586-024-07606-7; pmid: 39048816
7. M. I. Gabitto *et al.*, Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* **27**, 2366–2383 (2024). doi: 10.1038/s41593-024-01774-5; pmid: 39402379
8. P. Wang *et al.*, Molecular pathways and diagnosis in spatially resolved Alzheimer's hippocampal atlas. *Neuron* **113**, 2123–2140.e9 (2025). doi: 10.1016/j.neuron.2025.03.002; pmid: 40168986
9. V. Gazestani *et al.*, Early Alzheimer's disease pathology in human cortex involves transient cell states. *Cell* **186**, 4438–4453.e23 (2023). doi: 10.1016/j.cell.2023.08.005; pmid: 37774681
10. H. Mathys *et al.*, Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385.e27 (2023). doi: 10.1016/j.cell.2023.08.039; pmid: 37774677
11. D. A. Bennett *et al.*, Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* **64**, S161–S189 (2018). doi: 10.3233/JAD-179939; pmid: 29865057
12. G. S. Green *et al.*, Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* **633**, 634–645 (2024). doi: 10.1038/s41586-024-07871-6; pmid: 39198642
13. X. Xiong *et al.*, Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. *Cell* **186**, 4422–4437.e21 (2023). doi: 10.1016/j.cell.2023.08.040; pmid: 37774680
14. R. Nativio *et al.*, The chromatin conformation landscape of Alzheimer's disease. bioRxiv 2024.04.09.588722 [Preprint] (2024); https://doi.org/10.1101/2024.04.09.588722.
15. H. Zheng, W. Xie, The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019). doi: 10.1038/s41580-019-0132-4; pmid: 31197269
16. Y. Zhang *et al.*, Computational methods for analysing multiscale 3D genome organization. *Nat. Rev. Genet.* **25**, 123–141 (2024). doi: 10.1038/s41576-023-00638-1; pmid: 37673975
17. J. Dekker *et al.*, The 4D nucleome project. *Nature* **549**, 219–226 (2017). doi: 10.1038/nature23884; pmid: 28905911
18. J. H. Sun *et al.*, Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* **175**, 224–238.e15 (2018). doi: 10.1016/j.cell.2018.08.005; pmid: 30173918
19. K. Girdhar *et al.*, Chromatin domain alterations linked to 3D genome organization in a large cohort of schizophrenia and bipolar disorder brains. *Nat. Neurosci.* **25**, 474–483 (2022). doi: 10.1038/s41593-022-01032-6; pmid: 35332326
20. J. Bendl *et al.*, The three-dimensional landscape of cortical chromatin accessibility in Alzheimer's disease. *Nat. Neurosci.* **25**, 1366–1378 (2022). doi: 10.1038/s41593-022-01166-7; pmid: 36171428
21. Y. Fujita, S. R. Pather, G.-L. Ming, H. Song, 3D spatial genome organization in the nervous system: From development and plasticity to disease. *Neuron* **110**, 2902–2915 (2022). doi: 10.1016/j.neuron.2022.06.004; pmid: 35777365
22. J.-H. Yang *et al.*, Loss of epigenetic information as a cause of mammalian aging. *Cell* **186**, 305–326.e27 (2023). doi: 10.1016/j.cell.2022.12.027; pmid: 38428398
23. A. Krumm, Z. Duan, Understanding the 3D genome: Emerging impacts on human disease. *Semin. Cell Dev. Biol.* **90**, 62–77 (2019). doi: 10.1016/j.semcdb.2018.07.004; pmid: 29990539
24. T. Zhou *et al.*, GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat. Genet.* **56**, 1701–1711 (2024). doi: 10.1038/s41588-024-01745-3; pmid: 38744973
25. F. J. Garcia *et al.*, Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–899 (2022). doi: 10.1038/s41586-022-04521-7; pmid: 35158371
26. A. C. Yang *et al.*, A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022). doi: 10.1038/s41586-021-04369-3; pmid: 35165441
27. R. Zhang, T. Zhou, J. Ma, Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. *Cell Syst.* **13**, 798–807.e6 (2022). doi: 10.1016/j.cels.2022.09.004; pmid: 36265466
28. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000). doi: 10.1093/nar/28.1.27; pmid: 10592173
29. M. Milacic *et al.*, The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* **52**, D672–D678 (2024). doi: 10.1093/nar/gkad1025; pmid: 37941124
30. J. S. Sadick *et al.*, Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer's disease. *Neuron* **110**, 1788–1805.e10 (2022). doi: 10.1016/j.neuron.2022.03.008; pmid: 35381189
31. S. H. Tan *et al.*, Emerging pathways to neurodegeneration: Dissecting the critical molecular mechanisms in Alzheimer's disease, Parkinson's disease. *Biomed. Pharmacother.* **111**, 765–777 (2019). doi: 10.1016/j.biopha.2018.12.101; pmid: 30612001
32. J. Kelly, R. Moyeed, C. Carroll, D. Albani, X. Li, Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease. *Mol. Brain* **12**, 16 (2019). doi: 10.1186/s13041-019-0436-5; pmid: 30819229
33. E. M. Blalock *et al.*, Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2173–2178 (2004). doi: 10.1073/pnas.0308512100; pmid: 14769913
34. F. Yin, Lipid metabolism and Alzheimer's disease: Clinical evidence, mechanistic link and therapeutic promise. *FEBS J.* **290**, 1420–1453 (2023). doi: 10.1111/febs.16344; pmid: 34997690
35. M. J. Kipanyula, W. H. Kimaro, P. F. Seke Etet, The emerging roles of the calcineurin-nuclear factor of activated T-lymphocytes pathway in nervous system functions and diseases. *J. Aging Res.* **2016**, 5081021 (2016). doi: 10.1155/2016/5081021; pmid: 27597899
36. H. M. Abdul *et al.*, Cognitive decline in Alzheimer's disease is associated with selective changes in calcineurin/NFAT signaling. *J. Neurosci.* **29**, 12957–12969 (2009). doi: 10.1523/JNEUROSCI.1064-09.2009; pmid: 19828810
37. K. E. Prater *et al.*, Human microglia show unique transcriptional changes in Alzheimer's disease. *Nat. Aging* **3**, 894–907 (2023). doi: 10.1038/s43587-023-00424-y; pmid: 37248328
38. X. Li *et al.*, Transcriptional and epigenetic decoding of the microglial aging process. *Nat. Aging* **3**, 1288–1311 (2023). doi: 10.1038/s43587-023-00479-x; pmid: 37697166
39. N. Rachmian *et al.*, Identification of senescent, TREM2-expressing microglia in aging and Alzheimer's disease model mouse brain. *Nat. Neurosci.* **27**, 1116–1124 (2024). doi: 10.1038/s41593-024-01620-8; pmid: 38637622
40. V. Lau, L. Ramer, M.-È. Tremblay, An aging, pathology burden, and glial senescence build-up hypothesis for late onset Alzheimer's disease. *Nat. Commun.* **14**, 1670 (2023). doi: 10.1038/s41467-023-37304-3; pmid: 36966157
41. N. N. Fancy *et al.*, Characterisation of premature cell senescence in Alzheimer's disease using single nuclear transcriptomics. *Acta Neuropathol.* **147**, 78 (2024). doi: 10.1007/s00401-024-02727-9; pmid: 38695952
42. S. K. Dehkordi *et al.*, Profiling senescent cells in human brains reveals neurons with CDKN2D/p19 and tau neuropathology. *Nat. Aging* **1**, 1107–1116 (2021). doi: 10.1038/s43587-021-00142-3; pmid: 35531351
43. P. S. Gross *et al.*, Senescent-like microglia limit remyelination through the senescence associated secretory phenotype. *Nat. Commun.* **16**, 2283 (2025). doi: 10.1038/s41467-025-57632-w; pmid: 40055369
44. K. Jin *et al.*, Brain-wide cell-type-specific transcriptomic signatures of healthy ageing in mice. *Nature* **638**, 182–196 (2025). doi: 10.1038/s41586-024-08350-8; pmid: 39743592
45. Y. Hu *et al.*, Replicative senescence dictates the emergence of disease-associated microglia and contributes to Aβ pathology. *Cell Rep.* **35**, 109228 (2021). doi: 10.1016/j.celrep.2021.109228; pmid: 34107254
46. J. R. Herdy *et al.*, Increased post-mitotic senescence in aged human neurons is a pathological feature of Alzheimer's disease. *Cell Stem Cell* **29**, 1637–1652.e6 (2022). doi: 10.1016/j.stem.2022.11.010; pmid: 36459967
47. C. N. Byrns *et al.*, Senescent glia link mitochondrial dysfunction and lipid accumulation. *Nature* **630**, 475–483 (2024). doi: 10.1038/s41586-024-07516-8; pmid: 38839958
48. L. Tan *et al.*, Lifelong restructuring of 3D genome architecture in cerebellar granule cells. *Science* **381**, 1112–1119 (2023). doi: 10.1126/science.adh3253; pmid: 37676945
49. K. Xiong, R. Zhang, J. Ma, scGHOST: Identifying single-cell 3D genome subcompartments. *Nat. Methods* **21**, 814–822 (2024). doi: 10.1038/s41592-024-02230-9; pmid: 38589516

50. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
51. K. Xiong, J. Ma, Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. *Nat. Commun.* **10**, 5069 (2019). doi: 10.1038/s41467-019-12954-4; pmid: 31699985
52. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706
53. Ž. Avsec *et al.*, Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021). doi: 10.1038/s41592-021-01252-x; pmid: 34608324
54. N. Sun *et al.*, Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386–4403.e29 (2023). doi: 10.1016/j.cell.2023.08.037; pmid: 37774678
55. J. C. Udeochu *et al.*, Tau activation of microglial cGAS–IFN reduces MEF2C-mediated cognitive resilience. *Nat. Neurosci.* **26**, 737–750 (2023). doi: 10.1038/s41593-023-01315-6; pmid: 37095396
56. H. Keren-Shaul *et al.*, A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017). doi: 10.1016/j.cell.2017.05.018; pmid: 28602351
57. M. Lachen-Montes, M. V. Zelaya, V. Segura, J. Fernández-Irigoyen, E. Santamaría, Progressive modulation of the human olfactory bulb transcriptome during Alzheimer´s disease evolution: Novel insights into the olfactory signaling across proteinopathies. *Oncotarget* **8**, 69663–69679 (2017). doi: 10.18632/oncotarget.18193; pmid: 29050232
58. S. Morabito *et al.*, Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021). doi: 10.1038/s41588-021-00894-z; pmid: 34239132
59. J. Wu *et al.*, Requirement of brain interleukin33 for aquaporin4 expression in astrocytes and glymphatic drainage of abnormal tau. *Mol. Psychiatry* **26**, 5912–5924 (2021). doi: 10.1038/s41380-020-00992-0; pmid: 33432186
60. S. R. Rainey-Smith *et al.*, Genetic variation in Aquaporin-4 moderates the relationship between sleep and brain Aβ-amyloid burden. *Transl. Psychiatry* **8**, 47 (2018). doi: 10.1038/s41398-018-0094-x; pmid: 29479071
61. A. Serrano-Pozo *et al.*, Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's disease. *Nat. Neurosci.* **27**, 2384–2400 (2024). doi: 10.1038/s41593-024-01791-4; pmid: 39528672
62. Y. Gong *et al.*, Stereo-seq of the prefrontal cortex in aging and Alzheimer's disease. *Nat. Commun.* **16**, 482 (2025). doi: 10.1038/s41467-024-54715-y; pmid: 39779708
63. E. Miyoshi *et al.*, Spatial and single-nucleus transcriptomic analysis of genetic and sporadic forms of Alzheimer's disease. *Nat. Genet.* **56**, 2704–2717 (2024). doi: 10.1038/s41588-024-01961-x; pmid: 39578645
64. K. R. Maynard *et al.*, Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021). doi: 10.1038/s41593-020-00787-0; pmid: 33558695
65. S. Alam *et al.*, POPARI: Modeling multisample variation in spatial transcriptomics. bioRxiv 2025.05.08.652741 [Preprint] (2025); https://doi.org/10.1101/2025.05.08.652741.
66. B. Chidester, T. Zhou, S. Alam, J. Ma, SPICEMIX enables integrative single-cell spatial modeling of cell identity. *Nat. Genet.* **55**, 78–88 (2023). doi: 10.1038/s41588-022-01256-z; pmid: 36624346
67. C. Y. Lee *et al.*, Characterizing dysregulations via cell-cell communications in Alzheimer's brains using single-cell transcriptomes. *BMC Neurosci.* **25**, 24 (2024). doi: 10.1186/s12868-024-00867-y; pmid: 38741048
68. A. Liu, B. S. Fernandes, C. Citu, Z. Zhao, Unraveling the intercellular communication disruption and key pathways in Alzheimer's disease: An integrative study of single-nucleus transcriptomes and genetic association. *Alzheimers Res. Ther.* **16**, 3 (2024). doi: 10.1186/s13195-023-01372-w; pmid: 38167548
69. S. Jin *et al.*, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021). doi: 10.1038/s41467-021-21246-9; pmid: 33597522
70. U. Rutishauser, A. Acheson, A. K. Hall, D. M. Mann, J. Sunshine, The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. *Science* **240**, 53–57 (1988). doi: 10.1126/science.3281256; pmid: 3281256
71. V. Sytnyk, I. Leshchyns'ka, M. Schachner, Neural cell adhesion molecules of the immunoglobulin superfamily regulate synapse formation, maintenance, and function. *Trends Neurosci.* **40**, 295–308 (2017). doi: 10.1016/j.tins.2017.03.003; pmid: 28359630
72. Y. Wang, T. Ohshima, Unraveling the nexus: The role of collapsin response mediator protein 2 phosphorylation in neurodegeneration and neuroregeneration. *Neuromolecular Med.* **26**, 45 (2024). doi: 10.1007/s12017-024-08814-0; pmid: 39532785
73. M. Ayalew *et al.*, Convergent functional genomics of schizophrenia: From comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* **17**, 887–905 (2012). doi: 10.1038/mp.2012.37; pmid: 22584867
74. A. Zuko *et al.*, Contactins in the neurobiology of autism. *Eur. J. Pharmacol.* **719**, 63–74 (2013). doi: 10.1016/j.ejphar.2013.07.016; pmid: 23872404
75. F. Desprez, D. C. Ung, P. Vourc'h, M. Jeanne, F. Laumonnier, Contribution of the dihydropyrimidinase-like proteins family in synaptic physiology and in neurodevelopmental disorders. *Front. Neurosci.* **17**, 1154446 (2023). doi: 10.3389/fnins.2023.1154446; pmid: 37144098
76. Y. Zhang *et al.*, Single-cell multiomics connects 3d genome and transcriptome alterations in Alzheimer's disease [software], version 1, Zenodo (2026); https://doi.org/10.5281/zenodo.19320621.

ACKNOWLEDGMENTS

The authors thank members of the Ma lab for discussions and T. Zhou for contributions to the initial development of Hicformer. The authors also thank the study participants and staff of the Rush Alzheimer's Disease Center. **Funding:** This work was supported in part by the following grants: National Institutes of Health grants R01HG012303 (J.M., Z.D.), R56HG013092 (Z.D., J.M.), R01HG007352 (J.M.), R21DA061481 (J.M.), R61DA047010 (Z.D.), P30AG10161 (D.A.B.), P30AG72975 (D.A.B.), R01AG17917 (D.A.B.), R01AG015819 (D.A.B.), U01AG072572 (D.A.B.), and U01AG046152 (D.A.B.); National Institutes of Health Common Fund 4DN Program grants UM1HG011593 (J.M.) and UM1HG011586 (Z.D.); and National Institutes of Health Common Fund SenNet Program grant UH3CA268202 (J.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** Conceptualization: H.M., Z.D., J.M.; Methodology: Y.Z., X.L., A.K.K., S.A., J.T., R.Z., H.Z., H.M., Z.D., J.M.; Investigation: Y.Z., X.L., A.K.K., S.A., J.T., R.Z., Shike.W., J.B., W.I., D.J., Sahar.G., Sahel.G., Shihan.W., H.M., Z.D., J.M.; Funding acquisition: H.M., D.A.B., Z.D., J.M.; Supervision: H.M., Z.D., J.M.; Writing – original draft: Y.Z., X.L., A.K.K., S.A., J.T., J.M.; Writing – review & editing: Y.Z., X.L., A.K.K., S.A., J.T., R.Z., Shike.W., D.A.B., H.M., Z.D., J.M. **Competing interests:** Z.D. is an inventor on a provisional patent application (18/733,640) submitted by the University of Washington that covers the GAGE-seq experimental protocol. The other authors declare that they have no competing interests. **Data, code, and materials availability:** GAGE-seq data are available from Synapse in coordination with the ROSMAP project under accession code syn66400203. The data are available under controlled use conditions set by human privacy regulations. To access the data, a data use agreement is required. This registration is in place solely to ensure anonymity of the ROSMAP study participants. This data use agreement can be arranged with either Rush University Medical Center (RUMC) or Sage Bionetworks, which maintains Synapse. The snATAC-seq data used in this study were obtained from https://www.synapse.org/Synapse:syn52293417 (*13*). All code used in this manuscript is available on GitHub at https://github.com/ma-compbio/AD-multiome and through the Zenodo online repository (*76*). ROSMAP resources can be requested at http://www.radc.rush.edu and http://www.synapse.org. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adz1652
Materials and Methods; Supplementary Text; Figs. S1 to S24; Tables S1 to S6; References (*77*–*105*); MDAR Reproducibility Checklist

Submitted 23 May 2025; accepted 5 April 2026

10.1126/science.adz1652

RESEARCH ARTICLE SUMMARY

Epigenetic and 3D genome reprogramming during the aging of human hippocampus

Nathan R. Zemke *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adt8307

INTRODUCTION: Aging is the strongest risk factor for neurodegenerative disease and is accompanied by memory decline and chronic inflammation in the brain. The hippocampus, a region essential for learning and memory, is particularly vulnerable to aging. Prior studies have identified age-related shifts in gene expression, including increased inflammatory signaling and reduced synaptic function, implicating transcriptional dysregulation in brain aging. However, gene expression alone provides an incomplete view. It remains unclear how epigenetic regulation and higher-order genome organization change across cell types and how these changes drive hippocampal dysfunction. Defining these molecular alterations is critical for understanding cognitive decline and disease susceptibility.

RATIONALE: To address this, we profiled gene expression and multiple layers of epigenetic regulation, including chromatin accessibility, DNA methylation, and three-dimensional (3D) genome organization at single-cell resolution across the adult lifespan. This integrated multiomic approach enables direct assessment of how epigenetic and structural genome changes shape transcriptional programs in aging brain cells, and reveals coordinated, cell type–specific mechanisms not captured by single-modality analyses.

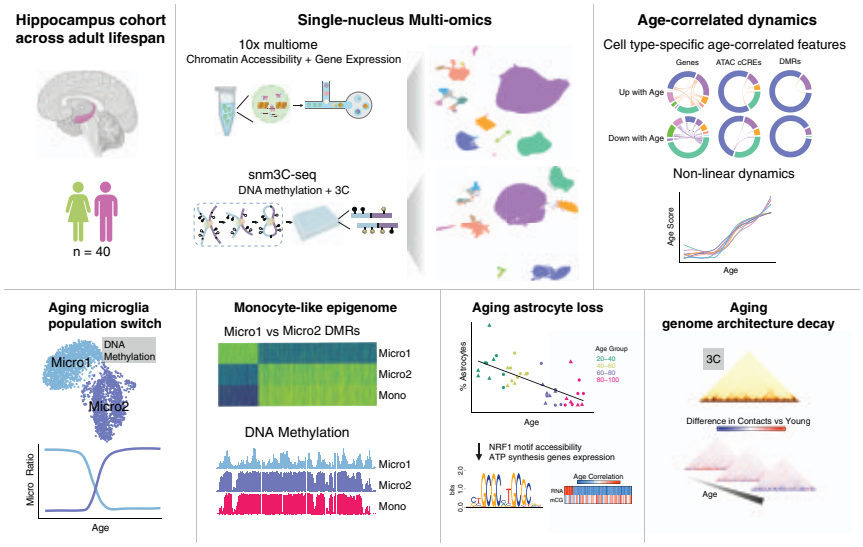
RESULTS: Aging is associated with coordinated and often nonlinear changes in gene regulation across cell types, with a major transition occurring around midlife. A notable finding was a remodeling of the brain's immune cell landscape. Microglia underwent a nonlinear transition in which embryonically derived, brain-resident microglial cells were progressively replaced by a population with epigenetic features resembling blood-circulating monocytes. This transition was not readily detectable using gene expression alone but was revealed by DNA methylation, which preserves cellular lineage. These monocyte-like microglia exhibited epigenetic, transcriptional, and 3D genome features associated with proinflammatory programs, suggesting a potential driver of age-related neuroinflammation.

In parallel, there was a marked decline in astrocytes, including those that support synaptic signaling and maintain the blood-brain barrier. This loss was accompanied by reduced expression of genes involved in mitochondrial energy production and increased cellular stress signatures, indicating metabolic dysfunction as a potential contributor to astrocyte attrition.

At the genome architecture level, aging was marked with a global weakening of 3D genome organization across cell types, including reduced integrity of topologically associating domains and increased trans chromosomal interactions. These structural changes coincided with epigenetic alterations and shifts in gene expression, reflecting widespread rewiring of regulatory programs.

CONCLUSION: Aging reprograms both the cellular composition and regulatory architecture of the human hippocampus. Replacement of embryonically derived microglia with a proinflammatory, monocyte-like population, together with astrocyte loss and global alterations of genome organization, defines key hallmarks of brain aging. These coordinated changes provide a mechanistic framework for age-related memory decline and increased vulnerability to neurodegenerative disease. □

Corresponding authors: Bing Ren (bren@nygenome.org); Xiangmin Xu (xiangmix@hs.uci.edu); Nathan R. Zemke (nzemke@health.ucsd.edu) Cite this article as: N. R. Zemke *et al.*, *Science* **393**, eadt8307 (2026). DOI: 10.1126/science.adt8307



Multimodal single-cell profiling of the aging human hippocampus. Single-nucleus analysis of gene expression, chromatin accessibility, DNA methylation, and 3D genome architecture across 40 adults reveals cell type–specific regulatory genes, nonlinear aging dynamics, and a shift from embryonic-derived to monocyte-like microglia. Aging is also marked by astrocyte loss, reduced mitochondrial function, and a global erosion of 3D genome organization.

RESEARCH ARTICLE

Epigenetic and 3D genome reprogramming during the aging of the human hippocampus

Nathan R. Zemke[1,2]†*, Seoyeon Lee[1]†, Sainath Mamde[1]†, Bing Yang[1,2], Nicole Berchtold[3,4], B. Maximiliano Garduño[3], Hannah S. Indralingam[1,2], Weronika M. Bartosik[1,2], Pik Ki Lau[1,2], Keyi Dong[1,2], Emily Hsu[2,5], Amanda Yang[1], Yasmine Tani[1], Chumo Chen[1], Qiurui Zeng[1], Varun Ajith[3], Liqi Tong[3], Chanrung Seng[6], Daofeng Li[6], Ting Wang[6,7], Jingtian Zhou[8,9], Joseph R. Ecker[9,10], Christopher K. Glass[1,11], Carl W. Cotman[12,13], Xiangmin Xu[3,14]*, Bing Ren[1,15,16]*

Changes in gene expression have been observed in the aging human brain, but our understanding of the underlying regulatory mechanisms remains limited. To unravel these complexities, we analyzed single-nucleus gene expression, chromatin accessibility, DNA methylation, and three-dimensional (3D) chromatin architecture from human hippocampal tissues spanning the adult lifespan. We identified both linear and nonlinear dynamic gene regulatory programs during aging. Between the ages of 50 to 75, embryonic yolk sac–derived microglia were depleted and replaced by cells resembling peripheral blood monocyte–derived microglia. Hippocampal astrocytes decreased substantially with age, including those regulating synaptic transmission. Across cell types, 3D genome architecture underwent global erosion. Our analysis provides insights for how altered gene regulatory programs promote cell type–specific aging phenotypes in the human brain.

Aging is the greatest risk factor for neurodegenerative diseases (*1*). As the US population older than 65 is projected to increase from 62 million in 2025 to 84 million by 2054 (*2*), understanding how the normal aging process promotes disease has become increasingly urgent. Aging in the human brain is accompanied by physiological and cognitive decline (*3*, *4*), often diminishing quality of life.

The hippocampus plays a critical role in learning, episodic memory, emotional regulation, and spatial navigation—functions that decline with age and are associated with hippocampal atrophy (*3*–*5*). Transcriptomic studies of the aging human brain have consistently reported increased expression of immune response genes, suggesting chronic neuroinflammation, alongside decreased expression of genes involved in synaptic connectivity and plasticity (*6*–*10*), implying that transcriptional dysregulation may drive age-related synaptic dysfunction rather than neuronal loss. Single-cell studies in mammals further reveal cell type–specific aging effects, including altered inflammatory and neuronal gene expression, epigenetic remodeling, and three-dimensional (3D) genome restructuring (*11*–*16*). Despite the central role of the hippocampus in cognition, a detailed investigation of single-cell epigenomics and 3D genome architecture in the aging human hippocampus is still lacking.

We present a comprehensive multiomic analysis of the aging human hippocampus. We uncovered age-associated changes that were regulated in a coordinated manner across different layers of epigenetic mechanisms, as well as changes only captured by individual modalities. Our data showed that age-associated epigenomic remodeling strongly influences transcriptional activity. DNA methylome profiling revealed that, with age, embryonically derived microglia were largely replaced by a population of microglia that epigenetically resemble blood monocytes. These potential monocyte-derived microglia possessed 3D genome and epigenomic features associated with inflammation. Concurrently, astrocytes and endothelial cells, critical for blood-brain barrier integrity, declined with age. Our data showed that age-related alterations in the epigenome and 3D genome architecture considerably affect the aging transcriptome, with erosion of genome topology accompanied by a decline in DNA-binding of CTCF, which together with cohesin defines the topologically associating domains (TADs) through the loop extrusion process (*17*–*20*). Our findings provide insights into regulatory mechanisms contributing to age-related neuroinflammation and synaptic dysfunction and provide a framework for dissecting complex aging phenotypes at single-cell resolution.

Integrative single-nucleus multiomic analysis identifies cell type–specific aging gene regulatory programs

We carried out our study using hippocampal tissues from 40 postmortem neurotypical human brains (Fig. 1A and table S1). To achieve an age- and sex-balanced cohort across the adult human lifespan, we included five males and five females, from each of four age groups: 20 to 40, 40 to 60, 60 to 80, and 80 to 100 (Fig. 1A and table S1). To identify age-associated dynamics in the transcriptome, epigenome, and 3D genome organization of each cell type, we utilized two single nucleus (sn) multiomic sequencing (seq) assays: 10x multiome (10x Genomics) (Fig. 1B and fig. S1A) and sn methyl-3C sequencing (snm3C-seq) (*21*), also known as Methyl-HiC (*22*) (Fig. 1C and fig. S1B). 10x multiome generates both snRNA-seq and sn assay for transposase-accessible chromatin with high-throughput sequencing (snATAC-seq) data, whereas snm3C-seq generates both DNA methylome and chromosome conformation capture (3C) data from single nuclei.

After filtering, we retained 295,033 quality nuclei for in-depth analysis of their transcriptomes and chromatin accessibility (Fig. 1B and fig. S1, A, C to F). Using known marker gene expression (fig. S2A) and reference mapping to a published dataset (*23*), we annotated clusters corresponding to 18 cell types (Fig. 1B and fig. S2, B and C), including eight non-neuronal, four excitatory neuronal, and six inhibitory neuronal types (table S2). Compared with non-neurons, neurons had more transcripts detected (mean = 16,419 versus 4459), genes expressed (mean = 4606 versus 1800), and accessible chromatin fragments (mean = 16,394 versus 8352) (fig. S2, G to I). In parallel, we performed snm3C-seq on 40 donors, with 39 being the same as those used in 10x multiome experiments (table S1). We next mapped DNA methylomes and 3D genome organization from 22,240 nuclei (fig. S1, B, G and H) and annotated 13 cell types (Fig. 1C and table S2) using fraction of methylation of CG dinucleotides (mCG) at known marker genes and integration with our 10x multiome data (fig. S2, D to F). Consistent with other reports (*24*, *25*), we observed higher levels of non-CG methylation

[1]Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, La Jolla, CA, USA. [2]Center for Epigenomics, University of California, San Diego School of Medicine, La Jolla, CA, USA. [3]Department of Anatomy and Neurobiology, University of California, Irvine School of Medicine, Irvine, CA, USA. [4]Immunis Inc., Irvine, CA, USA. [5]Department of Biochemistry and Molecular Medicine and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [6]Department of Genetics, The Edison Family Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO, USA. [7]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA. [8]Arc Institute, Palo Alto, CA, USA. [9]Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA. [10]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA, USA. [11]Department of Medicine, University of California, School of Medicine, La Jolla, CA, USA. [12]Department of Neurobiology and Behavior, University of California, Irvine, CA, USA. [13]Department of Neurology, University of California, Irvine School of Medicine, Irvine, CA, USA. [14]The Center for Neural Circuit Mapping, University of California, Irvine, CA, USA. [15] New York Genome Center, New York, NY, USA. [16]Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, USA. ***Corresponding author. Email: bren@nygenome.org (B.R.); xiangmix@hs.uci.edu (X.X.); nzemke@health.ucsd.edu (N.R.Z.)** †These authors contributed equally to this work.

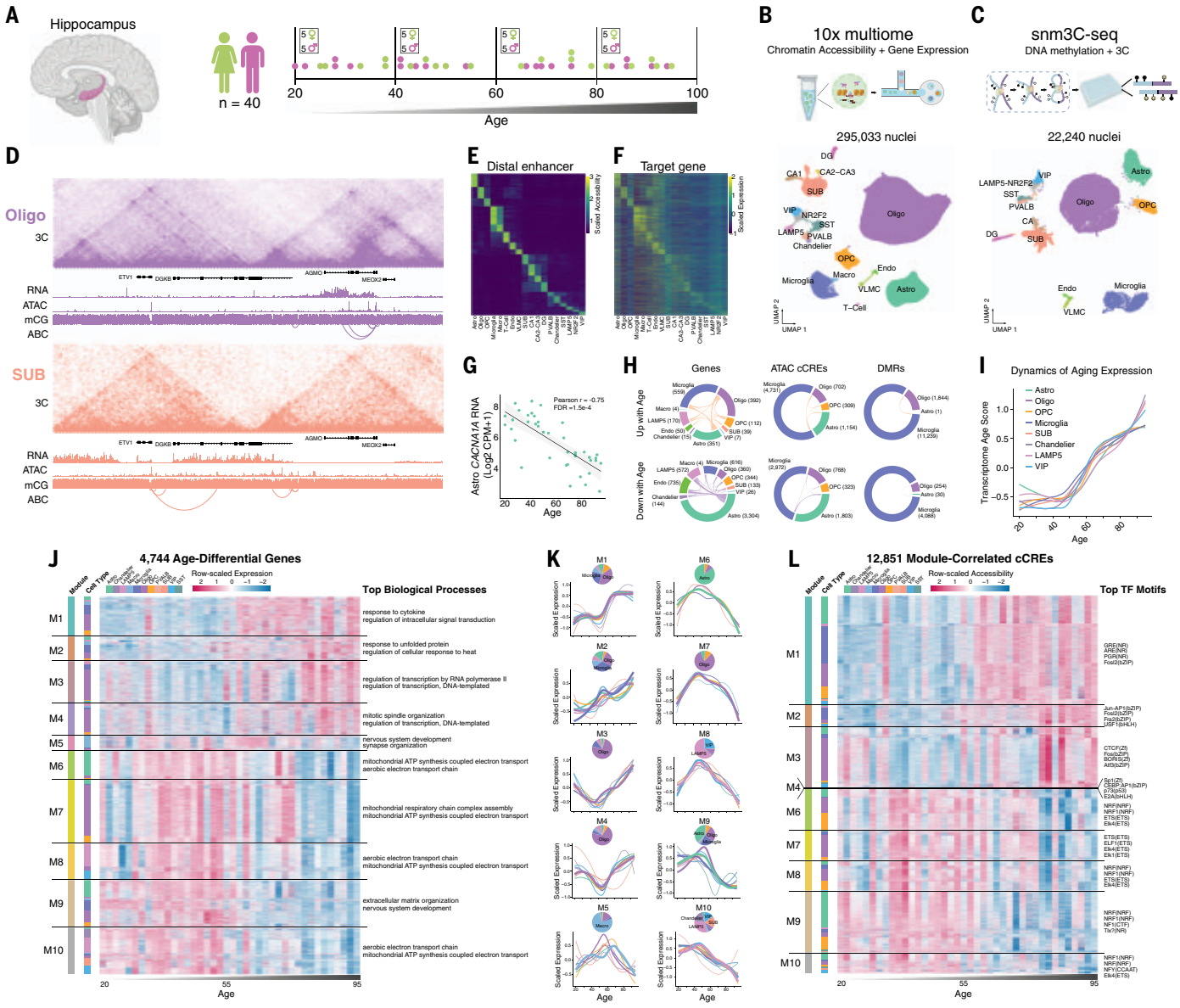


Fig. 1. Cell type–specific gene regulatory programs in the aging human hippocampus. (**A**) Diagram of the human hippocampus, generated with BioRender. (**B** and **C**) Workflow diagrams (made with BioRender) and UMAP embeddings of 10x multiome RNA-seq (B) and snm3C-seq DNA methylation (C) clustering, colored and annotated by cell type. DG, dentate gyrus; CA, cornu ammonis; SUB, subiculum; OPC, oligodendrocyte precursor cells; VLMC, vascular leptomeningeal cells. (**D**) WashU Epigenome Browser screenshots displaying 3C contact matrices, RNA expression, chromatin accessibility (ATAC), DNA methylation (mCG), and ABC-predicted enhancer-promoter links (ABC) for Oligo (top) and SUB (bottom). (**E** and **F**) Heatmaps of row-scaled Log2(CPM+1) for chromatin accessibility (E) or gene expression (F) for all ABC-predicted enhancers. For each predicted enhancer, the target gene with the highest ABC score is displayed. Rows are clustered by cell type with the highest CPM value of chromatin accessibility. (**G**) Scatter plot of *CACNA1A* expression versus donor age in astrocytes. (**H**) Circular plots showing the number of significant (FDR < 0.1) age-correlated genes (left), chromatin-accessible cCREs (center), and DMRs (right) in each subclass. Each segment represents the proportion of features in a cell subclass. Inner connecting lines indicate features shared between subclasses. (**I**) Transcriptome age score is average z-scores of CPM of age-correlated genes across donors for indicated cell types. Negatively age-correlated genes had z-scores multiplied by −1 before averaging. The line represents LOESS fit curve. (**J** and **K**) Heatmap showing scaled expression of 4744 age-correlated genes grouped into 10 modules identified by k-means clustering across donors and cell types. Top enriched biological GO terms are indicated for each module (J). Stacked bars in (J) and the pie chart in (K) represent the number of significantly age-correlated genes contributed by each cell type to each module. The line plot shows mean scaled expression for each cell type across age; thickness represents relative cell type contribution to the module (K). (**L**) Heatmap showing scaled chromatin accessibility of cCREs grouped into 10 modules across donors and cell types. Stacked bars represent module assignment and contributions from each cell type. Top enriched TF motifs are indicated for each module.

(mCH) from neurons than from glia (fig. S2K). Together, these data provided cell type–resolved transcriptomes, chromatin-accessible profiles, DNA methylomes, and 3D genome profiles in hippocampal cells across the adult lifespan (fig. S2, G to M). These data are available for visualizing on the WashU Epigenome Browser (*26*): https://epigenome.wustl.edu/seahorse/.

For each cell type, we identified accessible chromatin regions as candidate cis-regulatory elements (cCREs) and differentially methylated regions (DMRs), revealing a total of 472,859 cCREs and 940,687 DMRs (tables S3 and S4). Overall, 58% of the accessible cCREs overlapped with a DMR, with the greatest overlap observed within the same cell type (fig. S2N). To predict target genes for putative enhancers, we

applied chromatin accessibility and chromatin contact data from each cell type to the activity-by-contact (ABC) model (*27*). This yielded 75,177 pairs of putative enhancer-target genes (table S5) that display largely matching cell type–specific activities across cell types (Fig. 1, D to F).

Age-correlated gene regulatory programs

We next asked which genes and epigenomic features showed alterations during aging. We calculated Pearson correlation coefficients (PCC) for gene or epigenetic activity versus donor age. As one example, expression of the calcium channel subunit gene, *CACNA1A*, was negatively correlated with astrocyte aging (Fig. 1G). We identified genes (6478 unique), chromatin-accessible cCREs (12,474 unique), and DMRs (17,355 unique) that had either positive or negative age-correlated activity in each cell type (Fig. 1H and tables S6, S7, and S8). The age-correlated genes, cCREs, and DMRs showed similar nonlinear dynamics across cell types during aging, with gene expression and chromatin accessibility changing around age 50 in both sexes (Fig. 1I and fig. S3, A to C), an inflection point also reported in a recent study on the aging proteome in humans (*28*). Enriched biological processes for age-correlated genes were mostly cell type–specific (fig. S3, D and E). For example, genes that were positively age-correlated in microglia showed enrichment for immune response terms, including regulation of phagocytosis, response to cytokines, and response to interferon-gamma (fig. S3D), and likely contribute to neuroinflammation of the aging hippocampus (*26*). Notably, subiculum (SUB) excitatory neurons, LAMP5-expressing inhibitory neurons, Chandelier inhibitory neurons, and astrocytes showed reduced expression of mitochondrial ATP synthesis genes with age (fig. S3E). Endothelial cells had reduced expression of genes involved in the maintenance of the blood-brain barrier, such as *CDH5* and *CLDN5*, which may contribute to age-related blood-brain barrier permeability (*29*). Though subtle, genes and cCREs had considerably stronger correlation with age in male donors than female donors in almost every cell type (fig. S3, F and G), suggesting possible differential aging hippocampal phenotypes between sexes.

Nonlinear dynamics of age-related gene regulatory programs

As recent reports have highlighted the nonlinearity of aging dynamics (*28*, *30*), we took an unbiased approach to capture all major patterns of age-related gene expression. We performed differential gene expression analysis between every pairwise combination of the four 20-year age groups, and organized these genes (n = 4744, table S9) into 10 modules using k-means clustering for expression patterns across donors (Fig. 1J). These modules demonstrated widespread nonlinear dynamics of aging gene expression (Fig. 1K). Most modules had contributions from multiple cell types, although the cell type contributions proportionally varied between modules (Fig. 1, J and K), highlighting cell type–specificity of age-related gene expression. modules 1 to 4 contained genes with higher activation in aged donors and were enriched for cytokine response and response to unfolded proteins (Fig. 1J). Modules containing genes down-regulated in the aged donors were enriched for mitochondrial ATP synthesis (Fig. 1J), consistent with age-correlated genes (fig. S3E).

To identify the putative transcriptional drivers of the genes in each module (Fig. 1, J and K), we analyzed cCREs with age-matched chromatin accessibility patterns (Fig. 1L and table S10). The transcription factor (TF) motifs most enriched in cCREs of module 1 corresponded to nuclear receptors (NRs), such as the glucocorticoid response element (GRE) (Fig. 1L). This suggests that during aging, NRs may have activated genes involved in cytokine responses (Fig. 1J). The module 2 genes were enriched for responses to unfolded proteins (Fig. 1J) and were putatively regulated by AP-1 family TFs (Fig. 1L), known effectors of the unfolded protein–induced stress response (*31*). Notably, accessibility at cCREs containing AP-1 family member FOS motifs was age-correlated in multiple cell types (fig. S4A). FOS is known to contribute to neuroinflammation through activation of cytokines (*32*) and can also function as a synaptic immediate early gene in specific neuronal populations to promote learning and memory (*33*). In glia and excitatory neurons, cCREs with FOS motifs had increased accessibility with age, whereas cCREs with FOS motifs in inhibitory neurons had decreased accessibility with age, most notably in somatostatin-expressing (SST) inhibitory neurons (fig. S4A). Consistent with FOS-triggering neuroinflammation, the target genes for putative enhancers with FOS motifs in astrocytes were enriched for cytokine stimulus (fig. S4B). By contrast, putative enhancers with FOS motifs in SST neurons targeted genes enriched for neuronal ion channel clustering and receptor internalization (fig. S4B), such as *ATP1B1* and *RASGRF2*, which had decreased expression with age (fig. S4C). These results suggest that during aging, increased activity of FOS may contribute to neuroinflammation from certain cell types such as astrocytes whereas decreased activity of FOS in inhibitory neurons may contribute to synaptic dysfunction.

Nuclear respiratory factor (NRF) motifs were enriched in cCREs from modules containing ATP synthesis genes that were down-regulated with age (Fig. 1, J to L). Consistent with this finding, NRF1 is a known activator of nuclear genes with mitochondrial functions (*34*). Our analysis demonstrates widespread nonlinear changes of gene regulatory programs in the aging human hippocampus.

Aging microglia epigenome promotes neuroinflammation

Chronic activation of microglia is a major contributor of age-related neuroinflammation (*35*). Consistently, we found genes activated with age in microglia to be enriched for immune response terms such as regulation of phagocytosis, response to cytokines, and response to interferon-gamma (Fig. 2A). In addition, microglia of aged donors had larger somas (Fig. 2, B and C, and fig. S5), a morphological trait of activated microglia (*36*). Microglia also had relatively large numbers of age-correlated accessible cCREs and DMRs (Fig. 1H), demonstrating their highly dynamic epigenomes during aging. The considerable correspondence between age-correlated accessibility at putative enhancers with expression of their target genes (Fig. 2D and fig. S6A) suggests that the epigenome strongly influences the aging transcriptome. At age-correlated cCREs, mCG inversely correlated with chromatin accessibility (Fig. 2E and fig. S6B), highlighting both modalities as sources of epigenetic regulation during aging. Additionally, microglia showed strong correlation between their DNA methylation clock (*37*) and predicted biological age and chronological age, whereas oligodendrocyte precursor cell (OPC) and inhibitory neurons showed very weak correlations (fig. S6C).

We found several TF motifs enriched at cCREs that gained accessibility with age (fig. S6, D and E). Enrichments were highest for bZIP and ETS family members and included TFs that mediate proinflammatory responses in microglia such as CEBP, AP-1, and STAT families (fig. S6, D and E) (*38*, *39*). To estimate the contribution of every TF motif to the aging epigenome and transcriptome, we calculated a scaled average age-correlation for chromatin accessibility, DNA methylation, and target gene expression for all cCREs containing a given TF motif (table S11). Aging accessibility and DNA methylation across TF motifs were negatively correlated (PCC = −0.63, P-value = 1.1×10^{-46}) (Fig. 2F), whereas aging accessibility and expression of target genes were positively correlated (PCC = 0.24, P-value = 6.0×10^{-8}) (Fig. 2G). CEBP and FOS family motifs had the strongest age-correlated accessibility and DNA hypomethylation. In addition, CEBP and FOS target genes were among those with the largest age-correlated gain in expression (Fig. 2G). By contrast, EGR1 and EGR2 motifs tended to have negative aging accessibility and target gene expression. Although most age-correlated cCREs were promoter distal (>1 kb from a TSS) (Fig. 2, F and G, and fig. S6F), EGR1 and EGR2 motifs were found mainly in promoter-proximal cCREs, where DNA methylation was relatively stable with age (Fig. 2F). Our analysis of age-related TF motif activities identifies likely regulators of the aging epigenome and

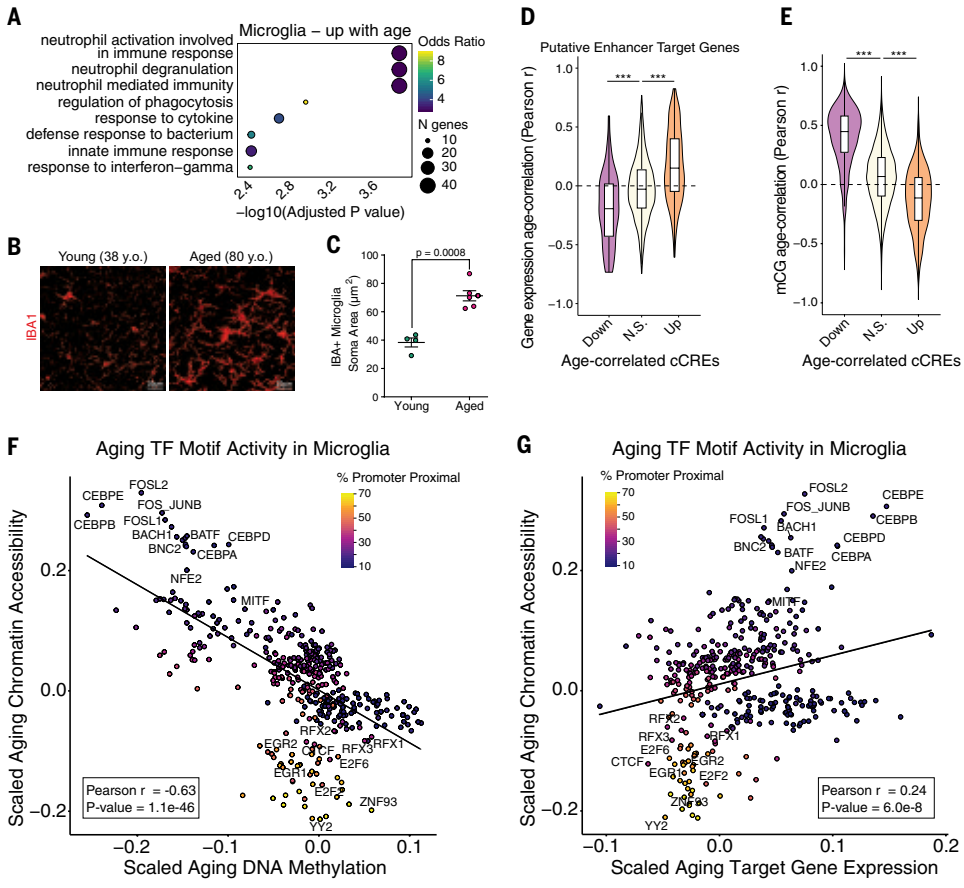


Fig. 2. Aging gene regulatory programs in microglia. (**A**) Dot plot of GO biological processes enriched in genes up-regulated with age in microglia. Dot size shows the number of genes, color indicates the odds ratio, and the *x*-axis shows −log10 adjusted *P*-value. (**B**) Representative micrographs of IBA1-immunostained hippocampal CA3 region from a 38-year-old (left) and an 80-year-old (right) human donor. (**C**) Graph showing quantification of microglia soma area between young ($n = 3$; ages 20 to 38) and aged ($n = 6$; ages 80+). *P*-value from an LME. (**D** and **E**) Violin plots showing the distribution of PCC of age-correlation of target gene expression (D) for age-correlated putative enhancers or mCG fractions of DMRs (E) overlapping age-correlated cCREs categorized as down-regulated (down), not significant (n.s.), or up-regulated (up) with age. (**F** and **G**) Scatter plots of scaled mean age-correlation for chromatin accessibility versus scaled mean age-correlation of DNA methylation at cCREs containing indicated TF motif (F), and chromatin accessibility versus target gene expression (G), across TF motifs. Dots represent TFs, colored by percentage of cCREs in promoter-proximal regions (<1 kb).

transcriptome, revealing coordinated activity across molecular modalities with age.

Microglia harboring epigenomic features of peripheral blood monocytes becomes dominant with age

We further explored the consequence of aging on microglial DNA methylomes and identified two robust sn mCG microglia subclusters labeled "Micro1" and "Micro2" (Fig. 3A). These were consistently detected with or without batch corrective methods (Fig. 3A and fig. S7A). We found a nonlinear age-dependence for the proportions of Micro1 and Micro2 cells (Fig. 3B). Donors aged 20 to 50 years old had mostly Micro1, whereas the switch to dominantly Micro2 occurred between ages 50 and 75. Donors above age 79 had almost exclusively Micro2 (Fig. 3B).

As with chromatin accessibility, microglia had relatively high numbers of age-correlated DMRs detected (Fig. 1H and table S8). Of these, 73% (11,239/15,327) became hypermethylated with age. The DNA methylation levels of age-correlated DMRs matched the ratio of Micro1 and Micro2 cells for each donor (Fig. 3B and table S12), demonstrating that most age-correlated DNA methylation is driven by Micro1 and Micro2 proportions. As the total microglia cell proportion did not increase with age and Micro2 cells had more mCH than Micro1 (fig. S7, B to D), which must occur de novo following cell division (*40*), it is unlikely that the increase in Micro2 results from an age-dependent clonal expansion.

To further characterize these age-associated microglia populations, we called DMRs between Micro1 and Micro2 (table S13). DMRs hypomethylated in Micro1 ($n = 23{,}580$) were enriched for the binding motif of ZNF281 (Zfp281 in mice) (fig. S7E), a TF that regulates DNA methylation in early development (*41*). These DMRs were proximal to genes with developmental functions (Fig. 3C, top), including those not expressed and inaccessible in the adult hippocampus, but showed hypermethylation in Micro2. For example, the Homeobox A (HOXA) locus contained 16 DMRs hypermethylated in Micro2 (fig. S8, A and B), and this age-dependent hypermethylation was specific to the microglia cell type (fig. S8, C and D). By contrast, DMRs hypomethylated in Micro2 ($n = 8{,}027$) were near genes involved in immune responses (Fig. 3C, bottom), such as cytokine receptor gene *IL1R1* (fig. S8A, bottom), and contained motifs of the bZIP and ETS families such as AP-1 and FLI1 (fig. S7E).

Microglia arise from myeloid progenitors, which migrate from the embryonic yolk sac to the developing brain (*42*). Although it is widely believed that this microglia population is maintained through adulthood, rodent studies have found that monocytes can repopulate the CNS niche, upon microglia depletion or brain engraftment, where they differentiate into microglia-like cells (*43*, *44*). Although monocytes are major contributors to the resident macrophages of several organs, the blood-brain barrier regulates monocyte entry into the brain parenchyma (*45*). However, rodent studies have shown that circulating monocytes will infiltrate upon injury, contributing to injury-induced CNS inflammation (*46*), but the extent to which monocytes contribute to the microglia pool in the human brain is thought to be limited to specific regions under certain pathological conditions (*47*–*50*). To check whether the DNA methylation that differentiates Micro1 and Micro2 could be traced to peripheral blood monocytes, we compared their DNA methylation profiles by leveraging an existing sn DNA methylation peripheral blood dataset (*51*). The monocyte DNA methylation patterns closely matched that of Micro2 at the DMRs between Micro1 and Micro2 (Fig. 3D). This included hypermethylation in Micro2 at the HOXA locus, which resembled the methylation observed in monocytes (Fig. 3E). We found a similar trend at *CLEC2B*, a ligand for natural killer (NK) gene complex receptor NKp80 (*52*), where Micro2 DNA methylation matched more closely with monocytes (fig. S8E).

We next clustered our hippocampus sn DNA methylation data together with existing data from select cells from peripheral blood, skin, and muscle (*51*). We also included sn DNA methylation microglia from

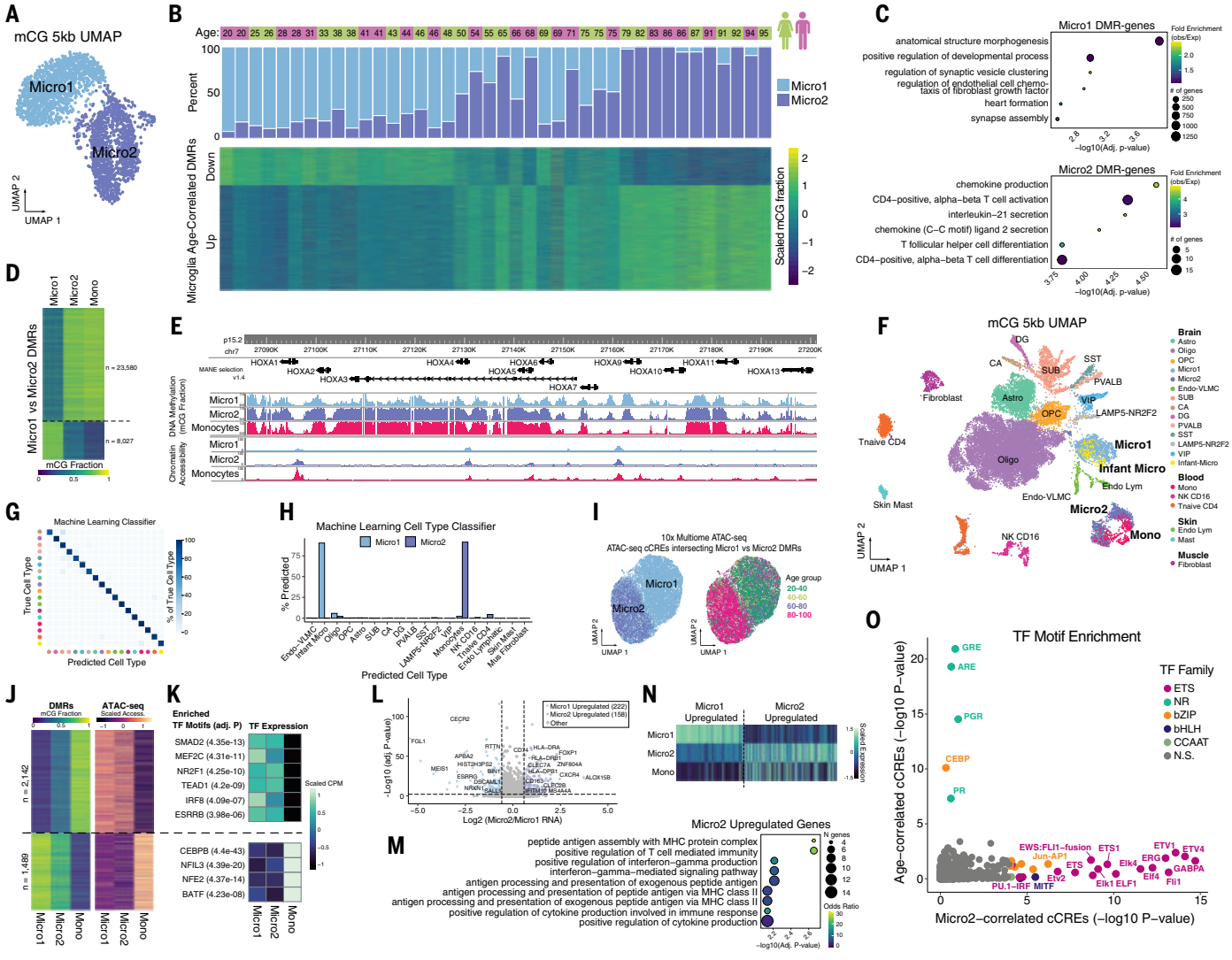


Fig. 3. An age-associated microglia population possesses DNA methylation profiles resembling peripheral blood monocytes. (**A**) UMAP plot of microglia from snm3C-seq, annotated with two DNA methylation microglia subclusters. (**B**) Stacked bar plots displaying the percentage of Micro1 and Micro2 cells recovered per donor from snm3C-seq experiments ordered by age (top), and heatmap of age-correlated DMRs showing row-scaled mCG fraction for each donor (bottom). (**C**) Enriched GO biological processes among nearest genes (closest transcription start site) associated with DMRs hypomethylated in Micro1 (top) and Micro2 (bottom). Dot size denotes gene count, and dot color indicates fold enrichment. (**D**) mCG fraction of DMRs between Micro1 and Micro2. Mono are peripheral blood monocytes from (*51*). (**E**) WashU Epigenome Browser screenshots of the HOXA cluster showing mCG fraction and chromatin accessibility in Micro1, Micro2, and peripheral blood monocytes. (**F**) UMAP of sn DNA methylation 5 kb mCG profiles from hippocampus, and select cell types of peripheral blood, skin, and muscle fibroblasts (*51*), and infant microglia (*53*). (**G**) Heatmap of cross-validation accuracy of a machine learning classifier trained on principal components of DNA methylation at 5 kb bins, excluding Micro1 and Micro2. (**H**) Predicted cell type assignments for Micro1 and Micro2 cells using the classifier from (G). (**I**) UMAP showing microglia from 10x multiome data clustered based on chromatin accessibility at microglia cCREs overlapping Micro1 or Micro2 differential DMRs ($n = 3405$). Cells are colored by Micro1 or Micro2 annotation (left) or by donor age group (right). (**J**) Heatmaps showing mCG fraction (left) and scaled chromatin accessibility (right) at regions hypomethylated in Micro1 ($n = 2142$) or monocytes ($n = 1489$). (**K**) Enriched TF motifs at the DMRs shown in (J) and scaled expression of the corresponding TFs. (**L**) Volcano plot of differential gene expression between Micro2 and Micro1, with select genes labeled. (**M**) Enriched GO biological process terms in genes up-regulated in Micro2 compared with Micro1, based on consensus clusters from 10x multiome data. (**N**) Heatmap showing scaled expression of genes up-regulated in Micro1 or Micro2 [identified in (L)] across Micro1, Micro2, and monocytes. (**O**) Scatter plot displaying TF motif enrichment for cCREs either Micro2 percentage correlated ("Micro2-correlated cCREs," fig. S13B) or age-correlated (fig. S13C) in 16 donors. Only TF motifs with q-value < 0.05 are colored by their TF family.

4- and 7-month-old infant brains (*53*). Micro1 cells clustered together with the infant microglia, whereas Micro2 clustered closely with monocytes (Fig. 3F). However, when clustering with gene expression or chromatin accessibility, microglia form a single cluster and remain completely distinct from monocytes (fig. S9, A to D). The differential DNA methylation patterns likely represent vestigial regulatory elements that were active in a prior developmental state (*54*). Although these elements retain stable DNA hypomethylation, they exist in inactive, closed chromatin. This epigenetic memory encoded in DNA methylation enables its use to accurately track a cell's developmental history (*55*). The relationships of DNA methylomes between adult hippocampal microglia, infant microglia, and blood monocytes highlights the Micro1 resemblance to infant yolk sac–derived microglia (YSDM), whereas Micro2 DNA methylomes suggest a potential monocyte origin. As such, our findings point to a possible replacement of YSDM by monocyte-derived microglia (MDM) between the ages of 50 and 75 (Fig. 3B).

To further assess the potential developmental origins of Micro1 and Micro2, we built a machine learning classifier to unbiasedly predict cell types of individual cells based on DNA methylation. The model was trained using principal components of methylation fractions at 5-kb genomic bins, with Micro1 and Micro2 excluded from the training (methods). Cross-validation demonstrated the model to be highly accurate at predicting cell type identity of unseen cells, with 97% overall accuracy (Fig. 3G). The classifier labeled the majority of Micro1 cells as infant microglia and Micro2 as monocytes (Fig. 3H), providing additional support for Micro1 and Micro2 potentially representing YSDM and MDM, respectively. Similar results were obtained from a model trained directly on mCG at DMRs (fig. S10).

Given the general correspondence between DNA methylation and chromatin accessibility at age-correlated cCREs in microglia (Fig. 2E, fig. S6B, and fig. S9, E and F), we reasoned that chromatin-accessible cCREs that overlap Micro1 and Micro2 DMRs ($n = 3405$) may separate Micro1 and Micro2 cells into distinct clusters in our 10x multiome data, whereas RNA and global chromatin accessibility failed to resolve these states (fig. S9, C and D). Indeed, when using these cCREs as features, two distinct, age-dependent clusters were apparent (Fig. 3I), and the proportion of cells between these clusters closely matched donor proportions obtained from snm3C-seq microglia cells (PCC = 0.88, *P*-value = 1.9×10^{-13}) (fig. S11A).

We assessed whether these features were effective in identifying Micro1 and Micro2 clusters in other brain regions by utilizing two external snATAC-seq datasets (*15*, *56*). Microglia separated into two major age-dependent clusters resembling Micro1 and Micro2 from these studies, demonstrating that epigenetic activity of these elements can faithfully resolve these microglial populations in multiple brain regions, and that the age-dependent increase of Micro2 is not restricted to the hippocampus (fig. S11, B to G).

Although DNA methylation matched closely between Micro2 and monocytes, we observed 4631 DMRs in which Micro2 resembled Micro1 more closely than monocytes, but levels were intermediate between Micro1 and monocytes (2142 Micro1-hypomethylated; 1489 monocyte-hypomethylated) (Fig. 3J). We hypothesized that these patterns are the result of DNA methylation remodeling driven by changes in TF activities during the transition from monocyte to microglia. Consistent with this hypothesis, we found that chromatin accessibility at these DMRs was similar between Micro1 and Micro2 but distinct from monocytes, and inversely corresponded to DNA methylation (Fig. 3J). We identified enriched TF motifs at these DMRs for TFs with expression that matched their motif accessibility patterns (Fig. 3K). These included TFs such as SMAD2, MEF2C, NR2F1, CEBPB, NFIL3, and NFE2 (Fig. 3K), which may promote epigenome remodeling during monocyte differentiation to microglia. Consistent with this model, Micro2 adopted the chromatin-accessible landscape of Micro1, but because of its higher stability DNA methylation retained patterns reminiscent of monocytes.

We further characterized Micro1 and Micro2 by identifying their differentially expressed genes ($n = 380$) using donor age as a latent variable (Fig. 3L and table S14). Genes that were up-regulated in Micro2 ($n = 222$) were enriched for major histocompatibility complex class II (MHCII), encoded by the human leukocyte antigen genes (e.g., *HLA-DRB1*, *HLA-DRA*, *HLA-DPB1*), and interferon-gamma signaling (Fig. 3M and fig. S12, A and B). MHCII gene expression was higher in monocytes than that in microglia (fig. S12C); therefore, their elevated expression in Micro2 may be inherited from their monocyte origin. In addition, many of the DEGs between Micro1 and Micro2 had expression patterns matching more closely to that between Micro2 and monocytes than Micro2 and Micro1 (Fig. 3N).

To characterize Micro2-specific chromatin activity, we identified cCREs that were more accessible in Micro2 compared with Micro1 ($n =$ 12,522) (table S15). These cCREs were enriched for ETS-domain TF motifs (fig. S12D). To distinguish accessibility changes specific to Micro2 from other age-related changes, we correlated accessibility with the percentage of Micro2 cells in 16 donors, aged 46 to 82 (fig. S13), in which age did not significantly correlate with Micro2 percentage (*P*-value = 0.082). We identified 858 cCREs with accessibility positively correlated with Micro2 percentages ("Micro2-correlated cCREs," PCC > 0.6) (fig. S13B and table S16), and 2118 cCREs that positively correlated with donor age, despite Micro1 and Micro2 proportions (fig. S13C and table S17). TF motif enrichment in Micro2-correlated cCREs were most significant for the ETS family, and age-correlated cCREs had highest enrichment for hormone-responsive elements of NRs (Fig. 3O and table S18), including glucocorticoid receptor (GR), which binds DNA at GR elements (GRE) in the presence of cortisol (*57*). Our results identify the ETS family as putative drivers of Micro2-specific chromatin accessibility, whereas binding sites for hormone-regulated NRs had age-correlated activity independent of the switch from Micro1 to Micro2.

Characterizing the 3D genome of age-dependent microglia populations

Further leveraging our snm3C-seq data, we investigated 3D genome organization specific to Micro2 (table S19). We found that the chromatin contacts that were stronger in Micro2 compared with Micro1 (Fig. 4A) had enrichment for promoters of type II interferon response genes (Fig. 4B). This included *IL15*, a proinflammatory cytokine involved in microglia activation (*58*), that was up-regulated during aging [PCC = 0.79, false discovery rate (FDR) = 4.2×10^{-6}] (Fig. 4C). Micro2-strengthened contacts were observed between the *IL15* promoter and an upstream cluster of cCREs with age-correlated chromatin accessibility and DNA hypomethylation (Fig. 4D), suggesting that these upstream cCREs may be distal enhancers that drive expression of *IL15*. As with DNA methylation, the Micro2-specific 3D genome conformations may be remnants of a monocyte origin. Leveraging an existing Hi-C dataset from human blood monocytes (*59*), we indeed observed similar chromatin contact frequencies between Micro2 and monocytes at the Micro2-stronger and Micro2-weaker contact bins (Fig. 4E). For example, monocytes possessed a contact similar to Micro2 between the *IL15* promoter and the upstream accessible putative enhancer (fig. S14A). Another example was observed at a region centered over the promoter of *ZNF804A*, a schizophrenia risk gene (*60*) with age-correlated expression in microglia (fig. S14B). Micro2 displayed several adjacent contacts, forming a local interaction domain absent in Micro1 (fig. S14C). At this locus, monocytes had a genome conformation resembling that of Micro2 (fig. S14C).

We found enrichment for cCREs gaining accessibility with age in Micro2-stronger contacts, whereas Micro2-weaker contacts were enriched for cCREs losing accessibility with age (Fig. 4F). We observed the same pattern of enrichment for age-correlated gene expression (Fig. 4G), demonstrating that the Micro2-specific 3D genome coincides with age-related epigenomic and transcriptomic changes. We also observed correspondence between the age-related 3D genome, epigenome, and transcriptome changes in additional glial types (fig. S15, A and B).

Human hippocampal astrocyte abundance proportionally declines during aging

We next identified cell types that either gained or lost abundance during aging (Fig. 5A, fig. S16, and table S20). Astrocytes displayed the greatest age-correlated decline (PCC = −0.69, *P*-value = 8.1×10^{-7};18.4% at age 20 with an average decline of 0.215% per year for forward ages) (Fig. 5B and fig. S16A). In addition, both OPC and endothelial cell abundance significantly decreased as donor age increased (OPC: PCC = −0.52, *P*-value = 4.8×10^{-4}; Endo: PCC = −0.53, *P*-value = 4.6×10^{-4}) (fig. S16C). Significant age-correlated decline in the number of astrocytes, OPCs, and endothelial cells was corroborated in our snm3C-seq data (fig. S16D and table S21). Notably, both astrocytes and endothelial cells are critical for blood-brain barrier integrity, and the loss of these cells in the aging hippocampus may contribute to the documented breakdown of the blood-brain barrier (*61*).

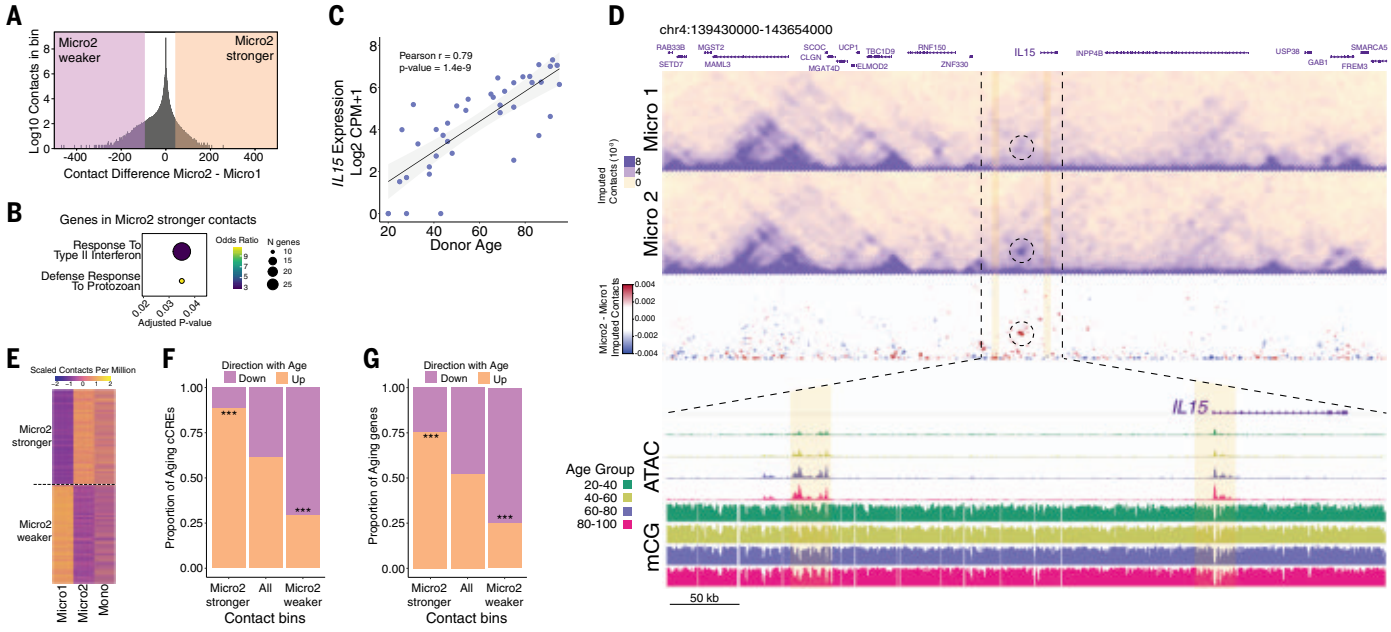


Fig. 4. Correspondence between age-related 3D genome reorganization with the epigenome and transcriptome. (**A**) Density histogram of contact differences between Micro1 and Micro2 across 100-kb genomic bins, subtracting Micro2 from Micro1 contact frequencies after sampling equal number of total contacts for each group. Left side indicates bins with weaker contacts in Micro2 than Micro1 ("Micro2 weaker"), and right side indicates stronger contacts ("Micro2 stronger"). Bins in the top ±0.0005 percentile are highlighted. (**B**) GO biological process terms (q-value < 0.05) enriched among genes with transcription start sites located within Micro2 stronger contact bins. (**C**) Scatter plot of *IL15* expression versus donor age. (**D**) Heatmaps showing imputed contact frequencies in Micro1 and Micro2, with a Micro2-Micro1 difference track. A dotted circle highlights the contact between the *IL15* promoter and an upstream cluster of cCREs. WashU Epigenome Browser screenshots of the region within dashed lines displaying chromatin accessibility and mCG fraction in microglia across age groups. (**E**) Heatmap displaying scaled contact frequencies (per million) across Micro1, Micro2, and monocytes for the Micro2-weaker and Micro2-stronger contacts identified in (A). (**F** and **G**) Stacked bar plots showing proportions of age-correlated cCREs (F) and genes (G) within Micro2 stronger or weaker contact bins. *P*-values computed using Pearson's Chi-square test comparing Micro2 stronger Up with All or Micro2 weaker Down with All, ***$P < 0$.

The observed decline in astrocyte number with age was particularly intriguing considering their indispensable roles in synaptic signaling in the adult brain through fine-tuning interstitial glutamate concentrations and preventing aberrant neurotransmission to neighboring synapses (*62*). We confirmed the age-related loss of astrocytes using confocal imaging of human hippocampal CA3 tissue and immunostaining for astrocytic marker ALDH1L1 using tissue from 11 donors (Fig. 5, C and D), seven of which were distinct from those used in sequencing experiments (table S1). Quantification of ALDH1L1-positive astrocytes, normalized to total nuclei number, confirmed a substantial reduction of astrocyte proportion in hippocampal CA3 tissue from aged donors (80+ years old) compared with young donors (20 to 38 years old) (Fig. 5D).

Age-dependent epigenetic regulation of mitochondrial ATP synthesis genes in aging astrocytes

We first checked whether age-induced senescence corresponded to loss of astrocytes by calculating the proportion of cells positive for expression of senescent marker genes, including *CDKN2A* (p16) (*63*). While OPC, oligodendrocytes, and macrophages had a considerable increase in p16-positive cells with age, astrocytes did not (fig. S17), suggesting that their proportional decline is unlikely to be due to age-related senescence. We next sought to identify gene regulatory processes that could underlie the age-related loss of astrocytes in the hippocampus. Gene ontology (GO) analysis revealed that genes with expression negatively correlated with age were highly enriched for those required for mitochondrial ATP synthesis (Fig. 5E), including 23 genes encoding subunits of the electron transport chain respiratory complex I (e.g., *NDUFA1*, *NDUFB2*, *NDUFC1*), and 16 encoding the subunits of ATP synthase (e.g., *ATP5F1A*, *ATP5MF*, *ATP5PF*). As part of mitochondrial dysfunction surveillance, cells respond to depleted ATP by activating lysosomal autophagy (*64*, *65*). Consistent with this interplay, genes with increased expression in aged astrocytes were enriched for lysosomal microautophagy (Fig. 5F), potentially leading to autophagy-dependent cell death (*66*). The DNA methylation dynamics at these genes suggest that their expression was epigenetically regulated through DNA methylation during aging (Fig. 5G).

To identify the regulators of dynamic gene expression programs in aging astrocytes, we performed TF motif enrichment analysis at cCREs with age-correlated chromatin accessibility (Fig. 5H and table S22). At cCREs that became less accessible with age in astrocytes, we observed enrichment for NRF1 motifs (Fig. 5H), known to activate nuclear genes with mitochondrial functions (*34*). Consistently, DNA methylation at NRF1 motifs increased during aging (Fig. 5I). As more than 50 genes with mitochondrial functions were down-regulated in astrocytes with age (Fig. 5E and table S6), we hypothesized that reduced DNA binding of NRF1 may be responsible. 86% of cCREs with an NRF1 motif and reduced accessibility with age were at promoters ($n = 354$, table S23) (Fig. 5J). Using a human neuroblastoma ChIP-seq dataset (*67*), we confirmed NRF1 enrichment at these sites (Fig. 5K). Genes with NRF1 motifs in their promoters that lost accessibility with age (table S24) were enriched for ATPase components (Fig. 5L), such as *ATP5ME* and *ATP5PF*, crucial for mitochondrial energy production. These lines of evidence suggest a potential mechanism in which reduced NRF1 DNA binding leads to reduced ATP synthesis and increased autophagy-dependent cell death in aging human hippocampal astrocytes. However, *NRF1* mRNA levels were unaltered during aging (PCC = −0.05), suggesting that NRF1 DNA binding may be inhibited by the observed increase in methylation at its binding sites (Fig. 5I), or through other posttranslational regulation, such as through interaction with its co-activator PGC-1 (*68*).

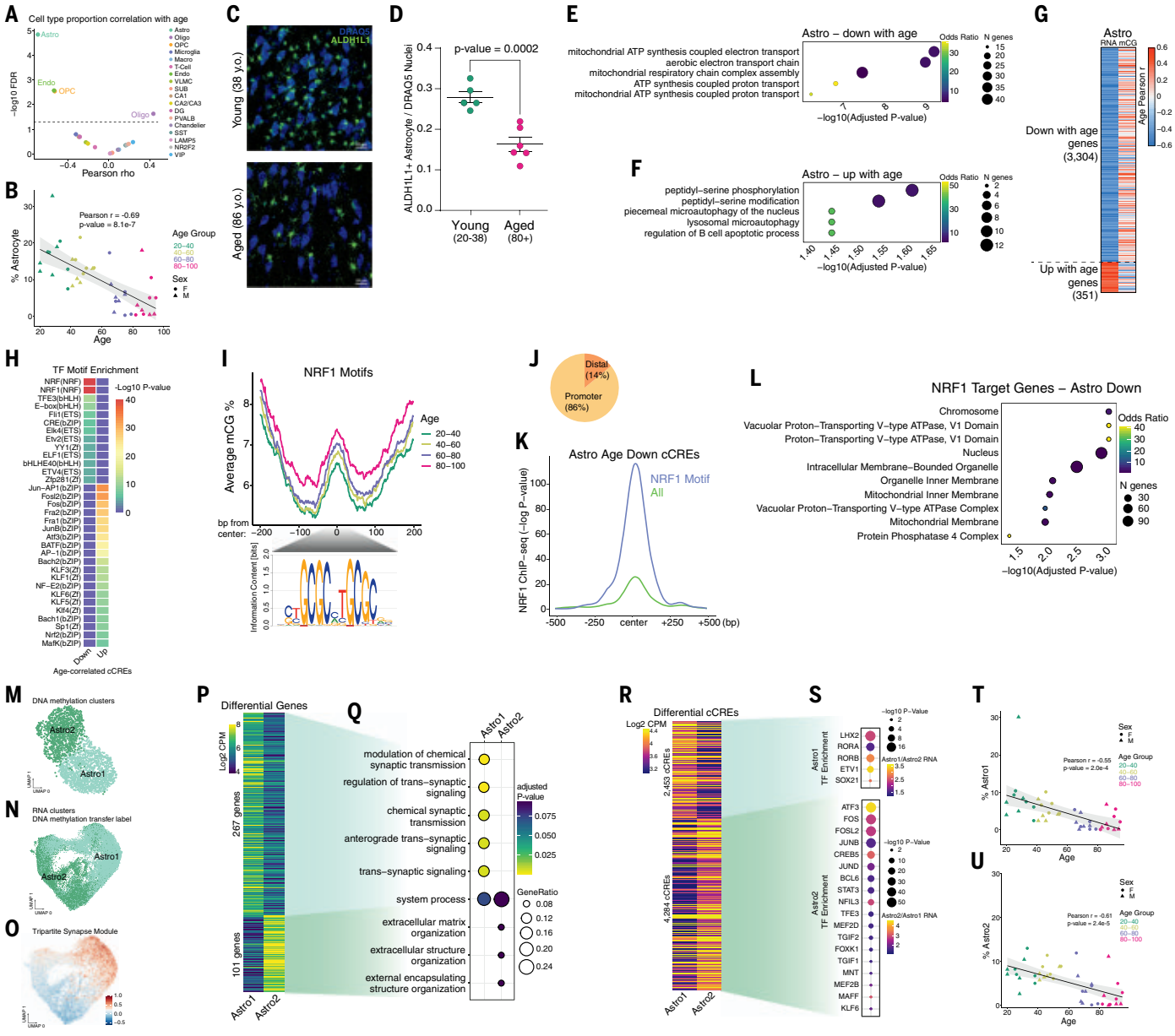


Fig. 5. Cellular composition and gene regulatory changes in aging astrocytes. (**A**) Scatter plot showing the PCC (rho) of age versus −log10 FDR. Each dot represents a cell type. (**B**) Scatter plots showing the percentage of recovered astrocytes (% = 22.573 ±0.215 × Age) out of total cells versus donor age from 10x multiome experiments. (**C**) Confocal imaging of hippocampal CA3 tissue immuno-stained for the astrocytic marker protein ALDH1L1 (green) and nuclei marker DRAQ5 (blue). (**D**) Ratio of ALDH1L1-positive astrocytes to DRAQ5-stained nuclei in young donors (n = 5, 20 to 8 years old) and aged donors (n = 6, 80+ years old). P-value = 0.0002 from an unpaired Welch's t test. (**E** and **F**) Enriched GO biological processes for age-correlated genes that decrease with age (E) or increase with age (F) in astrocytes. (**G**) Heatmap showing PCCs for genes with expression decreasing (n = 3304) or increasing (n = 351) with age, and mCG levels in their gene body regions ±2 kb in astrocytes. (**H**) TF motif enrichment heatmap for significant enrichments, q-value < 0.05. "Up" indicates chromatin accessibility increases with age and "Down" indicates chromatin accessibility decreases with age. (**I**) Average mCG fractions at NRF1 motifs in each age group in astrocytes. (**J**) Ratio of cCREs with negatively age-correlated accessibility in astrocytes with an NRF1 TF motif that are either promoter-proximal (<1 kb) or promoter distal (>1 kb) from a protein-coding transcription start site. (**K**) Average −Log10 P-value NRF1 ChIP-seq signal in SK-N-SH cell line (ENCODE Accession ENCFF762LEE) at all astrocyte cCREs that decrease accessibility with age (n = 1803, green line) or the subset that contain an NRF1 TF motif (n = 354, blue line). (**L**) GO cellular component terms significantly enriched terms (q-value < 0.05) for genes that have negative age-correlated accessibility in astrocytes with an NRF1 TF motif at their promoter. (**M**) UMAP plot of astrocytes from snm3C-seq DNA methylation data, annotated with two astrocyte subclusters. (**N**) UMAP plot of astrocytes from 10x multiome RNA data, annotated with two DNA methylation astrocyte subclusters by label transfer using hypomethylation gene scores and RNA integration. (**O**) UMAP showing the average module score (methods) of 18 tripartite synaptic genes (*71*). (**P**) Heatmap displaying Log2 RNA CPM of Astro1 and Astro2 marker genes. (**Q**) GO terms for enriched biological processes of Astro1 and Astro2 marker genes. (**R**) Heatmap displaying Log2 ATAC-seq fragments CPM of Astro1 and Astro2 marker cCREs. (**S**) Enriched TF motifs for Astro1 and Astro2 marker cCREs that also exhibited up-regulated TF expression in the indicated subcluster. (**T** and **U**) Scatter plots showing the percentage of recovered Astro1 or Astro2 cells identified in (N) out of total cells versus donor age from 10x multiome experiments.

Gene regulatory programs in tripartite astrocytes

From our sn DNA methylation clustering, astrocytes resolved into two major clusters ("Astro1" and "Astro2"), demonstrating two distinct DNA methylome states (Fig. 5M). We integrated DNA methylation with RNA using hypomethylation gene scores (methods) (Fig. 5N) and obtained consensus cluster labels for nuclei from both assays (fig. S18, A and B). As a subset of astrocytes participate in tripartite synapses by clearing neurotransmitters in the synaptic cleft and signaling to proximal neurons through gliotransmitters (*69*, *70*), we checked expression of 18 tripartite astrocyte gene markers (*NRXN1*, *NLGN3*, *CADM1*; see methods for full list) that can be used to distinguish tripartite from nontripartite astrocytes (*71*). Cells in the Astro1 subcluster had relatively high expression of tripartite marker genes (Fig. 5O) and lower methylation within their gene bodies compared with cells in the Astro2 subcluster (fig. S18, C and D), suggesting that Astro1 were tripartite astrocytes.

We identified genes differentially expressed between Astro1 and Astro2 (Fig. 5P and table S25) and observed GO–enriched terms related to synaptic signaling for Astro1, and extracellular matrix organization for Astro2 (Fig. 5Q). We also identified 6737 cCREs that were differentially accessible between Astro1 and Astro2 (Fig. 5R and table S26). The Astro1 cCREs were enriched for motifs corresponding to LHX2, RORA, RORB, ETV1, and SOX21 binding (Fig. 5S), suggesting a potential role for these TFs in promoting tripartite astrocyte identity. Both subclusters, including the synaptic-signaling tripartite astrocytes, proportionally declined with age (Fig. 5, T and U), a potential source of synaptic dysfunction in the aged human hippocampus. Consistent with a recent study of cortical astrocytes (*72*), genes involved in the synaptic neuron and astrocyte program (SNAP) showed decreased expression in hippocampal astrocytes during aging (fig. S18E).

Age-dependent decay of 3D genome organization and decreased CTCF DNA-binding

Chromatin contacts between different chromosomes (trans contacts) were previously reported to change with age in cerebellar granule cells (*11*). We observed a global increase in the number of trans contacts in the 80 to 100 age group for different cell types (fig. S19A), including microglia (median difference 3.8%, P-value = 1.1×10^{-23}) and oligodendrocytes (median difference 1.8%, P-value = 1.3×10^{-9}) (Fig. 6, A to F). While most cell types exhibited increased trans contacts, the change was cell type–dependent. For example, SUB excitatory neurons had increased trans contacts whereas CA and DG excitatory neurons had decreased trans contacts (fig. S19A).

Within chromosomes, chromatin is organized into self-interacting territories known as TADs (*18*), where regulatory elements are largely constrained to regulate transcription within the same TAD. We found that contacts within TADs were greatly diminished in the 80 to 100 age group (Fig. 6, G to J, fig. S19B, and table S27; microglia, median difference 1.6%, P-value = 3.0×10^{-20}; oligodendrocytes, median difference

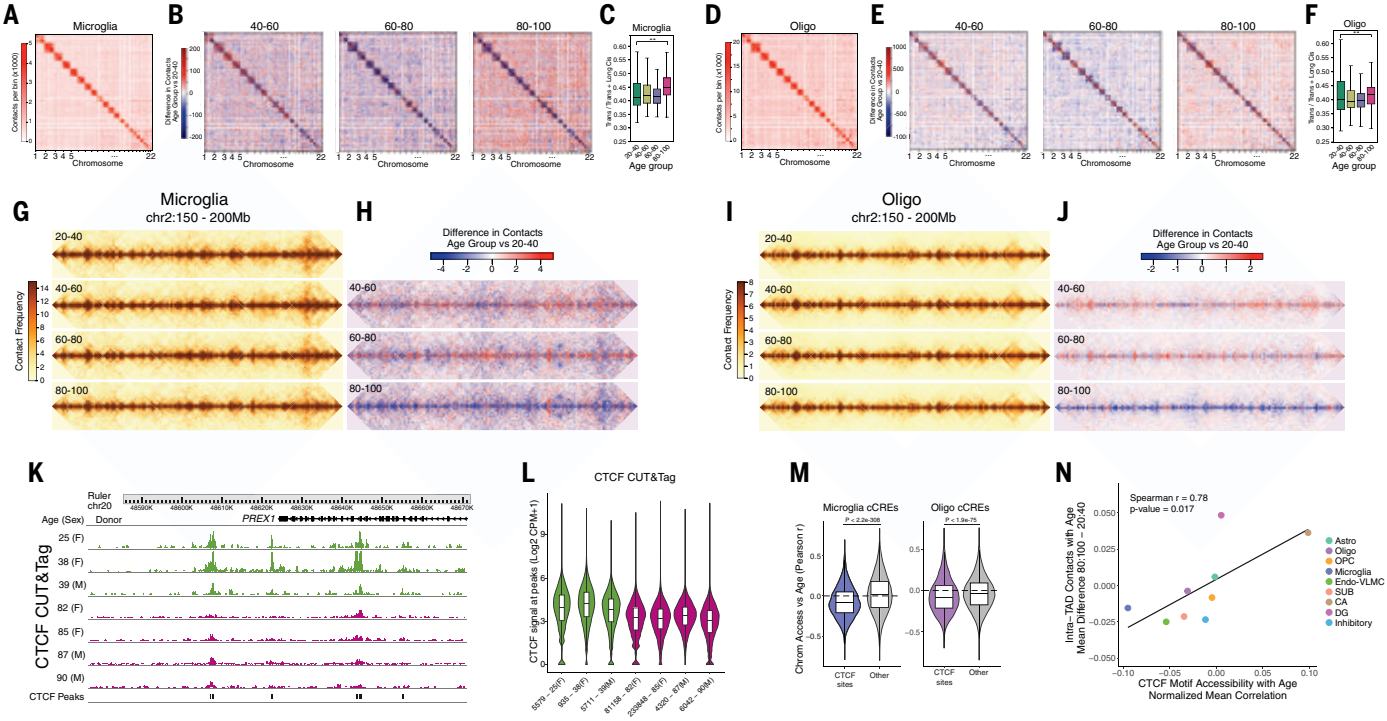


Fig. 6. Global loss of homeostatic 3D genome architecture and CTCF binding in aging hippocampal cells. (**A** to **F**) Aggregated raw interchromosomal contact maps at 6-Mb resolution of microglia (A) and oligodendrocytes (D). Heatmaps showing contact frequency differences between the 40 to 60, 60 to 80, and 80 to 100 age groups compared with the 20 to 40 age group in microglia (B) and oligodendrocytes (E). Box plots displaying the ratio of trans contacts to the sum of cis long-range and trans contacts across age groups in microglia (C) and oligodendrocytes (F). P-values from two-sided unpaired Wilcoxon rank sum test comparing 20 to 40 with 80 to 100 age groups, $^{**}P$-value < 0.001. (**G** to **J**) Aggregated contact maps of microglia (G) and oligodendrocytes (I) across age groups in the chr2:150,000,000 to 200,000,000 region at 100-kb resolution. Heatmaps showing contact frequency differences for the 40 to 60, 60 to 80, and 80 to 100 age groups compared with the 20 to 40 age group within the same region and resolution in microglia (H) and oligodendrocytes (J). (**K**) WashU Epigenome Browser screenshot of CTCF CUT&Tag from hippocampi of three young donors (ages 25, 38, 39) and four aged donors (ages 82, 85, 87, 90) at the *PREX1* gene locus. The bottom track shows CTCF peak locations. Donor age and sex (F or M) are indicated. (**L**) Distribution of CTCF signal (Log2 CPM+1) at CTCF peaks across donors. (**M**) Violin plots showing the distribution of age correlations for chromatin accessibility at cCREs overlapping CTCF motifs versus cCREs not overlapping CTCF motifs ("other") in microglia (left) and oligodendrocytes (right). P-values derived from unpaired, two-sided Wilcoxon rank sum test. (**N**) Scatter plot showing the difference in median intra-TAD contacts between the 80 to 100 and 20 to 40 age groups versus the difference between the median Pearson age-correlation of chromatin accessibility at CTCF motifs and that of all cCREs for each cell type.

1.7%, P-value = 6.9×10^{-31}), specifically for cell types where trans contacts increased. This shift from intra-TAD contacts toward trans chromosomal contacts highlights a widespread loss in homeostatic 3D genome organization in aging hippocampal cells.

As CCCTC-binding factor (CTCF), a DNA-binding factor, is critical for maintaining TADs (*19*), we asked whether the decay in TAD structures is a result of reduced CTCF DNA-binding in aged hippocampal cells. To this end, we performed CUT&Tag (*73*) from hippocampi of three young (ages 25, 38, and 39) and four aged (ages 82, 85, 87, and 90) donors. Compared with young donors, CTCF binding was globally reduced in aged donors (Fig. 6, K and L), consistent with age-associated TAD decay. To assess CTCF DNA-binding in each cell type, we checked accessibility at CTCF binding sites. In microglia and oligodendrocytes, chromatin accessibility at cCREs containing CTCF motifs was more negatively correlated with age compared with cCREs not containing CTCF motifs (Fig. 6M). We next asked whether CTCF DNA binding could explain cell type–dependent decay of TADs by correlating loss of chromatin accessibility at CTCF motifs with TAD decay across cell types. Indeed, we found a strong correlation (PCC = 0.74, P-value = 0.024) between the change in intra-TAD contacts with chromatin accessibility at CTCF motifs with age (Fig. 6N). Although DNA methylation is known to block CTCF from binding (*74*), we did not observe an appreciable change in methylation at CTCF motifs with age (fig. S20, A and B), suggesting that an alternative mechanism is responsible for the age-associated change in CTCF motif accessibility.

At the megabase scale, the genome is compartmentalized into A and B compartments, where chromatin contacts are largely constrained within each compartment (*75*). We observed a general weakening of compartmentalization in Micro2 compared with Micro1 cells and in the 80 to 100 age group (fig. S21, A to C). Because compartment A is enriched for euchromatin whereas compartment B is enriched for heterochromatin (*75*), we checked for correspondence between age-related changes in compartmentalization with the epigenome. We identified genomic regions with age-differential compartments in microglia as well as astrocytes and oligodendrocytes (table S28), which were enriched for corresponding changes in accessibility with age (fig. S21D). Our results identify widespread age-related 3D genome reorganization coincident with age-correlated changes in the epigenome and transcriptome.

Discussion

In this study we present a comprehensive, multiomic single-cell investigation of gene regulatory programs in the aging human hippocampus, providing a foundation for understanding how age reshapes cellular identity and regulatory architecture. This work provides a foundational resource for future studies promoting healthy brain aging and targeting age-associated neurodegenerative disease.

In donors under 50, yolk sac–derived microglia predominated, whereas in older individuals they were largely replaced by microglia with DNA methylation patterns resembling peripheral monocytes. This shift follows a nonlinear, age-dependent transition rather than a slow, continuous replacement across the lifespan. Because these two populations are nearly indistinguishable at the levels of gene expression and chromatin accessibility, previous studies lacked the resolution to separate them. Only DNA methylation, an epigenetic mark that preserves aspects of a cell's regulatory history (*54*), revealed the presence of these distinct microglial lineages. By defining their distinct epigenetic signatures, it will now be possible to investigate their relative associations with neurodegenerative disease processes within the human brain. Importantly, the transcriptomic and 3D genomic features specific to the monocyte-like microglia population were enriched for inflammatory signatures, suggesting that this microglia population contributes to age-related chronic neuroinflammation.

A recent study identified an increasing number of MDM in the aging mouse brain yet remained a small fraction of the total microglia pool (*76*). The extent to which infiltrating monocytes contribute to microglia in the human brain remains unanswered. Our analysis suggests that infiltrating monocytes may be the primary source of microglia in aged humans. Consistent with our results, a previous study showed that, in carriers of clonal hematopoiesis of indeterminate potential (CHIP), blood-derived somatic mutations were detected in brain nuclei, and the mutant-carrying cells exhibited a microglia-like identity, indicating that blood-derived cells may infiltrate the brain and adopt a microglial phenotype (*77*). Whether these results were due to the CHIP mutations conferring a phenotype facilitating entry into the brain and subsequent proliferation could not be determined.

Further investigation is needed to determine the mechanistic triggers that drive the emergence and expansion of this monocyte-like microglial population during aging. As the total number of human microglia remains constant with age, we speculate that the entry of monocytes into the brain results from a nonlinear age-dependent loss of the self-renewal capacity of embryonic microglia. This would explain their gradual decline, which in turn would create an open niche that would be expected to invite the entry of monocytes. Recent studies based on incorporation of atmospheric C^{14} estimated that human microglia slowly turn over at a yearly median rate of 28%, with high interindividual variation (*78*). These studies did not find evidence for a large quiescent subpopulation, indicating that most microglia must renew throughout life or die. Variable loss in self-renewal capability would be consistent with the variable onset of the increase in the Micro2 population observed between 50 and 75 years of age. Although our data strongly suggest that MDM constitute most hippocampal microglia in aged adults, we cannot conclusively trace these cells back to a monocyte origin. However, we note that the complete separation of the clusters defining Micro1 and Micro2, without evidence of transitional intermediates, reinforces the conclusion that they represent cells of distinct developmental origins, with the most plausible candidate for the origin of Micro2 being monocytes. Taken together with emerging evidence from other studies, our findings challenge the long-standing model that yolk sac–derived microglia maintain the microglial population across the entire human lifespan.

Previous studies across species have generally reported no substantial loss of astrocyte numbers with aging (*79*, *80*). However, results have varied depending on brain region and methodology, and a clear consensus has yet to be established. We observed a considerable loss of astrocyte proportion in the aging human hippocampus, supported by multiomics and immunostaining with a pan-astrocyte marker. Aging astrocytes had reduced expression of genes involved in synthesizing ATP, a potential source for astrocyte death, as ATP depletion can induce necrosis (*81*) and immune cell engulfment (*82*, *83*). These findings warrant future studies into the triggers of hippocampal astrocyte attrition in aging.

Materials and methods

Sample inclusion and feasibility testing

Human postmortem hippocampal tissue was obtained through the Center for Neural Circuit Mapping (CNCM) Brain Procurement Program at the University of California, Irvine and the NIH NeuroBioBank (table S1), under protocols approved by the UCI Institutional Review Board (IRB). Informed consent for tissue donation was obtained from donors or their legally authorized representatives, and for deceased donors, authorization was obtained from next-of-kin in accordance with institutional and HIPAA guidelines. All specimens were de-identified prior to distribution and analysis, and their use is consistent with federal regulations governing nonhuman subjects research. We required postmortem interval to $\leq$ 20 hours and Braak stage of IV or less in aged donors used in sequencing experiments. To assess chromatin quality of samples before performing sn experiments, we performed bulk ATAC-seq as described (*84*) from 5mg of tissue. We calculated a signal to noise ratio from qPCR with ATAC-seq libraries using LightCycler® 480 SYBR

Green I Master Mix (Roche) along with custom primers for the promoter of human *GAPDH* (5′-CATCTCAGTCGTTCCCAAAGT-3′, 5′-TTCCCAGGACTGGACTGT-3′) and a heterochromatic gene desert region (5′-CCCAAACTCTGAGAGGCTTATT-3′, 5′-GAGCCATCATCTAGACACCTTC-3′). We required a *GAPDH* promoter signal fold enrichment of > 10 compared with the gene desert region and RNA integrity (RIN) value of > 6 to be included in our study.

Nuclei preparation from frozen brain tissue for Chromium Single-Cell Multiome ATAC + Gene Expression (10x Genomics multiome)

Tissues were pulverized in liquid nitrogen using a mortar and pestle on dry ice. Around 30mg of pulverized brain tissue was added to a ice-cold dounce homogenizer containing 1mL of chilled NIM-DP-L buffer (0.25M sucrose, 25mM KCl, 5mM MgCl2, 10mM Tris-HCl pH 7.5, 1mM DTT, 1X Protease Inhibitor (Pierce), 1U/μL Recombinant RNase inhibitor (Promega, PAN2515), and 0.1% Triton X-100). Tissue was dounce homogenized with a loose pestle (5-10 strokes) to shear large pieces, followed by a tight pestle (15-25 strokes) until the solution was uniform. Homogenate was filtered through a 30μm CellTrics filter (Sysmex, 04-0042-2316) into a LoBind tube (Eppendorf, 22431021) and pelleted (1000 rcf, 10 min at 4°C) (Eppendorf, 5920 R). The pellet was resuspended in 1mL NIM-DP buffer (0.25M sucrose, 25mM KCl, 5mM MgCl2, 10mM Tris-HCl pH 7.5, 1mM DTT, 1X Protease Inhibitor, 1U/μL Recombinant RNase inhibitor) and pelleted (1000 rcf, 10 min at 4°C). The pellet was resuspended in 400uL Sort Buffer (1mM EDTA, 1U/μL Recombinant RNase inhibitor, 1X Protease Inhibitor, 1% fatty acid-free BSA in phosphate buffered saline (PBS)) with 2μM 7-AAD (Invitrogen, A1310). 120,000 7-AAD-positive nuclei were sorted (Sony, SH800S) into a LoBind tube containing Collection Buffer (5U/uL Recombinant RNase inhibitor, 1X Protease Inhibitor, 5% fatty acid-free BSA in PBS). Final collection volumes were determined and 5X Permeabilization Buffer (50mM Tris-HCl pH 7.4, 50mM NaCl, 15mM MgCl2, 0.05% Tween-20, 0.05% IGEPAL, 0.005% Digitonin, 5% fatty acid-free BSA in PBS, 5mM DTT, 1U/μL Recombinant RNase inhibitor, 5X Protease Inhibitor) was added to achieve a 1X solution. Nuclei were incubated on ice for 1 min, then centrifuged (500 rcf, 5min at 4C) in a swinging bucket centrifuge. Supernatant was discarded and 650uL of Wash Buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1%.Tween-20, 1% fatty acid-free BSA in PBS, 1mM DTT, 1U/μL Recombinant RNase inhibitor, 1X Protease Inhibitor) was added without disturbing the pellet followed by centrifuging (500 rcf, 5 min at 4°C) in a swinging bucket centrifuge. Supernatant was removed, and the pellet was resuspended in 7uL of 1X Nuclei Buffer (Nuclei Buffer (10x Genomics), 1mM DTT, 1 U/μL Recombinant RNase inhibitor). 1μL was used for counting on a hemocytometer after staining with Trypan Blue (Invitrogen, T10282). 16-20k nuclei were used for tagmentation reaction and 10x Genomics controller loading. Libraries were then generated following the manufacturer's recommended protocol (https://www.10xgenomics.com/support/single-cell-multiome-atac-plus-gene-expression). 10x multiome ATAC-seq and RNA-seq libraries were paired-end sequenced on a NextSeq 2000 to assess data quality. If data quality was satisfactory, libraries were deep sequenced on a NovaSeq 6000 to a target depth of ~50,000 reads per cell for each modality.

10x multiome sequence data processing and clustering

Raw sequencing was processed using cellranger-arc v2.0.0 (10x Genomics), generating snRNA-seq UMI count matrices for intronic and exonic reads mapping in the sense direction of a gene for hg38 GRCh38 human genome assembly and GENCODE v32 gene annotations. We performed unsupervised clustering with RNA UMI counts using Seurat v5 (*85*) integrative analysis pipeline. First, cell barcodes from the raw cellranger matrices were filtered for low quality nuclei by requiring ≥ 500 ATAC fragments, ≥ 200 genes detected per nucleus, and < 20% mitochondrial RNA reads. We further required > 5 TSS enrichment (GRCh38 GENCODE v46) using SnapATAC2 (*86*). Counts were normalized using SCTransform and 2000 protein coding variable genes were determined and used for principal components analysis (PCA). Putative multiplets were predicted using DoubletFinder (*87*) software and 10% of cell barcodes were removed from each sample that had the highest doublet score. Batch correction across donors was performed using reciprocal principal component analysis (rPCA) on SCTransformed PCs. A k-nearest neighbors graph was built using PCs 1:30 and clusters were identified using leiden clustering. To visualize clusters, we performed the nonlinear dimension reduction technique, uniform manifold approximation and projection (UMAP) (*88*). We annotated subclass-level cells by known marker gene expression and reference mapping to a published hippocampus snRNA-seq dataset (*89*, *90*) using Seurat. Additional multiplet clusters were identified and removed for expressing multiple cell type–specific marker genes. We assessed batch effects of pseudo bulk gene expression with t-SNE clustering. The same 2,000 variable genes were used for principal component analysis on aggregated count CPM for each cell type by donor ($n = 554$). PCs 2:50 were used for t-SNE clustering with Rtsne program and parameters: dims = 2, perplexity = 30, max_iter = 500.

ATAC-seq peak calling and filtering

For each cell type, we divided the cells into four donor age groups, 20-40, 40-60, 60-80, and 80-100 year olds. For each cell type, we called peaks using ATAC-seq fragments of combining cells from all 40 donors, as well as in each of the four individual age groups, to capture any age-specific peaks. This gave us five lists of peaks for each cell type. We called peaks using MACS2 (*91*) software with the following command: macs2 callpeak–shift -75–ext 150–bdg -q 0.1 –call-summits -f BED. We extended each peak summit by 249 bp upstream and 250 bp downstream to achieve a uniform 500 bp span for each peak. Peaks overlapping with ENCODE blacklist regions for the hg38 genome assembly (https://mitra.stanford.edu/kundaje/akundaje/release/blacklists/) were removed. To account for variability in sequencing depths because of differences in cell type abundance, MACS2 peak scores (-log10(q-value)) were converted to 'score-per-million' (SPM). We retained peaks with a score-per-million ≥ 4 in any of the five peak sets for each cell type. We then created a union set of peaks across all cell types and age groups by merging all peaks that have any overlap, then resized to 500bp using the peak summit with the highest SPM from any individual peak set. From this union set, we filtered out peaks that did not meet the score-per-million ≥ 4 threshold in any cell type, ensuring that all retained peaks were robustly detected in at least one cell type. To reduce redundancy, coordinates for peaks of each cell type were replaced with coordinates from the refined union set of peaks, by intersecting the filtered peaks from each cell type with the refined union set using bedtools intersect (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html). To assess the overlap between DMRs and ATAC-seq cCREs (peaks) coordinates, we calculated Jaccard Index using bedtools jaccard function. Data are displayed in the heatmap after column (ATAC-seq cCREs) Z-score scaling the Jaccard indexes to show relative overlap with DMRs from each cell type.

Nuclei isolation and fluorescence-activated nuclei sorting (FANS) for snm3C-seq

Fresh frozen hippocampus tissues were pulverized in liquid nitrogen using a mortar and pestle on dry ice. Around 20mg of pulverized brain tissue was added to an ice-cold dounce homogenizer containing 1mL of chilled NIM-DP-L buffer (0.25M sucrose, 25mM KCl, 5mM MgCl2, 10mM Tris-HCl pH 7.5, 1mM DTT, 1X Protease Inhibitor (Pierce), 1U/μL Recombinant RNase inhibitor (Promega, PAN2515), and 0.1% Triton X-100). Tissue was dounce homogenized with a loose pestle (5-10 strokes) to shear large pieces, followed by a tight pestle (15-25 strokes) until the solution was uniform. Homogenate was filtered through a 30μm CellTrics filter (Sysmex, 04-0042-2316) into a LoBind tube (Eppendorf, 22431021) and pelleted (1000 rcf, 10 min at 4°C) (Eppendorf, 5920 R).

Pellets were resuspended in 2% formaldehyde in DPBS for 5 min at room temperature on a rotator. Glycine was added to obtain 0.2M glycine and tubes were incubated for 5 min on a rotator at room temperature. Lysate was centrifuged at 1,000 × g for 10 min at 4°C (in bucket rotor). Supernatant was removed and the pellet was resuspended in 1mL of cold DPBS. Lysate was centrifuged at 2,500 × g for 5 min at 4°C (in bucket rotor). Supernatant was removed and the pellets were used for in situ 3C following the steps in the Arima-3C Single-Cell Beta Kit (Arima Genomics), with some modifications. Briefly, 44uL of Conditioning Solution was added, and lysate was transferred to a 0.2 mL PCR tube, and mixed gently by pipetting. Tubes were incubated at 62°C for 30 min in a thermal cycler, with the lid temperature set to 85°C. 20µL of Stop Solution 2 was added, and mixed gently by pipetting. Tubes were incubated at 37°C for 15 min in a thermal cycler with a lid temperature set to 85°C. 28µL of the restriction enzyme digestion mix (9.8uL water, 9.2uL Buffer H, 4.5uL Enzyme H1, 4.5uL Enzyme H2) was added and gently pipette mixed. Tubes were incubated in a thermal cycler for 37°C for 60 min, 65°C for 20 min, then 25°C for 10 min. 82uL of ligation mix was added (70uL Buffer C, 12uL Enzyme C) and gently mixed by pipetting, then incubated at 25°C for 15 min. Lysate was transferred to a tube containing 850µL 1% BSA in DPBS on ice and filtered through a 30um Celltrics strainer to a new tube on ice. Lysate was centrifuged at 1,000 × g for 10 min at 4°C (in bucket rotor). Supernatant was removed and pellets were resuspended each sample in 800uL of 1% BSA in DPBS. Nuclei were stained with DRAQ7 at 1:200 and mixed gently with a pipette and incubated on ice for 5 min. One tube of Proteinase K (Zymo cat. no. D3001-2-D) was dissolved in 1mL of Proteinase K Resuspension Buffer (Zymo cat. no. D3001-2-B) then added to a mix of 15mL M-Digestion Buffer 2x (Zymo cat. no. D5021-9) and 14mL water to make pK Digestion Buffer. 2uL of pK Digestion Buffer were added to each well of an Eppendorf twin.tec® PCR Plate 384 LoBind®, skirted, 45 µL, PCR clean (Eppendorf cat. No. 0030129547). Single DRAQ7 positive nuclei were sorted into individual wells of the 384-well plate using a Sony SH800 cell sorter. Plates were sealed then incubated in a thermal cycler at 50°C for 20 min with a heated lid at 85°C for nuclei digestion.

Library preparation and Illumina sequencing for snm3C-seq

The snm3C-seq samples followed the library preparation protocol detailed previously (*21*, *92*). This protocol has been automated using the Beckman Coulter Biomek i7 liquid handlers for reactions in 384 and 96-well plates. The snm3C-seq libraries were shallow sequenced on the Illumina NextSeq 2000 to 2-10M reads to check quality. Successful libraries were deeply sequenced on the NovaSeq X Plus instrument, using one lane of 25B flow cell per 4 384-well plates and using 150 bp paired-end mode.

Mapping of snm3C-seq

The snm3C-seq FASTQ reads were demultiplexed and mapped using YAP (Yet Another Pipeline) software (cemba-data v1.6.9), as previously described (*92*). First, FASTQ files were demultiplexed for each cell barcode. Next, two-pass mapping was performed with bismark (v0.20), using bowtie2 (v2.3). BAM file processing and QC was performed using samtools (v1.9) and picard (v3.0.0). Chromatin contacts were called and methylome profiles were generated using Allcools (v1.0.23). All reads were mapped to the hg38 reference genome assembly.

Quality control of snm3C-seq

Low quality cells were removed by requiring Bismark mapping rate for both reads R1 and R2 > 0.5; total final reads > 100,000; mCCC level < 0.03; mCH level < 0.2; mCG level > 0.5. Potential multiplets were identified using the MethylScrublet function in AllCools software for each sample, and cells with a doublet score > 0.015 were removed. Further quality filtering involved removing a low-quality cluster. Specifically, a cluster consisting of 36 cells with hypomethylation of mixed cell type markers was excluded. Finally, 22,240 cells were included in the downstream analysis.

Methylome clustering analysis

For the cells passing quality control, we generated the methylome cell dataset (MCDS) from single-cell DNA methylome profiles of the snm3C-seq datasets using "allcools generate-dataset" command. This dataset comprises a cell-by-feature tensor that encompasses various genomic regions and types of DNA methylation contexts. We used hypomethylation probability (hypo-score) in CGN contexts across non-overlapping 5kb bins (chrom5kb) on the hg38 chromosome. And we used gene body regions ±2 kb for clustering annotation and integration with the 10x multiome dataset. We performed clustering by (1) binarizing the matrix; (2) excluding 5kb bins that had a nonzero methylation level in below 10% of the cells; (3) normalizing using term frequency–inverse document frequency (TF-IDF); and (4) reducing dimensionality using truncated singular value decomposition (SVD). Then, we applied batch balanced k-nearest neighbor (BBKNN) (*93*) to remove technical batch effects, followed by visualization using UMAP (*88*). Finally, we performed graph clustering using the Leiden algorithm (*94*).

Cluster-level DNA methylome analysis

After methylome clustering analysis, we merged the single-cell ALLC files into pseudo bulk level using the "allcools merge-allc" command. Next, we performed DMR calling using 'call_dms' and 'call_dmr' functions in AllCools. In brief, we calculated CpG differentially methylated sites (DMS) using a permutation-based root mean square test (*95*). The base calls of each pair of CpG sites were added before analysis. We then merged the DMS into DMR if they are (1) within 500 bp and (2) the minimum methylation difference was greater than or equal to 0.3 across samples. We applied the DMR calling framework across the cell clusters and cell clusters for each age group.

Chromatin contact matrix and imputation

After chromatin contact data were generated by the YAP pipeline, we used the cis long range contacts (contact anchors distance > 1,000 bp) and trans contacts to generate single-cell chromatin contact matrices without imputation (hereafter referred to as raw contact matrices) at three genome resolutions: chromosome 100-kb resolution for the chromatin compartment analysis; 25-kb bin resolution is for the chromatin domain boundary analysis; and 10-kb resolution. The raw cell-level contact matrices for all resolutions are stored in HDF5-based mcool format. We then used the scHiCluster package (v1.3.5) to perform contact matrix imputation. In brief, the scHiCluster imputes the sparse single-cell matrix in two steps: the first step is Gaussian convolution (pad = 1); the second step is to apply a random walk with restart algorithm on the convoluted matrix. The imputation is performed on each chromosome of each cell. For 100-kb matrices, the whole chromosome is imputed; for 25-kb matrices, we imputed contacts within 10.05 Mb; for 10-kb matrices, we imputed contacts with 5.05 Mb. The imputed matrices for each cell were stored in cool format. For most following analyses, cell matrices were aggregated into cell groups identified in the previous section and stored in cool format.

10x multiome and snm3C-seq single-cell integration

The mCG fraction of gene body regions ±2 kb ("hypomethylation gene score") was used for integration with the 10x multiome snRNA-seq dataset. The snRNA-seq data were normalized by dividing the counts of each cell by the total counts across all genes. And logarithmic transformation was applied to both datasets. Highly variable genes and gene body regions ±2 kb were identified independently from both snRNA-seq and snm3C-seq datasets, and the datasets were scaled. Before integration, the mCG fraction was multiplied by -1 to account for the negative correlation between DNA methylation level and gene expression. Following integration, dimensional reduction was performed

based on the shared set of highly variable genes. Harmony (*96*) was applied for a batch corrected embedding. Finally, cell type labels from the 10x multiome dataset were transferred to predict cell types in the snm3C-seq dataset using a Euclidean nearest neighbor descent approach implemented in PyNNDscent (*97*).

Age-correlation of genes, cCREs, and DMRs

To identify genes, cCREs, or DMRs with age-correlated activity in each cell type, we calculated PCC for pseudo bulk CPM (gene expression or chromatin accessibility) or mCG fractions (DMRs) versus donor age. Due to differences in cell numbers captured for each donor from a given cell type, we filtered donors and features, using different criteria for each modality. For gene expression, we removed donors that had less than 20,000 total RNA counts. For the remaining donors, we removed genes with low expression by requiring the total counts across remaining donors must be equal to or greater than 2 multiplied by the number of remaining donors. For example, in cell type n if there are 30 donors remaining after filtering, genes must have a total count of 60 or greater when counts are summed across these donors. For chromatin accessible cCREs, all cCREs identified in a given cell type were considered. Donors with less than an average of 1 count per number of cCREs, were removed from the correlation analysis. All DMRs identified in a given cell type were considered. Donors with less than an average of 1 count per number of DMRs, were removed from the correlation analysis. Pearson correlation *P*-values were converted to FDR adjusted *P*-values. Features with a FDR < 0.1 were considered for downstream analysis. We calculated an "Age score" for each donor using lists of either age-correlated genes, ATAC-seq cCREs, or DMRs. Age score was calculated by averaging the donor Z-score of CPM (genes or ATAC-seq cCREs, or mCG fraction). For negatively age-correlated features, we multiplied by -1 before averaging with positively correlated features. Data are displayed as a LOESS smooth curve using a span equal to 0.4.

Nonlinear aging analysis

To capture all potential nonlinear aging trajectories, we used edgeR function glmQLFit with default parameters to find differentially expressed genes, between all pairwise combinations of age groups: 20-40, 40-60, 60-80, 80-100. We summed RNA counts of each gene, within each donor, for each cell type separately. Genes and donors were filtered as described in the section "Age-correlation of genes, cCREs, and DMRs" and genes with *P*-adjusted < 0.05 for any comparison were combined and organized into 10 modules based on k-means clustering. To capture putative enhancers for genes of each module, we calculated Pearson correlation coefficients comparing chromatin accessibility of every cCRE across donors with the average expression of the genes for each donor for each module. This was repeated for each cell type separately. cCREs that were significantly correlated by FDR < 0.05 and Pearson rho > 0.6, were included as putative enhancers for the module being tested.

Quantifying TF motif activity

To assess potential contributions of TF motifs to age-correlated epigenomic activity, we identified the cCREs containing each TF motif from the JASPAR (*98*) CORE vertebrate database. We used FIMO (v.5.5.3) (*99*) to scan for all incidences of every chromatin accessible cCRE (500 bp fixed width). For each TF motif, we averaged the age-correlation Pearson correlation coefficients (described above) for chromatin accessibility, mCG fraction, or ABC target gene expression then z-score scaled those averages across all TF motifs in each cell type to obtain scaled TF motif activity values. We calculated FDR adjusted P-values after performing Wilcoxon rank sum test (unpaired, two-sided) for all PCC values of cCREs with a given motif versus a background set of cCREs. The background set was randomly selected from all cCREs in that cell type with the same ratio of being promoter-proximal (< 1kb from TSS) to promoter distal (> 1kb from TSS).

Astrocyte subclustering

We employed an iterative clustering approach to identify astrocyte subclusters. We subsetted the astrocytes and re-normalized using SCTransform, identifying 3,000 highly variable genes. After scaling, we performed PCA and used Harmony for batch correction. We then utilized the first 30 dimensions of the Harmony-corrected data to find neighbors and generate a UMAP visualization. To determine the optimal resolution for clustering, we implemented an iterative clustering method as described in Li *et al.* (*100*). We then used resolution 0.2 for sub clustering and found 7 subclusters. To identify putative tripartite astrocyte clusters, we calculated module scores using Seurat's AddModuleScore function for 18 tripartite synaptic genes (*71*) *NRXN1*, *NLGN3*, *ITGB1*, *ITGAV*, *CDH2*, *CADM3*, *CADM1*, *NCAM1*, *TGFB2*, *ATP1B2*, *EPHA4*, *GJB6*, *BCAN*, *CSPG5*, *SPARCL1*, *SDC2*, *SDC4*, *GPC1*.

Differential gene expression and chromatin accessibility analysis

We used MAST (*101*) on single-cell normalized counts to identify differentially expressed genes and chromatin accessible cCREs between cell type subclusters, namely Astro1 versus Astro2 and Micro1 versus Micro2. For Astro1 versus Astro2 differentially expressed genes, we used donor as a latent variable. We required 10% of cells to be expressing the gene, *P*-adjusted < 0.05, and log2 fold change > 1. For Astro1 versus Astro2 differential cCREs, we called peaks on astrocytes and then employed logistic regression as recommended in Signac for differential peak analysis. We used donor as a latent variable. We required 10% of cells to have the cCRE accessible, *P*-adjusted < 0.01. For Micro2 and Micro1 differential gene expression we used age as a latent variable. We considered genes expressed in 5% of cells and *P*-adj. < 0.01 and 1.5-fold difference. For Micro1 and Micro2 differential cCREs, we considered cCREs accessible with *P*-adj. < 10e-24 due to inflated *P*-values and to obtain an appropriate number of elements for downstream analysis.

Gene ontology enrichment analysis

We performed GO enrichment analysis using the GSEApy (*102*) and ClusterProfiler (*103*). For each gene set, we used GO biological process terms. We performed such analyses using the most appropriate background set: for astro and microglia subcluster marker genes, we used genes expressed in 10% of cells, and for age-correlated genes, we used genes tested for correlation after filtering low count genes (see above). For GO enrichment analysis for DMRs, we first identified DMRs using the 'call_dmr' function in AllCools and filtered the DMRs to include those with more than 3 DMSs. Next, we used GREAT (*104*) with default settings to assign DMRs to genes that were hypomethylated in either Micro1 or Micro2. GO enrichment analysis for Biological Process was performed using GREAT with a background gene set consisting of all DMRs in microglia.

Transcription factor motif enrichment analysis

For TF motif enrichment analysis, we utilized the HOMER database. We focused on peaks that correlated with age, using significant peaks with a FDR of 0.1. As background, we employed peaks accessible in the specific cell type of interest. The enrichment analysis was performed using these parameters, and we considered only statistically significant TFs in our results. This approach allowed us to identify age-related changes in TF binding patterns within our dataset. To assess TF activity at the single-cell level in excitatory or inhibitory neurons, we employed chromVAR (*105*). For identifying differential TF activity, we utilized the FindMarkers function from the Seurat package, applying it to the chromVAR z-scores. We set the mean.fxn parameter to rowMeans and fc.name to "avg_diff" for this analysis. To focus on the most relevant and robust results, we applied the following filtering criteria: 1) TFs showing activity in at least 25% of cells, 2) adjusted *P*-value < 0.05, 3) minimum difference of 10% in activity between old and young samples.

Immunofluorescent staining

Immunolabeling was conducted on 30μm-thick free-floating hippocampal coronal sections. Tissues underwent sodium citrate (BioLegend, San Diego, CA; #928502) antigen retrieval for 30 min at 90°C and were subsequently washed three times with 1x PBS for 5 min each after returning to room temperature. Sections were then bathed in lipofuscin autofluorescence blocking TrueBlack solution (Biotium, Fremont, CA; #23007) for 5 min, rinsed three times with 1xPBS, and incubated in blocking buffer containing 5% normal donkey serum (Jackson ImmunoResearch Laboratories, West Grove, PA; #017-000-121) and 0.075% (v/v) Triton X-100 in 1xPBS for 2 hours at room temperature. Brain slices were then incubated in blocking solution containing an antibody against astrocytic marker ALDH1L1 (LSBio, Shirley, MA; #LS-C796716; 1:200) for 24 hours at 4°C. Sections were then rinsed three times with 1xPBS and incubated for 2 hours in blocking solution containing Alexa Fluor 488-conjugated donkey anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories, West Grove, PA; #715-545-150) and DRAQ5 nuclear stain (ThermoFisher, Waltham, MA; #62254; 1:2,000). The tissues were then rinsed with 1xPBS and finally mounted and coverslipped with Fluoromount-G (SouthernBiotech, Birmingham, AL; #0100-01) for microscopy.

Brain section imaging and quantification

Hippocampal CA3 confocal micrographs were obtained using an Olympus FV3000 microscope. Serial optical sections were captured for each brain slice using a 20x objective with a 2x zoom and a 0.89 μm step interval for a total of 10 slices (z-stack). All images were acquired using identical conditions. Maximum intensity projections from z-stacks were processed using the Olympus FV31S-SW Fluoview software (Version 2.6, Olympus Life Science). The resulting 2D projection images (0.101 mm^2 ROI area for each) were analyzed using Photoshop. ALDH1L1-positive astrocytes were manually counted in hippocampal CA3 from young ($n = 5$; 20–38 y.o.) and aged ($n = 7$; 80–106 y.o.) donor groups. Values were normalized by the number of DRAQ5 nuclei and statistically evaluated using an unpaired Welch's t test and visualized using GraphPad Prism 9 (GraphPad Software, CA, USA.).

For microglia soma size quantification, three CA3 confocal micrographs were taken from each of the young ($n = 3$; 20-38 y.o.) and aged ($n = 6$; 80-106 y.o.) donor's hippocampus. IBA1+ microglia 2D projection images (0.0262 mm2 ROI area) were then analyzed using ImageJ/FIJI software. The polygon selection tool was used to manually outline microglia somas and later quantified using the measure tool. The resulting data was statistically analyzed using a linear mixed-effects model (LME) to account for correlated data arising from repeated measurements from single donors (*106*). LME analysis was performed on the average microglia soma size from each ROI (3 per donor) using the MATLAB "fitlme" function.

Identifying putative enhancer–gene pairs with the ABC model

We used the Activity-by-Contact (ABC) model (*27*) to identify putative enhancers and target genes in each cell type. In brief, the ABC model uses normalized contact frequencies from Hi-C data, along with a measure of enhancer activity, to predict putative enhancer–gene pairs. For each cell type, we provided normalized Hi-C matrices as bedpe files at 10 kb resolution, ATAC-seq BAM files for each cell type and a list of cCREs from ATAC-seq peaks identified in that same cell type. Only genes expressed with CPM 1 or greater were considered. Predictions were considered positive and used for downstream analyses if ABC score was 0.02 or greater.

Compartment analysis

We used "hicluster compartment" command in scHicluster v1.3.5 (*107*) to compute compartment score using the cluster-level imputed contact matrices at 100 kb resolution contact matrices obtained after downsampling cells. The downsampling was performed to ensure a comparable total number of contacts for pseudo bulk analysis, which was carried out per age group within each cell type.

Differential contacts and differential compartments

We used raw contact matrices to look for chromatin contacts in *cis* with large differences between age groups or subtypes. To remove the baseline bias introduced by differences in sequencing depth between conditions under comparison, we downsampled the total number of contacts in each condition to the same level and regenerated the raw contact matrices. For each comparison, we extracted the *cis* contacts within the top ±0.0005% percentile from the distribution of differences in coverage between the conditions.

To identify the differential compartments, we used dcHiC (*108*). First, raw PC values were generated for each chromosome and group, followed by quantile normalization of these PC values. Subsequently, for each 100 kb bin, Mahalanobis distance, a multivariate z-score measuring the degree to which each bin deviates as an outlier from the overall compartment score distribution across all datasets, was computed and used to determine statistical significance through FDR correction. A compartment was considered differential if its adjusted P-value was below 0.05.

Intra-TAD contact quantifications

We used scHicluster v1.3.5 (*107*) (hicluster domain) to identify TADs on downsampled raw contact matrices at 10 kb resolution. We counted the number of contacts by summing up the contacts within the boundaries of each TAD for every single cell. These intra-TAD contacts were divided by the number of cis long contacts + trans contacts for each cell to obtain the ratio of intra-TAD contacts in each age group or microglia subcluster.

DNA methylation aging clock

To predict biological age using DNA methylation, we calculated mCG fractions at the 353 CpG sites listed in the Horvath epigenetic clock (*37*) using "allcools allc-to-region-count" command. We employed the pyaging tool (*109*) for age prediction, utilizing KNN as an imputation strategy for missing CpG ratios. Biological age predictions were performed both for each cell type separately and for all cell types combined (whole tissue). We calculated Pearson correlation coefficients between chronological and predicted biological age to address data sparsity in cell types with lower abundance, we computed pseudo donor methylation ratios after grouping data from 2, 4, 5, or 8 donors with adjacent chronological ages, subsequently calculating their predicted biological age. Average chronological age for binned donors was used for correlation with predicted biological age. This approach allowed us to mitigate the impact of missing data points to still evaluate biological age predictions for cell types with lower abundance.

CTCF CUT&Tag

All buffers were prepared as described in Kaya-Okur *et al*., 2019 (*73*) without the addition of digitonin. 30mg of pulverized postmortem human hippocampus tissue was obtained from seven neurotypical donors from NIH Neurobiobank: 1) Donor ID 5579, female, age 25; 2) Donor ID 935, male, age 38; 3) Donor ID 5711, male, age 39; 4) Donor ID 81158, female, 82; 5) Donor ID 233848, female, 85; 6) Donor ID 4320, male, age 87; 7) Donor ID 6042, male, age 90. Tissue was dounced in 1mL NE1 Buffer with a loose-fitting pestle (5-10 strokes) followed by a tight-fitting pestle (15-25 strokes). Lysates were filtered through 30μM CellTrics filters (Sysmex, Cat#04-0042-2316). After centrifugation for 4 min at 1300 g at 4°C, nuclei were resuspended in PBS supplemented with Roche cOmplete EDTA-free protease inhibitor (Roche, Cat# 11873580001). 300K nuclei were cross-linked with formaldehyde (Thermo Scientific, Cat#PI28906) at a final concentration of 0.1% at room temperature for 2 min, followed by quenching with 1M glycine. Cross-linked nuclei were centrifuged for 4 min at 1300 g at 4°C and resuspended in 100-300uL of Wash Buffer.

10uL of ConA beads per sample (Bangs Labs, Cat# BP531) were washed 3x in 1mL Binding Buffer and resuspended in Binding Buffer at the original slurry volume. 10uL of beads were added to the nuclei and tubes were rotated for 5-10 min at room temperature before placing on a magnetic stand. Supernatant was removed and beads were resuspended in 100uL Antibody Buffer. CTCF antibody (CST, Cat#3418S) was added at a 1:100 dilution and samples were rotated overnight at 4°C.

The next morning, the supernatant was removed and the nuclei were resuspended in 300uL Was Buffer with secondary antibody (Novus, Cat#NBP1-72763) at a 1:100 dilution. Samples were rotated for 1 hour at room temperature, then washed 3x in 1mL Wash Buffer. After washing, pA-Tn5 (EpiCypher, Cat#15-1017) was added at a 1:50 dilution in 300uL of Dig-300 Buffer and samples were rotated for 1 hour at room temperature followed by washing 3x in Dig-300 Buffer.

After removing the supernatant, 300uL Tagmentation Buffer was added to the nuclei and samples were incubated at 37°C for 1 hour with shaking at 500rpm. Tagmentation was stopped by adding 10 uL 0.5 M EDTA, 3 uL 10% SDS, and 2.5 uL 20 mg/mL Proteinase K. Samples were vortexed and incubated for 1 hour at 55°C with shaking at 500rpm to digest and reverse cross-links. DNA was extracted using a standard Phenol-Chloroform extraction protocol. PCR was carried out using Nextera i5 and i7 primers and NEBNext HiFi 2x PCR MM (NEB, Cat#M0541S). PCR products were purified using VAHTS DNA Clean Beads (Vazyme, Cat#N411-01) at 1.3x volume and eluted in 10mM Tris-HCl pH 8.

FASTQ files were processed using cutadapt v5.0 103 and trim galore v0.6.10 104 with -q 20 to remove low quality ends and adaptor sequences, then aligned to hg38 using bowtie2 v2.5.4. Bam files were deduplicated using samtools fixmate -m then samtools markdup -r. RPKM normalized bigwig files were generated using bamCoverage –normalizeUsing RPKM -bs 10. Peaks were called after merging bam files from all samples with macs2 callpeak and–shift -75–ext 150 -g hs -q 0.05.

Subclustering Microglia snATAC-seq

To identify Micro1 and Micro2 subclusters in snATAC-seq datasets, we set variable features to microglia cCREs (ATAC-seq peaks; table S3) that overlapped marker DMRs in Micro1 and Micro2 snm3C-seq (table S13). This yielded 3,405 cCREs as features. For UMAP dimension reduction we included PCs 2:6. Nuclei separated into two major clusters and were subjectively labeled as Micro1 or Micro2 based on the proportional difference in their donor age makeup. We merged snATAC-seq data downloaded from NCBI GEO (see External dataset section) from two studies (*15*, *56*) that we reprocessed with cellranger-arc using all microglia cCREs identified in our study as count features. We merged all three datasets, and clustered using the same 3,405 features and method described above. For each donor, proportions of Micro1 and Micro2 cells were determined.

Machine learning-based classification of cell types

We used XGBoost to build the cell type classifier to identify potential development origins of Micro1 and Micro2. To train the classifier, we downloaded snm3C-seq allc files from GSE213950 for infant MGC-1 and snm3C-seq allc files from Zhou *et al*., 2025 (*51*) for peripheral blood cells, skin cells and muscle fibroblasts. We used two methods to preprocess the data as input for the XGBoost model. In the first method, we applied LSI on the CGN rate for all 5kb nonoverlapping genomic bins to generate the top principal components by the function lsi from allcools package, and then ran harmony correction (harmonypy version 0.0.10) to remove data variations across different assays. In the second method, we applied the trinarization method described in Tian *et al*., 2023 (*25*) to the CGN rate for all the DMRs between Micro1 and Micro2 previously identified. For both methods, we applied the XGBoost algorithm (pyboost version 2.1.1) to the pre-processed data by splitting the input into train, evaluation and test sets in the ratio of 7:2:1. We set the round of boosting to 3000 and early stopping round at 100 in the xgboost.train function. Other notable parameters set for the training function are: learning rate at 0.1, max tree depth at 5, sampling rate for each tree at 0.8. We selected the model with the lowest multivariate log loss on the evaluation set as the final prediction model and applied it to the test sets that were pre-processed in the same way as the model input. The prediction accuracies presented in this paper were generated on the test dataset.

External dataset

NRF1 ChIP-seq in neuroblastoma cell line SK-N-SH signal *P*-value bigwig (Accession ENCFF762LEE) was downloaded from the ENCODE portal (https://www.encodeproject.org/experiments/ENCSR000EHZ/). Raw sn ATAC-seq data from NCBI GEO accessions GSE147672 and GSE261983 were downloaded and analyzed for microglia clustering. Sn DNA methylation data from peripheral blood cells, skin cells, and muscle fibroblasts were downloaded from https://huggingface.co/datasets/zhoujt1994/HumanCellEpigenomeAtlas_sc_allc/tree/main/, and data from microglia of 4- and 7-month-old infant brains were obtained from GEO accession GSE213950. 10x multiome human PBMC data was obtained from 10x Genomics website at this link: https://www.10xgenomics.com/datasets/10-k-human-pbm-cs-multiome-v-1-0-chromium-x-1-standard-2-0-0. Primary human blood monocyte Hi-C file was obtained at this link: https://ftp.ebi.ac.uk/biostudies/fire/E-MTAB-/848/E-MTAB-10848/Files/HiCseq_MO_freshly.isolated.rep1_donorA.hic.

REFERENCES AND NOTES

1. Y. Hou *et al.*, Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019). doi: 10.1038/s41582-019-0244-7; pmid: 31501588
2. S. M. Butler, Caring for an aging US population-the good news and the bad news. *JAMA Health Forum* **5**, e241893 (2024). doi: 10.1001/jamahealthforum.2024.1893; pmid: 38780936
3. J. Golomb *et al.*, Hippocampal atrophy in normal aging. An association with recent memory impairment. *Arch. Neurol.* **50**, 967–973 (1993). doi: 10.1001/archneur.1993.00540090066012; pmid: 8363451
4. C. R. Jack Jr *et al.*, Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* **51**, 993–999 (1998). doi: 10.1212/WNL.51.4.993; pmid: 9781519
5. L. E. B. Bettio, L. Rajendran, J. Gil-Mohapel, The effects of aging in the hippocampus and cognitive decline. *Neurosci. Biobehav. Rev.* **79**, 66–86 (2017). doi: 10.1016/j.neubiorev.2017.04.030; pmid: 28476525
6. L. Erraji-Benchekroun *et al.*, Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. *Biol. Psychiatry* **57**, 549–558 (2005). doi: 10.1016/j.biopsych.2004.10.034; pmid: 15737671
7. N. C. Berchtold *et al.*, Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15605–15610 (2008). doi: 10.1073/pnas.0806883105; pmid: 18832152
8. J. M. Zullo *et al.*, Regulation of lifespan by neural excitation and REST. *Nature* **574**, 359–364 (2019). doi: 10.1038/s41586-019-1647-8; pmid: 31619788
9. S. Ham, S. V. Lee, Advances in transcriptome analysis of human brain aging. *Exp. Mol. Med.* **52**, 1787–1797 (2020). doi: 10.1038/s12276-020-00522-6; pmid: 33244150
10. L. Soreq *et al.*, ; North American Brain Expression Consortium, J. Rose, E. Soreq, J. Hardy, D. Trabzuni, M. R. Cookson, C. Smith, M. Ryten, R. Patani, J. Ule, Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. *Cell Rep.* **18**, 557–570 (2017). doi: 10.1016/j.celrep.2016.12.011; pmid: 28076797
11. L. Tan *et al.*, Lifelong restructuring of 3D genome architecture in cerebellar granule cells. *Science* **381**, 1112–1119 (2023). doi: 10.1126/science.adh3253; pmid: 37676945
12. W. E. Allen, T. R. Blosser, Z. A. Sullivan, C. Dulac, X. Zhuang, Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194–208.e18 (2023). doi: 10.1016/j.cell.2022.12.010; pmid: 36580914
13. K. Jin *et al.*, Brain-wide cell-type-specific transcriptomic signatures of healthy ageing in mice. *Nature* **638**, 182–196 (2025). doi: 10.1038/s41586-024-08350-8; pmid: 39743592
14. Y. Zhang *et al.*, Single-cell epigenome analysis reveals age-associated decay of heterochromatin domains in excitatory neurons in the mouse brain. *Cell Res.* **32**, 1008–1021 (2022). doi: 10.1038/s41422-022-00719-6; pmid: 36207411
15. P. S. Emani *et al.*, Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024). doi: 10.1126/science.adi5199; pmid: 38781369
16. J.-F. Chien *et al.*, Cell-type-specific effects of age and sex on human cortical neurons. *Neuron* **112**, 2524–2539.e5 (2024). doi: 10.1016/j.neuron.2024.05.013; pmid: 38838671

17. S. S. P. Rao *et al.*, Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017). doi: 10.1016/j.cell.2017.09.026; pmid: 28985562
18. J. R. Dixon *et al.*, Topological domains in Mamm. Genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012). doi: 10.1038/nature11082; pmid: 22495300
19. E. P. Nora *et al.*, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22 (2017). doi: 10.1016/j.cell.2017.05.004; pmid: 28525758
20. E. Alipour, J. F. Marko, Self-organization of domain structures by DNA-loo*P*-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012). doi: 10.1093/nar/gks925; pmid: 23074191
21. D.-S. Lee *et al.*, Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019). doi: 10.1038/s41592-019-0547-z; pmid: 31501549
22. G. Li *et al.*, Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019). doi: 10.1038/s41592-019-0502-z; pmid: 31384045
23. D. Franjic *et al.*, Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469.e14 (2022). doi: 10.1016/j.neuron.2021.10.036; pmid: 34798047
24. C. Luo *et al.*, Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017). doi: 10.1126/science.aan3351; pmid: 28798132
25. W. Tian *et al.*, Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382**, eadf5357 (2023). doi: 10.1126/science.adf5357; pmid: 37824674
26. D. Li *et al.*, Epigenome Browser update 2022. *Nucleic Acids Res.* **50**, W774–W781 (2022). doi: 10.1093/nar/gkac238; pmid: 35412637
27. C. P. Fulco *et al.*, Activity-by-contact model of enhancer-promoter regulation from thoUSA.nds of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019). doi: 10.1038/s41588-019-0538-0; pmid: 31784727
28. Y. Ding *et al.*, Comprehensive human proteome profiles across a 50-year lifespan reveal aging trajectories and signatures. *Cell* **188**, 5763–5784.e26 (2025). doi: 10.1016/j.cell.2025.06.047; pmid: 40713952
29. A. Montagne *et al.*, Blood-brain barrier breakdown in the aging human hippocampus. *Neuron* **85**, 296–302 (2015). doi: 10.1016/j.neuron.2014.12.032; pmid: 25611508
30. X. Shen *et al.*, Nonlinear dynamics of multiomics profiles during human aging. *Nat. Aging* **4**, 1619–1634 (2024). doi: 10.1038/s43587-024-00692-2; pmid: 39143318
31. J. A. Smith, Regulation of cytokine production by the unfolded protein response; Implications for infection and autoimmunity. *Front. Immunol.* **9**, 422 (2018). doi: 10.3389/fimmu.2018.00422; pmid: 29556237
32. C.-J. Wei *et al.*, Inhibition of activator protein 1 attenuates neuroinflammation and brain injury after experimental intracerebral hemorrhage. *CNS Neurosci. Ther.* **25**, 1182–1188 (2019). doi: 10.1111/cns.13206; pmid: 31392841
33. F. T. Gallo, C. Katche, J. F. Morici, J. H. Medina, N. V. Weisstaub, Immediate Early Genes, memory and psychiatric disorders: Focus on c-Fos, Egr1 and arc. *Front. Behav. Neurosci.* **12**, 79 (2018). doi: 10.3389/fnbeh.2018.00079; pmid: 29755331
34. R. C. Scarpulla, Nuclear activators and coactivators in mammalian mitochondrial biogenesis. *Biochim. Biophys. Acta* **1576**, 1–14 (2002). doi: 10.1016/S0167-4781(02)00343-3; pmid: 12031478
35. R. M. Barrientos, M. M. Kitt, L. R. Watkins, S. F. Maier, Neuroinflammation in the normal aging hippocampus. *Neuroscience* **309**, 84–99 (2015). doi: 10.1016/j.neuroscience.2015.03.007; pmid: 25772789
36. S. C. Woodburn, J. L. Bollinger, E. S. Wohleb, The semantics of microglia activation: Neuroinflammation, homeostasis, and stress. *J. Neuroinflammation* **18**, 258 (2021). doi: 10.1186/s12974-021-02309-6; pmid: 34742308
37. S. Horvath, DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013). doi: 10.1186/gb-2013-14-10-r115; pmid: 24138928
38. I. R. Holtman, D. Skola, C. K. Glass, Transcriptional control of microglia phenotypes in health and disease. *J. Clin. Invest.* **127**, 3220–3229 (2017). doi: 10.1172/JCI90604; pmid: 28758903
39. S. E. van der Krieken, H. E. Popeijus, R. P. Mensink, J. Plat, CCAAT/enhancer binding protein β in relation to ER stress, inflammation, and metabolic disturbances. *BioMed Res. Int.* **2015**, 324815 (2015). doi: 10.1155/2015/324815; pmid: 25699273
40. Y. He, J. R. Ecker, Non-CG methylation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015). doi: 10.1146/annurev-genom-090413-025437; pmid: 26077819
41. X. Huang *et al.*, ZFP281 controls transcriptional and epigenetic changes promoting mouse pluripotent state transitions via DNMT3 and TET1. *Dev. Cell* **59**, 465–481.e6 (2024). doi: 10.1016/j.devcel.2023.12.018; pmid: 38237590
42. F. Ginhoux, S. Lim, G. Hoeffel, D. Low, T. Huber, Origin and differentiation of microglia. *Front. Cell. Neurosci.* **7**, 45 (2013). doi: 10.3389/fncel.2013.00045; pmid: 23616747
43. J. Bastos *et al.*, Monocytes can efficiently replace all brain macrophages and fetal liver monocytes can generate bona fide SALL1^{+} microglia. *Immunity* **58**, 1269–1288.e12 (2025). doi: 10.1016/j.immuni.2025.04.006; pmid: 40311613
44. H. Lund *et al.*, Competitive repopulation of an empty microglial niche yields functionally distinct subsets of microglia-like cells. *Nat. Commun.* **9**, 4845 (2018). doi: 10.1038/s41467-018-07295-7; pmid: 30451869
45. G. Zhang *et al.*, Infiltration by monocytes of the central nervous system and its role in multiple sclerosis: Reflections on therapeutic strategies. *Neural Regen. Res.* **20**, 779–793 (2025). doi: 10.4103/NRR.NRR-D-23-01508; pmid: 38886942
46. E. Kim, S. Cho, Microglia and monocyte-derived macrophages in stroke. *Neurotherapeutics* **13**, 702–718 (2016). doi: 10.1007/s13311-016-0463-1; pmid: 27485238
47. C. Muñoz-Castro *et al.*, Monocyte-derived cells invade brain parenchyma and amyloid plaques in human Alzheimer's disease hippocampus. *Acta Neuropathol. Commun.* **11**, 31 (2023). doi: 10.1186/s40478-023-01530-z; pmid: 36855152
48. N. H. Varvel *et al.*, Infiltrating monocytes promote brain inflammation and exacerbate neuronal damage after status epilepticus. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5665–E5674 (2016). doi: 10.1073/pnas.1604263113; pmid: 27601660
49. J. Dietrich *et al.*, Bone marrow drives central nervous system regeneration after radiation injury. *J. Clin. Invest.* **128**, 281–293 (2018). doi: 10.1172/JCI90647; pmid: 29202481
50. A. Mildner *et al.*, Microglia in the adult brain arise from Ly-6ChiCCR2+ monocytes only under defined host conditions. *Nat. Neurosci.* **10**, 1544–1553 (2007). doi: 10.1038/nn2015; pmid: 18026096
51. J. Zhou *et al.*, Human body single-cell atlas of 3D genome organization and DNA methylation. bioRxivorg 2025.03.23.644697 [Preprint] (2025); https://doi.org/10.1101/2025.03.23.644697.
52. S. Welte, S. Kuttruff, I. Waldhauer, A. Steinle, Mutual activation of natural killer cells and monocytes mediated by NKp80-AICL interaction. *Nat. Immunol.* **7**, 1334–1342 (2006). doi: 10.1038/ni1402; pmid: 17057721
53. M. G. Heffel *et al.*, Temporally distinct 3D multiomic dynamics in the developing human brain. *Nature* **635**, 481–489 (2024). doi: 10.1038/s41586-024-08030-7; pmid: 39385032
54. G. C. Hon *et al.*, Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013). doi: 10.1038/ng.2746; pmid: 23995138
55. N. Loyfer *et al.*, A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023). doi: 10.1038/s41586-022-05580-6; pmid: 36599988
56. M. R. Corces *et al.*, Single-cell epigenomic analyses implicate candidate caUSA.l variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020). doi: 10.1038/s41588-020-00721-x; pmid: 33106633
57. N. C. Nicolaides, G. Chrousos, T. Kino, *Glucocorticoid Receptor*, MDText.com, Inc., (2020); https://www.ncbi.nlm.nih.gov/books/NBK279171/
58. D. Gomez-Nicola, B. Valle-Argos, M. Nieto-Sampedro, Blockade of IL-15 activity inhibits microglial activation through the NFkappaB, p38, and ERK1/2 pathways, reducing cytokine and chemokine release. *Glia* **58**, 264–276 (2010). doi: 10.1002/glia.20920; pmid: 19610094
59. J. Minderjahn *et al.*, Postmitotic differentiation of human monocytes requires cohesin-structured chromatin. *Nat. Commun.* **13**, 4301 (2022). doi: 10.1038/s41467-022-31892-2; pmid: 35879286
60. M. C. O'Donovan *et al.*, Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053–1055 (2008). doi: 10.1038/ng.201; pmid: 18677311
61. B. Hussain, C. Fang, J. Chang, Blood-Brain Barrier Breakdown: An Emerging Biomarker of Cognitive Impairment in Normal Aging and Dementia. *Front. Neurosci.* **15**, 688090 (2021). doi: 10.3389/fnins.2021.688090; pmid: 34489623
62. N. Chalmers, E. Masouti, R. Beckervordersandforth, Astrocytes in the adult dentate gyrus-balance between adult and developmental tasks. *Mol. Psychiatry* **29**, 982–991 (2024). doi: 10.1038/s41380-023-02386-4; pmid: 38177351
63. S. He, N. E. Sharpless, Senescence in health and disease. *Cell* **169**, 1000–1011 (2017). doi: 10.1016/j.cell.2017.05.015; pmid: 28575665
64. N. Plotegher, M. R. Duchen, Mitochondrial Dysfunction and Neurodegeneration in Lysosomal Storage Disorders. *Trends Mol. Med.* **23**, 116–134 (2017). doi: 10.1016/j.molmed.2016.12.003; pmid: 28111024
65. D. Glick, S. Barth, K. F. Macleod, Autophagy: Cellular and molecular mechanisms. *J. Pathol.* **221**, 3–12 (2010). doi: 10.1002/path.2697; pmid: 20225336
66. D. Denton, S. Kumar, Autophagy-dependent cell death. *Cell Death Differ.* **26**, 605–616 (2019). doi: 10.1038/s41418-018-0252-y; pmid: 30568239
67. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
68. Z. Wu *et al.*, Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999). doi: 10.1016/S0092-8674(00)80611-X; pmid: 10412986
69. J.-M. Bouteiller, T. W. Berger, "Tripartite Synapse (Neuron–Astrocyte Interactions), Conductance Models" in *Encyclopedia of Computational Neuroscience*, D. Jaeger, R. Jung, Eds. (Springer, 2014), pp. 1–4.
70. I. Farhy-Tselnicker, N. J. Allen, Astrocytes, neurons, synapses: A tripartite view on cortical circuit development. *Neural Dev.* **13**, 7 (2018). doi: 10.1186/s13064-018-0104-y; pmid: 29712572

71. A. E. J. Hillen, J. P. H. Burbach, E. M. Hol, Cell adhesion and matricellular support by astrocytes of the tripartite synapse. *Prog. Neurobiol.* **165-167**, 66–86 (2018). doi: 10.1016/j.pneurobio.2018.02.002; pmid: 29444459
72. E. Ling *et al.*, A concerted neuron-astrocyte program declines in ageing and schizophrenia. *Nature* **627**, 604–611 (2024). doi: 10.1038/s41586-024-07109-5; pmid: 38448582
73. H. S. Kaya-Okur *et al.*, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019). doi: 10.1038/s41467-019-09982-5; pmid: 31036827
74. A. C. Bell, G. Felsenfeld, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000). doi: 10.1038/35013100; pmid: 10839546
75. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
76. J.-S. Kim *et al.*, Clonal hematopoiesis-associated motoric deficits caused by monocyte-derived microglia accumulating in aging mice. *Cell Rep.* **44**, 115609 (2025). doi: 10.1016/j.celrep.2025.115609; pmid: 40279248
77. H. Bouzid *et al.*, Clonal hematopoiesis is associated with protection from Alzheimer's disease. *Nat. Med.* **29**, 1662–1670 (2023). doi: 10.1038/s41591-023-02397-2; pmid: 37322115
78. P. Réu *et al.*, The lifespan and turnover of microglia in the human brain. *Cell Rep.* **20**, 779–784 (2017). doi: 10.1016/j.celrep.2017.07.004; pmid: 28746864
79. A. Verkhratsky *et al.*, Astroglial asthenia and loss of function, rather than reactivity, contribute to the ageing of the brain. *Pflugers Arch.* **473**, 753–774 (2021). doi: 10.1007/s00424-020-02465-3; pmid: 32979108
80. L. Labarta-Bajo, N. J. Allen, Astrocytes in aging. *Neuron* **113**, 109–126 (2025). doi: 10.1016/j.neuron.2024.12.010; pmid: 39788083
81. W. Lieberthal, S. A. Menza, J. S. Levine, Graded ATP depletion can cause necrosis or apoptosis of cultured mouse proximal tubular cells. *Am. J. Physiol.* **274**, F315–F327 (1998). pmid: 9486226
82. V. Shacham-Silverberg *et al.*, Phosphatidylserine is a marker for axonal debris engulfment but its exposure can be decoupled from degeneration. *Cell Death Dis.* **9**, 1116 (2018). doi: 10.1038/s41419-018-1155-z; pmid: 30389906
83. L. Magtanong, P. J. Ko, S. J. Dixon, Emerging roles for lipids in non-apoptotic cell death. *Cell Death Differ.* **23**, 1099–1109 (2016). doi: 10.1038/cdd.2016.25; pmid: 26967968
84. M. R. Corces *et al.*, The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018). doi: 10.1126/science.aav1898; pmid: 30361341
85. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024). doi: 10.1038/s41587-023-01767-y; pmid: 37231261
86. K. Zhang, N. R. Zemke, E. J. Armand, B. Ren, A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**, 217–227 (2024). doi: 10.1038/s41592-023-02139-9; pmid: 38191932
87. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019). doi: 10.1016/j.cels.2019.03.003; pmid: 30954475
88. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018). doi: 10.21105/joss.00861
89. Z. Yao *et al.*, A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021). doi: 10.1038/s41586-021-03500-8; pmid: 34616066
90. T. E. Bakken *et al.*, Author Correction: Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **604**, E8 (2022). doi: 10.1038/s41586-022-04562-y; pmid: 35319013
91. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
92. H. Liu *et al.*, DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021). doi: 10.1038/s41586-020-03182-8; pmid: 34616061
93. K. Polański *et al.*, BBKNN: Fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020). doi: 10.1093/bioinformatics/btz625; pmid: 31400197
94. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019). doi: 10.1038/s41598-019-41695-z; pmid: 30914743
95. Y. He *et al.*, Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature* **583**, 752–759 (2020). doi: 10.1038/s41586-020-2119-x; pmid: 32728242
96. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: 10.1038/s41592-019-0619-0; pmid: 31740819
97. W. Dong, C. Moses, K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures" in *Proceedings of the 20th International Conference on World Wide Web* (Association for Computing Machinery, 2011), pp. 577–586. doi: 10.1145/1963405.1963487
98. A. Sandelin, W. Alkema, P. Engström, W. W. Wasserman, B. Lenhard, JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004). doi: 10.1093/nar/gkh012; pmid: 14681366
99. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). doi: 10.1093/bioinformatics/btr064; pmid: 21330290
100. Y. E. Li *et al.*, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021). doi: 10.1038/s41586-021-03604-1; pmid: 34616068
101. G. Finak *et al.*, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015). doi: 10.1186/s13059-015-0844-5; pmid: 26653891
102. Z. Fang, X. Liu, G. Peltz, GSEApy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023). doi: 10.1093/bioinformatics/btac757; pmid: 36426870
103. T. Wu *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021). doi: 10.1016/j.xinn.2021.100141; pmid: 34557778
104. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010). doi: 10.1038/nbt.1630; pmid: 20436461
105. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706
106. Z. Yu *et al.*, Beyond t test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron* **110**, 21–35 (2022). doi: 10.1016/j.neuron.2021.10.030; pmid: 34784504
107. J. Zhou *et al.*, Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14011–14018 (2019). doi: 10.1073/pnas.1901423116; pmid: 31235599
108. A. Chakraborty, J. G. Wang, F. Ay, dcHiC detects differential compartments across multiple Hi-C datasets. *Nat. Commun.* **13**, 6827 (2022). doi: 10.1038/s41467-022-34626-6; pmid: 36369226
109. L. P. de Lima Camillo; de L. Camillo, pyaging: A Python-based compendium of GPU-optimized aging clocks. *Bioinformatics* **40**, btae200 (2024). doi: 10.1093/bioinformatics/btae200; pmid: 38603598
110. N. R. Zemke, S. Lee, S. Mamde, B. Yang, nrzemke/aging_human_hippocampus: aging_human_hippocampus, Version v1.0, Zenodo (2026); https://doi.org/10.5281/zenodo.19391233.

ACKNOWLEDGMENTS

This publication is part of the 4D Nucleome consortium. We thank the Center for Neural Circuit Mapping (CNCM) Brain Procurement Program at the University of California, Irvine, and the NIH NeuroBioBank for outstanding technical and administrative support for tissue collection, preservation, and characterization; we also thank the brain donors and their loved ones, without whom this research would be impossible. We acknowledge all members of the Ren and Xu labs for their helpful input. We thank L. Stern for the datahub logo. This publication includes data generated at the UC San Diego IGM Genomics Center using an Illumina NovaSeq 6000 that was purchased with funding from the National Institutes of Health (SIG grant S10 OD026929 to K.J.). **Funding:** This work was supported by the National Institutes of Health Common Fund 4D Nucleome Program grant 1U01DA052769 and the National Institute on Aging (NIA) grants R01AG067153, R01AG082127, and NIH AG083977, as well as Alzheimer's Association grant ADSF-21-816672. **Author contributions:** Conceptualization: B.R., X.X., and C.W.C. Investigation: N.R.Z., N.B., H.S.I., W.M.B., P.K.L., K.D., B.M.G., E.H., A.Y., Y.T., C.C., Q.Z., V.A., and L.T. Formal analysis: N.R.Z., S.L., S.M., B.Y., B.M.G. Visualization: N.R.Z., S.L., S.M., B.Y., B.M.G., C.S., D.L., T.W. Funding acquisition: B.R., X.X., C.W.C. Writing – original draft: N.R.Z., S.L., S.M., B.R., C.K.G. Writing – review & editing: All authors contributed. **Competing interests:** B.R. is a cofounder and consultant of Arima Genomics and cofounder of Epigenome Technologies. C.K.G. is a cofounder, equity holder and member of the Scientific Advisory Board of Asteroid Therapeutics. J.R.E. is a scientific adviser for Zymo Research Inc. and Ionis Pharmaceuticals. **Data, code, and materials availability:** Data produced in this study are available at the NCBI GEO under accession number GSE278576 for 10x multiome, GSE299139 for snm3C-seq, and GSE299899 for CTCF CUT&Tag. Data are also available at the 4DN portal (https://data.4dnucleome.org/ren-lab-hippocampus-single-cell-analysis). Genome browser visualization is available at the WashU Epigenome Browser datahub link: https://epigenome.wustl.edu/seahorse/. Scripts used for data processing and analysis are available at Zenodo (*110*) and on github: https://github.com/nrzemke/aging_human_hippocampus. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the Author Accepted Manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt8307
Figs. S1 to S21; Tables S1 to S28; MDAR Reproducibility Checklist

Submitted 14 October 2024; resubmitted 7 December 2025; accepted 6 April 2026
10.1126/science.adt8307

RESEARCH ARTICLE SUMMARY

Human body single-cell atlas of three-dimensional genome organization and DNA methylation

Jingtian Zhou†, Yue Wu† *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adx0673

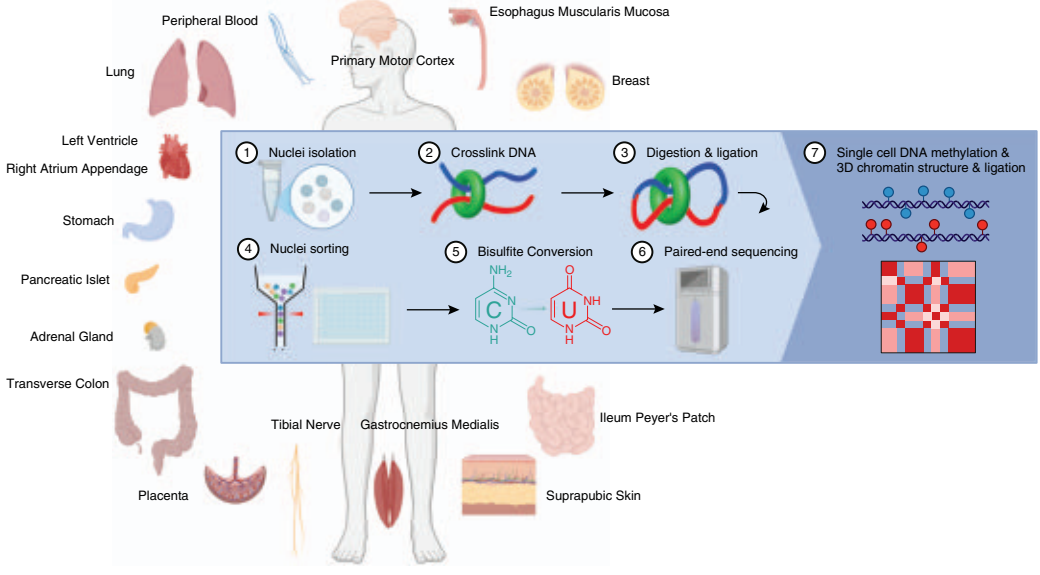
INTRODUCTION: Each cell type in the human body maintains a distinct identity through a specific program of gene expression. Two key features that regulate this program are DNA methylation and the three-dimensional (3D) folding of chromosomes inside the nucleus; these features influence which genes are active or silent in a given cell. However, how DNA methylation and 3D genome architecture vary across hundreds of distinct cell types in the human body has remained largely uncharacterized.

RATIONALE: Previous efforts to map DNA methylation and 3D genome structure have focused on bulk tissue samples, which average signals across cell types and obscure cell type–specific patterns. More recent work using single-cell methods has begun to characterize DNA methylation and 3D genome structure in complex tissues. Still, these studies have been applied only to limited tissues or studied in isolation. To overcome these limitations, we used a method called single-nucleus methyl-3C sequencing (snm3C-seq), which simultaneously measures both DNA methylation and 3D chromatin contacts in individual nuclei. Applying snm3C-seq to a panel of adult human tissues, we constructed a comprehensive atlas of DNA methylation and 3D genome structure across a broad range of cell types, enabling direct comparison of how these epigenomic features relate to one another across lineages.

RESULTS: We profiled 86,689 single nuclei from 16 human tissues, identifying 35 major cell types and 206 subtypes defined by their DNA methylation and 3D genome signatures. We found that canonical CG methylation varies widely across cell types, with certain lineages, including trophoblast cells, memory B cells, and acinar epithelial cells, showing distinctive domains of partial methylation that correspond to repressive genomic regions. We further found that low-level non-CG methylation, previously thought to be largely confined to neurons, is present across a broad range of cell types, retaining information about cell identity and active regulatory elements. Characterizing 3D chromatin architecture at previously unattained cellular resolution, we identified hundreds of thousands of chromatin loops that vary between cell types and are correlated with cell type–specific gene expression. We also found that different cell lineages exhibit distinct modes of 3D genome organization, with some, such as neurons, characterized by strong local chromatin domains. By contrast, other cell types, such as hematopoietic cells, are dominated by long-range compartment interactions. Comparing DNA methylation and 3D genome structure, we observe that the two epigenomic modalities together provide a richer view of cellular diversity than either feature alone and that in some lineages, such as skeletal muscle fibers and Schwann cells, the two features identify overlapping but distinct cell states, potentially reflecting cells in transition between differentiation states.

CONCLUSION: This study provides a comprehensive atlas of DNA methylation and 3D genome architecture across human cell types. It reveals that both features carry distinct, complementary information about cell identity and that the relationship between them varies across lineages in ways that reflect underlying differences in chromatin regulation and cell biology. The dataset, accessible through an interactive browser, offers a resource for investigating the epigenomic basis of gene regulation, cell differentiation, and human disease. □

Corresponding authors: Jingtian Zhou (jingtian.zhou@arcinstitute.org); Jesse R. Dixon (jedixon@salk.edu); Joseph R. Ecker (ecker@salk.edu) †These authors contributed equally to this work. **Cite this article as J. Zhou *et al.*, *Science* 393, eadx0673 (2026). DOI: 10.1126/science.adx0673**



Single-cell atlas of DNA methylation and 3D genome organization across the human body. Simultaneous profiling of DNA methylation and chromatin folding in 86,689 single nuclei from 16 human tissues reveals 35 major cell types and 206 cell subtypes. The two epigenomic layers carry distinct, lineage-specific information about cell identity and gene regulation, providing a comprehensive reference for human biology and disease. [Figure created with BioRender.com; J. Zhou (2026), https://BioRender.com/23dupwc]

RESEARCH ARTICLE

Human body single-cell atlas of three-dimensional genome organization and DNA methylation

Jingtian Zhou[1,2,3]*†, Yue Wu[2,4]†, Hanqing Liu[5,2], Wei Tian[2], Rosa G. Castanon[2], Anna Bartlett[2], Zuolong Zhang[6], Guocong Yao[7], Dengxiaoyu Shi[2], Ben Clock[8], Samantha Marcotte[8], Joseph R. Nery[2], Michelle Liem[9], Naomi Claffey[9], Lara Boggeman[9], Cesar Barragan[2], Rafael Arrojo e Drigo[10,11,12], Annika K. Weimer[13,14], Minyi Shi[13], Johnathan Cooper-Knock[15], Sai Zhang[16,17], Michael P. Snyder[13], Sebastian Preissl[18,19,20,21], Bing Ren[18,22], Carolyn O'Connor[9], Shengbo Chen[23], Chongyuan Luo[24], Jesse R. Dixon[8]*, Joseph R. Ecker[2,25]*

Higher-order chromatin structure and DNA methylation are critical for gene regulation, but how these vary across the human body remains unclear. We performed multiomic profiling of three-dimensional (3D) genome structure and DNA methylation for 86,689 single nuclei across 16 tissues, identifying 35 major and 206 cell subtypes. We revealed extensive changes in CG and non-CG methylation across cell types and characterized 3D chromatin structure at an unprecedented cellular resolution. Extensive discrepancies exist between cell types delineated by DNA methylation and genome structure, which indicates that the role of distinct epigenomic features in maintaining cell identity may vary by lineage. This study expands our understanding of the diversity of DNA methylation and chromatin structure and offers a reference for exploring gene regulation in human health and disease.

Distinct cellular identities are established by cooperative gene regulatory programs that give rise to cell type–specific gene expression patterns. Proper patterns of lineage-specific gene expression require coordination between distinct chromatin regulatory processes, including open chromatin, transcription factor (TF) binding, histone modifications, DNA methylation, and higher-order three-dimensional (3D) chromatin structure (*1*). The development of single-cell technologies to profile gene expression (*2*) or aspects of the chromatin regulatory landscape (*3*–*7*) has led to efforts to comprehensively profile these features across human cell types (*8*–*11*), identifying previously unknown human cell states (*12*) and regulatory elements contributing to human disease risk (*13*–*17*). However, single-cell methods that analyze other critical mechanisms of gene regulation, such as DNA methylation (*6*, *18*, *19*) or 3D genome architecture (*20*–*23*), have only been applied to limited tissues and cell types.

Previous efforts to broadly characterize these epigenomic features across human cell types have focused on bulk tissue profiling of DNA methylation (*24*, *25*) or 3D genome organization (*26*) as well as cultured primary or cancer cell lines (*27*–*31*). These studies have revealed distinct differences in DNA methylation and 3D genome architecture, including non-CG methylation (mCH, where H is A, T, or C) in human brain samples and partially methylated domains (PMDs) in cancer cell lines and some tissues, but these methods have not been comprehensively applied to distinguish human cell types from in vivo samples. More recently, DNA methylation has been profiled across broad human cell types using cell sorting to isolate specific populations (*32*). However, some subpopulations—such as neuronal subtypes—cannot be isolated by cell sorting from human tissue. Moreover, most previous efforts have studied these features in isolation, leaving relationships among distinct chromatin regulatory mechanisms in single cells unexplored.

A single-cell atlas of DNA methylation and 3D genome structure across human tissues

We used the single-nucleus methyl-3C sequencing (snm3C-seq) assay (*22*) to simultaneously analyze DNA methylome and 3D genome structure within single cells across 16 human tissues from at least two human donors (table S1) obtaining 86,689 cells after quality control (Fig. 1, B and C). We generated 195 billion nonclonal methylation reads and 18 billion long-range chromatin contacts across all cells (Fig. 1A). For 12/16 tissues, the samples were previously profiled as part of the ENTEx project (*33*), with associated single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) (*14*). The remaining four tissues—primary motor cortex, placenta, isolated pancreatic islets, and peripheral blood—broadened the diversity of cell types represented.

We developed methods to remove artifactual contacts (fig. S1), improve single-cell clustering (fig. S2, A and B), and integrate data across donors (fig. S2, C and D). Using combined DNA methylation and 3D genome information, we classified cells into 35 major types (Fig. 1D and fig. S2E). Some major types were exclusive to single tissues (e.g., excitatory neurons), and others were shared across tissues (e.g., memory T cells) (Fig. 1E). We further subclassified cells into 206 subtypes (Fig. 1F) using both modalities and integration with single-cell RNA sequencing (scRNA-seq) and snATAC-seq data (figs. S3 to S5; see materials and methods for details). Validating the data quality, identical cell types across donors were strongly correlated (fig. S6). DNA methylation profiles showed strong correlation with previous sorted-cell and bulk-tissue methylomes (fig. S7), and chromatin loops matched the specificity of bulk Hi-C data from ENCODE (fig. S8). To facilitate exploration of the data, we have developed an interactive cell browser (https://humancellepigenomeatlas.arcinstitute.org) to display both DNA methylation and chromatin contact data across tissues, major types, and subtypes.

Variability of CG methylation across cell types

Cytosine DNA methylation occurs in a CG dinucleotide context across diverse cell types. We observed a wide range of average mCG levels (Fig. 2A), ranging from 59% [epithelial trophoblast cells (Epi TPB)] to 81% [inhibitory neurons (Neu Inh)]. Most cell types show a bimodal distribution of mCG, where cytosines are largely unmethylated (<10%) or

[1]Arc Institute, Palo Alto, CA, USA. [2]Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA. [3]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA, USA. [4]Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA. [5]Society of Fellows, Harvard University, Cambridge, MA, USA. [6]School of Software, Henan University, Kaifeng, Henan, China. [7]School of Computer and Information Engineering, Henan University, Kaifeng, Henan, China. [8]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA. [9]Flow Cytometry Core Facility, Salk Institute for Biological Studies, La Jolla, CA, USA. [10]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA. [11]Center for Computational Systems Biology, Vanderbilt University, Nashville, TN, USA. [12]Diabetes Research and Training Center (DRTC), Vanderbilt University Medical Center, Nashville, TN, USA. [13]Department of Genetics, Stanford School of Medicine, Stanford, CA, USA. [14]Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [15]Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK. [16]Department of Biomedical Informatics & Data Science, Yale School of Medicine, New Haven, CT, USA. [17]Department of Epidemiology, University of Florida, Gainesville, FL, USA. [18]Center for Epigenomics, University of California San Diego, La Jolla, CA, USA. [19]Institute of Experimental and Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [20]Department of Pharmacology and Toxicology, Institute of Pharmaceutical Sciences, University of Graz, Graz, Austria. [21]Field of Excellence BioHealth, University of Graz, Graz, Austria. [22]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [23]School of Software, Nanchang University, Nanchang, Jiangxi, China. [24]Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA. [25]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA, USA. ***Corresponding author. Email: jingtian.zhou@arcinstitute.org (J.Z.); jedixon@salk.edu (J.R.D.); ecker@salk.edu (J.R.E.)** †These authors contributed equally to this work.

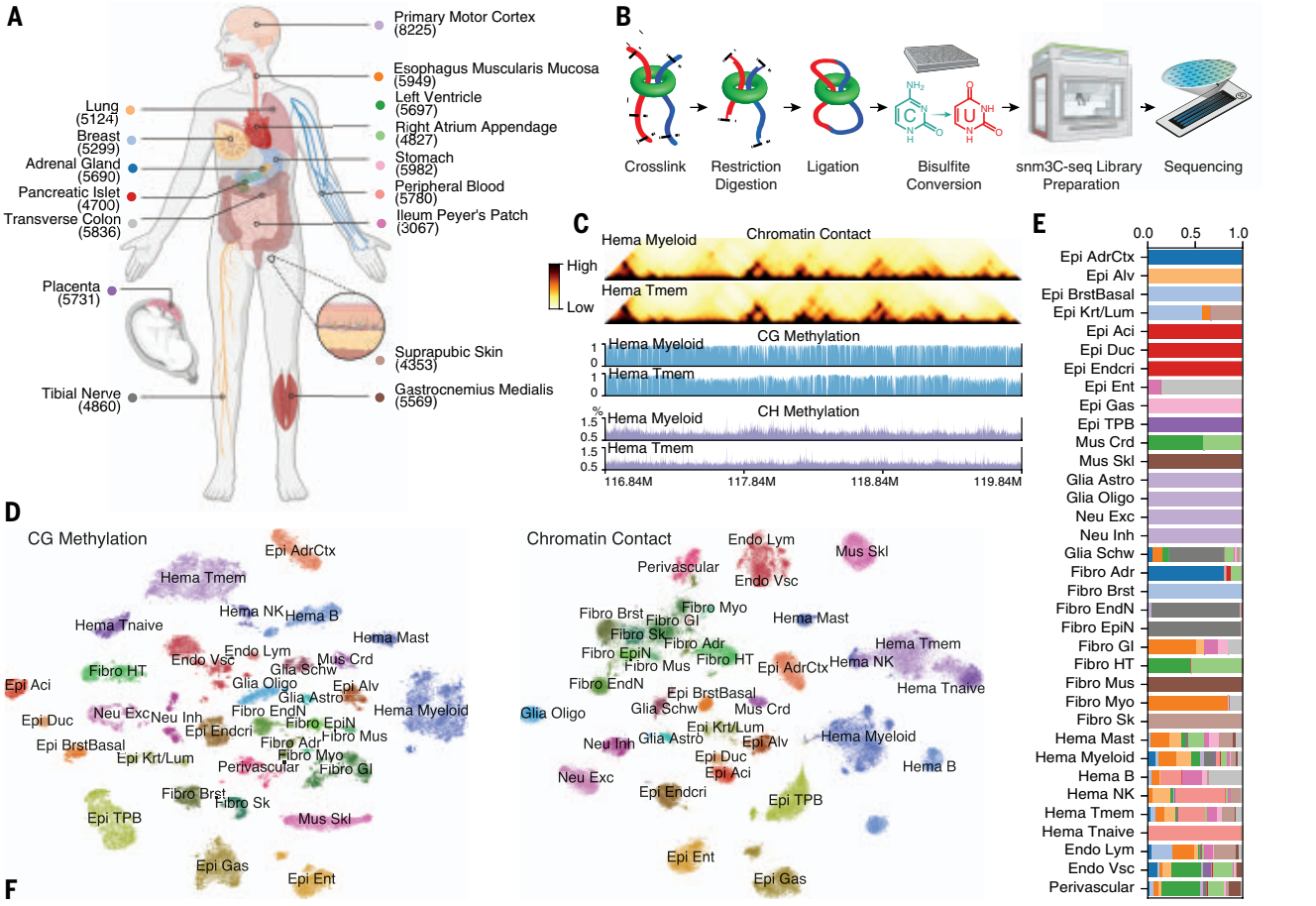


Fig. 1. Single-cell atlas of human DNA methylation and 3D genome structure. (**A**) Schematic showing 16 human tissues profiled by the snm3C-seq assay. Below each tissue, the number of single cells passing quality control filters is shown. Each tissue is assigned a specific color that is used throughout the figures. (**B**) Workflow diagram of the snm3C-seq assay. (**C**) Browser shot of DNA methylation and chromatin contacts in myeloid (Hema Myeloid) or T memory cells (Hema Tmem) after clustering and merging single

cells into pseudobulk cell types. (**D**) t-distributed stochastic neighbor embedding (t-SNE) of all single cells (n = 86,689) using either CpG cytosine methylation (mCG; left) or chromatin contact (right) data colored by major types. The same color palette is used for major types throughout the figures. (**E**) The proportion of cells derived from each tissue for the 35 major cell types. The colors for each tissue are labeled in (A). (**F**) Dendrogram of major cell types (top) and subtypes (bottom). The x-axis coordinate of each major type is manually aligned to the center of the highest branch in the subtype dendrogram of the major type.

are fully methylated (>70%) (Fig. 2B). A subset of major cell types, including trophoblasts (Epi TPB), memory B cells (Hema Bmem), and acinar cells (Epi Aci), were enriched for partially methylated CGs (Fig. 2B) found in large discrete domains (Fig. 2C), previously called PMDs (*31*). PMDs have been described as a prominent feature in cancer cell lines and cultured fibroblasts (*34*), and previous studies have also found PMDs in bulk DNA methylation profiling from placenta tissue (*35*).

To characterize PMDs across human cell types, we clustered 10-kb genomic bins on the basis of the distribution of methylation across CG sites (Fig. 2, D and E). The method accurately identifies PMDs across diverse lineages compared with existing tools (*36*) (fig. S9A), with the partially methylated CGs enriched in PMDs and depleted from non-PMD regions (Fig. 2F and fig. S9B). Examining mCG levels over PMD regions identified from lineages with clear PMDs (e.g., Hema Bmem) in related non-PMD types (e.g., Hema Bnaive) revealed lower methylation compared with non-PMD regions (fig. S9C), which suggests that the genome may be divided into "DNA methylation compartments" even in non-PMD cell types. Applying the clustering framework to all cell types, we found three compartments: two major compartments showing full (pink) or partial (blue) cytosine methylation and one minor compartment with a complete absence of mCG (green) (Fig. 2G and fig. S9D). Compartmentalization is highly reproducible, with ~95% of bins assigned to the same compartment across donors (fig. S10, A to C). The partially methylated compartments show greater mCG reduction in lineages with apparent PMDs compared with related lineages without (fig. S10, D and E), which indicates that mCG depletion within the PMD-like compartment lies on a continuum across cell types.

The overall distribution of the three compartments is similar across major cell types (fig. S9D). However, some lineages show an expansion of the partially methylated compartment (Fig. 2H). The partially and hypermethylated compartments occupy 37 to 75% and 21 to 58% of the genome, respectively (95 to 96% in total; Fig. 2H, top). The hypocompartment (green) is strongly enriched for transcription start sites (TSSs) (Fig. 2H, middle) and candidate cis-regulatory elements (cCREs) (Fig. 2H, bottom) and is associated with higher gene expression (Fig. 2I and fig. S11A). Highly expressed genes have lower promoter mCG and higher gene body mCG compared with lowly expressed genes (Fig. 2J and fig. S11B). Genes with the most variable expression are more likely to reside in the partially methylated compartments (Fig. 2, K and L, and fig. S11, C to F), suggesting association with mechanisms of dynamic gene regulation.

Consistent with previous studies (*31*, *37*, *38*), repressive markers H3K9me3 and H3K27me3 were enriched within the partially methylated compartment, whereas active markers H3K27ac, H3K4me3, H3K4me1, and H3K36me3 were enriched in the hypermethylated compartment (fig. S11, G and H). The promoter mark H3K4me3 shows a higher signal only in the hypomethylated compartment (fig. S11H). A small portion of the hypomethylated compartment also shows a high H3K27me3 signal, which could represent bivalent promoter regions (fig. S11H).

In total, we identified 1,364,566 differentially methylated regions (DMRs) across all cell types, with an average length of 377 base pairs (bp), covering 17.9% of the genome. Of these DMRs, 51.8% are differentially methylated between major types, of which 76.8% also exhibited differential methylation between subtypes within major types. We identified 95.4% of DMRs from previous bulk tissue methylome profiling, along with 587,648 additional DMRs (Fig. 2M). A total of 1,249,253 (91.5%) of DMRs are >2 kb from TSSs, and DMRs are enriched near TSSs but depleted from a core promoter region of ±1 kb surrounding TSSs (fig. S12A). DMRs are usually hypomethylated in specific cell types, with only 1% of DMRs hypomethylated in >90% of subtypes and 96% of DMRs hypomethylated in <10% of subtypes (fig. S12, B and C).

Within a given major type, 48% of ATAC-seq peaks overlapped DMRs, whereas 33% of DMRs overlapped ATAC-seq peaks (fig. S13A). Overall, 91.9% of ATAC-seq peaks overlapped with DMRs across all tissues, whereas 61.9% of DMRs overlapped with ATAC-seq peaks (Fig. 2N). DMR methylation levels clearly distinguish different lineages (fig. S13B). This was similar for DMRs that overlapped ATAC-seq peaks (peak DMRs) and those that did not (nonpeak DMRs; fig. S13, B to D), which suggests that both classes of DMRs contain information about the cis-regulatory landscape of cells.

We analyzed cell type–specific DMRs to identify TFs contributing to lineage-specific gene regulation (table S2). For example, TBX20, a conserved regulator of cardiac cell function (*39*), was enriched in cardiac muscle cells, heart fibroblasts, and perivascular cells. IRF4, FOXO1, and BCL6 motifs are enriched in the immune cell lineage (*40*–*42*), and PAX7, MYOG, and MYOD1, important players in myogenesis (*43*), were identified in skeletal muscle (Fig. 2O). These enrichments were similar for peak and nonpeak DMRs (fig. S13E). Overall, 559 motifs were shared between peak and nonpeak DMRs, with 514 motifs specific to peak DMRs and 1249 motifs specific to nonpeak DMRs. Notably, many TF motifs regulating cell development and function were found only in nonpeak DMRs, including POU domain family TFs (POU3F3 and POU5F2) in neuronal lineages (*44*) and zinc finger family TFs (ZNF620, ZNF334, and ZNF665) in epithelial cells, fibroblasts, and neuronal cells (Fig. 2O).

Across all lineages, mCG and mCH were depleted at ATAC-seq peaks, with greater depletion in the cell types in which the peaks were called (fig. S14, A and B). A small fraction of peaks were hypermethylated (fig. S14C), and motif enrichment analyses between hyper- and hypomethylated peaks identified TFs reported to preferentially bind methylated motifs (fig. S14, D and E). Finally, many repeat element categories, including DNA transposons, long interspersed nuclear elements (LINEs), and long terminal repeats (LTRs), showed cell type–specific mCG patterns sufficient to separate most major cell types, whereas shorter regions [e.g., short interspersed nuclear elements (SINEs), satellite DNA, and retroposons] showed lower clustering performance (Fig. 2P and fig. S15). Even after excluding methylation over genes or ATAC peaks, most major types could still be separated (Fig. 2P and fig. S15), which suggests that mCG specificity outside of cCREs and genes retains information about cell identity.

Patterns of non-CG methylation across cell types

mCH has been reported to occur in neurons, glia, and stem cells and at low or undetectable levels (<1%) in other human tissues (*45*). Previous studies have shown that mCH frequently occurs in a CAC context in neurons versus in a CAG context in stem cells (*45*). Using our single-cell data, mCH was enriched in nearly all base contexts, above that observed for spike-in control lambda phage DNA (fig. S16A). Neurons had the highest levels of mCH (Fig. 3A, rightmost panel), with glial populations and muscle-associated cells showing modest mCH levels (fig. S16B). Examining the trinucleotide contexts for mCH across major cell types, mCH most frequently occurs in a CAC context. Four of the five most frequent mCH contexts are in a CA context, whereas the lowest seven trinucleotide contexts contain CC or CT, which suggests that the presence of a purine after cytosine may increase the likelihood of mCH (Fig. 3B).

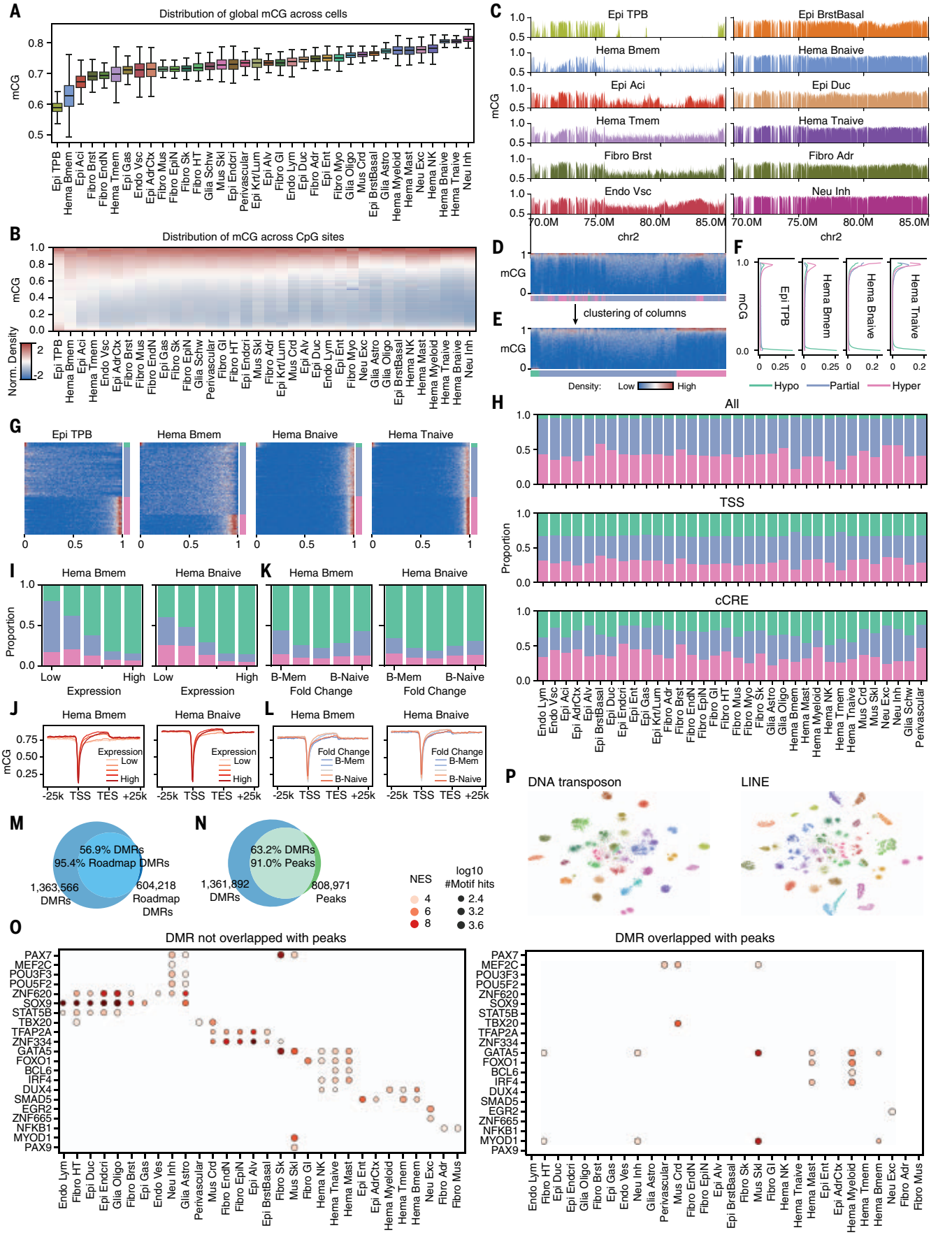

A
Distribution of global mCG across cells
mCG
B
Distribution of mCG across CpG sites
Norm. Density
C
Epi TPB
Epi BrstBasal
Hema Bmem
Hema Bnaive
Epi Aci
Epi Duc
Hema Tmem
Hema Tnaive
Fibro Brst
Fibro Adr
Endo Vsc
Neu Inh
chr2
D
E
clustering of columns
Density: Low High
F
Hypo
Partial
Hyper
G
H
All
TSS
cCRE
Proportion
I
Expression
K
Fold Change
J
L
M
56.9% DMRs
95.4% Roadmap DMRs
1,363,566 DMRs
604,218 Roadmap DMRs
N
63.2% DMRs
91.0% Peaks
1,361,892 DMRs
808,971 Peaks
NES
log10 #Motif hits
P
DNA transposon
LINE
O
DMR not overlapped with peaks
DMR overlapped with peaks

Fig. 2. CG methylation across cell types. (**A**) Boxplot showing the distribution of average genome-wide mCG of single cells grouped by major types. For all boxplots throughout the figures, the center line denotes the median, box limits denote first and third quartiles, and whiskers denote 1.5× the interquartile range. (**B**) The distribution of mCG of individual CpGs in each major type. (**C**) Genome browser views of mCG at chromosome 2: 70,000,000 to 88,000,000, showing the presence of PMDs. (**D** and **E**) Distribution of mCG in Endo Vsc over CpGs for 10-kb bins (columns of top heatmap) and methylation compartment (bottom) at the same region as in (C) ordered by genome coordinate (D) or methylation compartment (E). All methylation compartment plots in all figures share the same color bar. (**F**) Distribution of mCG in different methylation compartments (colors) and major types (panels). (**G**) Distribution of mCG over CpGs for 10-kb bins (row) ordered by methylation compartments. Color bars from top to bottom: fully unmethylated, bimodal, partially methylated, and fully methylated. The same color palette is used for methylation compartments throughout the figures. (**H**) Fraction of the genome associated with each methylation compartment for all 10-kb bins (top), TSSs (middle), or cCREs (bottom). (**I** to **L**) Fraction of genes associated with each methylation compartment [(I) and (K)] or gene body mCG [(J) and (L)] in memory (Hema Bmem) or naïve (Hema Bnaive) B cells stratified by gene expression level [(I) and (J)] or fold-change between naïve and memory B cells [(K) and (L)]. TES, transcription end site. (**M**) Overlap between all DMRs from this study and DMRs from Schultz *et al*. (*24*). (**N**) Overlap between DMRs identified in this study and previous scATAC-seq peaks from matching cell types. This DMR set excludes trophoblast and naïve B cells owing to a lack of matching ATAC-seq data. (**O**) Enrichment of selected TF binding motifs in major cell types for nonpeak DMRs (left) and peak DMRs (right). Normalized enrichment scores (NESs) are row-wise *z*-scored. A full list of enriched motifs is shown in table S2. (**P**) t-SNE of downsampled cells ($n = 26{,}423$) using mCG at DNA transposons (left) and LINEs (right), colored by major types.

We also examined whether patterns of mCH in the genome were nonrandom by testing the correlation of mCH across different trinucleotide contexts. At small bin sizes (50 bp), mCG methylation showed correlations across contexts (figs. S17 and S18). However, only at kilobase bin sizes, strong correlations were apparent for different CH nucleotide contexts in multiple lineages (figs. S17 and S18). We observed six groups of major cell types with similar patterns of CH correlation across base contexts (fig. S17A). mCH was correlated across a wide variety of trinucleotide contexts in neuronal cell types (fig. S17). Outside of neurons, mCH was correlated largely in the CTC or CAN trinucleotide contexts (fig. S17), which are the most frequently enriched trinucleotides for mCH across major cell types (Fig. 3B), including in muscles, fibroblasts, and endothelial cells (fig. S17A, orange, green, red, gray, and purple groups). On the contrary, mCH was less well correlated in a subset of major types (fig. S17A, blue group), including some hematopoietic (Hema B, Hema Tmem, and Hema NK) or epithelial lineages (Epi Gas, Epi Alv, Epi Ent, and Epi Krt/Lum). There is a relatively weak correlation for mCH outside of the CTC or CAN trinucleotide context across most major cell types.

We further quantified the genomic scale of mCH signatures by correlating methylation between genomic bins separated by a given distance. In most major types, mCAN or mCTC show higher correlations over longer distances compared with mCG (Fig. 3C and fig. S19A). mCTD shows high correlation at longer distances in neurons, glia, muscle, endocrine, and trophoblast cells (fig. S19A). Trophoblast shows high correlation over long distances for mCG, mCA, and mCT (fig. S19, A and B), likely due in part to strong PMD signatures. In neurons, all CH contexts (including CC) show correlations over longer distances compared with CG (fig. S19, A and B), consistent with the observation that neurons have longer mCH features than mCG (*25*, *46*). Some major types, including the luminal, alveolar, and gastrointestinal epithelial cells; memory T cells; and B cells, show weak correlations in all CH contexts (fig. S19A), which aligns with the observation that the global mCH is lower in these cell types (Fig. 3B). These cell types also show lower expression of DNMT3A ($P = 1.768 \times 10^{-3}$, Wilcoxon rank sum test), the writer of mCH in the brain, but not the other DNA methyltransferases (fig. S20), which suggests that DNMT3A could also be the primary enzyme depositing mCH in the different cell types.

Similar to mCG, mCH is depleted at regulatory elements across major cell types (Fig. 3, D and E). However, major cell types exhibit distinct patterns of mCH across promoters and gene bodies. For example, neurons showed relatively low levels of mCH across exons and introns, whereas skeletal muscle showed higher levels of mCH across these regions (Fig. 3D). mCH depletion over cCREs was observed only when those cCREs are active in the specific major cell type (Fig. 3E), indicating cell type–specific information content. We also observed higher mCH in hypermethylated compartments and lower mCH in partially methylated compartments, with a shift at compartment boundaries in all major types except intestinal epithelial cells (Fig. 3F).

Although promoters, cCREs, and partially methylated compartments show similar patterns of mCG and mCH, we next investigated whether some genomic regions exhibit different signatures for mCG versus mCH. Comparing across 1-kb bins, regions with low mCG and low mCH are enriched near TSSs, 5′ untranslated regions (5′UTRs), and exons, whereas bins with low mCG but high mCA are mostly in introns, 3′UTRs, and intergenic regions (fig. S21A). CTCF (BORIS) motifs are enriched in the high-mCA but low-mCG regions across 28 of the 35 major types ($P < 1 \times 10^{-5}$). Other motifs show cell type–specific enrichment, including known lineage regulators, such as POU motifs in excitatory neurons, LHX motifs in inhibitory neurons, MEF motifs in endocrine cells, and GLI motifs in skeletal muscle (fig. S21B). This suggests that the binding of these factors could be affected not only by mCG but also by mCA. By contrast, the low-mCA regions showed many more shared motifs across cell types (fig. S21B).

To further test the cell type specificity of mCH, we clustered cells by mCH and observed clusters for most major cell types (Fig. 3G and fig. S22). Similarly, we trained logistic regression models to predict cell type from mCH levels (fig. S23A). Some lineages, such as fibroblasts, were globally distinguished by mCH, but specific subsets of fibroblasts were not resolved. In other lineages, mCH could distinguish relevant cell populations, such as pancreatic islets or skeletal muscle (Fig. 3H and fig. S23A). Taken together, these results suggest that low levels of mCH are present across a wide variety of lineages and retain information regarding regulatory elements and cell identity.

Previous studies have shown that mCH in neurons is anticorrelated with gene expression levels, but whether this holds for other lineages remains unclear. For both inhibitory and excitatory neurons, mCH and gene expression were anticorrelated at both TSS-proximal regions and within the gene body (fig. S23B). By contrast, in the endocrine pancreas, mCH in regions downstream of the TSS were anticorrelated with gene expression but positively correlated upstream (Fig. 3I) while showing only a weak correlation with gene expression over the entire gene body. Similarly, gene body mCH was mildly anticorrelated with gene expression within skeletal muscle cells but positively correlated in natural killer (NK) cell subtypes (fig. S23B). This suggests that the mechanisms relating to mCH and gene activity may vary between lineages and is an important area for future exploration.

Variability in 3D genome architecture across cell types

Using our single-cell atlas, we analyzed multiple features of 3D genome architecture, including chromatin contact decay patterns, chromatin loops, and chromatin compartments. Examining contact decay patterns, there are two prominent "peaks" of enrichment of chromatin interactions: one from 100 kb to 1 Mb and another greater than 10 Mb (Fig. 4A). This pattern in contact decay plots has been described previously (*18*, *47*) and is attributed to the presence of chromatin loops and topologically associating domains (TADs) (100 kb to 1 Mb) and to A/B chromatin compartments (>10 Mb). Across major cell types, some

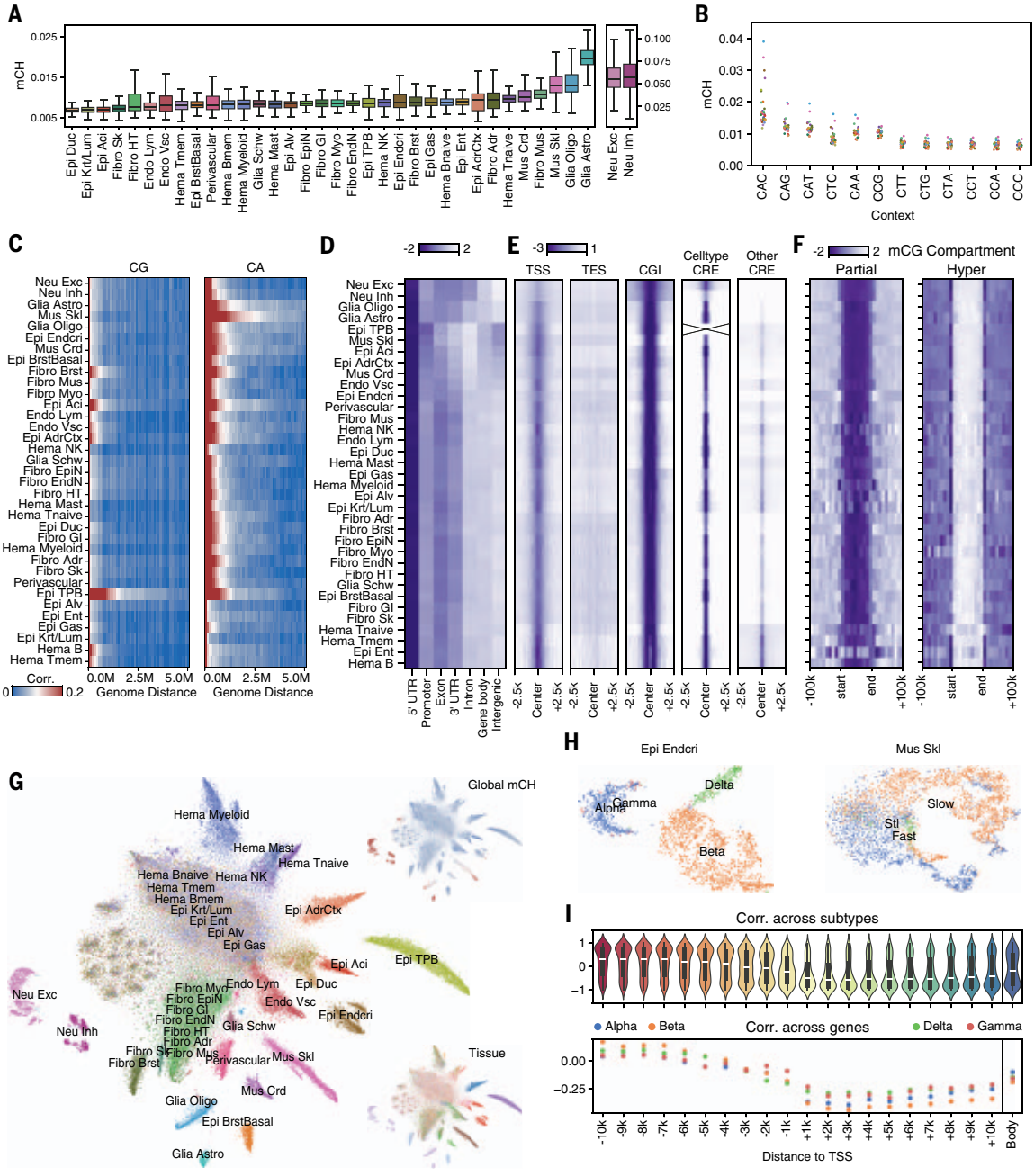


Fig. 3. Non-CG methylation across cell types. (**A**) Distribution of the average non-CG methylation (mCH) level in single cells grouped by major type. The plot is split to allow the scale to show both neuronal (far-right panels) and nonneuronal cell types. (**B**) Average mCH of major cell types excluding neuronal cells grouped by trinucleotide context. (**C**) Correlation between methylation levels at certain genomic distances (*x* axis) in different major types (*y* axis) at different contexts (panels). (**D**) mCH at different genomic features across major cell types. Values are row-wise *z*-scored. (**E**) mCH surrounding different elements. Values are row-wise *z*-scored across all five plots together. The last two columns show ATAC-seq–defined cCREs present in the given cell type or cCREs remaining after excluding the cCREs of the cell type from cCREs of all cell types. (**F**) mCH surrounding mCG compartments longer than 100 kb. Values are row-wise *z*-scored within each plot. (**G**) t-SNE of all cells ($n = 86{,}689$) using mCH colored by major types (center), global mCH (top right), or tissue source (bottom right). (**H**) t-SNE of endocrine pancreas epithelial cells (Epi Endocri; $n = 2324$; left) or skeletal muscle cells (Mus Skl; $n = 3973$; right) using mCH colored by subtypes. (**I**) Correlations between mCH and expression of differentially expressed genes (DEGs) in Epi Endcri across subtypes for all DEGs (top) or across all DEGs for each cell subtype (bottom). The correlations are calculated using mCH from regions spanning from a TSS to different distances on either side of the TSS (*x* axis) or across the entire gene body (right). Colors in the violin plot are a continuous palette from TSSs upstream to downstream.

lineages show a "domain-dominant" phenotype, primarily of short-range interactions, whereas others show a "compartment-dominant" phenotype, primarily of long-range interactions, or a mixture of the two (Fig. 4B). Neuronal cell types are among the most domain-dominant cells, whereas five of the six most compartment-dominant lineages are hematopoietic lineages (Fig. 4B). Apart from the enrichment of hematopoietic or neuronal lineages, there is no apparent enrichment of other major cell types for compartment- or domain-dominant phenotypes.

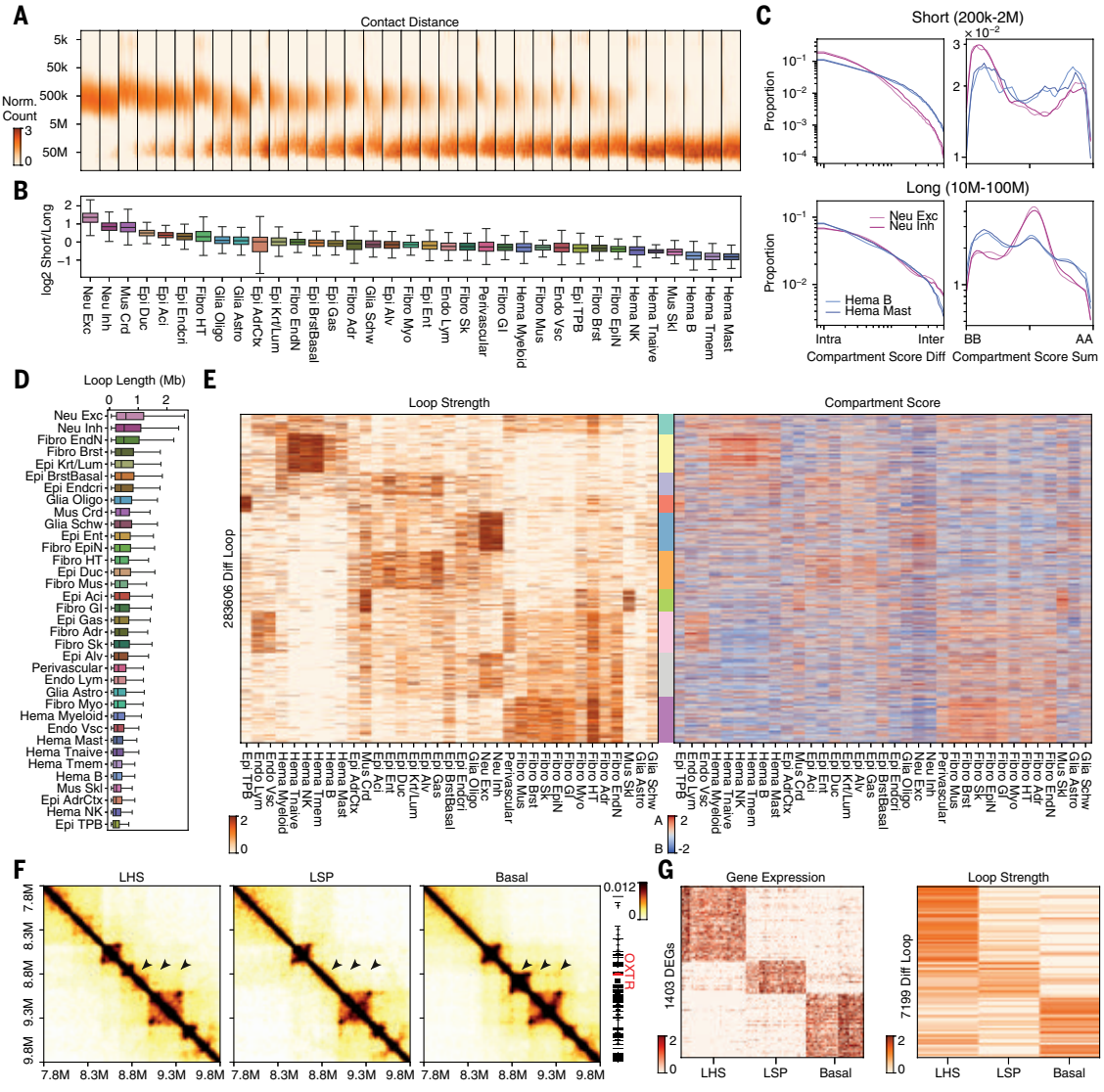


Fig. 4. Chromatin contacts across cell types. (**A**) Distribution of contact distances (*y* axis) across single cells ($n = 86{,}689$; *x* axis) grouped by major types. (**B**) Ratio of short-range (200 kb to 2 Mb) to long-range (10 to 100 Mb) chromatin contacts among major cell types. Order and labels are the same between (A) and (B). (**C**) Proportion of short-range (top) or long-range (bottom) contacts stratified by the differences (left) or sums (right) of compartment scores at the two interacting regions. (**D**) Distribution of loop lengths by major cell type. (**E**) Interaction strength of differential loops between major types (left) and the compartment scores of loop anchors (right) across major types. Both values are row-wise *z*-scored. Color bar (middle) shows *k*-means clusters of differential loops. (**F**) Browser shot showing subtype-specific chromatin loops in breast epithelial cells (LHS, luminal hormone sensitive; LSP, luminal secretory precursor; Basal, basal myoepithelial) at chromosome 3: 7,800,000 to 9,800,000, surrounding basal cell marker *OXTR*. (**G**) Expression level of DEGs whose TSSs are within 2 kb of either anchor of the differential loops (left) and interaction strength of differential loops between breast epithelial subtypes (right) across cell type samples. Values are row-wise *z*-scored. Left and right heatmaps share the row orders. When a loop overlaps multiple genes, the loop is repeated in the left heatmap; similarly, when a gene overlaps multiple loops, it is repeated in the right heatmap.

To directly test whether the compartment-dominant interactions are due to A/B compartments, we compared chromatin compartment interactions as a function of genomic distance (fig. S24A). At short distances, the domain-dominant neuronal lineages have relatively more intracompartment contacts (Fig. 4C, upper left) for both A-A and B-B compartment interactions (Fig. 4C, upper right). By contrast, at longer distances, the compartment-dominant hematopoietic cells show stronger within-compartment interactions for intra–B-B and A-A interactions (Fig. 4C, bottom). Across major lineages in the longer-range (>10 Mb), domain-dominant lineages (higher short/long contact ratios) showed weaker within-compartment interactions (AA or BB), whereas compartment-dominant lineages (lower short/long ratios) showed stronger AA or BB compartment strengths (fig. S24B).

We also examined the presence of chromatin loops across cell types. Chromatin loops varied in both number (fig. S25A) and size (Fig. 4D) across lineages. Neurons showed the most and longest chromatin loops. More compartment-dominant hematopoietic lineages showed shorter and fewer chromatin loops (Fig. 4D). We also identified chromatin domains across lineages and observed notably more consistency in domain number (fig. S25B) and size (fig. S25C) across major lineages. Comparing loops between cells, we identified 283,606 differential chromatin loops across major cell types (Fig. 4E and fig. S25D). We observe that many differential chromatin loops are shared among related cell types (Fig. 4E), in particular among fibroblasts, epithelial cells, neurons, and hematopoietic cells (Fig. 4E). Smaller subsets of lineage-specific loops are specific to trophoblast epithelial cells (Epi TPB)

or to muscle lineages (Mus Crd and Mus Skl) (Fig. 4E). The presence of loops in a specific lineage also corresponds with a relative enrichment of those lineages in the A compartment regions (Fig. 4E, right), consistent with these being more active regions of the genome.

In addition to the major types, we also identified differential loops between cell subtypes (Fig. 4F), which allowed us to compare chromatin looping differences with associated gene expression changes. For example, we identified differentially expressed genes between breast epithelial subtypes and showed that these are frequently associated with subtype-specific chromatin looping (Fig. 4G). Taken together, these results indicate an abundance of variable chromatin loops that differ across major cell type and lineage subtypes and are strongly associated with differential gene expression.

Association between chromatin conformation and DNA methylation

We examined the relationship between DNA methylation and chromatin contacts across cell types. For most cell types, A/B compartment scores show a negative correlation with mCG, with A compartment regions depleted of mCG (Fig. 5A and fig. S26A). By contrast, seven cell types showed a positive correlation between mCG and compartment score, all of which have strong PMDs (Fig. 5A and fig. S26A), consistent with the association of PMDs with repressive chromatin. mCH showed a negative correlation with compartment scores only in the neuronal lineage (Fig. 5B and fig. S26B), consistent with active genes having lower mCH in neurons.

A/B compartments were well correlated with DNA methylation compartments across cells. We found that the partially methylated compartment corresponded with the B compartment, and the fully methylated compartment matched the A compartment (Fig. 5C). At high resolutions, the methylation compartment provides a more fine-grained segregation of the genome than chromatin contact–defined A/B compartments (Fig. 5D). Most regions of the genome were stable in their A/B compartment and DNA methylation compartment patterns across major cell types, (Fig. 5E). A subset of genomic regions shows concordant lineage-specific switches of A/B compartments and DNA methylated compartments (Fig. 5E). The one notable exception is in neurons and glial cells, where there was not an obvious correlation between B compartment regions and the partially methylated compartments. Except for neural lineages, these results show that the B compartments highly resemble the partially methylated compartments (fig. S9D), which suggests that B compartment regions may have distinct mechanisms for maintaining DNA methylation.

Across nearly all lineages, domain boundaries demarcate regions showing enrichment and depletion of mCG (Fig. 5F). Further, most lineages show a relative focal depletion of mCG at the domain boundary (Fig. 5F and fig. S26C). When considering domain boundaries that separate A and B compartments, the association of mCG with A/B compartments varies between lineages (Fig. 5F and fig. S26C). Lineages that show evidence of PMDs, including trophoblast and some hematopoietic lineages, show lower levels of mCG in the B compartment side adjacent to domain boundaries (Fig. 5F and fig. S26C), whereas other lineages, including neurons and glia, show a depletion of mCG within the A compartment side (Fig. 5F and fig. S26C). By contrast, nearly all lineages show enrichment of mCH in A compartment regions except for neuronal lineages, which show enrichment of mCH in B compartment regions (Fig. 5G and fig. S26D). Across all lineages, mCH is depleted at domain boundaries separating compartments of similar type (Fig. 5G). This is consistent with the depletion of mCH over cCREs and hypermethylated compartments (Fig. 3, E and F). The difference in the distribution of mCH between neurons and nonneuronal lineages is perplexing because neurons show much higher levels of mCH than most lineages. Why mCH shows distinct associations with compartments and gene expression in neuronal versus nonneuronal lineages remains unclear.

To investigate the association between the two modalities at higher genomic resolution, we compared chromatin loops with DMRs (fig. S27A). Examining DMRs associated with loop anchors against those not associated with loops, the CTCF motif is enriched across almost all major types, supporting its crucial role in the regulation of higher-order chromosome structure (Fig. 5H). Notably, we also identified other lineage-specific TFs, including IRF2 and ISRE in the immune lineage (*40*), NR5A2 in pancreatic acinar cells (*48*), and GATA factors in trophoblasts (*49*) (Fig. 5H). This suggests that these TFs could also contribute to cell type–specific 3D genome structure, possibly by modulating CTCF accessibility to chromatin.

In most major types, DMRs are enriched at loops that show higher variability across subtypes (fig. S27B). Comparing DNA methylation and chromatin contact across subtypes, some lineages, such as neurons, showed strong anticorrelations between chromatin loops and mCG at loop anchors. By contrast, other lineages showed little clear relationship between DNA methylation and chromatin looping or, in some cases, bimodal distributions (Fig. 5I). This may be due to the presence of PMDs because the lineages that show the least correlation between DMRs and chromatin looping (Hema Tmem, Epi TPB, and Epi Aci) also show the strongest presence of PMDs. This is not a product of the clustering method or data modality because the same lineages show consistently little correlation whether we cluster by DNA methylation, by 3D genome organization, or jointly across both data modalities (fig. S28).

To determine whether DMRs and chromatin loops could help map disease-associated variants, genes, and cell types, we used stratified linkage disequilibrium score regression (LDSC) (*50*) to estimate heritability enrichment in DMRs for each major type. We used the DMRs at loop anchors and identified 135 significant major type–phenotype pairs (fig. S29A). The associations include blood glucose with endocrine cells, balding with skin fibroblasts, atrial fibrillation with cardiomyocytes, and bipolar and schizophrenia with excitatory and inhibitory neurons. We observed significantly higher heritability enrichment in loop DMRs than in all DMRs among the associated major type–phenotype pairs ($P = 1.49 \times 10^{-25}$, Wilcoxon signed-rank test), which suggests that chromatin loops help refine and prioritize single-nucleotide polymorphisms (SNPs) that are potentially functional (fig. S29B).

Discrepancies between DNA methylation and chromatin contact

Overall, we observed frequent correlations between DNA methylation and 3D chromatin architecture across major cell types. However, we also observed distinct differences, particularly in the clustering of major cell types into subtypes (Fig. 1F and figs. S30 and S31). To better understand these differences, we compared the degree of similarity between DNA methylation and chromatin contact subtype clusters. Some lineages showed strong agreement between the two modalities, including neurons, endothelial vascular cells, and distinct subsets of fibroblast, epithelial, and hematopoietic cells (Fig. 6A). By contrast, other lineages showed little to no agreement in clusters between the two modalities (Fig. 6A). These results suggest that the association of DNA methylation and 3D genome organization with cellular identity may vary across lineages.

One lineage with poor clustering agreement was skeletal muscle (Mus Skl). Both DNA methylation and chromatin contacts identify clusters corresponding to satellite cells (muscle stem cells; blue), and slow- (green) or fast-twitch (orange) fibers (Fig. 6B). Chromatin contact data showed the presence of three distinct clusters without obvious DNA methylation gradients (Fig. 6C), whereas the clusters from DNA methylation appear to separate along a continuum of mCG levels (Fig. 6C). The two modalities also showed mismatches for subtype assignment (Fig. 6D). Subsets of methylation-annotated stem cells were assigned to mature cell populations with chromatin contacts. However, almost no cells annotated as mature by mCG were annotated

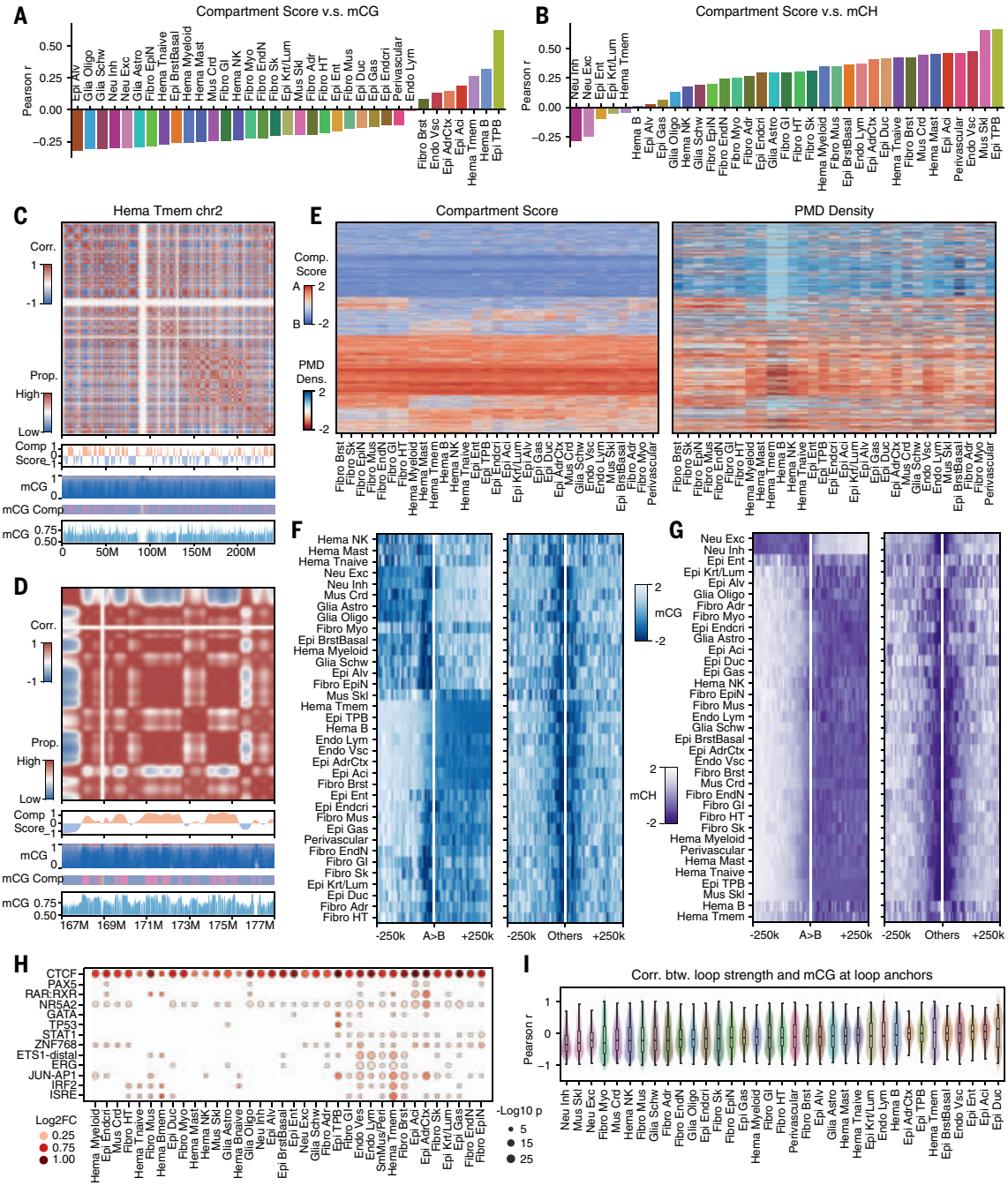


Fig. 5. Comparison of DNA methylation and 3D genome structure across cell types. (**A** and **B**) Correlation between chromatin compartments and mCG (A) or mCH (B) across 100-kb bins for all major cell types. (**C** and **D**) From top to bottom: correlation matrix of distance-normalized contact maps of Hema Tmem, compartment score at 100-kb resolution, distribution of mCG of individual CpG, methylation compartment, and mCG at 10-kb resolution for whole chromosome 2 (C) or chromosome 2: 167,000,000 to 177,000,000 (D). (**E**) Chromatin compartment score (left) and proportion of partially methylated compartment (right) (see materials and methods) at 100-kb resolution across major types. Neurons and glia in the cortex show weak correlations between the two scores and are not shown. (**F** and **G**) mCG (F) or mCH (G) surrounding chromatin domain boundaries that separate A/B compartments (left) or not (right). (**H**) Enrichment of motifs at loop DMRs versus nonloop DMRs for each major type. The hue represents the $\log_2$ fold change of the proportion of motif-containing DMRs. (**I**) Distribution of correlations between DMRs and differential loops across cell clusters within each major type.

as stem cells by the 3D genome (Fig. 6D). This indicates that there is a population of muscle fibers that, at the 3D chromatin contact level, resembles mature myofibers but, at the DNA methylation level, resembles stem cells, possibly reflecting recently differentiated myofibers. This is consistent with our previous study of the developing human brain, in which 3D genome structures typically precede DNA methylation changes during developmental transitions (*47*).

Additionally, some mature myofiber cells were clustered differently by the two modalities (Fig. 6, B and D). At the MYH7 locus, a slow-twitch muscle marker gene, cells assigned as slow twitch by mCG but fast twitch by chromatin contacts (slow/fast) show weaker chromatin loops between the gene promoter and upstream regions but a stronger depletion of mCG (Fig. 6E). To analyze this more systematically, we identified differential loops and DMRs between the satellite, slow-twitch,

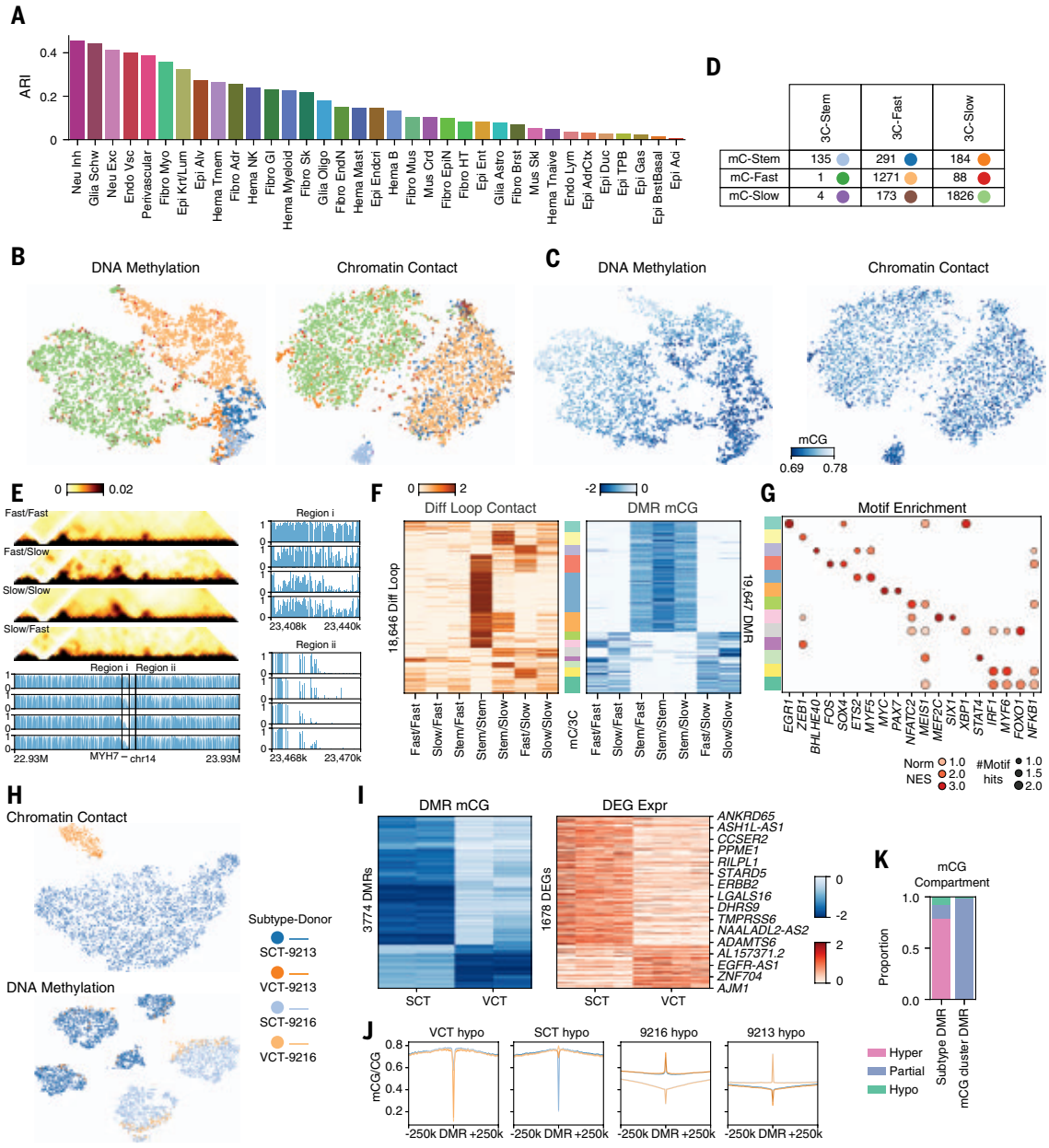


Fig. 6. Differences between DNA methylation– and 3D genome architecture–defined cell types. (**A**) Adjusted Rand index (ARI) quantifying the consistency between mCG and chromatin contact–based clustering of each major type into cell subtypes. Higher ARI indicates greater agreement in clustering. (**B** and **C**) t-SNE of Mus Skl ($n = 3973$) using mCG (left) or chromatin contact (right), colored by cell groups (B) or global mCG (C). (**D**) Confusion matrix of subtypes defined with mCG (mC-) versus chromatin contacts (3C-). Cell group colors are shared in (C) and (D). (**E**) Browser shot of chromatin contacts and mCG with the same cell group orders at chromosome 14: 22,930,000 to 23,930,000 (left), surrounding slow-twitch fiber marker *MYH7*, and zoomed-in view of DMRs (right) of the regions in the boxes. (**F**) Interaction strength of differential loops between cell groups (left) and mCG of DMRs at either anchor of the differential loops (right) across cell groups. Values are *z*-score normalized within each row. Color bar (middle) shows *k*-means clusters of loop-DMR pairs. (**G**) Motif enrichment of DMRs in each group of loop-DMR pairs in (F). A full list of enriched motifs is shown in table S3. (**H**) t-SNE of Epi-TPB ($n = 5228$) using chromatin contact (top) or DNA methylation (bottom) colored by subtype-donors (VCT, villous cytotrophoblast; SCT, syncytiotrophoblast) (top). (**I**) mCG of DMRs between subtypes (left) and expression of DEGs whose TSSs are within 2 kb of DMRs (right) across subtype-donors. Values are row-wise *z*-scored. (**J**) mCG in subtype-donors at flanking regions of subtype DMRs (left) or donor DMRs (right). Colors are shared between (H) and (J). (**K**) Fraction of the DMRs between subtypes (chromatin contact–based clusters) or mCG-based clusters in each methylation compartment.

and fast-twitch fibers (Fig. 6F and fig. S32A). In the cells that differ between chromatin contacts and DNA methylation (Stem/Fast or Stem/Slow), they have lost the stem cell chromatin loops and begin to display chromatin loops reminiscent of differentiated cells while retaining the methylation signatures of the stem cell populations (Fig. 6F, compare left and right panels for Stem/Fast or Stem/Slow cells).

The hypo-DMRs in the differentiated muscle types (groups 0, 2, 4, 7, 9, 10, and 11) are more enriched in TFs involved in muscle cell regeneration and adaptation to external stimuli. For example, PAX7 is a marker for stem cell population (*51*, *52*), MYC is increased after exercise (*51*, *52*), and CEBPB is induced on injury (*53*). By contrast, hypo-DMRs in the methylation-defined stem cell clusters (groups 1, 3, 5, 6,

and 12) are enriched for motifs related to muscle fiber differentiation, such as *MYF5* (*51*) (Fig. 6G, fig. S32A, and table S3). Notably, we found *FOXO1* and *NFKB1* in group 8, which contains stem cell hypo-DMRs overlapping stem cell loops. Both TFs are negative regulators of muscle differentiation (*51*, *52*, *54*, *55*), which suggests that cluster 8 may be critical for maintaining the muscle stem cell population. These observations may reflect differences in the dynamics of 3D chromatin organization and DNA methylation during terminal differentiation in some lineages.

Cases where the cell type classification is inconsistent across modalities are also observed among other cell subtypes, including nonmyelinated versus myelinated Schwann cells, memory B cells versus plasma cells, and alveolar type 1 versus type 2 cells in the lungs. In Schwann cells, we observed relatively continuous changes from nonmyelinated cells to myelinated cells in the mCG embedding (fig. S32B, left), whereas two discrete clusters of subtypes were observed in the chromatin contact embedding (fig. S32B, right). A cluster of cells (c21-c7) clustered with myelinated cells based on mCG, whereas it clustered with nonmyelinated cells with chromatin contacts (fig. S32, B and C). Further examination of the DMRs and differential loops between the clusters revealed that for the same genome locus harboring DMRs and differential loops, cluster c21-c7 showed methylation patterns similar to those of myelinated cells, whereas chromatin looping patterns were similar to those of nonmyelinated cells (fig. S32D).

In addition, some lineages show almost no correlation between DNA methylation– and chromatin contact–defined clusters. For example, within placenta epithelial cells, we observe two chromatin contact clusters and six DNA methylation clusters (Fig. 6H). Notably, the DMRs between clusters from chromatin contact–based embedding were centered around the marker genes between villous cytotrophoblast (VCT) and syncytiotrophoblast (SCT) (Fig. 6I). These DMRs have high mCG at their flanking regions (Fig. 6J) and are usually found in the fully methylated compartment (Fig. 6K). Contrarily, the DMRs between methylation-based clustering show lower flanking mCG (Fig. 6J) and are more likely to be located in the partially methylated compartment (Fig. 6K). Taken together, these indicated that the 3D genome embedding reflected cell type differences more precisely in the trophoblast. Similarly, the same pattern was observed in epithelial cells in the stomach and intestine (fig. S32, E to H). Together, global patterns of DNA methylation may lack cell type information, whereas local methylation patterns within specific genes may be informative. This may suggest that DNA methylation still plays a critical role in regulating these genes and lineages but that in some lineages, global variation in DNA methylation across single cells is higher and not primarily defined by cell type.

Discussion

DNA methylation and 3D chromatin architecture are critical for proper gene regulation, but we currently lack comprehensive information on how these features vary across human cell types. To address this, we used snm3C-seq (*22*) to simultaneously profile DNA methylation and 3D chromatin architecture, characterizing the landscape of DNA methylation and 3D chromatin architecture at an unprecedented scale and cell type resolution, providing a comprehensive resource to the community.

In examining mCG across cell types, we have identified features that we call DNA methylation compartments related to patterns of PMDs (*31*). The patterns of DNA methylation compartments are highly correlated with patterns of A/B compartments (*56*), with the PMD-like compartments strongly associated with the repressive, heterochromatin B compartment regions of the genome (*56*, *57*). Notably, the methylation compartment patterns are more fine-grained than those from chromatin contact–based analysis, consistent with recent ultradeep Hi-C observations that A/B compartments may be smaller than previously realized (*58*). DNA methylation compartments may therefore reveal fine-grained patterns of chromatin compartmentalization more cost-effectively.

The association between PMD-like compartments and B compartment regions suggests that mechanisms maintaining DNA methylation vary by genomic context, particularly in heterochromatin. Previous studies have shown that late-replicating regions, also associated with the B compartment, progressively lose DNA methylation over cell divisions to form PMDs (*59*). These observations are consistent with a model in which PMD-like compartments are "primed" to become PMDs when cells divide rapidly, as may occur between naïve and mature memory lymphoid cells.

In addition to canonical mCG, mCH has previously been described as prominent in neurons and embryonic stem cells (*45*). In our data, most cells, apart from neuronal lineages, show low levels of mCH. Despite this, mCH patterns alone can distinguish many of the major cell types that we identified in this study, which suggests that mCH patterns reflect cell identity. Our data do not provide direct evidence of whether mCH is actively regulated through a different mechanism than mCG or whether it could regulate downstream gene expression. Within neurons, high mCH levels over genes influence binding of the repressor MECP2 (*60*), but whether lower mCH levels play a similar role across diverse cell types remains unclear. Future studies on mCH deposition and interpretation in nonbrain tissues are required to understand its functional role, as in recent brain studies (*61*, *62*).

We also characterized 3D genome organization across an unprecedented number of cell types. Distinct lineages exhibit compartment-dominant or domain-dominant chromatin phenotypes, similar to prior single-cell 3D genome studies in the human brain (*47*), with the degree of domain dominance and compartment dominance being anticorrelated. The basis for this is unclear but is reminiscent of the anticorrelated relationship between chromatin loops and compartments revealed on depletion of the cohesin complex (*63*–*65*). It is tempting to speculate that compartment- or domain-dominant phenotypes may result from differences in global rates of chromatin loop extrusion. Recent evidence suggests that loop extrusion may be preferentially initiated in euchromatic regions and arrested in B compartment regions in lymphocytes (*66*). Therefore, differences in the type or degree of heterochromatin within a lineage may be a contributing factor.

Our data show that DNA methylation and 3D genome structure can yield different pictures of the cellular composition within a tissue. Notably, when we cluster cells either separately or jointly based on DNA methylation and chromatin contacts, we observe some cell types with divergent classifications between the two features. Discordant signatures between epigenomic modalities have been described in previous single-cell RNA-ATAC and RNA–Hi-C studies and quantified using chromatin potential analysis (*67*–*69*). Our previous work has shown a discrepancy in human brain development where 3D genome changes precede DNA methylation and define radial glial cells precommitted to the astrocyte lineage (*47*). In this work, we identified multiple adult cell types with discordant annotations, including skeletal muscle myofibers and myelinated and nonmyelinated Schwann cells. Because these adult cell types also transit between each other (*70*–*72*), these observations could reflect different dynamic rates across modalities during cell state transitions and might be used to study such transitions in adult tissues. Reducing data sparsity in future single-cell multiomic technologies could further clarify the nature of discordant cell types, and understanding the mechanisms that give rise to divergent epigenomic patterns will be an important area for future study.

An important limitation of this work is that bisulfite sequencing cannot distinguish between 5mC and 5hmC. However, because 5hmC is enriched at regulatory elements whereas we observe mCH and mCG depletion there, the signals should not be dominated by 5hmC. This limitation should nonetheless be kept in mind when interpreting the data.

Materials and methods are available in the supplementary materials.

REFERENCES AND NOTES

1. C. D. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016). doi: 10.1038/nrg.2016.59; pmid: 27346641
2. F. Tang *et al.*, mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009). doi: 10.1038/nmeth.1315; pmid: 19349980
3. J. D. Buenrostro *et al.*, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015). doi: 10.1038/nature14590; pmid: 26083756
4. D. A. Cusanovich *et al.*, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015). doi: 10.1126/science.aab1601; pmid: 25953818
5. H. S. Kaya-Okur *et al.*, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019). doi: 10.1038/s41467-019-09982-5; pmid: 31036827
6. C. Luo *et al.*, Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017). doi: 10.1126/science.aan3351; pmid: 28798132
7. S. A. Smallwood *et al.*, Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014). doi: 10.1038/nmeth.3035; pmid: 25042786
8. A. Regev *et al.*, Science Forum: The Human Cell Atlas. *eLife* **6**, e27041 (2017). doi: 10.7554/eLife.27041
9. The Tabula Sapiens Consortium, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022). doi: 10.1126/science.abl4896
10. X. Han *et al.*, Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020). doi: 10.1038/s41586-020-2157-4; pmid: 32214235
11. S. He *et al.*, Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020). doi: 10.1186/s13059-020-02210-0; pmid: 33287869
12. L. W. Plasschaert *et al.*, A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018). doi: 10.1038/s41586-018-0394-6; pmid: 30069046
13. Y. E. Li *et al.*, A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023). doi: 10.1126/science.adf7044; pmid: 37824643
14. K. Zhang *et al.*, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021). doi: 10.1016/j.cell.2021.10.024; pmid: 34774128
15. J. C. Ulirsch *et al.*, Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019). doi: 10.1038/s41588-019-0362-6; pmid: 30858613
16. M. R. Corces *et al.*, Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020). doi: 10.1038/s41588-020-00721-x; pmid: 33106633
17. S. Domcke *et al.*, A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020). doi: 10.1126/science.aba7612; pmid: 33184180
18. W. Tian *et al.*, Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382**, eadf5357 (2023). doi: 10.1126/science.adf5357; pmid: 37824674
19. R. V. Nichols *et al.*, High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nat. Commun.* **13**, 7627 (2022). doi: 10.1038/s41467-022-35374-3; pmid: 36494343
20. T. Nagano *et al.*, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013). doi: 10.1038/nature12593; pmid: 24067610
21. L. Tan *et al.*, Lifelong restructuring of 3D genome architecture in cerebellar granule cells. *Science* **381**, 1112–1119 (2023). doi: 10.1126/science.adh3253; pmid: 37676945
22. D.-S. Lee *et al.*, Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019). doi: 10.1038/s41592-019-0547-z; pmid: 31501549
23. N. Rappoport *et al.*, Single cell Hi-C identifies plastic chromosome conformations underlying the gastrulation enhancer landscape. *Nat. Commun.* **14**, 3844 (2023). doi: 10.1038/s41467-023-39549-4; pmid: 37386027
24. M. D. Schultz *et al.*, Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015). doi: 10.1038/nature14465; pmid: 26030523
25. R. Lister *et al.*, Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013). doi: 10.1126/science.1237905; pmid: 23828890
26. A. D. Schmitt *et al.*, A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016). doi: 10.1016/j.celrep.2016.10.061; pmid: 27851967
27. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
28. J. R. Dixon *et al.*, Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015). doi: 10.1038/nature14222; pmid: 25693564
29. Z. Xu *et al.*, Structural variants drive context-dependent oncogene activation in cancer. *Nature* **612**, 564–572 (2022). doi: 10.1038/s41586-022-05504-4; pmid: 36477537
30. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
31. R. Lister *et al.*, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009). doi: 10.1038/nature08514; pmid: 19829295
32. N. Loyfer *et al.*, A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023). doi: 10.1038/s41586-022-05580-6; pmid: 36599988
33. J. Rozowsky *et al.*, The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models. *Cell* **186**, 1493–1511.e40 (2023). doi: 10.1016/j.cell.2023.02.018; pmid: 37001506
34. K. D. Hansen *et al.*, Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43**, 768–775 (2011). doi: 10.1038/ng.865; pmid: 21706001
35. D. I. Schroeder *et al.*, The human placenta methylome. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6037–6042 (2013). doi: 10.1073/pnas.1215145110; pmid: 23530188
36. Q. Song *et al.*, A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLOS ONE* **8**, e81148 (2013). doi: 10.1371/journal.pone.0081148; pmid: 24324667
37. G. C. Hon *et al.*, Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22**, 246–258 (2012). doi: 10.1101/gr.125872.111; pmid: 22156296
38. A. Nishiyama, M. Nakanishi, Navigating the DNA methylation landscape of cancer. *Trends Genet.* **37**, 1012–1027 (2021). doi: 10.1016/j.tig.2021.05.002; pmid: 34120771
39. Y. Chen *et al.*, The role of *Tbx20* in cardiovascular development and function. *Front. Cell Dev. Biol.* **9**, 638542 (2021). doi: 10.3389/fcell.2021.638542; pmid: 33585493
40. L. Wang *et al.*, The multiple roles of interferon regulatory factor family in health and disease. *Signal Transduct. Target. Ther.* **9**, 282 (2024). doi: 10.1038/s41392-024-01980-4; pmid: 39384770
41. S. M. Hedrick, R. H. Michelini, A. L. Doedens, A. W. Goldrath, E. L. Stone, FOXO transcription factors throughout T cell biology. *Nat. Rev. Immunol.* **12**, 649–661 (2012). doi: 10.1038/nri3278; pmid: 22918467
42. T. McLachlan *et al.*, B-cell lymphoma 6 (BCL6): From master regulator of humoral immunity to oncogenic driver in pediatric cancers. *Mol. Cancer Res.* **20**, 1711–1723 (2022). doi: 10.1158/1541-7786.MCR-22-0567; pmid: 36166198
43. A. E. Almada, A. J. Wagers, Molecular circuitry of stem cell fate in skeletal muscle regeneration, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **17**, 267–279 (2016). doi: 10.1038/nrm.2016.7; pmid: 26956195
44. M. H. Dominguez, A. E. Ayoub, P. Rakic, POU-III transcription factors (Brn1, Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex. *Cereb. Cortex* **23**, 2632–2643 (2013). doi: 10.1093/cercor/bhs252; pmid: 22892427
45. Y. He, J. R. Ecker, Non-CG methylation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015). doi: 10.1146/annurev-genom-090413-025437; pmid: 26077819
46. A. Mo *et al.*, Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015). doi: 10.1016/j.neuron.2015.05.018; pmid: 26087164
47. M. G. Heffel *et al.*, Temporally distinct 3D multi-omic dynamics in the developing human brain. *Nature* **635**, 481–489 (2024). doi: 10.1038/s41586-024-08030-7; pmid: 39385032
48. M. A. Hale *et al.*, The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. *Development* **141**, 3123–3133 (2014). doi: 10.1242/dev.109405; pmid: 25063451
49. H. Bai, T. Sakurai, J. D. Godkin, K. Imakawa, Expression and potential role of GATA factors in trophoblast development. *J. Reprod. Dev.* **59**, 1–6 (2013). doi: 10.1262/jrd.2012-100; pmid: 23428586
50. H. K. Finucane *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678
51. F. Le Grand, M. A. Rudnicki, Skeletal muscle satellite cells and adult myogenesis. *Curr. Opin. Cell Biol.* **19**, 628–633 (2007). doi: 10.1016/j.ceb.2007.09.012; pmid: 17996437
52. R. G. Jones 3rd, F. von Walden, K. A. Murach, Exercise-induced MYC as an epigenetic reprogramming factor that combats skeletal muscle aging. *Exerc. Sport Sci. Rev.* **52**, 63–67 (2024). doi: 10.1249/JES.0000000000000333; pmid: 38391187
53. F. Marchildon, D. Fu, N. Lala-Tabbert, N. Wiper-Bergeron, CCAAT/enhancer binding protein beta protects muscle satellite cells from apoptosis after injury and in cancer cachexia. *Cell Death Dis.* **7**, e2109 (2016). doi: 10.1038/cddis.2016.4; pmid: 26913600
54. M. Xu, X. Chen, D. Chen, B. Yu, Z. Huang, FoxO1: A novel insight into its molecular mechanisms in the regulation of skeletal muscle differentiation and fiber type specification. *Oncotarget* **8**, 10662–10674 (2017). doi: 10.18632/oncotarget.12891; pmid: 27793012
55. A. Lu *et al.*, NF-κB negatively impacts the myogenic potential of muscle-derived stem cells. *Mol. Ther.* **20**, 661–668 (2012). doi: 10.1038/mt.2011.261; pmid: 22158056
56. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
57. T. Ryba *et al.*, Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010). doi: 10.1101/gr.099655.109; pmid: 20430782

58. H. L. Harris *et al.*, Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.* **14**, 3303 (2023). doi: 10.1038/s41467-023-38429-1; pmid: 37280210
59. J. L. Endicott, P. A. Nolte, H. Shen, P. W. Laird, Cell division drives DNA methylation loss in late-replicating domains in primary human cells. *Nat. Commun.* **13**, 6659 (2022). doi: 10.1038/s41467-022-34268-8; pmid: 36347867
60. R. Tillotson *et al.*, Neuronal non-CG methylation is an essential target for MeCP2 function. *Mol. Cell* **81**, 1260–1275.e12 (2021). doi: 10.1016/j.molcel.2021.01.011; pmid: 33561390
61. A. W. Clemens *et al.*, MeCP2 represses enhancers through chromosome topology-associated DNA methylation. *Mol. Cell* **77**, 279–293.e8 (2020). doi: 10.1016/j.molcel.2019.10.033; pmid: 31784360
62. N. Hamagami *et al.*, NSD1 deposits histone H3 lysine 36 dimethylation to pattern non-CG DNA methylation in neurons. *Mol. Cell* **83**, 1412–1428.e7 (2023). doi: 10.1016/j.molcel.2023.04.001; pmid: 37098340
63. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24 (2017). doi: 10.1016/j.cell.2017.09.026; pmid: 28985562
64. W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017). doi: 10.1038/nature24281; pmid: 29094699
65. J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L. A. Mirny, Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6697–E6706 (2018). doi: 10.1073/pnas.1717730115; pmid: 29967174
66. Y. Guo *et al.*, Chromatin jets define the properties of cohesin-driven *in vivo* loop extrusion. *Mol. Cell* **82**, 3769–3780.e5 (2022). doi: 10.1016/j.molcel.2022.09.003; pmid: 36182691
67. S. Ma *et al.*, Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 (2020). doi: 10.1016/j.cell.2020.09.056; pmid: 33098772
68. Z. Liu *et al.*, Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science* **380**, 1070–1076 (2023). doi: 10.1126/science.adg3797; pmid: 37289875
69. S. Mitra *et al.*, Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. *Nat. Genet.* **56**, 627–636 (2024). doi: 10.1038/s41588-024-01689-8; pmid: 38514783
70. J. M. Wilson *et al.*, The effects of endurance, strength, and power training on muscle fiber type shifting. *J. Strength Cond. Res.* **26**, 1724–1729 (2012). doi: 10.1519/JSC.0b013e318234eb6f; pmid: 21912291
71. D. L. Plotkin, M. D. Roberts, C. T. Haun, B. J. Schoenfeld, Muscle fiber type transitions with exercise training: Shifting perspectives. *Sports* **9**, 127 (2021). doi: 10.3390/sports9090127; pmid: 34564332
72. J. L. Salzer, Switching myelination on and off. *J. Cell Biol.* **181**, 575–577 (2008). doi: 10.1083/jcb.200804136; pmid: 18490509

ACKNOWLEDGMENTS

We thank R. Ilango, S. Newins, N. D. Youngblut, and D. P. Burke for helping embed the web application into Arc domain names. We thank B. Plosky and J. Caputo for assistance with manuscript preparation. We thank R. Zhang and S. He for insightful discussions. **Funding:** This work was supported by the National Human Genome Research Institute (NHGRI) (grant 5R01HG010634 to J.R.E. and J.R.D.), the National Cancer Institute (NCI) (grant 5U01CA260700 to J.R.D.), and the Wellcome Trust (216596/Z/19/Z to J.C.-K.). J.Z. is supported by funding from the Arc Institute. J.R.D. is also supported as a Pew Biomedical Scholar and a Rita Allen Foundation Scholar. The Flow Cytometry Core Facility of the Salk Institute is supported by funding from the National Cancer Institute of the National Institutes of Health [Cancer Center Support Grant P30 014195 and Shared Instrumentation Grant S10-OD023689 (Aria Fusion cell sorter) and S10 OD034268 (Thermo Fisher Bigfoot)]. J.R.E. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** Conceptualization: J.Z., C.L., J.R.D., J.R.E.; Data analysis: J.Z., Y.W., H.L., W.T., D.S.; Data production: R.G.C., A.B., J.R.N., B.C., S.M., M.L., N.C., L.B., C.O.; Sample source: C.B., R.A.e.D., A.K.W., M.S., J.C.-K., S.Z., M.P.S., S.P., B.R.; Visualization tool: Z.Z., G.Y., J.Z., S.C.; Writing – original draft: J.R.D., J.Z., Y.W.; Writing – review & editing: all authors. **Competing interests:** J.R.E. is a scientific adviser for Zymo Research, Inc.; Ionis Pharmaceuticals; and Guardant Health. B.R. is a cofounder and consultant for Arima Genomics, Inc., and cofounder of Epigenome Technologies. M.P.S. is a cofounder and the scientific advisory board member of Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, Iollo, Marble Therapeutics, Crosshair Therapeutics, NextThought, and Mirvie. He is a scientific advisor of Jupiter, Neuvivo, Swaza, Mitrix, Yuvan, TranscribeGlass, and Applied Cognition. The authors declare no other competing interests. **Data, code, and materials availability:** Single-cell DNA methylation data, chromatin contact data, and cell metadata generated in this study have been deposited at https://huggingface.co/zhoujt1994/datasets in tsv format. Raw data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE326618. The analysis code is available at https://github.com/zhoujt1994/HumanCellEpigenomeAtlas.git and Zenodo (*106*). Interactive browser and download links for more processed files can be found at https://humancellepigenomeatlas.arcinstitute.org. All materials are available on request.

License information: Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. This research was also funded in whole or in part by the Wellcome Trust (216596/Z/19/Z), a cOAlition S organization; as required, the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx0673
Materials and Methods; Figs. S1 to S33; Tables S1 to S6; References (*73*–*106*); MDAR Reproducibility Checklist

Submitted 23 March 2025; resubmitted 11 January 2026; accepted 7 April 2026
10.1126/science.adx0673



AAAS.ORG/COMMUNITY







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

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

INVASIVE SPECIES

Underestimated impacts

Non-native invasive species have outsized environmental and economic impacts. More invasive species are recognized in the Global North, where research money and effort are higher, but this bias may underestimate invasive species impacts in the Global South. Bacher *et al.* compared the average reported severity of invasive species impacts across countries. They found higher impact severity in the Global South, particularly in smaller countries, but also in large, high-income developing economies, where economic activity may offer opportunities for species introductions. Impact severity was negatively related to governance, and Global South nations, on average, had lower capacity to prevent and respond to invasive species. These findings highlight the need for increased research and management of invasive species in the Global South.
—Bianca Lopez *Science* p. 376, 10.1126/science.aed0603

Ecosystems in the Global South face substantial threats from invasive species such as water hyacinth, pictured here completely covering a lake in Madagascar.

LANGUAGE DIVERSITY

A historical context for language loss

More than four languages are lost every year, and this rate is predicted to increase. Blasi *et al.* investigated how language diversity has changed over the Holocene using a Bayesian modeling approach, ethnographic data, and estimates of the total human population size over time. They assumed a maximum size of ethnolinguistic groups and that the number of different ethnolinguistic groups is related to the number of languages. Their findings suggested that language loss is not a new phenomenon, and that language diversity has been decreasing rapidly over the past one to three millennia following a peak of an order of magnitude higher diversity than we see today.
—Bianca Lopez
Science p. 387, 10.1126/science.adx4343

PHYSICS

Active film-substrate interactions

Traditionally, underlying substrates have been considered static, passive participants, with certain exceptions such as piezoelectric or flexible materials. However, whether dynamic changes in the film could propagate into the substrate and if such substrate changes could, in turn, influence film behavior remains an open question. Using combined x-ray and electron microscopy, Kisiel *et al.* demonstrated that thin films of VO_2 not only induced significant strain deep into the underlying Al_2O_3 substrate under equilibrium conditions, but also that this strain could reciprocally modify film properties during localized phase transitions. Their findings challenge conventional assumptions and could open new avenues for thin-film physics, device engineering, neuromorphic computing, and x-ray imaging.
—Yury Suleymanov
Science p. 392, 10.1126/science.adt9347

EXOPLANETS

The magnetic field of an exoplanet

The magnetic fields of extrasolar planets are far too small to be observed directly. However, if an exoplanet orbits sufficiently close to its host star, then its magnetic field could induce detectable effects on the star's activity level. Revilla *et al.* analyzed long-term spectroscopic monitoring of a red dwarf star closely orbited by a Neptune-sized planet. Using spectral lines that are sensitive to the star's atmosphere, they identified a periodic increase in activity at frequencies related

PHOTO: NICK GARBUTT / NPL / MINDEN PICTURES

to the exoplanet's orbit. The strength of this star-planet interaction allowed the authors to determine the magnetic field of the exoplanet. —Keith T. Smith

Science p. 403, 10.1126/science.adv3075

NANOMATERIALS

Taking steps to control graphene

Phase-pure rhombohedral graphene, which has an ABC layer stacking, has been grown through step geometry–guided epitaxy. Zhao *et al.* used a copper-nickel–coated sapphire surface with a stepped geometry chosen so that interlayer slip and orientation angles of the graphene layers favored rhombohedral stacking. The layer-antiferromagnetic state and the quantum anomalous Hall effect for this phase were determined from electronic transport measurements. —Phil Szuromi

Science p. 422, 10.1126/science.aed9202

DRUG DESIGN

Versatile and just reactive enough

Covalent inhibitors are a class of molecules that has gained increased attention as approaches to improve the target selectivity of these drugs have advanced. However, off-target reactions can occur if the drug molecule's electrophilic warhead is too reactive. Shultz *et al.* identified a synthetic pathway to generate sulfonyl and sulfonimidoyl bicyclobutanes from late-stage drug precursors containing a primary or secondary amine (see the Perspective by Kraft and Gehringer). This approach permits the installation of a selectively reactive electrophile into complex molecules, taking the place of an acrylamide moiety. The researchers tested several drug molecules and found that they could be converted into covalent inhibitors with favorable activity and pharmacodynamic profiles. —Michael A. Funk

Science p. 408, 10.1126/science.adx7219; see also p. 357, 10.1126/science.aej4566

CANCER THERAPY

Lethal combinations with NDRG1

DNA damage repair is essential to cell survival, making it an attractive target for cancer therapy. Using high-throughput screens, Mkrtchyan *et al.* identified a therapeutically exploitable role for the scaffolding protein NDRG1 in the DNA damage response. High *NDRG1* expression in cancer cells correlated with sensitivity to the antimalarial drug quinacrine, which impaired the ubiquitylation-dependent recruitment of a critical protein to sites of DNA damage. Colorectal cancer cells with mutations in two DNA repair–associated genes were particularly sensitive to quinacrine or NDRG1 loss. These findings suggest that therapy may be informed by *NDRG1* expression. —Leslie K. Ferrarelli

Sci. Signal. (2026) 10.1126/scisignal.adv4272

MACROPHAGES

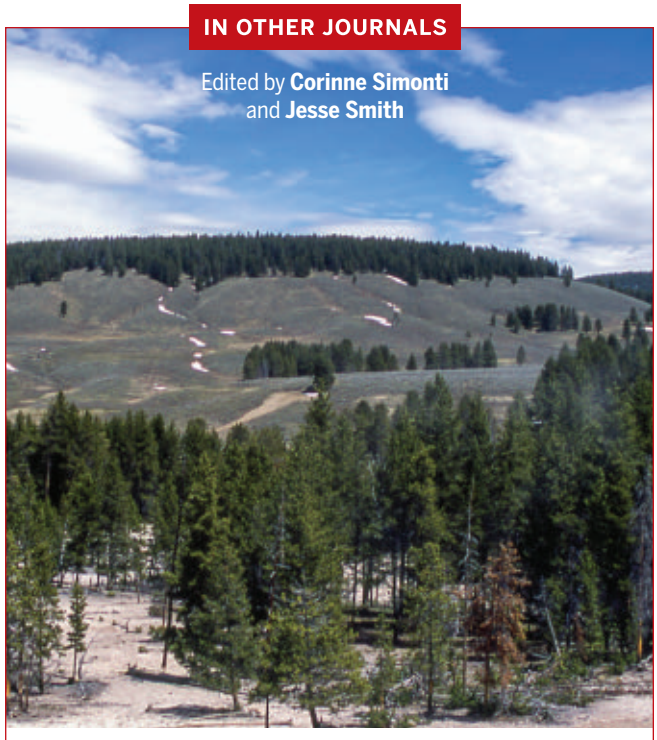
Oxysterol organization of lung immunity

Although fungal pathogens strongly induce type 2 immunity, these same responses can also impair fungal clearance. Using a mouse model of *Cryptococcus neoformans* infection, Zheng *et al.* found that cholesterol-derived oxysterols produced by lung macrophages position T helper 2 (T_H2) cells within fungal granulomas. The G protein–coupled receptor GPR183 mediated sensing of chemotactic oxysterols by T_H2 cells, suppressed macrophage responsiveness to interferon-γ, and promoted *C. neoformans* persistence. These findings identify lung macrophages as key organizers of pulmonary type 2 immunity through oxysterol-mediated positioning of T_H2 cells. —Claire Olingy

Sci. Immunol. (2026) 10.1126/sciimmunol.aee0990

IN OTHER JOURNALS

Edited by **Corinne Simonti** and **Jesse Smith**



YELLOWSTONE CALDERA

Explosive new details

The Lava Creek Tuff (LCT) is the widespread ash pile left behind during Yellowstone's youngest supereruption 631,000 years ago. However, the presumed two-stage eruption process and subsequent resurgent uplift have recently been called into question. Myers *et al.* undertook focused field mapping and mineralogical assessment of LCT layers within the Sour Creek Dome, a site of heightened deformation, fluid efflux, and shallow magma activity. They found that the LCT event involved at least five stages, with each eruption originating from distinct magma chambers that vented well inside the mapped caldera boundary. —Angela Hessler *Geosphere* (2026) 10.1130/GES02939.1

The Lava Creek Tuff, deposited during emplacement of the Sour Creek Dome in Yellowstone National Park, accumulated over at least five stages.

PHOTO: NPS/JO SUDERMAN

PLANT PHYSIOLOGY

How plant wounds recruit fuel

Photosynthetic sugars are transported through the phloem to fuel plant growth, but how they are redirected to wounded tissue for repair and regeneration is not clear. Combining root tip regeneration assays with reporter lines in the model plant *Arabidopsis*, Matosevich *et al*. found that sucrose is restricted from the site of injury. At the same time, wounding-induced enzymes hydrolyze sucrose into hexoses, and hexose transporters enable neighboring cells to rapidly take up glucose from the cell wall space. The resulting sugar gradient drives sucrose import toward the wound area. Activation of similar transporter genes in other wounding contexts indicates that this is a general mechanism used by plants to direct resources after injury. —Unnati Sonawala

Proc. Natl. Acad. Sci. U.S.A. (2026) 10.1073/pnas.2535587123

CELL BIOLOGY

Getting into a lipid droplet

Lipid droplets are important cellular hubs for a variety of metabolic functions. They are composed of a phospholipid monolayer packed with neutral lipids, and their surfaces contain membrane proteins that are delivered from the endoplasmic reticulum (ER). Exactly how this delivery is achieved remains unclear. Mizrak *et al.* showed that lipid droplet proteins diffusing in the ER membrane become confined in nanodomains within the lipid droplet membrane where they accumulate. Single-molecule live-cell imaging revealed lateral transport of droplet-destined proteins across seipin-containing ER-lipid droplet bridges. Thus, the targeting and accumulation of proteins in the lipid droplet involves selective protein accumulation and membrane bridges. —Stella M. Hurtley

Nat. Cell Biol. (2026) 10.1038/s41556-026-01963-3

IMMUNOLOGY

Clues in early immune responses to TB

Most people who are infected with *Mycobacterium tuberculosis* do not progress to active disease. Branchett *et al.* used single-cell sequencing approaches to profile immune cells in the lungs of individuals who were recent household contacts for patients with tuberculosis (TB). Contacts who went on to develop TB exhibited elevated neutrophils and dysfunctional T cells in the lung. By contrast, the immune responses in people who did not progress toward disease were dominated by protective T cells with a quiescent or stem-like state. These findings indicate that the early immune response of an individual to *M. tuberculosis* may dictate the outcome of the infection. —Sarah H. Ross

Nat. Immunol. (2026) 10.1038/s41590-026-02544-0

ADDITIVE MANUFACTURING

Simple 3D ceramic printing

A standard route to three-dimensional (3D) printing of ceramics makes use of polymer precursors that can be precisely printed using two-photon techniques and then converted into a ceramic at elevated temperature. Luo *et al.* developed a process that allows the fabrication of complex silicon oxycarbide shapes with excellent mechanical properties using simple precursors. They made a resin from polycarbonsilane and combined it with ethyl methyl ketone and xylene as a photoinitiator. A key step in the process is the heating of the resin in air at 90°C to induce some cross-linking and increase the viscosity. Control over the high-temperature conversion process allows for tuning of the final structures. The authors demonstrated the potential of their technique by printing truss networks and hollow microneedles. —Marc S. Lavine

ACS Nano (2026) 10.1021/acsnano.6c02930

MACHINE LEARNING

ML accelerates polymer blend discovery

Using accurate experimental training data remains a central challenge for integrating machine learning (ML) into many scientific disciplines, a gap that automated high-throughput experimentation can help bridge by generating the large-scale datasets that ML requires. Wang *et al.* developed a data-driven workflow integrating modular polymer synthesis, robotic blending, automated atomic force microscopy characterization, and ML-guided phase behavior prediction for accelerated discovery of supramolecular polymer blends (SPBs). From just 33 homopolymer precursors, 260 SPBs were fabricated and fully characterized in weeks rather than years of manual labor, yielding 2340 high-quality morphology datasets. This provides a foundation sufficient for developing predictive ML models to probe structure-property relationships and facilitate morphology prediction and inverse design of SPBs. —Yury Suleymanov

JACS Au (2026) 10.1021/jacsau.6c00041



PHOTO: SHIGEKI IIMURA/NATURE PRODUCTION/MINDEN PICTURES

AGRICULTURE

Morning blooming protects from heat

Global warming is becoming a serious threat to crop production. Rice is especially sensitive to high temperatures during flowering, which can reduce fertilization and grain yield. Ishizaki *et al.* investigated whether changing the timing of flower opening could help rice avoid heat damage. They identified a gene called *EMF3* that controls flower opening time through a previously unknown mechanism in the anther. Rice plants carrying early-flowering alleles of *EMF3* flowered during the cooler morning hours and maintained a higher seed-setting rate under heat stress. These results provide a new strategy for breeding rice varieties that are better adapted to global warming and for improving hybrid rice seed production. —Di Jiang

Plant Biotechnol. J. (2026) 10.1111/pbi.70653

Gene editing rice for earlier flowering time may help protect against climate change–induced heat stress.

Edited by **Michael Funk**

BIRDSONG

Environment shapes birdsong

Birds sing to attract mates and ward off rivals. The form of these songs can vary from single, repeated notes to complex melodies. Bacquelé *et al.* explored the relationships among song, environment, and latitude from birds across the world (see the Perspective by Briefer). They found that in regions where forests are thick, and thus sound transmission is slower, bird songs are simpler and more effective for transmission across vegetation-impeded landscapes. By contrast, in temperate regions, where sound is less impeded and sexual selection is especially strong due to seasonal limitations, birdsong is much more complex, likely being a better indicator of singer condition. —Sacha Vignieri

Science p. 381, 10.1126/science.aee6239; see also p. 354, 10.1126/science.aej0087

COMPARATIVE PHYSIOLOGY

Mighty larvae

Insects have colonized nearly every environment on earth, including the poles, high elevations, and fresh waters. They are conspicuously absent from the deep sea, however. One hypothesis for this is that their tracheal system would collapse at depth during vertical migrations to avoid predators. McKenzie *et al.* studied the larvae of a fly that lives in Lake Malawi and found that it has evolved an air-filled tracheal system that resists crushing at depth. This facilitates diel vertical migrations within the lake, with tests showing resistance at depths of half a kilometer (see the Perspective by Harrison and Woods). These results may suggest that the sea awaits eventual insect colonization. —Sacha Vignieri

Science p. 398, 10.1126/science.aed0667; see also p. 356, 10.1126/science.aei9743

SURFACE CHEMISTRY

Controlling collision angles

The reaction of projectile CF_2 molecules with an organic target molecule on a copper (110) surface was studied with control over reactant trajectories and alignment. Timm *et al.* used a scanning tunneling microscope to initiate and image this reaction, in which energetic CF_2 molecules were created by electron-induced dissociation of absorbed CF_3 groups (see the Perspective by Björk). The addition reaction with dibromoterfluorene, which had numerous adsorption alignments on the surface, occurred only along a ±3° range of target alignments. Theoretical calculations attributed this narrow cone of reaction to steric effects around the reactive carbon atom. —Phil Szuromi

Science p. 417, 10.1126/science.aec7913; see also p. 359, 10.1126/science.aej5890

WEARABLES

Monitoring movement

Evaluation of infant spontaneous motor activity currently requires clinic visits and trained observers. Airaksinen *et al.* used a wearable motion-sensing jumpsuit to monitor infant movement and postures at home during free play. Deep-learning classifiers objectively characterized 220 motor metrics in 92 typically developing infants aged 4 to 19 months, generating motor growth charts of emerging postures and their dynamics. Approximately 91% of these motor metrics generalized to an independent clinical cohort of infants including both typical and atypical neurodevelopment, and 55% of the metrics statistically distinguished the two groups. These findings suggest that wearable sensors could provide objective, home-based quantitative measures of motor development dynamics to support routine care or to better understand developmental patterns. —Molly Ogle

Sci. Transl. Med. (2026) 10.1126/scitranslmed.adz7035

LET YOUR KNOWLEDGE BLOSSOM

Visit the *Science* Custom Publishing sites today and grow your knowledge with a wide selection of booklets, podcasts, posters, sponsored features, and webinars!

Booklets | **Podcasts** | **Posters** | **Sponsored Features** | **Webinars**



Brought to you by the *Science*/AAAS Custom Publishing Office.

Science AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

PUT YOUR RESEARCH OUT IN FRONT



Submit your research:
cts.ScienceMag.org

Science Signaling
AAAS

Science.org/journal/signaling

INVASIVE SPECIES

Invasive species' environmental impacts are more severe in the Global South

Sven Bacher[1]†, Petr Pyšek[2,3], Hanno Seebens[4,5]†*

Invasive alien species (IAS) are a major threat to biodiversity, yet their impact distribution remains poorly understood. Global biodiversity assessments suggest the highest impacts in wealthy countries of the Global North, but this is only inferred from IAS numbers, reflecting research bias. Using a new global database of standardized impact measures, we calculated average impact severity per country and found that it is higher in the Global South despite more than twice as many reports in the Global North. Weak governance and limited management capacity are the main drivers of high impact severity. Emerging economies with rapid economic growth but poor governance are particularly vulnerable. Failure to recognize the Global South as facing the highest IAS impacts diverts attention away from the most threatened regions.

The introduction of species into regions outside their native range—a process known as biological invasion—represents one of the major drivers of biodiversity loss (*1*). Some of these species become invasive and pose harm to the environment, society, or economy, with far-reaching consequences for biodiversity and human well-being. Up to now, these invasive alien species (IAS) contributed to 60% of all documented global species extinctions and caused economic costs exceeding USD 400 billion in 2019 (*2*). The ongoing increase in the number of newly introduced species (*3*, *4*) and associated rises in environmental and economic impacts (*5*, *6*) indicate that current policy and management efforts are insufficient to halt biological invasions globally. Effective and targeted management requires a robust understanding of the geographic distribution and magnitude of IAS impacts.

Although the spatial distribution of alien species has recently received much attention (*7*–*10*), the global geographical distribution of environmental impacts of IAS remains unknown, although this is key information to guide policy and management. In global biodiversity assessments and studies on major biodiversity threats, the impacts of IAS are often inferred indirectly from the numbers of reported IAS (*1*) or from proxies such as global trade volume or traffic intensity (*11*, *12*). These studies concluded that the impacts of IAS are highest in countries of the Global North, where IAS numbers are also highest (*8*, *13*, *14*). This conclusion is fueled by fewer impacts reported from countries of the Global South (*5*, *15*).

The Global South, as defined by the United Nations Conference on Trade and Development (UNCTAD) (*16*), is a political designation that refers to developing countries—primarily located in the southern hemisphere—that generally experience lower income levels and face a range of structural problems, such as high poverty, poor infrastructure, and deficient governance (*17*), in contrast to the mostly developed countries comprising the Global North. High potential for economic growth (*18*) in emerging economies, paired with high levels of corruption (*19*), potentially lead to high rates of IAS introductions in the Global South through trade (*13*, *20*), while simultaneously limiting governance and thus the capacity to implement effective IAS management strategies (*21*). This combination should result in high impacts of IAS in countries of the Global South, which contradicts existing biodiversity assessments and studies (*1*, *11*, *12*). However, conclusions based simply on numbers of IAS are biased because they conflate species presence with impact magnitude and often reflect disparities in research effort, which is likely greater in the Global North (*2*, *22*).

Until recently, a global analysis of the distribution of impacts was not possible owing to the lack of (i) an impact metric independent of research effort and (ii) a global, standardized cross-taxonomic dataset of impacts and their severity. The latter limitation has now been addressed with the publication of the new comprehensive Global Impacts Dataset of IAS (GIDIAS), which provides standardized information on the local magnitudes of >22,000 impacts for >3300 IAS (*15*). GIDIAS is based on >6700 sources (scientific publications and gray literature) and contains documented impacts of plants, vertebrates, invertebrates, and microorganisms from 172 countries. In accordance with the definition of the International Union for Conservation of Nature (IUCN) (*23*), we consider species with documented impacts as listed in GIDIAS as "invasive" and use this term in the paper. For comparison to all alien species (not only those for which impacts have been documented), we used the dataset Standardising and Integrating Alien Species (SInAS), which provides information about the worldwide distributions of >39,700 alien species across all taxonomic groups (*24*). This dataset is based on published checklists for countries or subnational units (such as states or islands far away from the mainland country) called "regions" in the following text.

Combining GIDIAS with datasets of alien species distributions, we compared the geographic patterns of IAS numbers from various taxonomic groups with the distribution of their negative environmental impacts across countries worldwide. The assessment of impact magnitudes follows the global standard for Environmental Impact Classification for Alien Taxa (EICAT) (*25*) of the IUCN (*23*). Under EICAT, impacts are classified as negative when a native species is harmed by the introduction of an IAS (*25*) and as severe when it results in a native population decline, local extirpation, or global extinction (*25*–*27*). As the number of species and their documented environmental impacts depend on the number of reports from a region, we explored both their distributions and relationship in absolute and relative terms. As a relative measure for documented impacts, we suggest a new metric, "average impact severity," which is independent of research effort and calculated as the ratio of documented severe impacts to all reported environmental impacts in a region; it can be interpreted as the regional risk of severe impacts from introduced IAS.

We tested the following hypotheses: (i) The numbers and magnitudes of documented impacts vary across regions owing to the number of IAS present in that region and the impacts of individual IAS, but they also depend on the capacity and governance to manage IAS and their impacts. (ii) The numbers of reported IAS and documented impacts are affected by research effort, leading to an overrepresentation of both metrics in countries of the Global North, where research effort is higher (*22*). (iii) The average impact severity is highest in countries with high introduction rates of IAS and low capacity to govern. We expect that wealthier countries, although experiencing high introduction rates of IAS, have the capacity to mitigate impact severity, whereas poorer countries with less capacity face more severe impacts. Lastly, we compared the importance of governance relative to other variables that may influence impact severity, such as introduction

[1]Department of Biology, University of Fribourg, Fribourg, Switzerland. [2]Czech Academy of Sciences, Institute of Botany, Průhonice, Czech Republic. [3]Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic. [4]Faculty of Biology and Chemistry, Justus-Liebig-University Giessen, Giessen, Germany. [5]Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany. ***Corresponding author: hanno.seebens@allzool.bio.uni-giessen.de** †These authors contributed equally to this work.

rates of IAS (e.g., through economic activity) and environmental conditions (e.g., country area, climatic conditions, island status, and anthropogenic disturbance).

Distribution of IAS and impacts

The documented numbers of alien species per country are unevenly distributed globally, with most recorded in North America, Europe, and Australasia (Fig. 1A). On other continents, high alien species numbers were found in, e.g., Japan, India, Russia, China, South Africa, Argentina, and Brazil. A similar spatial pattern was observed for the total number of documented impacts (Fig. 1B), IAS (Fig. 1C), and severe impacts (Fig. 1D). In relative terms, the pattern changes: The highest proportions of IAS (Fig. 1E) and proportions of severe impacts per country (i.e., average impact severity, Fig. 1F) occur in the countries of South America, Africa, and Asia. This pattern is particularly pronounced for average impact severity, for which high values were also found in Central America, the Caribbean, and on oceanic islands.

The distribution of the total numbers of alien species and numbers of documented impacts is highly correlated [Pearson's correlation coefficient (r) = 0.89, df = 138, $P < 0.001$; Fig. 2A]. Aggregating the values at the continental scale shows that Europe and North America have high alien species numbers and high numbers of impacts, whereas Central America and the Pacific have the lowest (fig. S1B). Numbers of IAS and severe impacts are highly correlated at the regional level (Pearson's $r = 0.9$, df = 138, $P < 0.001$; fig. S1C). At the continental level, the correlation is weaker but still significant (Pearson's $r = 0.68$, df = 7, $P < 0.05$; fig. S1D). By contrast, the proportion of IAS and the average impact severity are neither correlated at regional (fig. S1E) nor continental level (fig. S1F).

Absolute numbers of all alien species, IAS, all documented impacts, and severe documented impacts were highly correlated to research effort to address alien species [measured as the number of studies per region providing checklists of alien species (*24*)] (Pearson's $r > 0.6$, $P < 0.001$ for all four variables; fig. S2). By contrast, relative numbers, i.e., the proportion of IAS and average impact severity, were uncorrelated with research effort (Pearson's $r < 0.12$, $P > 0.2$ for both variables; fig. S2). It can be argued that severe impacts might be reported earlier than weaker impacts because severe impacts are more apparent or disturbing. This would result in an over-representation of severe impacts in regions with a low number of impact reports, which are mostly small countries located in the Global South. However, as in a

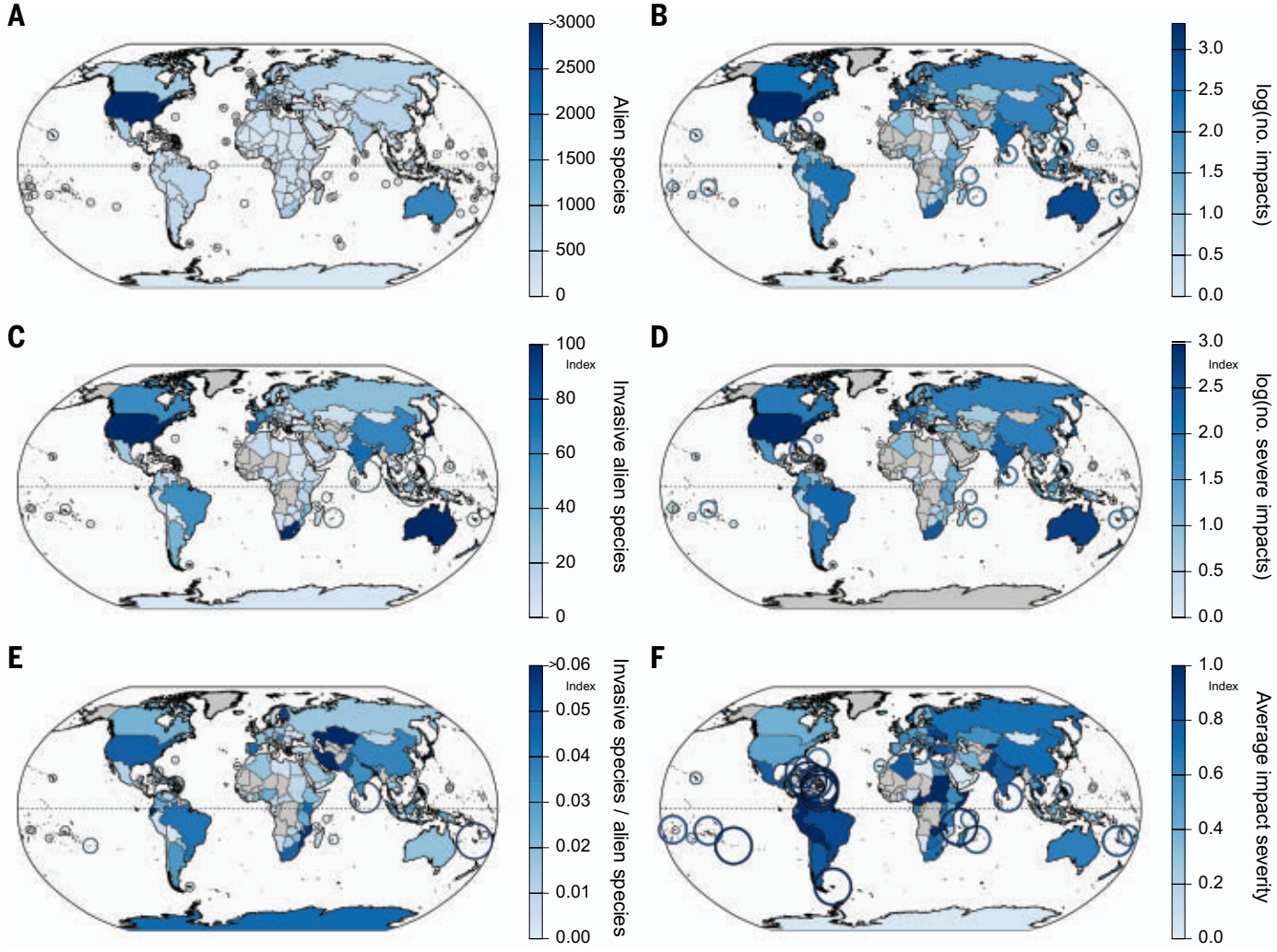


Fig. 1. Maps of documented numbers of alien species (left) and impacts (right) per region. (**A**) Number of all alien species. (**B**) Number of all impacts. (**C**) Number of invasive alien species (IAS). (**D**) Number of severe impacts. (**E**) Proportion of invasive to alien species. (**F**) Average impact severity (i.e., proportion of severe impacts to all impacts). All regions with at least one documented impact are included. Missing information is indicated by gray color. Islands are highlighted by circles, and circle sizes indicate the value for the respective island.

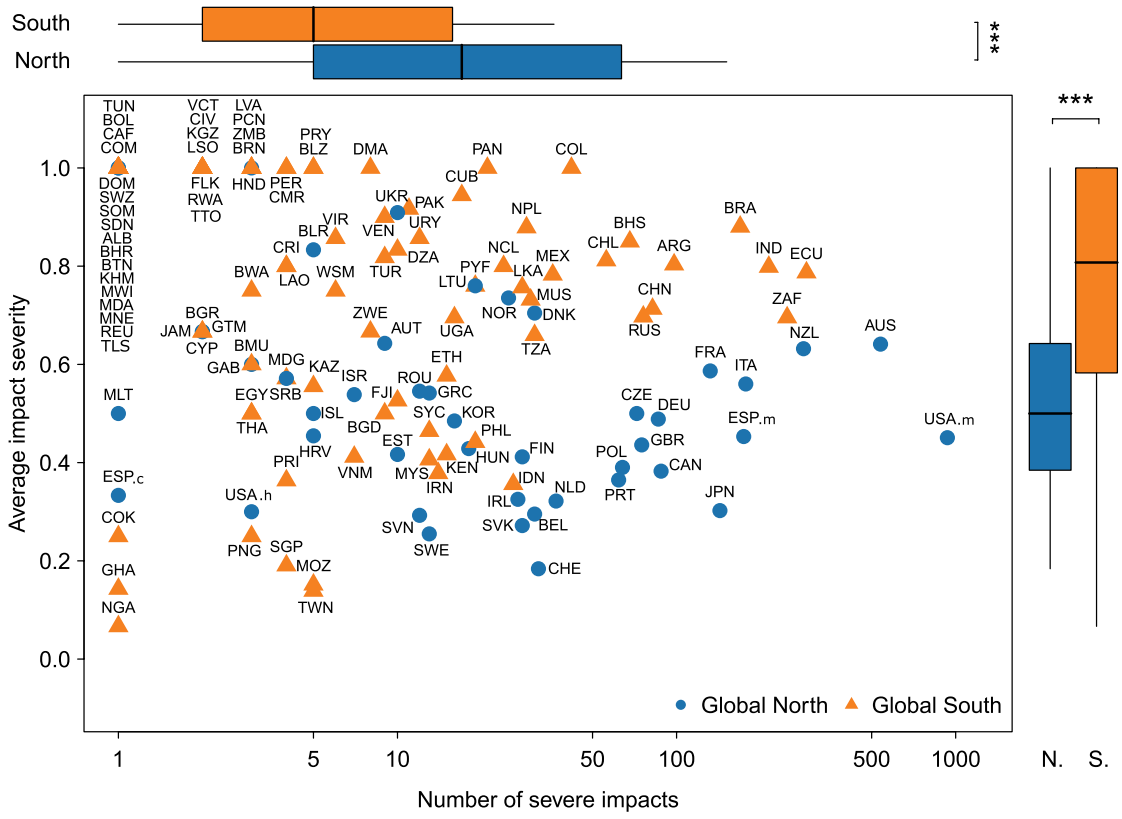


Fig. 2. Relationship between the number of severe impacts and the proportion of severe impacts (average impact severity) per region. Abbreviations follow International Organization for Standardization (ISO) standard 3166-1. Affiliations to Global South or Global North are indicated by color and shape. Symbols may overlay. For subnational units, we extended the ISO codes for distinct identification (ESP.c: Canary Islands; ESP.m: mainland Spain; USA.h: Hawaiian Islands; USA.m: mainland USA) (see table S1 for a full list). Box plots indicate variation of data using interquartile ranges (IQR) (box) and 1.5 times IQR (whiskers). Asterisks show significant differences between data subsets using *t* tests (***$P < 0.005$).

previous study on global impacts of introduced large mammalian herbivores (*27*), we did not find any temporal pattern in the probability to report severe impacts within regions and continents (generalized linear mixed model: $z_{Year} = 0.31$, $p_{Year} = 0.76$; fig. S3).

Emerging economies experience the most severe impacts

Distributing countries according to their absolute and relative documented impacts shows that countries from the Global South report significantly lower numbers of severe impacts (*t* test: $t = 3.3$, df = 80, $P < 0.005$) but higher average impact severity ($t = 5.2$, df = 109, $P < 0.001$) than many countries from the Global North (Fig. 2 and fig. S4). Many large, high-income developing countries according to the World Bank (*28*)—such as Brazil, India, South Africa, Argentina, Russia, China, Chile, and Mexico (upper right corner of Fig. 2 and fig. S4)—report similarly high absolute numbers of documented impacts but distinctly higher average impact severity than large economies from the Global North (lower right corner of Fig. 2 and fig. S4). Regions in the left half of Fig. 2 have very low numbers of impact reports, often below five, which prevents making robust inferences about their impact severity.

Drivers of impact severity

We tested the influence of economic variables [economic growth, gross domestic product (GDP), and per capita GDP], geographic variables (island or mainland, area, annual precipitation, and annual temperature), proxies of disturbance (human footprint index and human population size), and governance indicators [World Governance Indicators (WGI), proactive and reactive capacity for IAS mitigation] on documented impact

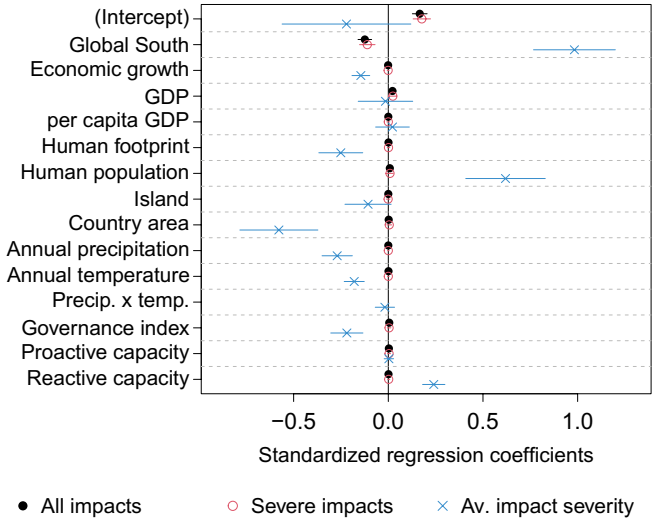


Fig. 3. Regression coefficients and their standard error intervals for the model predicting impact numbers (black and red symbols) and their average severity (blue crosses). Coefficients with intervals overlapping zero are considered as not significant. To visually compare the coefficients of linear mixed models (impact numbers) and generalized linear mixed models (impact severity) on the same scale, we standardized the response variables for all models between 0 and 1 by dividing all numbers by the maximum. This affects only the size of the coefficients but not the analysis results (table S1).

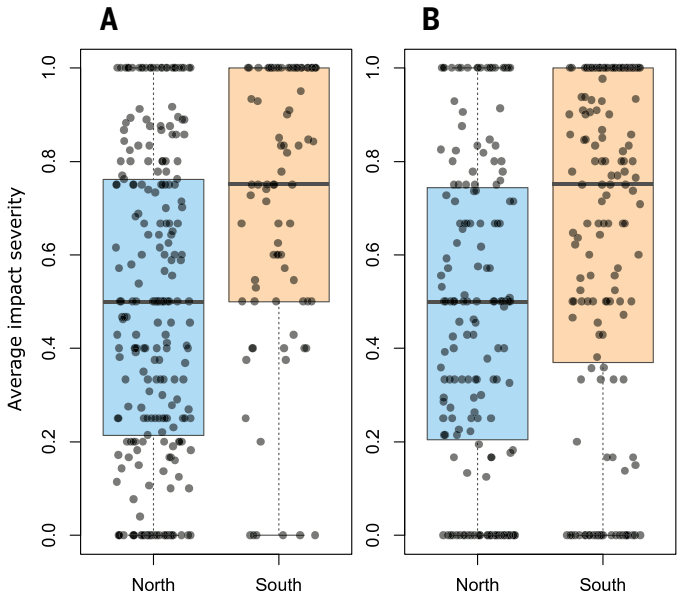


Fig. 4. Average impact severity per species. (**A**) Average impact severity for IAS having at least four recorded impacts exclusively in the Global North (blue, $n = 233$) or in the Global South (orange, $n = 85$) and (**B**) for the 164 IAS having at least four recorded impacts both in the Global North (blue) and Global South (orange). Box plots indicate variation of data using IQR (box) and 1.5 times IQR (whiskers).

numbers and average impact severity using (generalized) linear mixed models. Regression analyses and subsequent model selection and averaging confirmed that countries from the Global South report significantly fewer absolute numbers of impacts and fewer severe impacts than countries from the Global North (Fig. 3; table S2, A and B; and fig. S5). Inclusion in the Global South was also the variable with by far the highest explanatory power predicting the number of documented impacts (Fig. 3 and table S2). Yet the average impact severity was higher in countries of the Global South and increased with decreasing governance (Fig. 3, table S2C, and fig. S5).

Average impact severity also varied with environmental factors—such as the level of anthropogenic disturbance and the climatic conditions—and geographic variables. The more pristine the environment (judged by the human footprint index) and the higher the human population, the higher the impact severity. Smaller countries tended to have higher values of impact severity. Against our expectation (*27*), we did not find indications that impact severity was higher in island regions. However, we expected that the "island effect" would be more pronounced on small islands, which were excluded from our analysis because of missing data on governance and socioeconomic factors. Contrary to our expectation, economic growth was negatively related to impact severity (Fig. 4). However, economic growth is significantly higher in countries of the Global South compared with countries of the Global North (*t* test: $t = 4.0$, df = 154, $P < 0.001$) and, thus, the variable "Global South" likely accounts for this difference already. All the best-fitting models used for model averaging contained the factor "Global South," and thus the effect of other factors should be interpreted after accounting for the effect of "Global South." Another probably unexpected result is that a country's capacity to react to biological invasions ("reactive capacity") is positively related to average impact severity. This may also be influenced by the variable "Global South" because the reactive capacity is lower in countries of the Global South.

The regression analyses were repeated on subsets of the dataset for marine, terrestrial, and freshwater impacts, and for plants, vertebrates, invertebrates, and microorganisms (tables S3 to S9). The observed geographic pattern of higher impact severity in the Global South was consistent across taxonomic groups (plants, vertebrates, invertebrates, and microorganisms) and realms (terrestrial, freshwater, and marine) (figs. S6 to S16). In line with our general findings, numbers of documented total (tables S3A to S9A) and severe impacts (tables S3B to S9B) were lower—and average impact severity was higher—in the Global South for all realms and taxa, although not always significantly (tables S3C to S9C). Except for the freshwater realm (table S4C), the governance index was negatively associated with average impact severity across taxa and realms, indicating that good governance can mitigate severe impacts. The importance of other confounding factors varied among realms and taxa. Whereas economic factors appear to determine average documented impact severity in the terrestrial realm (table S3C), geographic factors seem more important in the freshwater realm (table S4C). Proactive management capacity of a country was related to impact severity in the marine realm (table S5C), where prevention is regarded as the only effective management option; whereas reactive measures such as control or eradication generally failed in marine environments (*2*). No factor explained variation in documented average impact severity for alien microorganisms (table S9C), likely owing to the low sample of countries with documented impact records. In general, splitting the data into realms or taxa reduces the sample sizes and, hence, the statistical power of the regression models. Results should therefore be interpreted with care. Nevertheless, the consistency of results demonstrates the robustness of our main conclusions and shows that the observed difference between the Global North and South is not restricted to certain taxa or realms.

To explore the role of individual species on the distribution of severe impacts, we calculated the global average impact severity for 482 (of 1284) IAS that caused at least four documented negative environmental impacts either in the Global South or in the Global North, or in both regions. The restriction allowed us to calculate the average impact severity per species with greater precision and avoid including 460 species for which only one impact record is documented. We found that the 85 species causing at least four impacts only in the Global South (ranging from four up to 39 reports of individual impacts per species) have higher average impact severity than the 233 IAS with at least four impacts (ranging from four to 190 impact records per species) only in the Global North (Wilcoxon rank-sum test: $W = 97{,}582$, $P < 0.0001$; Fig. 4A). This might indicate that IAS with higher impacts were more often introduced to the Global South; but the same pattern would also be observed if conditions in the Global South would facilitate more severe impacts. We also found that for 164 IAS causing impacts in both regions (ranging from four to 178 impact records per species), the average impact severity of the same species was 50% higher in the Global South (Wilcoxon signed-rank test for paired data: $V = 2939.5$, $P < 0.0001$; Fig. 4B), indicating that the local conditions in the Global South (including governance) facilitate more severe impacts.

Discussion

Our findings overturn the widely held assumption that the most severe impacts of IAS are concentrated in the Global North (*1*, *11*, *12*). We found the highest average impact severity in regions of the Global South, where research effort is generally lower (*2*, *22*). There is clear variation in the average impact severity among countries of the Global South, with some emerging economies exhibiting both the highest average impact severity and high absolute numbers of impacts. These countries are characterized by strong and accelerating economic activity, which results in increasing numbers of newly introduced alien species (*20*). At the same time, governance structures are often weakly developed, resulting in low capacities of managing biological invasions (*21*). The combination of high introduction rates through economic activity, low management capacities, and weak governance, typical for

emerging economies (*19*), results in comparatively high numbers of documented severe impacts of IAS. Emerging economies have also been identified as countries with the highest expected rise in alien species numbers in the near future as a result of strong economic development (*20*), likely resulting in increasing numbers of absolute impacts, many of which will probably be severe.

By contrast, wealthy countries of the Global North may have high introduction rates of alien species but also lower corruption and stronger governance as compared with the Global South, resulting in higher management capacities and better policy implementation (*19*, *21*) and thus in more effective management of IAS impacts. We show that, through good governance, the management of biological invasions is indeed able to effectively reduce the impacts of IAS for most taxa and realms irrespective of the region. This finding is supported by a recent study showing that IAS policies within the European Union and UK indeed had notable protective effects (*29*). However, current efforts and policies are insufficient to keep the rising tide of IAS and their impacts at bay (*2*).

Our conclusion that poor governance and the lack of appropriate management drive average impact severity is difficult to test quantitatively owing to the lack of standardized data on governance and management of IAS across countries. Information about the impacts of new legal regulations and the long-term effects of management efforts is usually not gathered, which hinders further analyses. For Europe, studies indicate that targeted regulations indeed slow introduction rates of IAS (*29*, *30*). However, whether they also reduce impact severity remains untested.

Mitigation of biological invasions is most efficiently achieved by prevention, and if not successful, by coordinated management thereafter (*2*). Following the recommendations on integrated governance by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) might provide the incentive to governments and institutions to strengthen prevention, international collaboration, and efficiency of responsible structures to reduce the impacts of biological invasions.

REFERENCES AND NOTES

1. IPBES, *Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES Secretariat, 2019); https://zenodo.org/doi/10.5281/zenodo.3831673.
2. IPBES, *Thematic Assessment Report on Invasive Alien Species and Their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES Secretariat, 2023); https://zenodo.org/doi/10.5281/zenodo.7430682.
3. H. Seebens *et al.*, *Nat. Commun.* **8**, 14435 (2017).
4. H. Seebens *et al.*, in *Thematic Assessment Report on Invasive Alien Species and Their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, H. E. Roy, A. Pauchard, P. Stoett, T. Renard Truong, Eds. (IPBES Secretariat, 2023), p. 190.
5. S. Bacher *et al.*, in *Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, H. E. Roy, A. Pauchard, P. Stoett, T. Renard Truong, Eds. (IPBES Secretariat, 2023); https://doi.org/10.5281/zenodo.7430731.
6. C. Diagne *et al.*, *Nature* **592**, 571–576 (2021).
7. W. Dawson *et al.*, *Nat. Ecol. Evol.* **1**, 0186 (2017).
8. M. van Kleunen *et al.*, *Nature* **525**, 100–103 (2015).
9. E. E. Dyer *et al.*, *PLOS Biol.* **15**, e2000942 (2017).
10. S. Pagad *et al.*, *Sci. Data* **9**, 391 (2022).
11. B. S. Halpern *et al.*, *Science* **319**, 948–952 (2008).
12. D. E. Bowler *et al.*, *People Nat.* **2**, 380–394 (2020).
13. P. Pyšek *et al.*, *Glob. Ecol. Biogeogr.* **28**, 779–783 (2019).
14. H. Seebens *et al.*, *Biol. Rev. Camb. Philos. Soc.* **100**, 2542–2583 (2025).
15. S. Bacher *et al.*, *Sci. Data* **12**, 832 (2025).
16. United Nations Conference on Trade and Development (UNCTAD), Data Hub (2026); https://unctadstat.unctad.org/EN/Index.html.
17. A. M. Kowalski, in *No Poverty: Encyclopedia of the UN Sustainable Development Goals*, W. Leal Filho, A. Azul, L. Brandli, A. Lange Salvia, P. Özuyar, T. Wall, Eds. (Springer, 2020), pp. 1–12.
18. R. E. Hoskisson, L. Eden, C. M. Lau, M. Wright, *Acad. Manage. J.* **43**, 249–267 (2000).
19. B. Warf, *GeoJournal* **81**, 657–669 (2016).
20. H. Seebens *et al.*, *Glob. Change Biol.* **21**, 4128–4140 (2015).
21. R. Early *et al.*, *Nat. Commun.* **7**, 12485 (2016).
22. L. J. Martin, B. Blossey, E. Ellis, *Front. Ecol. Environ.* **10**, 195–201 (2012). .
23. IUCN Species Survival Commission (SSC), *IUCN EICAT Categories and Criteria: The Environmental Impact Classification for Alien Taxa (EICAT)* (IUCN, ed. 1, 2020); https://portals.iucn.org/library/node/49101.
24. H. Seebens, SinAS: A global dataset of native and alien distributions of alien species, version 2.5, Zenodo (2023); https://doi.org/10.5281/zenodo.10038256.
25. T. M. Blackburn *et al.*, *PLOS Biol.* **12**, e1001850 (2014).
26. L. Forgione, S. Bacher, G. Vimercati, *Divers. Distrib.* **28**, 1832–1849 (2022).
27. Z. Bescond-Michel, S. Bacher, G. Vimercati, *Nat. Commun.* **16**, 8260 (2025).
28. World Bank, "How does the World Bank classify countries?" (2025); https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries.
29. Q. Canelles *et al.*, *One Earth* **8**, 101355 (2025).
30. C. Garcia-Lozano *et al.*, *Glob. Change Biol.* **31**, e70028 (2025).
31. S. Bacher, P. Pyšek, H. Seebens, Reproducible workflows for data analysis and visualisation, *Dryad* (2026); https://doi.org/10.5061/dryad.70rxwdcct.

ACKNOWLEDGMENTS

We are grateful for comments by H. Kreft, P. Hulme, J. Wilson, and an anonymous reviewer on earlier versions of the manuscript. **Funding:** This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant 521529463; Czech Science Foundation grant 26-21350S; Czech Academy of Sciences long-term research development project RVO 67985939; and Swiss National Science Foundation grants IC00I0-231475 and 31003A_179491. **Author contributions:** Conceptualization: S.B., H.S., P.P.; Investigation: S.B., H.S., P.P.; Methodology: S.B., H.S.; Writing – original draft: S.B., H.S.; Writing – review & editing: S.B., H.S., P.P. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** The GIDIAS dataset of IAS impacts (*15*) is available online at Figshare and the SinAS dataset on alien species checklists is accessible online at Zenodo (*24*). The R scripts to process the datasets and to produce the figures, tables, and analyses are provided in an open repository at Dryad (*31*). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aed0603
Materials and Methods; Figs. S1 to S16; Tables S1 to S9; References (*32*–*50*)

Submitted 14 October 2025; accepted 4 June 2026

10.1126/science.aed0603

BIRDSONG

The global biogeography of passerine songs

Quentin Bacquelé[1,2]*, Jean-Yves Barnagaud[2]†, Cyrille Violle[2], Frédéric Theunissen[3], Nicolas Mathevon[1,4,5]†

Although birdsongs are classic models for understanding the evolution of vocal communication, their global diversity has long made the development of a unifying framework challenging. By analyzing the acoustic architecture of songs from more than 3000 passerine species worldwide, we show that this acoustic space can be structured around eight elemental motifs. The differential use of these motifs is driven by a combination of species' biological traits (social organization, morphology, and mating system) and the physics of sound propagation. In tropical rainforests, environmental filtering for transmission efficiency favors structurally simple motifs, such as flat whistles. Conversely, in temperate regions, where high population densities facilitate close-range communication and short breeding seasons intensify sexual selection, the balance shifts toward complex, information-rich motifs, such as ultrafast trills, despite their susceptibility to acoustic degradation. Ultimately, the global geography of birdsong reflects a spatially varying equilibrium between physical environmental constraints and the biological drive for complex communication.

Acoustic communication is a cornerstone of animal behavior, having evolved into a spectacular diversity of sound signals that mediate survival and reproduction (*1*–*3*). This diversity is iconic in birds, whose songs are shaped by a complex interplay of evolutionary trade-offs and mediate social interactions among individuals (*4*–*7*). Songs must travel effectively through the environment (*8*–*10*), encode specific information for mates or rivals (*2*–*4*, *11*), and remain within the singer's physical capabilities (*12*, *13*), all within the constraints imposed by evolutionary history (*14*). Although the conjunction of these factors has sculpted a global mosaic of bird acoustics (*15*, *16*), a unifying framework explaining its complexity and worldwide distribution is missing. Specifically, it remains unknown how much the global bird soundscape is a historical consequence of phylogenetic radiation relative to shaping by the physics of the biosphere. In this work, we explore song diversity in passerine birds, testing whether their songs converge on acoustic motifs aligned with environmental gradients.

With nearly 6500 species, passerines represent more than 60% of all birds and have radiated over the past 50 million years, colonizing nearly all terrestrial biomes (*17*–*19*). Their global pervasiveness and acoustic richness make them an ideal system for uncovering biogeographic patterns of communication. However, previous large-scale analyses have been hampered by an oversimplified quantification of acoustic complexity, often reducing songs to a few metrics, such as average pitch and frequency range (*15*, *16*, *20*, *21*). These approaches do not fully capture the rich informational content embedded in a song's intricate spectro-temporal architecture—the very patterns of frequency and amplitude modulations that carry its meaning (*22*, *23*).

[1]ENES Bioacoustics Research Lab, CRNL, University of Saint-Etienne, CNRS, Inserm, Saint-Etienne, France. [2]CEFE, University of Montpellier, CNRS, EPHE-PSL University, IRD, Montpellier, France. [3]Neuroscience Department, Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [4]École Pratique des Hautes Études - PSL, CHArt Lab, University Paris-Sciences-Lettres, Paris, France. [5]Institut Universitaire de France, Paris, France. ***Corresponding author. Email: qbacquele@gmail.com** †These authors contributed equally to this work.

In this work, we explicitly treat birdsongs as vectors of information to map the global diversity of acoustic communication and uncover its environmental drivers.

We use a holistic framework to understand the composition and worldwide structuring of birdsongs. Our study thus integrates the quantification of the relative influence of phylogeny, morphological traits, sexual selection, and social traits on song structure. This is combined with mapping of global song diversity, analysis of associations with environmental gradients (vegetation density, climate, topography, and human footprint), and modeling of song transmission through the environment. We show that the seemingly tremendous acoustic diversity of birdsongs can be partitioned into a limited set of elementary "acoustic motifs" that birds combine to construct their songs. We show that the geographic distribution of these motifs follows environmental gradients that structure the composition of regional passerine species assemblages, thereby connecting songbird biogeography and acoustic communication. By providing the geographic map of passerine songs, this work establishes an explicit link between the evolutionary mechanisms shaping communication and the macroecological patterns of global bioacoustic diversity.

Eight acoustic motifs structure the global passerine song diversity

We analyzed the complete spectro-temporal architecture of 116,792 song bouts from 3160 passerine species, encompassing more than half of all extant species and more than 80% of passerine families (fig. S1; see materials and methods for details on automatic song segmentation). We characterized each song bout by its modulation power spectrum, a sound representation that quantifies the modulations of sound energy in time (e.g., trill rates), frequency (e.g., tonality), and joint time-frequency (e.g., sweep curvature) (Fig. 1A and materials and methods) (*24*). A modulation power spectrum captures a sound's spectro-temporal texture in a manner that is both biologically salient, because these modulations are key to how birds encode information, and analytically robust for comparing song patterns across thousands of species (materials and methods and figs. S2 and S3).

We explored the variability of these modulation power spectra in a 37-dimensional acoustic space where proximity between song bouts reflects their spectro-temporal similarity (Fig. 1B and figs. S4 and S5). Using unsupervised classification, we partitioned this high-dimensional volume into eight distinct acoustic motifs. Although the acoustic space of birdsongs is continuous, song bouts concentrate around eight poles, each captured by a motif (fig. S6). Motif assignment is unambiguous: 75% of the song bouts are classified as belonging exclusively to one of the eight motifs with low ambiguity (maximum posterior probability > 0.9; mean = 0.92, median = 0.99). These motifs form elementary building blocks for passerine songs (Fig. 1C, fig. S7, and materials and methods). They include three types of trills distinguished by their speed, three types of whistles based on their pitch modulation, and two more complex structures: atonal chaotic notes and harmonic stacks (Fig. 1C, fig. S8, and the interactive web application in the supplementary materials). On average, each species displays 4.6 ± 1.8 (SD) different motifs in its song repertoire; 65% of all analyzed songs are composites that combine two to eight motifs (mean = 2.23 ± 1.25 motifs). Notably, most species (65.3%) display their dominant motif in more than 80% of their songs (fig. S9). In the other species (34.7%), the dominant motif is used to a lesser extent, which allows for greater song variability. Given that passerines share a keen ability to resolve complex spectro-temporal modulations, these eight motifs are certainly perceptually distinguishable by all passerine species (*25*).

Acoustic motifs are shaped by life history beyond shared ancestry

We first tested whether the dominant motif of each species reflects its phylogenetic history. We built a circular phylogram reconstructing the

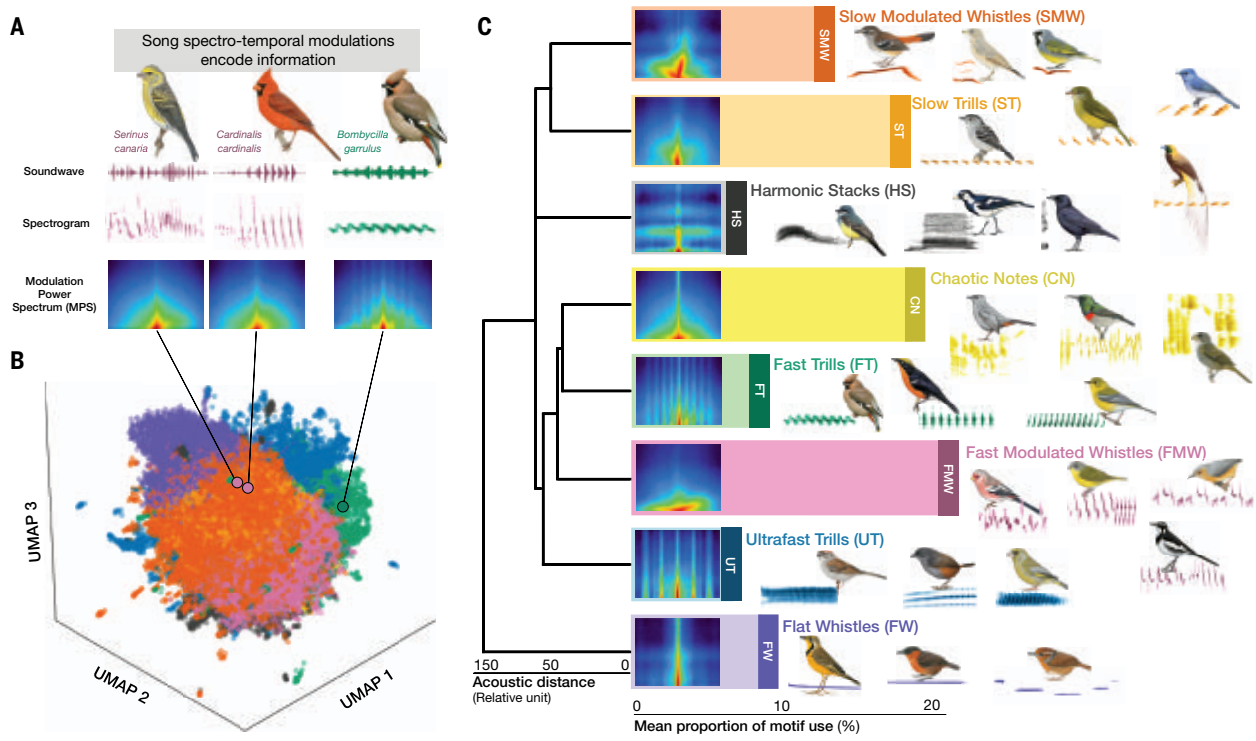


Fig. 1. Eight distinct acoustic motifs structure the global passerine song repertoire. (**A**) Song feature extraction: soundwave, spectrogram, and modulation power spectrum (MPS) (see materials and methods). The MPS quantifies spectro-temporal modulations encoding information such as species identity, individual identity, and singer's quality. Two of the three species shown as examples exhibit songs with similar spectro-temporal patterns (colored in purple) that contrast with the pattern found in the vocalization of the third species (colored in green). (**B**) Global acoustic space derived from 116,792 passerine vocalizations (3160 bird species). Visualization uses a three-dimensional uniform manifold approximation and projection (UMAP) of the 37-dimensional MPS space (materials and methods). Colors represent acoustic motifs. (**C**) The eight acoustic motifs. Dendrogram shows acoustic similarities (Euclidean distances between MPS profiles). Bar plots indicate the proportion of each motif across all passerine songs. Representative MPSs (song bout with probability = 1 for that motif) and spectrograms exemplify each acoustic motif. See also fig. S8 and the interactive web application in the supplementary materials, which visualize and play the global vocal repertoire of passerines. [Bird illustrations: © Lynx Nature Books and courtesy of the Cornell Lab of Ornithology (supplementary materials, list of artists)]

ancestral states of these motifs (Fig. 2A). Our analysis revealed a weak phylogenetic signal (Pagel's $\lambda = 0.17$, Blomberg's $K = 0.08$), consistent with closely related species frequently using different dominant motifs. However, phylogenetic conservatism varied substantially across specific motifs: The simplest tonal elements (flat whistles and slow trills) exhibit the strongest phylogenetic signal ($\lambda = 0.88$ and 0.84, respectively), whereas fast trills show the weakest ($\lambda = 0.25$), with intermediate values for other motifs. Consistent with prior knowledge (*4*, *26*–*29*), the phylogenetic signal is weaker in oscine passerines, whose songs are culturally transmitted (*3*–*5*), compared with suboscine passerines, where vocal production is genetically driven (*3*–*5*) (dominant acoustic motif split by oscines versus nonoscines: $\lambda_{oscines} = 0.09$, $\lambda_{nonoscines} = 0.25$, Wilcoxon test $P < 0.001$).

Phylogeny accounted for 82% of the explained variation in motif use, leaving 18% for biological traits (multivariate variance partitioning; see materials and methods). Within this trait-based component, social organization accounted for 41% of the variance followed by morphology (34%) and mating system (25%). Given this trait signal, we used a Dirichlet regression to quantify each trait's effect on motif composition (materials and methods and fig. S10).

Social organization, reflecting the propensity of individuals to form long-term associations and produce long-range vocal signals (*30*), drives a marked compositional shift in song motifs (Fig. 2B, top panel, and fig. S11). Species engaging in long-range communication produce a higher proportion of flat whistles and slow trills, reallocating ~5% [4.9%, 95% credible interval (CI): 3.4 to 6.6%] of their repertoire to these motifs compared with species communicating at shorter ranges. By contrast, species signaling at close range favor faster and more complex motifs, such as harmonic stacks and ultrafast trills.

Biophysical constraints imposed by body size and beak shape drive a comparable shift (4.8%, 95% CI: 3.3 to 6.5%) (Fig. 2B, middle panel, and fig. S12). Consistent with prior research in specific avian groups (*31*, *32*), simpler sounds (flat whistles and slow trills) are more represented in larger birds. Conversely, fast modulated whistles, a challenging motif to produce, are predominantly found in smaller species, which may take advantage of their neuromuscular agility to generate complex signals (*13*).

Finally, the mating system drives a smaller but consistent shift in motif composition (3.6%, 95% CI: 2.4 to 4.9%) (Fig. 2B, bottom, and fig. S13). Complex sounds, such as fast modulated whistles, occur more often in polygynous species, a result consistent with the evolution of physiologically costly signals as honest indicators of emitter quality when sexual selection is intense (*2*, *6*, *33*). Conversely, simpler and less demanding motifs, such as flat whistles and slow trills, are more common in monogamous species subject to weaker sexual selection pressures.

The global map of birdsongs echoes biogeographic history and biome patterns

We investigated whether the diversity of acoustic motifs used by birds living in the same area varies across the planet by quantifying this diversity within geographical assemblages of bird species found in

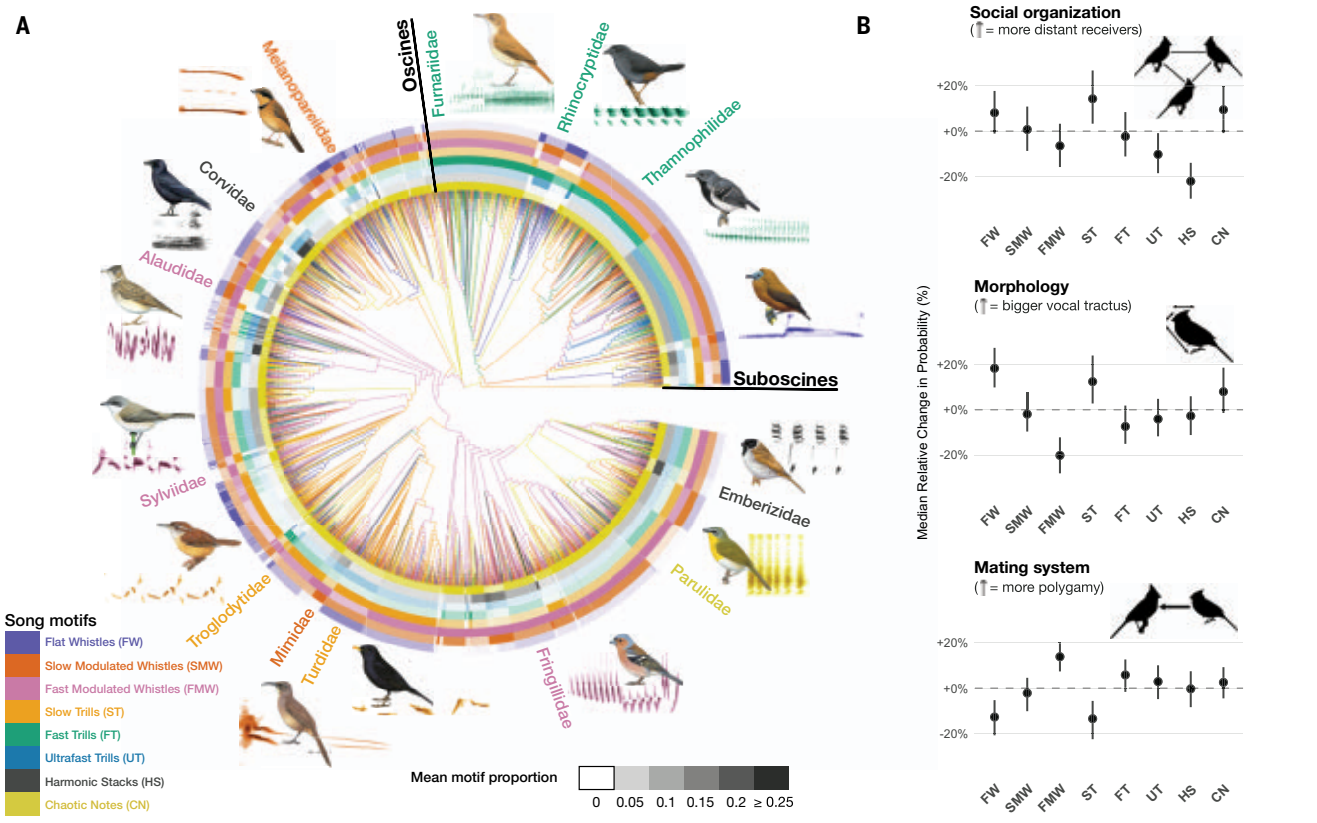


Fig. 2. Phylogenetic distribution and trait correlates of acoustic motifs in passerine birds. (**A**) Passerine phylogeny showing dominant acoustic motifs (ancestral states reconstructed) along branches. Outer heatmap indicates family-aggregated mean of species' posterior probabilities for a given motif. Representative families, species, and spectrograms are shown. [Bird illustrations: © Lynx Nature Books and courtesy of the Cornell Lab of Ornithology (supplementary materials, list of artists)] (**B**) Species traits predict the motifs composing birdsongs. Dots show the relative change in each motif's predicted share of the repertoire when comparing species at the 10th versus 90th percentile of each trait. For example, +10% means that a motif becomes 10% more prevalent in the repertoire of high-trait versus low-trait species. FW, flat whistles; SMW, slow modulated whistles; FMW, fast modulated whistles; ST, slow trills; FT, fast trills; UT, ultrafast trills; HS, harmonic stacks; CN, chaotic notes. Error bars show 95% CIs (Dirichlet model with trait gradients as predictors; see materials and methods for details).

$1° \times 1°$ grid cells (Fig. 3A). We found that co-occurring species exhibit more similar acoustic motifs than expected by chance given local species richness [mean functional dispersion measured as a standardized effect size (SES) = -2.1 ± 1.7; Fig. 3A and fig. S14]. This acoustic "niche packing" (*34*) implies that underlying ecological or evolutionary processes drive acoustic convergence within assemblages. Notably, we found no tendency for birds to use a greater number of distinct song motifs than expected by chance in species-rich areas where competition for making oneself heard in acoustic space may be strongest. This lack of trait divergence between co-occurring species suggests that interspecies acoustic competition is insufficient to shape motif usage at broad scales, though it may still occur at finer spatial or temporal resolutions (*11*).

To further investigate how the eight song motifs are geographically structured, we mapped motif usage in every grid cell. More specifically, we counted the number of bird species displaying a given motif in more than 5% of their songs (fig. S15). This observed species count was compared with a null expectation model that randomly drew, for a grid cell, the same number of species from the pool of species occurring in the same biogeographic realm (fig. S16A). Observed counts often departed from the null expectation in geographical regions corresponding to Earth's major biomes (Fig. 3, B to I, and fig. S16B). Notably, song motifs characterized by constant or slowly varying sounds, such as flat whistles and slow trills, are disproportionately common in tropical rainforests. These overrepresented song motifs reach their highest relative prevalence in the Amazonian, Congo Basin, and Southeast Asian rainforests. Biome identity alone explains 47% of intercell variance in flat whistle prevalence and 43% in slow trill prevalence (table S1). Conversely, fast modulated whistles are underrepresented in tropical rainforests. Instead, they occur broadly across the Palearctic, largely independent of biome boundaries within this realm. The global map of the most overrepresented motif for each location (fig. S16C) confirms these observations. Only slow modulated whistles and harmonic stacks are largely independent of biome identity [coefficient of determination (R^2) = 0.18 and 0.19, respectively; table S1]. These patterns reveal that some motifs track broad habitat types across continents, which suggests that ecological constraints may shape bird acoustic signals at a planetary scale.

Global variation in climate and vegetation shapes the efficiency of song transmission

One of the most important ecological constraints that may explain the geographical distribution of acoustic motifs could be the effect of the environment on sound information degradation. The physical environment degrades sound waves during propagation, imposing strong selection on acoustic signals to maintain information transfer (*8*, *9*). However, propagation-induced constraints depend strongly on the properties of the transmission channel—e.g., presence and density of vegetation (*10*). We thus tested whether these environmental pressures contribute to shaping the global arrangement of acoustic motifs. We

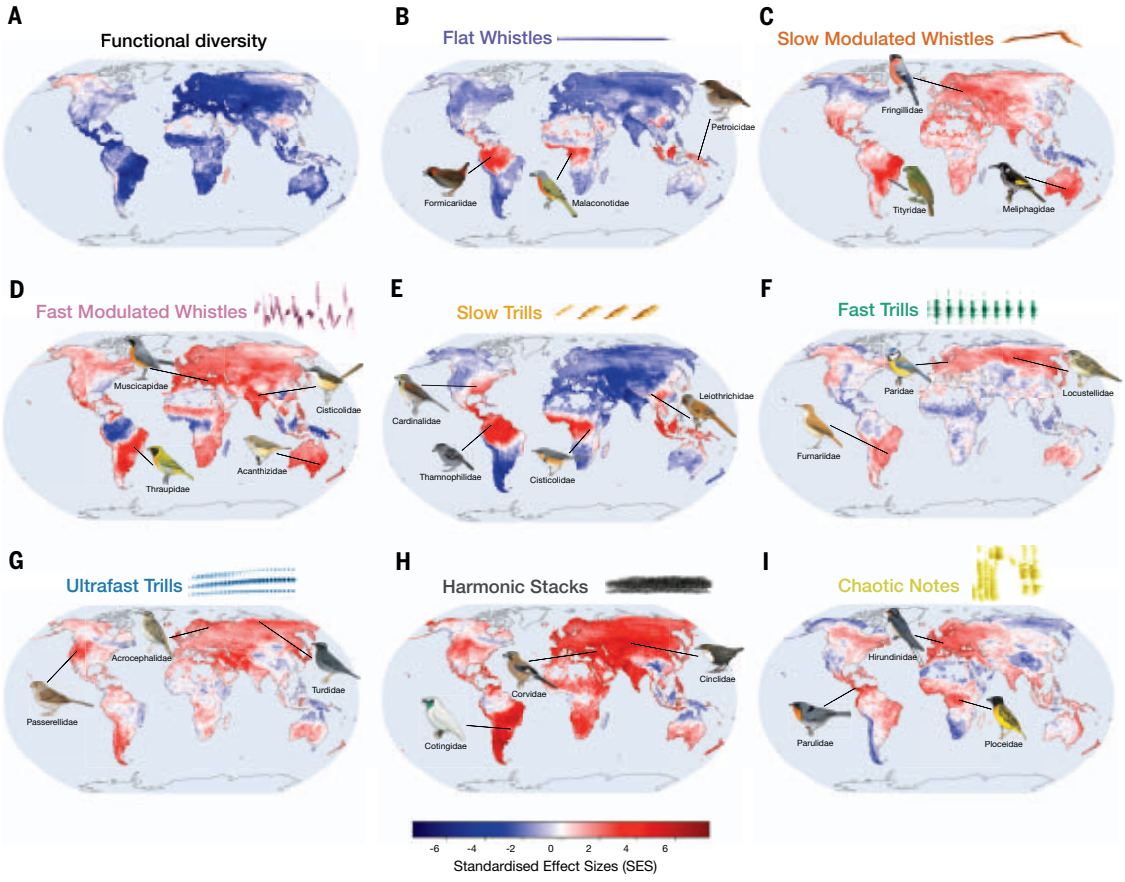


Fig. 3. The geographic structure of songbird acoustic motifs is partly consistent with terrestrial biomes. (**A**) Diversity of acoustic motifs within geographical assemblages of bird species [1° × 1° grid cell; diversity quantified by SESs of functional dispersion (FDis)]. Dark blue indicates less motif diversity in acoustic assemblage than expected by chance given local species richness, and red indicates overdispersed assemblages. (**B** to **I**) Deviation maps of motif-specific species richness per 1° × 1° grid cell. Values are SESs that quantify the deviation in motif species richness from a null expectation based on random sampling. A positive SES indicates an overrepresented motif given local species richness, whereas a negative SES indicates an underrepresented motif given local species richness. Representative bird species are shown in areas where a given motif largely exceeds its expected species richness (top 5% SES). [Bird illustrations: © Lynx Nature Books and courtesy of the Cornell Lab of Ornithology (supplementary materials, list of artists)]

developed a physics-based transmission efficiency index (TEI) to assess the ability of acoustic motifs to convey species identity information along environmental gradients affecting sound propagation (materials and methods). This index quantifies the persistence of a signal's informational content during propagation by modeling information degradation caused by key physical processes in a forest environment (fig. S17). These processes include attenuation (overall sound intensity loss), reverberation (echoes filling the internote silences and blurring the sound structure), frequency-dependent filtering (greater loss of signal energy for higher-pitched sounds), and ambient noise (background sounds that mask the signal) (figs. S17 and S18) (*3*, *4*, *10*). As expected, given their simpler acoustic structure (*4*), flat whistles and slow trills exhibit the highest TEI values, which highlights their resilience to propagation-induced information degradation (Fig. 4A and fig. S19). By contrast, spectro-temporally complex signals, such as harmonic stacks and chaotic notes, are least effective in transferring information over long ranges. To refine our understanding of environmental constraints on sound propagation, we further simulated propagation in an open habitat, where reverberation due to vegetation is absent but where atmospheric turbulence contributes to signal masking (*35*) (materials and methods). Under these open-habitat conditions, information transmission is more efficient, regardless of the song motif considered (mean TEI in open habitat = −0.36 versus −0.77 in closed habitat; TEI dispersion reduced by 41%; fig. S20). This result emphasizes that propagation constraints in densely vegetated environments constitute a differential selective pressure among birdsong motifs.

Mapping mean TEI values at a planetary scale reveals a distinct signature of rainforests (Fig. 4B). Grid cells across Amazonia, the Congo Basin, and Southeast Asia are significantly enriched in motifs associated with long-range information transmission and depleted in motifs characterized by a lower transmission index. To disentangle the environmental drivers of this geographic pattern, we correlated the TEI with standardized environmental predictors. The index shows strong positive correlations with vegetation structure [Pearson's correlation coefficient (r) = 0.50; $P < 0.001$], temperature ($r = 0.36$; $P < 0.001$), and relative humidity ($r = 0.31$; $P < 0.001$) (Fig. 4C). Overall, acoustic motifs most suitable for long-distance information transfer are thus used by more bird species than predicted by chance in densely vegetated environments with high humidity and temperature, mostly located in tropical rainforests. Using spatial multiple regressions including all environmental drivers alongside phylogenetic diversity, we found that humidity is the strongest positive driver of the TEI (Fig. 4D and table S2). A 1-SD increase in humidity raises the index by ∼+0.55 SD units (+0.28 accounting for phylogenetic diversity). Temperature also shows a positive effect (+0.10 without versus +0.38 with phylogeny). The effect of vegetation is smaller but still positive (+0.07 without

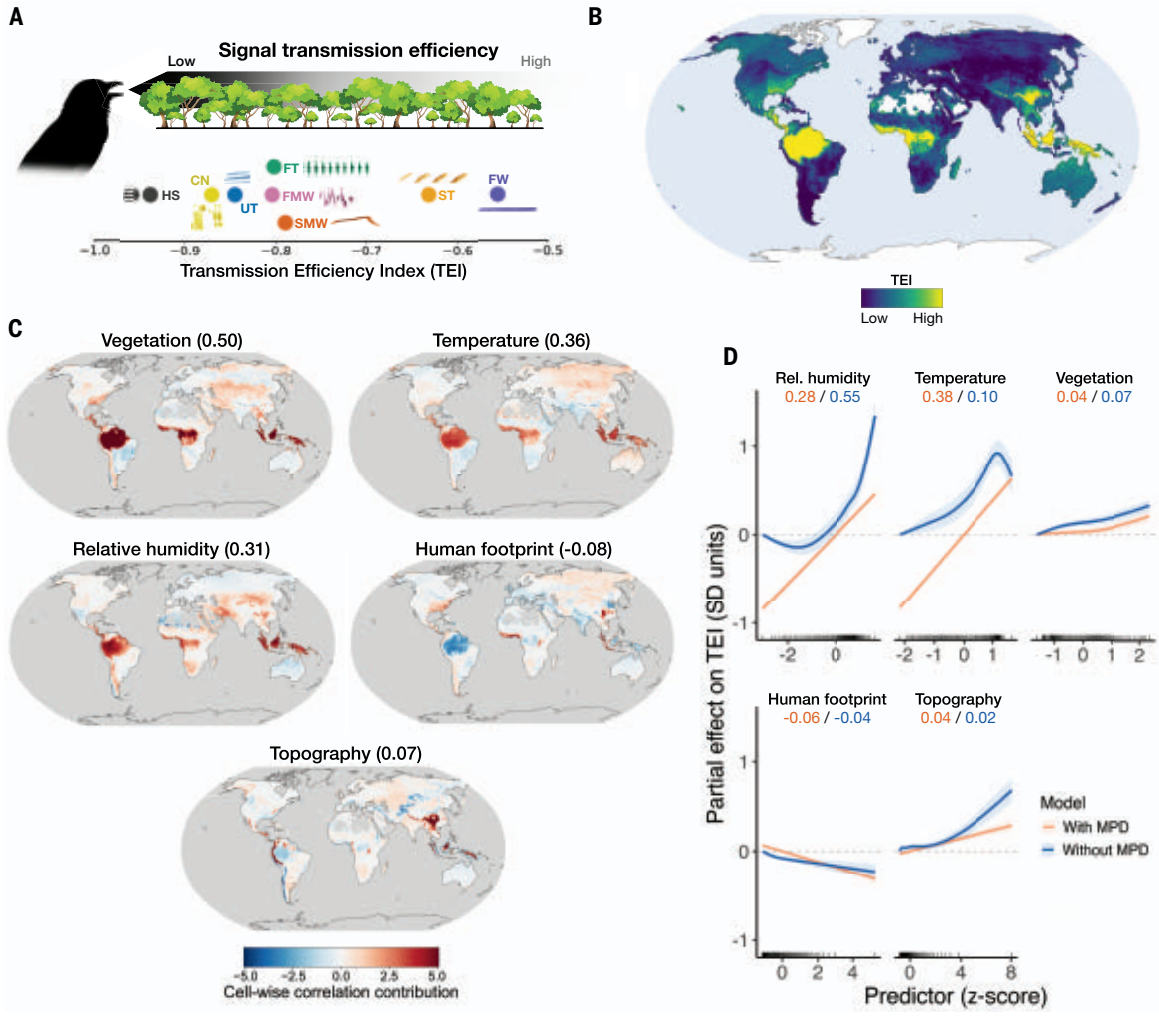


Fig. 4. Climate and vegetation modulate global acoustic transmission efficiency. (**A**) The TEI quantifies an acoustic motif's robustness to information loss during propagation through the environment. Acoustic motifs, such as flat whistles, preserve species identity over farther distances compared with harmonic stacks (see materials and methods). (**B**) Global distribution of TEI. Each 1° cell's value is the mean TEI in the species assemblage. White indicates no data. (**C**) Maps showing the covariation score *c* between TEI and each predictor *x* per 1° cell *i*: $c_i = z(\mathrm{TEI}_i)z(x_i)$. *c* values are positive where TEI and *x* are both above or below their means and negative where they move in opposite directions. The Pearson correlation *r* computed across all cells is shown in parentheses above each map. (**D**) The TEI depends on environmental characteristics. Each panel shows the effect of one *z*-scored predictor on TEI (in TEI SD units) with other predictors at their mean (*z* score = 0). Solid lines represent two models: blue excludes mean phylogenetic distance (MPD), and orange includes it. Shaded bands are 95% uncertainty intervals. Black tick marks along the bottom indicate observation locations. Values above each panel are standardized average marginal effects (average TEI change in SD units per 1-SD increase).

versus +0.04 with phylogeny). These results are consistent with a macroecological expression of the acoustic adaptation hypothesis, which proposes that the acoustic properties of animal vocalizations are adapted to habitat constraints facilitating information transfer between individuals (*8*, *36*–*38*). However, although our results support the acoustic adaptation hypothesis in tropical rainforests, where species preferentially use song motifs resilient to signal degradation, they reveal a contrasting regime in temperate latitudes. In these regions, more complex and degradation-prone motifs are favored. We propose that this shift is driven by shorter interindividual communication distances (*39*) and more intense sexual selection (*40*), which together outweigh the costs of propagation constraints.

Conversely, other variables have smaller effects, including topography (+0.04 with versus +0.02 without phylogeny) and phylogenetic diversity (+0.10). Notably, the human footprint [an indicator of the imprint of all human activities on land (*41*)] is associated with lower transmission efficiency (−0.06 with versus −0.04 without phylogeny), a finding consistent with previous studies reporting negative associations between humans and ecological functions at macroecological scales (*42*). This pattern is particularly prominent in regions with a long history of human influence, such as Europe and India (Fig. 4C) (*43*).

Conclusions

Our research unveils a simplicity underlying the spectacular diversity of birdsongs: Across thousands of lineages and biomes, passerine songs can be partitioned into a limited set of eight elementary acoustic motifs. We demonstrate that the global distribution of these motifs is structured by three main forces: biological traits that shape species-specific vocalizations; biogeographic history, which influences song diversity in local communities; and the physics of sound propagation, which favors specific motifs in specific environments. Furthermore, distinct motif usage across species with shared phylogeny indicates that the capacity to produce these extant acoustic motifs emerged early in the passerine radiation. Natural selection, potentially driven by the

physics of sound transmission, could have fostered rapid and convergent evolution.

Our framework provides a robust foundation for upscaling to a global level the evolutionary hypotheses on signal transmission that were previously formulated at local scales. These results align with functional biogeography's central tenet—that functional traits often converge on a few elementary combinations (*44*). The planetary-scale biogeographic structure of the birdsong fundamental acoustic motifs seems critically shaped by acoustic adaptation to long-distance information transfer (*8*, *36*–*38*, *45*). However, the adaptive benefits of propagation efficiency are counterbalanced by the demands of information encoding. In agreement with the predictions of the mathematical theory of communication (*46*), complex motifs, such as fast modulated whistles and fast trills, carry more information [e.g., species identity (fig. S21), group and individual identities (*47*–*50*), population dialects (*25*, *51*), and singer quality (*52*)], but they are highly susceptible to degradation. By contrast, simpler motifs, such as flat whistles and slow trills, convey less information (fig. S21) but are more resilient to propagation (*53*). This trade-off between encoding more information and transmitting it over long distances exhibits strong geographic variation. Climatic stability in tropical rainforests favors large territories and low encounter rates (*39*), requiring resilient, long-range signals. Contrastingly, the pronounced seasonality of temperate environments reduces both territory sizes and breeding periods, intensifying sexual selection (*20*, *40*). With individuals communicating at closer ranges and under strict receiver assessment, selection favors information-dense motifs that are inherently more prone to degradation during propagation. The biogeography of birdsong thus emerges as a spatially varying equilibrium between these two fundamental constraints.

By providing a conceptual framework akin to a "periodic table" of birdsong at a planetary scale (*54*), our work substantiates a general principle: The evolutionary trajectories of animal vocal signals are channeled along a few geographically structured pathways. Expanding this hypothesis beyond birdsongs, by testing it across other acoustic signals and taxonomic groups, will provide a solid basis for a theory of animal communication rooted in the interaction between information coding, rapid signal evolution, biogeography, and functional ecology (*55*–*57*).

REFERENCES AND NOTES

1. P. Marler, *Science* **157**, 769–774 (1967).
2. W. A. Searcy, S. Nowicki, *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems* (Princeton Univ. Press, 2005).
3. N. Mathevon, *The Voices of Nature: How and Why Animals Communicate* (Princeton Univ. Press, 2023).
4. P. Marler, H. Slabbekoorn, Eds., *Nature's Music: The Science of Birdsong* (Elsevier, 2004).
5. C. K. Catchpole, P. J. B. Slater, *Bird Song: Biological Themes and Variations* (Cambridge Univ. Press, ed. 2, 2010).
6. M. R. Wilkins, N. Seddon, R. J. Safran, *Trends Ecol. Evol.* **28**, 156–166 (2013).
7. J. Podos, M. S. Webster, *Curr. Biol.* **32**, R1100–R1104 (2022).
8. E. S. Morton, *Am. Nat.* **109**, 17–34 (1975).
9. R. H. Wiley, D. G. Richards, *Behav. Ecol. Sociobiol.* **3**, 69–94 (1978).
10. M. Wahlberg, O. N. Larsen, in *Comparative Bioacoustics: An Overview*, C. Brown, T. Reide, Eds. (Bentham Science Publishers, 2017), pp. 62–119.
11. M. Garcia *et al.*, *Nat. Commun.* **11**, 4970 (2020).
12. T. Riede, F. Goller, *Proc. Biol. Sci.* **281**, 20132306 (2014).
13. C. P. H. Elemans *et al.*, *Nat. Commun.* **6**, 8978 (2015).
14. A. N. G. Kirschel *et al.*, *Evolution* **65**, 3162–3174 (2011).
15. W. D. Pearse *et al.*, *Evolution* **72**, 944–960 (2018).
16. H. S. S. C. Sagar *et al.*, *Proc. Biol. Sci.* **291**, 20241908 (2024).
17. J. Fjeldså, L. Christidis, P. G. P. Ericson, *The Largest Avian Radiation: The Evolution of Perching Birds, or the Order Passeriformes* (Lynx Edicions, 2020).
18. C. J. Schmitt, S. V. Edwards, *Curr. Biol.* **32**, R1149–R1154 (2022).
19. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, *Nature* **491**, 444–448 (2012).
20. J. T. Weir, D. Wheatcroft, *Proc. Biol. Sci.* **278**, 1713–1720 (2011).
21. P. Mikula *et al.*, *Ecol. Lett.* **24**, 477–486 (2021).
22. S. M. N. Woolley, T. E. Fremouw, A. Hsu, F. E. Theunissen, *Nat. Neurosci.* **8**, 1371–1379 (2005).
23. F. E. Theunissen, J. E. Elie, *Nat. Rev. Neurosci.* **15**, 355–366 (2014).
24. N. C. Singh, F. E. Theunissen, *J. Acoust. Soc. Am.* **114**, 3394–3411 (2003).
25. O. N. Larsen *et al.*, in *Exploring Animal Behavior Through Sound*, vol. 2, C. Erbe, J. A. Thomas, Eds. (Springer, 2025), pp. 285–359.
26. P. Marler, M. Tamura, *Science* **146**, 1483–1486 (1964).
27. N. A. Mason *et al.*, *Evolution* **71**, 786–796 (2017).
28. J. Arato, W. T. Fitch, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200241 (2021).
29. M. Rivera, J. A. Edwards, M. E. Hauber, S. M. N. Woolley, *Sci. Rep.* **13**, 7076 (2023).
30. J. A. Tobias *et al.*, *Front. Ecol. Evol.* **4**, 74 (2016).
31. J. Podos, *Nature* **409**, 185–188 (2001).
32. E. P. Derryberry *et al.*, *Evolution* **66**, 2784–2797 (2012).
33. J. Podos, H.-C. Sung, in *The Neuroethology of Birdsong*, J. T. Sakata, S. C. Woolley, R. R. Fay, A. N. Popper, Eds., **vol. 71**, *Springer Handbook of Auditory Research* (Springer, 2020), pp. 245–268.
34. R. MacArthur, R. Levins, *Am. Nat.* **101**, 377–385 (1967).
35. A. Michelsen, O. N. Larsen, in *Neuroethology and Behavioral Physiology*, F. Huber, H. Markl, Eds. (Springer, 1983), pp. 321–331.
36. G. Boncoraglio, N. Saino, *Funct. Ecol.* **21**, 134–142 (2007).
37. E. Ey, J. Fischer, *Bioacoustics* **19**, 21–48 (2009).
38. B. Freitas, P. B. D'Amelio, B. Milá, C. Thébaud, T. Janicke, *Biol. Rev. Camb. Philos. Soc.* **100**, 815–833 (2025).
39. J.-M. Thiollay, *J. Trop. Ecol.* **10**, 449–481 (1994).
40. R. A. Barber *et al.*, *PLOS Biol.* **22**, e3002856 (2024).
41. H. Mu *et al.*, *Sci. Data* **9**, 176 (2022).
42. A. L. Šizling *et al.*, *Glob. Ecol. Biogeogr.* **25**, 1072–1084 (2016).
43. E. C. Ellis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7978–7985 (2013).
44. R. H. MacArthur, in *Challenging Biological Problems: Directions Toward Their Solution*, J. A. Behnke Ed. (Oxford Univ. Press, 1972), pp. 253–259.
45. H. Brumm, M. Naguib, in *Advances in the Study of Behavior*, **vol. 40**, M. Naguib, K. Zuberbuumlhler, N. S. Clayton, V. M. Janik, Eds. (Elsevier, 2009), pp. 1–33.
46. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379–423 (1948).
47. D. A. Nelson, *Anim. Behav.* **124**, 263–271 (2017).
48. C. Vignal, N. Mathevon, S. Mottin, *Nature* **430**, 448–451 (2004).
49. E. Briefer, T. Aubin, K. Lehongre, F. Rybak, *J. Exp. Biol.* **211**, 317–326 (2008).
50. N. Mathevon *et al.*, *PLOS ONE* **3**, e1580 (2008).
51. D. J. Mennill *et al.*, *Curr. Biol.* **28**, 3273–3278.e4 (2018).
52. D. Gil, M. Gahr, *Trends Ecol. Evol.* **17**, 133–141 (2002).
53. O. N. Larsen, in *Coding Strategies in Vertebrate Acoustic Communication*, T. Aubin, N. Mathevon, Eds., vol. 7 of *Animal Signals and Communication* (Springer, 2020), pp. 11–44.
54. E. R. Pianka, L. J. Vitt, N. Pelegrin, D. B. Fitzgerald, K. O. Winemiller, *Am. Nat.* **190**, 601–616 (2017).
55. J. Sueur, B. Krause, A. Farina, *Trends Ecol. Evol.* **34**, 971–973 (2019).
56. C. A. Morrison *et al.*, *Nat. Commun.* **12**, 6217 (2021).
57. J. H. Rasmussen, D. Stowell, E. F. Briefer, *Science* **385**, 138–140 (2024).

ACKNOWLEDGMENTS

We are deeply grateful to the XENO-CANTO citizen science database and all its contributors as well as to Lynx Nature Books, the Cornell Lab of Ornithology, and the artists for the bird drawings (see list of artists in the supplementary materials). We thank J. Elie, C. Girard-Buttoz, and M. Greenfield for their advice on a previous draft of the manuscript. N.M. thanks T. Aubin, who introduced him to bioacoustics more than 30 years ago by proposing that he work on the acoustic adaptation hypothesis. We also thank three anonymous referees who provided insightful comments. **Funding:** This research has been supported by the University of Saint-Etienne; École Pratique des Hautes Études - PSL; University Paris-Sciences-Lettres (PhD stipend to Q.B.); Labex CeLyA (N.M.); CNRS; Institut Universitaire de France (N.M.); and Acoucene project funded by the French Foundation for Research on Biodiversity (FRB) through its CESAB, the Ministry of the Ecological Transition (MTE), and the French Office of Biodiversity (OFB) (J.-Y.B.). **Author contributions:** Conceptualization: Q.B., J.-Y.B., N.M.; Data curation: Q.B.; Funding acquisition: J.-Y.B., N.M.; Investigation: Q.B., J.-Y.B., N.M.; Methodology: all authors; Project administration: J.-Y.B., N.M.; Supervision: J.-Y.B., N.M.; Visualization: Q.B.; Website: Q.B.; Writing – original draft: Q.B., J.-Y.B., N.M.; Writing – review & editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All acoustic data were acquired and processed using Python 3.11 (Python Software Foundation). Subsequent data extraction, feature engineering, and clustering analyses were performed using custom Python scripts. Statistical modeling and hypothesis testing were conducted with Python and R version 4.4.3 (R Core Team, 2024) using RStudio version 2024.12.1+563 (Posit Team, 2024). An interactive visualization of the data is available at https://acoustic-biogeography.vercel.app/ (see also interactive web application in the supplementary materials). Dataset (including audio files), data processing workflows, analytical scripts, and code for reproducibility are available in a permanent repository at Zenodo (*91*), https://zenodo.org/records/19916290. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aee6239
Materials and Methods; Supplementary Text; Figs. S1 to S23; Tables S1 and S2; References (*58*–*91*); MDAR Reproducibility Checklist

Submitted 15 December 2025; accepted 4 May 2026

10.1126/science.aee6239

LANGUAGE DIVERSITY

The rise and fall of language diversity through the Holocene

D. E. Blasi[1,2,3]*, M. J. Hamilton[4,5,6], R. D. Gray[7,8], C. L. Bowern[9]*

Characterizing the factors that have shaped linguistic diversity is fundamental for understanding human history, culture, and cognition. In this study, we combined statistical and social computational modeling, ethnographic data, and paleodemographic inference to model trajectories of global linguistic diversity. Before the onset of plant and animal domestication, the number of languages was smaller than it is today (4500 to 6000 compared with 7500). Subsequent increases in global population precipitated increased linguistic diversity. We uncovered a linguistic "golden age" with tens of thousands of languages 3000 to 1000 years ago. Great loss of linguistic diversity did not begin with recent colonial expansion but as multinational empires first spread along with their languages, pathogens, and cultures. Thus, extinction has likely played a much greater role in shaping linguistic and cultural diversity than previously thought.

There are ~7500 languages signed and spoken in the world today (*1*, *2*), and they have a wide range of grammatical structures, vocabularies, speech sounds, signs, and usage rules. The study of language diversity has brought distinctive insights into how human behavior and cognition are shaped by ecological, social, and communicative pressures (*3*–*8*). The observable limits of linguistic diversity—features that are either universally present or absent in the world's languages—have been at the core of what makes humans a distinctive species (*9*). Insights into language have also been used to understand other dimensions of human variation and homogeneity, including cultural and psychological traits. Over the past 500 years, all evidence points to a large systematic decrease in linguistic diversity, in parallel with the spread of a handful of behemoth languages, such as English, Spanish, Mandarin, Arabic, and Hindi. Nearly 5 billion people use at least 1 of the 10 most common languages, whereas half of the world's languages are now endangered, and about four become dormant every year (*10*). This results in the loss of cultural and scientific knowledge embedded in minority languages and contributes to the unequal development in science, technology, medicine, and education that predominantly benefits speakers of the world's largest languages (*11*–*13*).

In consequence, understanding past stages of linguistic diversity is fundamental for making sense of human cultural and cognitive history as well as for framing the current reduction in the world's languages. However, direct evidence for language diversity is absent before the first attestation of writing around 6000 years ago (*14*) and is sparse for most of the historical record, as language catalogs and systematic attempts at language documentation are recent (see supplementary text). Moreover, although methods from historical linguistics and evolutionary biology have been utilized to reconstruct prehistoric linguistic events, few languages were documented in antiquity (*15*). Most contemporary languages are derived from a handful of population expansions stemming from major technological and cultural Holocene developments: the domestication of plants and animals in the Fertile Crescent, the Yellow and Yangtze River Basins, the Andes, and Mesoamerica (*16*, *17*) and the development of groundbreaking technologies, such as seafaring in the case of the Austronesian world (*18*), among others.

In the absence of direct evidence, researchers have speculated on the global trajectory of the total number of languages over the Holocene (Fig. 1). The Megadiverse Paleolithic Hypothesis suggests that the total number of languages has been declining over the last 12,000 years (*19*), when agriculture-led language expansions began replacing hunter-gatherer language diversity. Although peak numbers are unknown, estimates range between 12,000 (*19*) and "tens of thousands" (*17*) of languages just before the advent of agriculture; such estimates extrapolate from the diversity now found only in New Guinea and Vanuatu. Under this view, proximate causes of the current linguistic diversity crisis (such as warfare, genocide, and the expansion of colonial powers along with their languages, cultures, and pathogens) have accelerated a process of language loss already in place throughout much of the Holocene.

By contrast, the Demographic Boost Hypothesis suggests that there were between 6000 and 10,000 languages in the terminal Pleistocene, based on assuming that there were 1000 to 2000 speakers per language and 10 million people in the preagricultural world. Plant and animal domestication increased the total number of languages by providing stable sources of subsistence and by triggering migrations, allowing languages to survive, spread, and diversify (*5*). Only the emergence of large states and their lingua francas in the past 4000 years increased rates of language loss. The evidence supporting this hypothesis comes from some small-scale agricultural regions of the world (such as the island of New Guinea) where dense mosaics of linguistic diversity still exist.

Lastly, the Premodern Equilibrium Hypothesis is inspired by observations that language diversification can be characterized by long periods of equilibrium, punctuated by brief bursts of change (*20*) associated with exogenous population shocks, such as ecological disasters or the introduction of new technology. The main factor predicting such events is the movement of groups associated with the peopling of large regions, most of which were already occupied by humans by the end of the Upper Paleolithic. Under this hypothesis, agriculture and later large states did not effectively alter this dynamic, and the total number of languages remained approximately uniform for most of the Holocene, until modern colonial expansions started

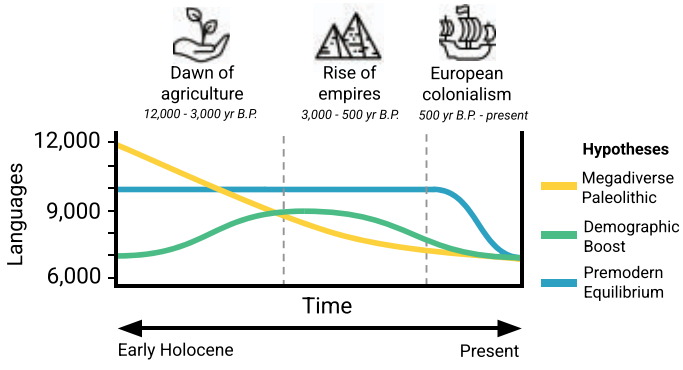


Fig. 1. Schematic representation of three competing hypotheses concerning the trajectory of the global number of languages over the Holocene (megadiverse paleolithic, demographic boost, and premodern equilibrium). Icons, from left to right, represent the main factors associated with changes in linguistic diversity across the periods of the dawn of agriculture, the rise of empires, and European colonialism. Time axis not to scale.

[1]Catalan Institute for Advanced Research and Study, Barcelona, Spain. [2]Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain. [3]Human Evolutionary Biology Department, Harvard University, Cambridge, MA, USA. [4]Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA. [5]School of Data Science, University of Texas at San Antonio, San Antonio, TX, USA. [6]Santa Fe Institute, Santa Fe, NM, USA. [7]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [8]School of Psychology, University of Auckland, Auckland, New Zealand. [9]Department of Linguistics, Yale University, New Haven, CT, USA. ***Corresponding author. Email: dblasi@fas.harvard.edu (D.E.B.); claire.bowern@yale.edu (C.L.B.)**

approximately 500 years ago. Within this framework, current rates of language loss are a catastrophic new development, and the premodern number of languages was larger than what it is today.

These hypotheses imply very different scenarios for the past of linguistic diversity and its correlates, and they provide very different contexts for assessing the magnitude of current rates of language loss and endangerment. In this study, we lay out a modeling strategy to identify the evolution of linguistic diversity over the Holocene based on the premises that mobility, fertility, and subsistence impose constraints on the number of individuals that can make up a human population and that the total number of such populations globally is associated with the total number of languages. Combined with paleodemographic reconstructions of global population estimates, this allows us to shed light on the history of linguistic diversity. Our inferences emerge from three sequential steps based on justified generalizations about the relationship between language numbers, population size, and how they change. These points are supported by evidence from ethnography, paleodemography, human biology, and history (Fig. 2).

Modeling the number of languages through the Holocene

We started with the assumption that distributions of forager group sizes are comparable over time and space. Human groups up to the early Holocene obtained subsistence through hunting, gathering, and fishing (*21*). Contemporary forager populations differ from early Holocene groups in key ways (*22*), and the archaeological record increasingly suggests a wide diversity of hunter-gatherer adaptations preceding the development of agriculture, much of which is not represented in the contemporary ethnographic and ethnohistorical record (*23*, *24*). Contemporary foragers also live in a much narrower range of environments than that which earlier groups occupied. Despite these differences, it is still possible to infer some fundamental traits of foraging populations back to the early Holocene (and points in between), assuming uniformitarian principles of human subsistence and organization (*25*).

Forager group sizes are constrained by several factors inherent to subsistence and mobility but are largely independent of ecological determinants. Forager population densities vary greatly across environments (*26*–*28*), yet variation in net primary production is not predictive of variation in group size (*29*–*31*). Thus, distributions of group sizes are probably comparable, even though contemporary forager populations occur in fewer environments. This independence stems from foraging imposing an upper limit on the range and cohesion of a hunter-gatherer population: Whereas energetic requirements increase with population size (*26*), the foraging radius is limited by mobility (*32*). Group sizes have a median of ~800 individuals approximately following a log-normal distribution (*33*), consistent with both the stochastic multiplicative growth of natural fertility in human populations (*30*) and fusion-fission dynamics (*34*), which suggest that large populations fission into smaller units and that populations that experience a crash can recover rapidly (*34*, *35*).

Thus, we derived the distribution of early Holocene group sizes using ethnographically recorded ethnolinguistic group sizes as a proxy for the population sizes of human groups 12,000 years before the present (yr B.P.) (see materials and methods). We only sourced data from the ethnographic literature (*21*, *36*) on hunter-gatherer-fisher forager groups whose subsistence and mobility at the time of the ethnographic assessment were not profoundly altered by contact with food-producing groups ($N = 171$; figs. S1 and S2). Based on these data, we modeled early Holocene group sizes as demographic distributions resulting from natural population growth with an upper bound on the maximum possible size of any ethnolinguistic group. We performed the inference on a Bayesian framework, including statistical controls for regional nonindependence and for data imbalance across regions. This resulted in an array of statistical models yielding similar distributions of early Holocene group sizes (fig. S3).

The second step in our model assumes that forager groups serve as proxies for languages in the early Holocene. The basic unit of description of hunter-gatherer populations is the ethnolinguistic group, i.e., a population sharing the same language and culture with frequent interactions among its members. In the ethnographic record, ethnolinguistic groups are often associated with one distinctive spoken language (*36*). Although multiple ethnolinguistic groups might share a language, a single ethnolinguistic group is unlikely to foster two or more mutually unintelligible spoken languages not present elsewhere. Taken together, this suggests that, in the early Holocene, the total number of ethnolinguistic groups was comparable to the number of languages. We included complementary analyses that relax this equivalence later in the inference process for completeness. This assumption does not necessarily hold later in the Holocene, when larger populations may share language but not culture and when groups grow so large that individuals no longer regularly interact with all other members.

Therefore, the distribution of the total number of languages in the early Holocene is simply the number of early Holocene groups that can be accommodated within the total human population in that period (see materials and methods). Formally, we derived this as the distribution of the number of repeated samples from the posterior predictive distribution required to match the total human population, estimates of which range from 4.4 million to 7 million in recent literature (table S3). We derived the corresponding distributions for the total number of languages in the early Holocene

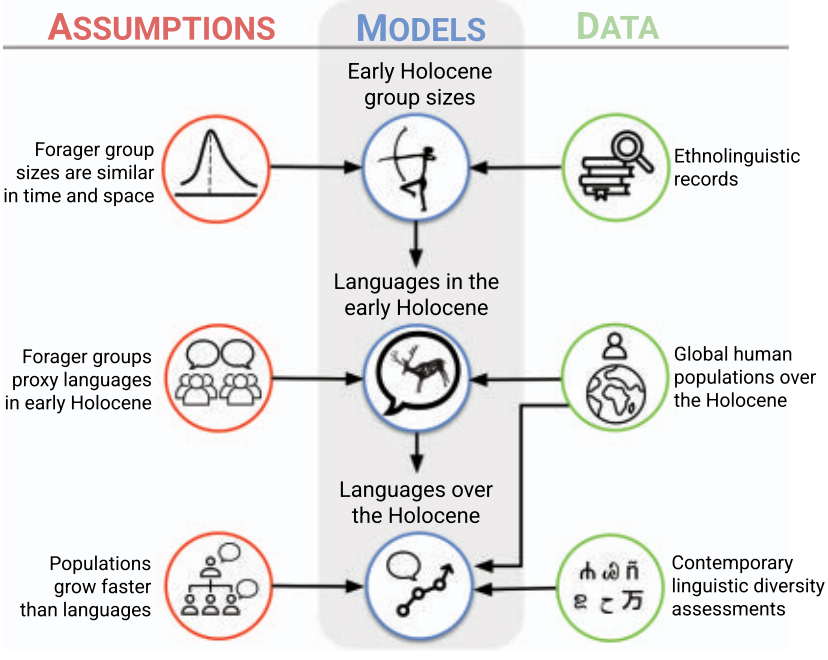


Fig. 2. Graphical summary of the data and assumptions guiding models of language diversity change over the Holocene. The flow from top to bottom represents three modeling stages and their respective outcomes: (i) We estimated the distribution of early Holocene group sizes, (ii) which allowed us to derive the number of languages in the early Holocene, resulting in (iii) the inference of plausible trajectories for the total number of languages over the Holocene.

(figs. S4 and S5). In addition, we determined that the model is robust in the presence of large, unobserved ethnolinguistic groups (fig. S6).

In our final modeling step, we assumed that Holocene developments caused population growth to outpace the number of languages over long timescales. A defining feature of the Holocene is the sustained rise in technologies and subsistence practices that allowed for larger populations (*21*). These larger carrying capacities paved the way for an increasingly larger number of individuals to share the same language. The increase in average group size was not monotonic, as pathogens, wars, and ecological factors have repeatedly decimated human populations (*37*); for example, the 14th Century witnessed the loss of 40% of Europe's population to the Black Death (*38*). However, human populations tend to rebound quickly from crashes, as observed throughout history and across subsistence types (*34*, *35*). Thus, when considering long timescales, the global language/individual ratio has decreased monotonically.

This decrease allowed us to estimate the total number of languages over the Holocene (see materials and methods). We achieved this indirectly, as explicitly modeling the convoluted region-specific dynamics that have shaped worldwide linguistic diversity over the Holocene is unfeasible. The basis of our strategy is the observation that, at any point during the Holocene, the total number of languages is the product of two global exogenous variables: (i) the ratio of the number of languages over individuals and (ii) the total human population. Global Holocene human population estimates are readily available from HYDE 3.2 (*39*) (fig. S7). The language/individual ratio becomes modelable under the assumption that it has decreased monotonically from the start of each millennium through the Holocene. This constraint reduces the challenge to a straightforward statistical interpolation problem with boundary conditions. More precisely, we obtained the language/individual ratios both at 12,000 yr B.P. (based on paleodemographic models) and the present [using contemporary assessments (*1*, *2*)] and then analyzed all monotonically decreasing discrete time series that go from the first to the latter estimate, sampling every 1000 years. We summarized these trajectories using functional principal components analysis (PCA; fig. S8). We also analyzed the effect of assuming larger early languages and find that the overall patterns hold robustly.

Lastly, our understanding of the Holocene strongly points to larger changes in the language/individual ratio happening closer to the present. It is unlikely that early in the Holocene, the language/individual ratio was more similar to that of the present than to that of the start of the epoch, as many of the changes that allowed for large-scale social coordination and stability developed after essential technologies and practices that emerged over the second half of the period. Instead of imposing further constraints in our model, we inspected specific subclasses of monotonic functions reflecting stricter and potentially more realistic scenarios of language diversity evolution, including different average ratios of language to ethnolinguistic group, accelerated change and fast transitions in specific time periods (Iron Age and Chalcolithic), and bounded diversification rates. The main findings of our model replicate in these more grounded cases (figs. S9 to S11).

Results

These models of early Holocene group sizes vary somewhat across models but provide consistent overall results of global language counts (Fig. 3). The early Holocene supported a comparable number or, more likely, a smaller number of languages than is found today, at odds with both the Megadiverse Late Paleolithic and the Premodern Equilibrium Hypotheses. Across all modeling conditions, the number of languages during the early Holocene ranges from about 3300 to 7800 (95th percentile interval), with a concentration around 4500 to 6200 (50th percentile interval). The upper bound of this interval overlaps with the hypothesized range in the Demographic Boost scenario.

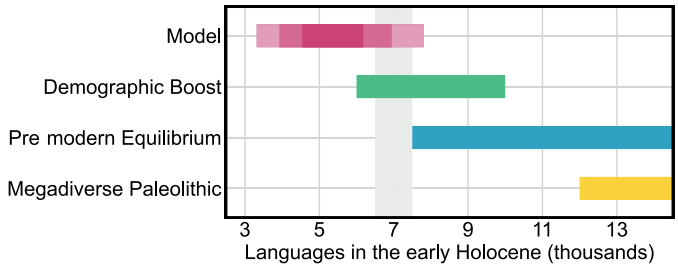


Fig. 3. Plausible intervals for the number of languages in the early Holocene for three hypotheses (Demographic Boost, Premodern Equilibrium, and Megadiverse Paleolithic) compared with those of our paleodemographic model. Shading in our model's estimates corresponds to the 95th, 80th, and 50th percentiles (from lighter to darker) aggregating across all conditions. The vertical gray shading corresponds to the current estimate of the number of extant languages.

From this early Holocene linguistic diversity, we derived the distribution of the total number of languages across the Holocene for each millennium (Fig. 4A). Although many theoretically possible demographic histories are compatible with the data, the inferred trajectories of linguistic diversity over the Holocene consistently show a handful of salient patterns: a functional PCA decomposition revealed that just three components are enough to account for over 96% of their variance (Fig. 4B) and that the majority of individual estimates share several properties (figs. S10 and S11). These trajectories reveal two facets of prehistoric linguistic diversity.

Typically, the total number of languages has been increasing for most of the Holocene (Fig. 4A), mimicking the global expansion of human populations. This suggests that the rise of large-scale regional powers 4000 years ago did not generally and immediately deter the growth of linguistic diversity. Instead, most trajectories reached a maximum of 1000 to 3000 yr B.P. Although there is wide uncertainty in our estimates, most trajectories display an approximately 10-fold increase in the number of languages between the early Holocene and this point (Fig. 4C). This period, harboring tens of thousands of languages, has escaped previous theorizing.

In our model, this "golden age" was followed by an extremely rapid decrease in linguistic diversity, especially over the last two millennia (Fig. 4D). The process unfolded so quickly that the rate of language loss must have been even faster than the current pessimistic forecasts for the year 2100 (*40*).

Implications for the history of languages and cultures

These results have several implications for our understanding of recent and contemporary linguistic and cultural diversity. The current number of languages is most likely an order of magnitude lower than that of the "golden age" of linguistic diversity 1000 to 3000 yr B.P. This surviving linguistic diversity is not an unbiased sample of past languages. Converging evidence suggests that premodern language loss was driven by the expansion of the linguistic groups that have survived to the present. The expanding language groups were receptive to genetic inflow from nonexpanding groups, and these groups, in turn, often adopted the expanding languages and lifestyles (*17*, *41*). This probably led to a rise in the number of nonnative speakers and multilingualism, both catalysts for language change (*42*).

Contact between populations tends to homogenize languages (*43*), and some linguistic outcomes of language contact are more likely than others (*42*). Therefore, a consequence of language reduction is that the world's languages were both locally and globally less diverse. However, languages belonging to expanding groups encountering different nonexpanding languages might have become less similar to each other and more similar to those nonexpanding languages. This is particularly

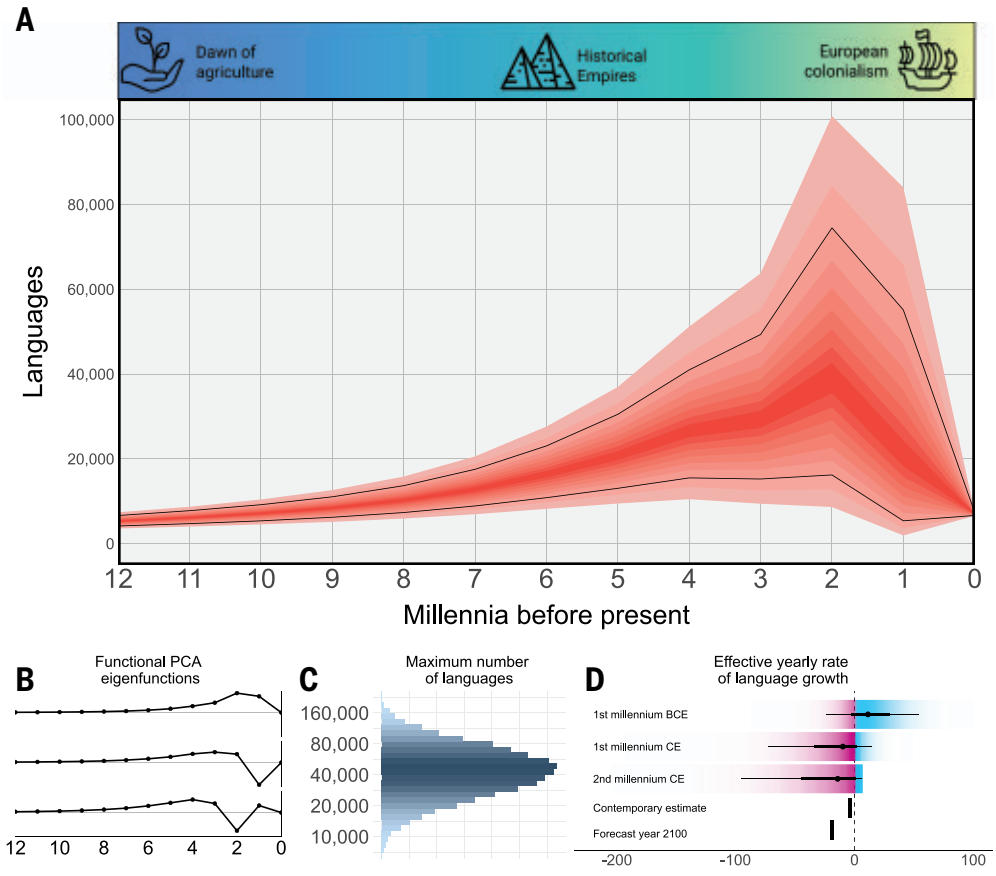


Fig. 4. Marginal results of our model aggregated across all conditions. (**A**) The number of languages through the Holocene. Lighter shades of the curve correspond to progressively wider intervals comprising 10 to 90% of the predictions. Black lines indicate the interval containing 70% of the predictions. (**B**) Leading eigenfunctions of the functional PCA decomposition of these curves, from top to bottom, cumulatively explaining 0.73, 0.90, and 0.96 of the total variance. (**C**) Distribution of the maximum number of languages observed. (**D**) Effective yearly language loss over the past 3000 years, plus a contemporary estimate of ~4.1 languages lost per year (*10*). BCE, before the common era; CE, common era.

relevant because the period of exceptional linguistic diversity (1000 to 3000 yr B.P.) is comparable to or even more recent than the characteristic time depths of language groups with uncontroversially identifiable common ancestry, such as Germanic or Bantu (*44*). Therefore, the splits and subdivisions within these groups might have experienced accelerated change (*45*). The conclusion is that recent rates of language change might have been higher than what has characterized most of human history (*46*, *47*).

It follows that this massive and biased reduction in linguistic diversity has shaped the relative abundance of different language traits. However, functional factors (such as learnability, efficient communication, or computational costs) are typically invoked instead to explain why some language traits are more common than others (*48*). But the intensity and recency of the linguistic diversity bottleneck make it a clear nonfunctional null hypothesis for understanding any skewed linguistic distribution.

Lastly, these results show that the current endangerment crisis has deep roots beyond the most recent colonial period. Moreover, this drastic reduction in diversity was not solely linguistic. Languages (and their histories) are regularly utilized as proxies for other units of culture (*49*) because language is often transmitted and acquired along with other cultural traits, including marriage patterns, religious practices, kinship systems, or modes of social organization. Therefore, cultural variants with transmission dynamics similar to those of languages have also undergone a recent mass extinction.

Our models offer a coarse view of prehistory that is agnostic about the mechanisms that curtailed linguistic diversity, and our assumptions preclude us from testing the causal role of factors specific to narrow time periods and regions. However, the common factor underlying the decrease in cultural diversity (shared by both early and modern empires) is rapid, large-scale interactions between human groups, accompanied by the movement of people, technology, diseases, and culture across vast territories. Thus, the survival of specific cultural and linguistic variants hitchhiked on a handful of population traits (e.g., military technology, political organization, or pathogen immunity) rather than on the variants themselves. This should caution against strong selectionist views of recent human cultural history and raise the possibility of a much larger role for extinction in structuring the patterns of cultural and linguistic diversity that we see today.

REFERENCES AND NOTES

1. H. Hammarström, R. Forkel, M. Haspelmath, S. Bank, Glottolog/Glottolog: Glottolog database 4.5, Zenodo (2021); https://doi.org/10.5281/ZENODO.5772642.
2. D. M. Eberhard, G. F. Simons, C. D. Fennig, Eds., Ethnologue: Languages of the world (2022); http://www.ethnologue.com.
3. M. C. Gavin *et al.*, *Bioscience* **63**, 524–535 (2013).
4. P. Epps, in *Language Dispersal, Diversification, and Contact*, M. Crevels, P. Muysken, Eds., 2020), pp. 275–290.
5. N. Evans, *Dying Words: Endangered Languages and What They Have to Tell Us* (Blackwell, 2010).

6. J. B. Axelsen, S. Manrubia, *Proc. Biol. Sci.* **281**, 20133029 (2014).
7. S. Lee, T. Hasegawa, *J. Evol. Biol.* **27**, 1905–1912 (2014).
8. D. Nettle, *Linguistic Diversity* (Oxford Univ. Press, 1999).
9. D. E. Brown, *Daedalus* **133**, 47–54 (2004).
10. L. Campbell, A. Belew, *Cataloguing the World's Endangered Languages* (Routledge, 2018).
11. D. E. Blasi, J. Henrich, E. Adamou, D. Kemmerer, A. Majid, *Trends Cogn. Sci.* **26**, 1153–1170 (2022).
12. A. M. García, J. de Leon, B. L. Tee, D. E. Blasi, M. L. Gorno-Tempini, *Brain* **146**, 4870–4879 (2023).
13. D. Blasi, A. Anastasopoulos, G. Neubig, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, A. Villavicencio, Eds. (Association for Computational Linguistics, 2022), pp. 5486–5505.
14. P. T. Daniels, W. Bright, *The World's Writing Systems* (Oxford Univ. Press, 1996).
15. C. Bowern, *Annu. Rev. Linguist.* **4**, 281–296 (2018).
16. P. Bellwood, *Annu. Rev. Anthropol.* **30**, 181–207 (2001).
17. J. M. Diamond, *Nature* **389**, 544–546 (1997).
18. R. Blust, *The Austronesian Languages* (Pacific Linguistics, 2013).
19. M. Pagel, in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy, J. Hurford, Eds. (Cambridge Univ. Press, ed. 1, 2000), pp. 391–416.
20. D. Nettle, S. Romaine, *Vanishing Voices: The Extinction of the World's Languages* (Oxford Univ. Press, 2000).
21. L. R. Binford, *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Ethnographic and Environmental Data Sets* (Univ. of California Press, 2019).
22. S. L. Kuhn, M. C. Stiner, in *Hunter-Gatherers: An Interdisciplinary Perspective*, C. Panter-Brick, R. Layton, P. Rowley-Conwy, Eds. (Cambridge Univ. Press, 2001), pp. 99–142.
23. A. H. Simmons, (Univ. of Arizona Press, 2007).
24. S. Nalawade-Chavan, J. McCullagh, R. Hedges, *PLOS ONE* **9**, e76896 (2014).
25. J. C. French, A. T. Chamberlain, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20190720 (2021).
26. M. J. Hamilton, B. T. Milne, R. S. Walker, J. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4765–4769 (2007).
27. M. Tallavaara, J. T. Eronen, M. Luoto, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1232–1237 (2018).
28. P. H. Kavanagh *et al.*, *Nat. Hum. Behav.* **2**, 478–484 (2018).
29. F. W. Marlowe, *Evol. Anthropol.* **14**, 54–67 (2005).
30. M. J. Hamilton, R. S. Walker, *Evol. Ecol. Res.* **19**, 85–102 (2018).
31. C. C. Porter, F. W. Marlowe, *J. Archaeol. Sci.* **34**, 59–68 (2007).
32. R. L. Kelly, in *Handbook on Evolution and Society*, A. Maryanski, R. Machalek, J. Turner, Eds. (Routledge, 2015), pp. 295–315.
33. M. J. Hamilton, B. T. Milne, R. S. Walker, O. Burger, J. H. Brown, *Proc. Biol. Sci.* **274**, 2195–2202 (2007).
34. M. J. Hamilton, R. S. Walker, D. C. Kesler, *Sci. Rep.* **4**, 4541 (2014).
35. M. D. Gurven, R. J. Davison, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12758–12766 (2019).
36. K. R. Kirby *et al.*, *PLOS ONE* **11**, e0158391 (2016).
37. M. Livi-Bacci, *A Concise History of World Population* (Wiley, 2017).
38. R. Jedwab, N. D. Johnson, M. Koyama, *J. Urban Econ.* **139**, 103628 (2024).
39. K. K. Goldewijk, A. Beusen, J. Doelman, E. Stehfest, *Earth Syst. Sci. Data* **9**, 927–953 (2017).
40. L. Bromham *et al.*, *Nat. Ecol. Evol.* **6**, 163–173 (2022).
41. J. Diamond, P. Bellwood, *Science* **300**, 597–603 (2003).
42. S. Thomason, in *The Handbook of Language Contact*, R. Hickey, Ed. (Wiley, 2020), pp. 31–49.
43. A. Graff *et al.*, *Sci. Adv.* **11**, eadv7521 (2025).
44. M. S. Dryer, *Linguist. Typol.* **4**, 334–356 (2000).
45. P. Trudgill, *J. Hist. Socioling.* **6**, 1–13 (2020).
46. D. Blasi *et al.*, *Science* **363**, eaav3218 (2019).
47. P. Trudgill, in *Language Dispersal, Diversification, and Contact*, E. Crevels, P. Muysken, Eds., 2020), pp. 44–57.
48. D. Dediu *et al.*, in *Cultural Evolution: Society, Technology, Language, and Religion*, P. Richerson, M. Christansen, Eds., pp. 303–324.
49. A. Mesoudi, *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (University of Chicago Press, 2011).
50. D. E. Blasi, M. J. Hamilton, R. D. Gray, C. L. Bowern, Code and data for "The rise and fall of language diversity over the Holocene." Open Science Foundation, OSF (2026); https://doi.org/10.17605/OSF.IO/NEUTV.

ACKNOWLEDGMENTS

We thank Q. Atkinson, M. Singh, and M. Holton Price for their useful comments on earlier versions of this manuscript. **Funding:** Branco Weiss Foundation (D.E.B.); Harvard University Data Science Initiative (D.E.B.); Max-Planck-Gesellschaft (R.D.G.); US National Science Foundation grant BCS-1423711 (C.L.B.); US National Science Foundation grant BCS-0844550 (C.L.B.). **Author contributions:** Conceptualization: D.E.B., with input from C.L.B. and M.J.H.; Methodology: D.E.B., with input from C.L.B.; Investigation: D.E.B., C.L.B., M.J.H., R.D.G.; Visualization: D.E.B.; Writing – original draft: D.E.B. and C.L.B., with input from M.J.H. and R.D.G.; Writing – review & editing: D.E.B., C.L.B., M.J.H., R.D.G. **Competing interests:** The authors declare no competing interests. **Data, code, and materials availability:** All data, code, and materials used in the analysis are available from Open Science Foundation (*50*).
License information: Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx4343
Materials and Methods; Supplementary Text; Figs. S1 to S11; Tables S1 to S6; References (*51–97*)

Submitted 13 March 2025; accepted 19 May 2026

10.1126/science.adx4343

PHYSICS

Dynamic asymmetric strain imprinted into substrates by an oxide thin film

Elliot Kisiel[1,2]*, Alexandre Pofelski[3], Pavel Salev[4], Erbin Qiu[1], Spencer Reisbick[3], Chuhang Liu[3], Andreas Glatz[5,6], Ishwor Poudyal[2,5], Wei He[1], Rourav Basak[1], Kaan Alp Yay[7,8,9], Junjie Li[1], Umeshkumar Patel[2], Zhan Zhang[2], Arndt Last[10], Yimei Zhu[3]*, Ivan K. Schuller[1], Zahir Islam[2]*, Alex Frano[1]*

In film-substrate systems, the substrate role is often considered to be limited to providing static mechanical constraints. Dynamic film-substrate interactions when a structural change in the film modifies the substrate are generally disregarded. Using combined x-ray and electron microscopies, we observed that an electrically induced filament in a vanadium dioxide film created strong asymmetric strain in an underlying sapphire substrate. This asymmetric substrate strain fed back into the film and defined the filament expansion direction, revealing the importance of film-substrate dynamic interactions in determining film functionality. Furthermore, the strain imprint propagated at least tens of micrometers deep into the substrate, exceeding the film thickness by more than 200-fold, potentially enabling substrate functionalization as an active mechanical coupling media in three-dimensional integrated microelectronic architectures.

Substrates define elastic constraints that determine the crystallinity of supported thin films. Choosing the substrate enables tuning the film's electronic, magnetic, and optical properties (*1*–*4*) and can stabilize phases that are inaccessible in bulk crystals (*5*–*9*). Substrates are widely regarded as static passive components [with the exception of piezoelectric or flexible materials (*10*, *11*)]. When external stimuli prompt structural changes in the film, such changes must be accommodated within the film, preserving the substrate-imposed constraints. For example, epitaxial strain imposed by a titanium dioxide substrate on a vanadium dioxide (VO_2) film leads to the preferential nucleation of metallic domains at the film-substrate interface and development of striped surface morphological features because of the lattice volume change during the coinciding electronic and structural transitions in the film (*12*, *13*). Alternatively, constraints imposed by a (001) sapphire (Al_2O_3) substrate on the corundum *c* plane in a vanadium(III) oxide film hinder the *c* plane's expansion required by the structural transition in the film, which in turn suppresses the electronic insulator-metal transition (IMT) (*14*, *15*). The possibility that static structural changes within a film may substantially modify an underlying substrate is in general regarded to be inconsequential. Furthermore, questions such as whether dynamic changes in the film could propagate into the substrate and, if so, whether such substrate changes in turn affect the film behavior remain open.

In this study, we performed experiments that challenge the common viewpoint about the passive, static role of substrates. We found that a structural transition in a film imprinted strain deep into its underlying substrate, which fed back into the film, dictating phase transition pathways at the core of film's functional behavior. The key phenomenon that enabled this coupled film-substrate strain field is the localization of the phase transition within the film, which we achieved by the electrical triggering of a coupled IMT and structural (monoclinic-to-rutile) transition in VO_2 thin-film devices on an Al_2O_3 substrate (*16*–*20*). When a rutile (R) metal filament formed in the insulating monoclinic (M1) matrix of VO_2, the underlaying regions in the Al_2O_3 developed an asymmetric strain profile with a magnitude of ±0.12%, which exceeded thermal expansion by threefold, and propagated down more than 70 μm, or >230 times the film's thickness. The induced strain on the substrate during electrical activation made the phase transition progression in the film highly anisotropic. Although Joule heating is the primary driver of the electrically induced phase transition in VO_2 film (*21*)—and a similar substrate response was expected during electrical and thermal cycling—we observed no measurable strain imprint in Al_2O_3 during the thermally driven transition. The lack of imprint during the thermal transition indicated that local, electrically triggered switching affects the substrate in a qualitatively different way than does global heating, despite both being thermally mediated. Our work reveals the complexity of dynamic film-substrate elastic interactions fundamental in the film's response to external stimuli, such as in the resistive switching process. Our results present an evolving understanding that may shed light on material behaviors involving functional dynamic changes (*22*–*24*) and point toward new research avenues to harness the substrate's active functionalities in next-generation electronics.

Structural modulation of Al_2O_3 substrate by VO_2 thin film

To begin, we examined the film-substrate interactions in equilibrium, that is, without applying voltage to VO_2 thin-film devices. Baseline characteristics of VO_2/Al_2O_3 samples are shown in figs. S1 to S6. We used dark-field x-ray microscopy (DFXM) (*25*–*31*), a full-field x-ray microscopy technique that uses an x-ray objective lens in the path of the diffracted beam to produce spatially resolved diffraction-selective images (see supplementary materials for more information and fig. S7). A diagram of DFXM operation is shown in Fig. 1A. We used weak-beam (WB) imaging, which utilizes the Bragg peak tails, making it sensitive to local lattice deformations (*25*, *29*). We performed the measurements near the (012) Al_2O_3 peak (Fig. 1C and fig. S1) on a device where the VO_2 film covered the substrate and imaged the Al_2O_3 substrate through the film (Fig. 1B, FOV-1). Figure 1, D to F, compares DFXM imaging under WB and strong-beam (SB) conditions. The SB image (Fig. 1D) shows a uniform structure because the strong diffraction intensity overshadows the subtle inhomogeneity contributions. In the WB images (Fig. 1, E and F), we observed an inhomogeneous, granular-like structure, corresponding to local orientation and strain deviations randomly spread throughout the substrate. We thus identified the conditions to explore structural inhomogeneities in the Al_2O_3 substrate.

We performed the WB imaging of the Al_2O_3 substrate (at the conditions identified in Fig. 1, C and F) in a region where VO_2 was etched to make an isolated 30-μm-diameter island (Fig. 1G, FOV-2). The island location was confirmed by SB DFXM at the monoclinic (M1) $(200)_{M1}$ VO_2 peak (Fig. 1H). We note that the DFXM images appear stretched because of the viewing angle (fig S8). The WB image of Al_2O_3 (Fig. 1I) showed that structural inhomogeneities in the substrate were only present in the regions covered by the film, and the regions where the film was etched exhibited a uniform structure (see figs. S9 and S10 for additional details). The one-to-one correspondence between the inhomogeneities in Al_2O_3 and location of VO_2 indicated that the film's

[1]Physics Department, University of California, San Diego, La Jolla, CA, USA. [2]X-ray Science Division, Argonne National Laboratory, Lemont, IL, USA. [3]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY, USA. [4]Department of Physics and Astronomy, University of Denver, Denver, CO, USA. [5]Materials Science Division, Argonne National Laboratory, Lemont, IL, USA. [6]Department of Physics, Northern Illinois University, DeKalb, IL, USA. [7]Department of Physics, Stanford University, Stanford, CA, USA. [8]Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA, USA. [9]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [10]Institute of Microstructure Technology, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany. ***Corresponding author. Email: ekisiel@anl.gov (E.K.); afrano@ucsd.edu (A.F.); zahir@anl.gov (Z.I.); zhu@bnl.gov (Y.Z.)**

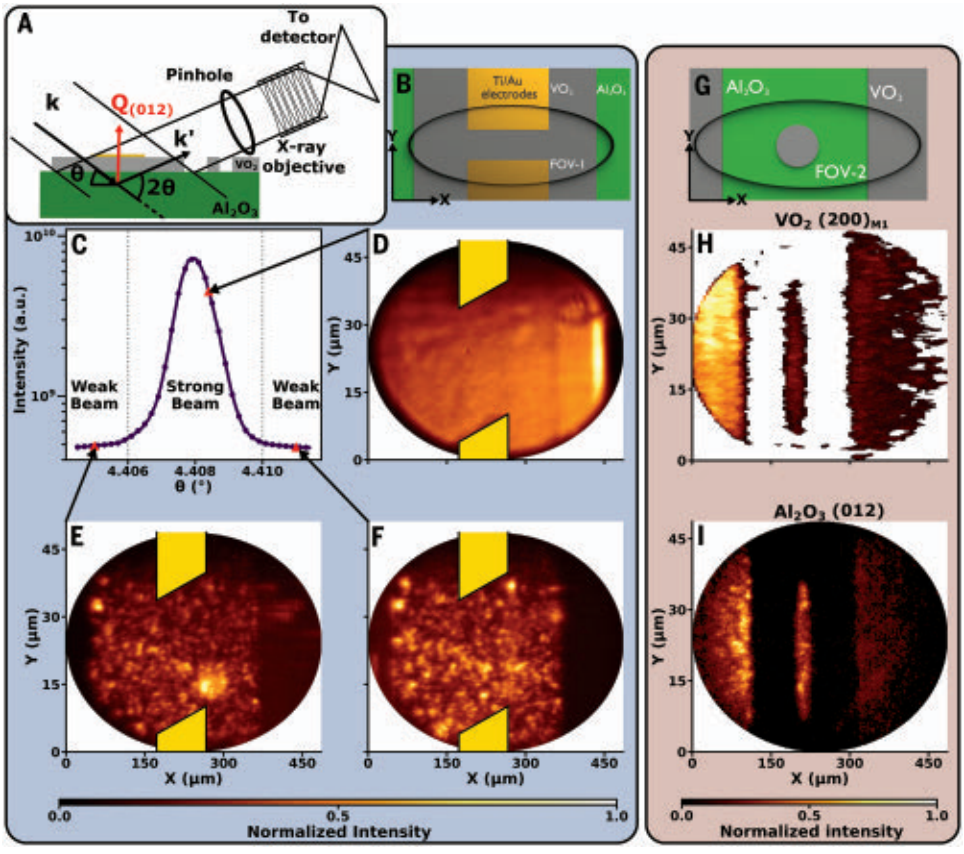


Fig. 1. DFXM imaging of structural inhomogeneities in the Al_2O_3 substrate induced by the VO_2 film. (**A**) DFXM setup schematic. The diffracted beam from the sample passes through the x-ray objective lens to produce a real-space image of the Bragg peak on the detector. (**B**) FOV schematic for (C) to (F). (**C**) Rocking curve of the (012) Al_2O_3 peak. SB and WB conditions are highlighted. Red triangles indicate image collection conditions. a.u., arbitrary units. (**D** to **F**) SB (D) and WB [(E) and (F)] DFXM images of Al_2O_3. Rectangular electrodes (highlighted in yellow) appear slanted because of the viewing angle. WB images reveal structural inhomogeneities in Al_2O_3. (**G**) FOV schematic for (H) and (I). (**H**) SB DFXM image of the VO_2 $(200)_{M1}$ film island region. (**I**) WB DFXM Al_2O_3 image of the (012) peak acquired in the same region as in (H). Structural inhomogeneities in Al_2O_3 appear only in the regions covered by VO_2 (fig. S10), revealing the film-substrate elastic interactions.

presence changes the substrate's structure. The absence of inhomogeneities in the Al_2O_3 regions where VO_2 was etched indicated that these inhomogeneities were not present in the substrate before the film deposition or were not induced during the high-temperature film growth. Overall, these experiments provided initial evidence that there are elastic interactions between the film and substrate that structurally modify the substrate.

Elastic response in the substrate imprinted by the local phase transition in the film

Next, we focused on the quantitative analysis of the strain imprinted into the Al_2O_3 substrate by the electrically triggered local phase transition in the VO_2 film (Fig. 2A). Although uniformly heating the film produced a gradual IMT with the coexistence of nanoscale metallic and insulating domains near 340 K (Fig. 2B) (*18*), applying voltage to VO_2 induced abrupt resistive switching (Fig. 2C), which occurred by the formation of a localized metallic phase filament within the insulating phase matrix (*17*, *27*). Because the electronic IMT coincided with the structural monoclinic-to-rutile transition (*19*, *27*), the voltage-induced filament had a distinct structure (figs. S2 and S6). We directly identified the filament location in our 30 μm by 120 μm device (Fig. 2D) using DFXM at the rutile $(011)_R$ VO_2 peak (fig. S6), similar to previous structural mapping experiments (*27*). For high-precision quantitative analysis, we then switched to the microdiffraction operation using a 10 μm by 10 μm beam. Figure 2, E and F, compares reciprocal space maps near the (012) Al_2O_3 peak at 0 and 15 V, acquired in the region beneath the filament in VO_2 (Fig. 2D, blue box). We observed that the filament in VO_2 prompted distortions of the substrate's Bragg peak, corresponding to an ~0.1° orientation change (Fig. 2F, peak splitting, white two-headed arrow) and an ~0.1% out-of-plane strain (Fig. 2F, peak shift, red arrow). This experiment thus revealed that the film-substrate elastic interactions can be dynamically actuated using voltage.

By scanning across the filament region (Fig. 2D, green arrow) using the (012) Al_2O_3 Bragg peak, we observed an unexpected asymmetric strain profile imprinted into the substrate by the filament in VO_2. Using microdiffraction (Fig. 2A), we found that on one side of the filament, the Al_2O_3 lattice is compressed, and on the other side, the Al_2O_3 lattice is expanded (Fig. 2G, green line and symbols). Because the switching in VO_2 is volatile (i.e., the film recovers the monoclinic insulating phase when voltage is turned off), the strain imprint into the substrate was also volatile: Once the filament in VO_2 disappeared, the substrate returned to its original state (Fig. 2H shows the maximum and minimum substrate strain measured during the voltage decrease). An identical asymmetric strain imprint in the Al_2O_3 substrate was also observed when the polarity of the voltage applied to the VO_2 device was flipped (fig. S11). Additional measurements showed that as the film thickness increased, the strained Al_2O_3 Bragg peak intensity increased, suggesting that a larger substrate volume was strained, but the strain magnitude remained unchanged, indicating the robustness of the strain imprint effect (fig. S12). Furthermore, we observed a substrate strain imprint in VO_2 grown on a Si substrate (fig. S13), indicating that dynamic film-substrate elastic interactions could be a general phenomenon across a wide range of IMT and substrate material combinations.

Because the filament in VO_2 is likely maintained by Joule heating (*21*), a local temperature increase in the substrate was the initially suspected cause for the strain imprint. The observed compressive-to-tensile strain reversal across the filament, however, contradicted the picture of the strain's thermal origin: It is implausible that the temperature increased on one side of the filament to produce a lattice expansion in Al_2O_3 and simultaneously decreased on the opposite side to make the lattice contract. Substrate strain mapping at zero voltage and 350 K (Fig. 2G, purple line and symbols, and fig. S14) showed no evidence of the strain asymmetry in Al_2O_3 when the entire VO_2 film underwent its first-order phase transition. The Al_2O_3 strain magnitudes in the voltage- and temperature-dependent experiments also differed by a factor of ~3, ±0.12% versus +0.04%, respectively, providing further evidence that thermal expansion alone cannot account for the observed strain. Moreover, infrared imaging and thermal diffusion

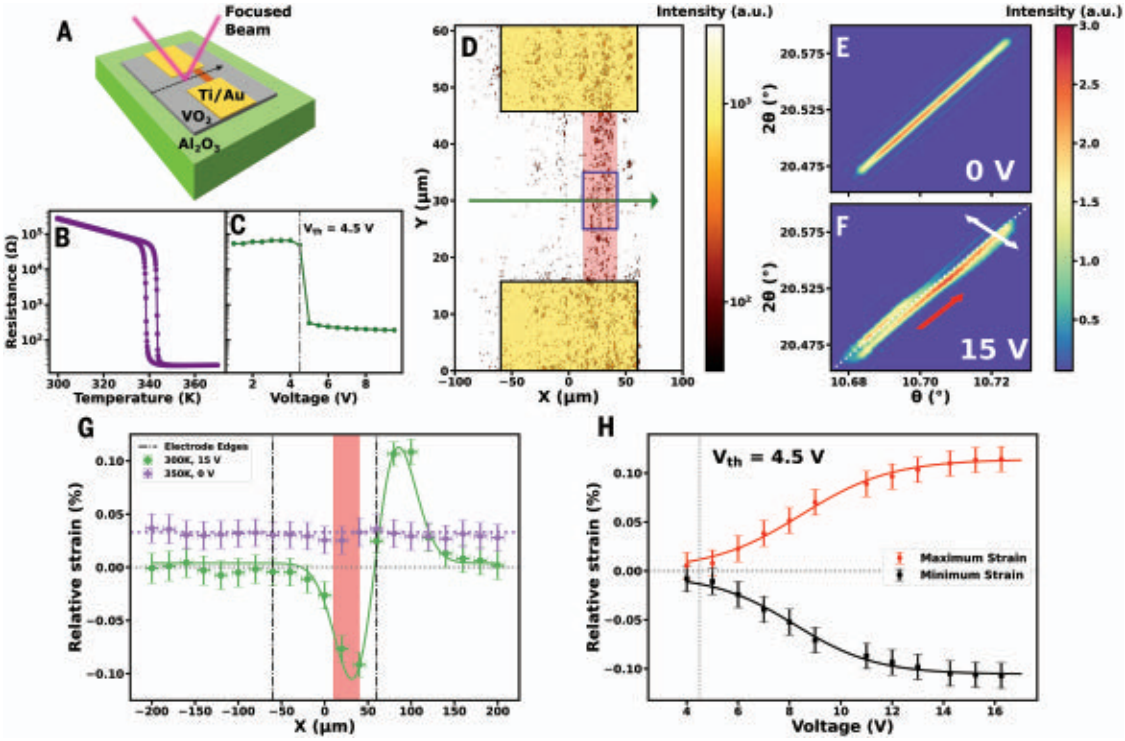


Fig. 2. Al_2O_3 substrate strain imprinted by filament formation in the VO_2 film. (**A**) Schematic of the VO_2/Al_2O_3 device. The black arrow indicates the microdiffraction scan direction across the filament region. (**B**) Resistance-temperature measurements showing the metal-insulator transition at ~340 K. (**C**) Resistance-voltage dependence showing the abrupt electrical switching in VO_2. The filament formation occurs at $V_{th} = 4.5$ V. (**D**) DFXM image of the voltage-induced filament in VO_2 using the (011) rutile VO_2 peak. The intensity in the figure shows rutile domains whose Bragg conditions are satisfied (see fig. S6). The electrodes are shaded in gold, and the filament is highlighted in red. The green arrow indicates the scanning path for (G), and the blue box shows the area probed in (E) and (F). (**E** and **F**) (012) Al_2O_3 peak acquired using microdiffraction at 0 V (E) and 15 V (F) applied to VO_2. The peak at 15 V splits (white two-headed arrow) and shifts (red arrow). The white dotted line in (F) indicates the zero-voltage peak location. (**G**) Microdiffraction mapping of the out-of-plane strain in Al_2O_3 during the voltage-induced filament formation in VO_2 at 300 K (green line and symbols) and at 350 K and 0 V (purple line and symbols). The filament location is highlighted in red, and the vertical dot-dash lines indicate the device electrodes. Unlike the spatially uniform phase transition in VO_2 at 350 K, voltage-induced localized phase transition in the film imprints large asymmetric out-of-plane strain in Al_2O_3. (**H**) Voltage dependence of the asymmetric out-of-plane strain imprint (peak and valley in red and black, respectively) in Al_2O_3. The imprinted out-of-plane strain in the substrate correlates with the switching in VO_2. Data points in (G) and (H) show the calculated strain, and error bars show a standard deviation.

simulations (*32*) showed that heating effects only produce a symmetric temperature profile across the filament, which cannot account for the strain asymmetry in Fig. 2G (figs. S15 and S16). We note that we observed a stable asymmetric strain imprint (both compressive and tensile) in Al_2O_3 under DC voltage application, in contrast to the recent report of transient substrate bulging (tensile strain only) attributed to the voltage-induced oxygen vacancy ionization (*33*). The IMT properties also remain unchanged after multiple voltage switching cycles (fig. S17), further indicating that oxygen migration was absent in our experiments (*34*). Overall, we found that the filament formation in VO_2 prompts an anomalous asymmetric compressive-to-tensile strain imprint in Al_2O_3 exceeding the thermal expansion by a magnitude of severalfold, revealing intricate film-substrate elastic interactions when the film undergoes a localized voltage-driven phase transition.

Filament growth anisotropy from the film-substrate strain feedback

Asymmetric filament growth in IMT switching devices, that is, preferential filament expansion in one direction under increasing voltage, has been reported in multiple studies (*35*–*38*), without a clear consensus about its origin. Our measurements on a VO_2 device (Fig. 1B) revealed that structural changes imprinted into the substrate by the local voltage-induced phase transition in the film can feed back into the film, affecting the direction of filament expansion. A schematic of the collocated VO_2 and Al_2O_3 imaging is shown in Fig. 3A. Above V_{th} (4.5 V), a rutile phase filament formed inside the monoclinic VO_2 (Fig. 3B). When the applied voltage increased, the filament became wider. We observed, however, that the filament expanded only in one direction, toward the left in the phase maps. Conversely, the filament edge on the right side remained at the same position at all voltages (Fig. 3B, white dashed line). Because the filament formed between the device electrode edges, filament expansion in both the left and right directions should be possible. Geometric constraints, thus, could not account for the anisotropic filament expansion.

We observed a correlation between the VO_2 filament expansion direction and the propagation of the compressively strained region in Al_2O_3. The DFXM image in Fig. 3C, i, showed that before the filament formation in VO_2, there was a small expansion in the substrate (~0.01%), well matched with the expected thermal expansion (fig. S14). Notably, there were no structural inhomogeneities in Al_2O_3 that may have enticed the filament formation in VO_2 at a particular location. Figure 3C, ii to iv, shows the Al_2O_3 strain maps collocated with the filament region in VO_2. In line with the microdiffraction results in Fig. 2, we observed an asymmetric strain in Al_2O_3 imprinted by the filament in VO_2. The strain imprint region in Al_2O_3 also exhibited the Bragg peak broadening indicative of local structural inhomogeneities (fig. S18). With increasing voltage, the compressive strain region in Al_2O_3 expanded toward the left, in the same direction as the filament

growth in VO_2. We suggest that the monoclinic insulator to rutile metal transition in VO_2 is favorable over the local regions of compressively strained Al_2O_3. Regions of the Al_2O_3 expansion likely promoted the stabilization of the monoclinic insulator VO_2 phase, in agreement with the temperature-pressure phase diagram (*7*, *35*–*38*). We therefore observed a subtle film-substrate interaction phenomenon: A local phase transition induced asymmetric strain in the substrate, which in turn determined the direction of the phase propagation front in the film.

Propagation depth of the substrate strain imprint

To investigate the strain imprint propagation depth, we used transmission mode DFXM (*30*) at the (102) Al_2O_3 peak (Fig. 4A). The transmission measurements were done using a 5 μm by 75 μm sheet beam (Fig. 4A), which allowed for quantitative mapping of the strain depth profile in the Al_2O_3 region directly under the VO_2 filament. Figure 4B shows the integrated diffraction scans of the (102) Al_2O_3 peak at several voltages applied to VO_2. Above the switching threshold, we observed the development of a shoulder at higher 2θ near the main Al_2O_3 Bragg peak, which corresponded to the compressive strain imprint in the substrate. The shoulder's intensity increased with increasing voltage, indicating that the strain imprint in the substrate spreads into a larger volume.

In the spatially resolved DFXM images, we observed small structural inhomogeneities in Al_2O_3 present without applying voltage to VO_2 (Fig. 4C), which is consistent with data in Fig. 1. To highlight the strain imprint effect in the substrate, we used the zero-voltage structural map to construct differential strain maps under applied voltage (Fig. 4D). Below the switching threshold (4 V), we observed a small tensile strain

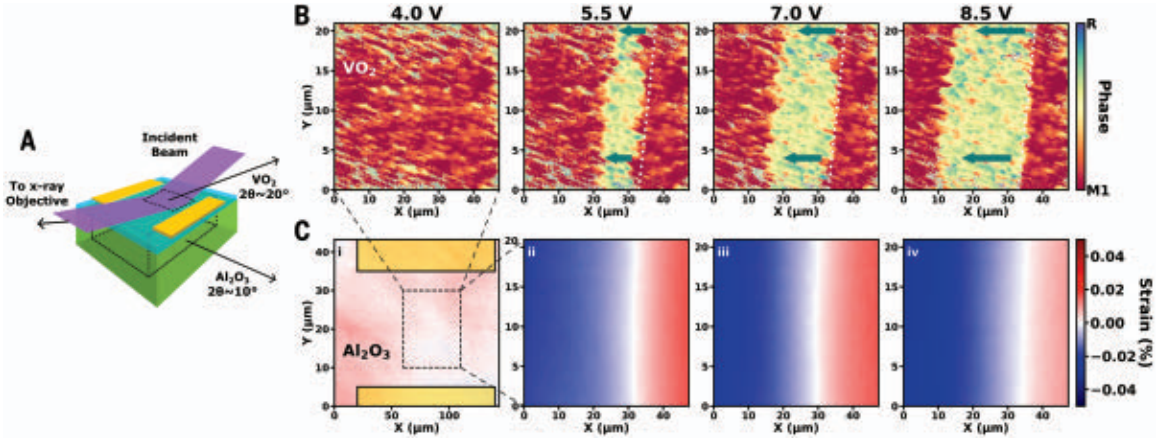


Fig. 3. Correlation between filament evolution in the film and substrate strain imprint. (**A**) Diagram showing the collocated VO_2 and Al_2O_3 DFXM imaging. (**B**) Monoclinic-rutile structural phase maps obtained using the VO_2 DFXM imaging (see supplementary materials). The images show a rutile phase filament formation above 4 V and its anisotropic expansion with increasing voltage. The filament expansion proceeds to the left (green arrows). The white dashed line is at a fixed location to show that the filament does not substantially expand to the right. (**C**) DFXM strain maps of Al_2O_3: (i) shows the strain induced by thermal expansion before the filament formation, and (ii) to (iv) show the asymmetric strain imprint. Comparing data in (B) and (C) reveals that the filament growth in VO_2 correlates with the compressive strain imprint in Al_2O_3.

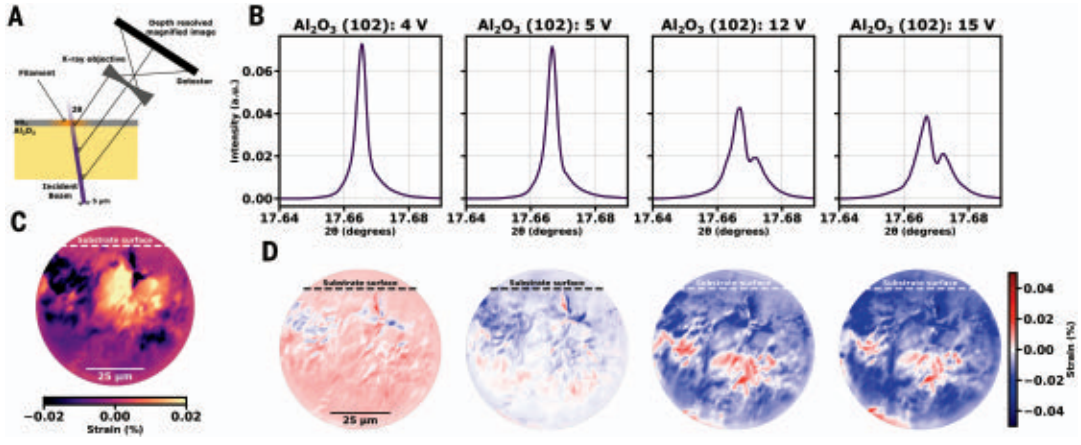


Fig. 4. DFXM imaging of the substrate strain imprint propagation depth. (**A**) Schematic of transmission DFXM. The sheet beam is shown here with the thickness of the beam ~5 μm wide in the scattering plane. The vertical beam size (out-of-the-page) matches the DFXM FOV of 75 μm. The device is oriented such that the filament extends out of the page. (**B**) Integrated diffraction scans showing the evolution of the (102) Al_2O_3 peak with voltage applied to the VO_2 device. As the voltage increases, the Bragg peak develops a shoulder at higher 2θ, indicating compressive strain imprint in the substrate. (**C**) Al_2O_3 DFXM strain map at 300 K and 0 V. The map reveals small structural inhomogeneities, consistent with data in Fig. 1. (**D**) Al_2O_3 differential strain maps at several voltages applied to VO_2 [referenced off the map in (C)]. Below the switching threshold of VO_2 (4 V), the tensile strain in the substrate can be attributed to thermal expansion. Above the switching voltage (≥5 V), the strain in Al_2O_3 becomes mainly compressive, and the strain propagation depth is tens of micrometers.

of ~0.015%, which could be attributed to the thermal expansion resulting from the heat exchange between the substrate and electrothermally heated VO_2. After the filament formed in VO_2, the strain imprint in Al_2O_3 became compressive, meaning that thermal effects can no longer account for the strain sign. We note that because the (102) Al_2O_3 peak is sensitive to both out-of-plane and in-plane strains, the strain magnitude in these measurements differs from the strain reported in Fig. 2 (0.05% versus 0.1%, respectively), suggesting that the strain imprint in the substrate was anisotropic. At voltages close to the switching threshold (5 V), the compressed region extended ~15 μm deep into the substrate. At higher voltages (15 V), the compressed region propagated at least 70 μm deep, extending beyond the DFXM field of view (FOV). The tens-of-micrometers strain imprint propagation depth in the substrate was especially notable considering that the thickness of the VO_2 film was only 300 nm and that Al_2O_3 is a highly rigid material.

Because the strain imprint deep within a substrate due to the localized phase transition in the overlaying thin film was unexpected, we conducted in situ bright-field transmission electron microscopy (TEM) imaging using a different electrical protocol to corroborate our x-ray microscopy results. A 150-nm-thick lamella was cut out of the VO_2/Al_2O_3 to make an electron-transparent sample for cross-sectional TEM. To trigger the switching in VO_2, we used voltage pulses with T_{on} = 0.05...0.9 μs width and T = 1 μs repetition rate (Fig. 5A) and 1.2 V amplitude (above the switching voltage of the TEM sample), which helped prevent premature device failures. Although the threshold voltages between the nanoscale TEM lamella and microscale devices used in the x-ray microscopy experiments differ, the power densities at the switching threshold in these devices were equivalent (5×10^{13} W/m^3). Collected TEM images were averages of the device pulse cycling over 100 ms (100,000 pulses), which further allowed us to investigate the cycle-to-cycle repeatability of the strain imprint effect. In TEM measurements, we tracked the motion of bending contours—the curved lines in TEM images in the Al_2O_3 and VO_2 regions. These contours arise from the electron beam interactions with the crystal and are highly sensitive to distortions and/or rotations of the lattice. We used the bending contour displacement as a direct signature of the mechanical response of Al_2O_3 to the electrically induced transition in VO_2.

TEM confirmed the strain imprint effect in the Al_2O_3 substrate by the VO_2 film. Figure 5B shows a TEM frame acquired using T_{on}/T = 5% voltage pulses. Under such conditions, the VO_2 film remained mostly insulating during the acquisition time. Variations of the TEM image intensity (bending contours) in the Al_2O_3 region indicated structural distortions in the substrate, consistent with the DFXM imaging at zero voltage (Figs. 1 and 4C). Increasing the pulse duty cycle to T_{on}/T = 90% made the VO_2 film stay predominately in the metallic phase. The TEM image acquired at T_{on}/T = 90% (Fig. 5C) shows that the bending contours in Al_2O_3 were displaced upward in comparison with the image acquired using T_{on}/T = 5% (Fig. 5B). The bending contour motion indicated a structural response in Al_2O_3 induced by switching the VO_2 (movie S1), in qualitative agreement with the x-ray microscopy experiments.

To highlight the switching-induced structural changes in TEM images, we plotted differential intensity maps (Fig. 5, D to F) referenced off the T_{on}/T = 5% image. At T_{on}/T = 10%, the differential map showed no contrast either in VO_2 or Al_2O_3, indicating that VO_2 remains in the insulating state and, consequently, no strain is imprinted into Al_2O_3. As T_{on}/T increased, the differential images began to show contrast (movie S2). The contrast in VO_2 film is due to the monoclinic-rutile structural transition that accompanies the electrically triggered IMT. The contrast in Al_2O_3 corresponded to the strain imprint in the substrate. Thus, the TEM measurements independently confirmed that electrical switching in VO_2 prompted structural changes in the underlaying Al_2O_3. The structural changes in Al_2O_3 occurred in the entire TEM FOV, which confirmed the DFXM observation that a thin VO_2 film imprinted strain into the Al_2O_3 substrate over many micrometers in depth, greatly exceeding the film thickness.

TEM measurements also demonstrated the high cycle-to-cycle repeatability of the strain imprint in the substrate. As discussed earlier, each TEM image represented the average over 100,000 voltage pulses. Because the features in the raw and differential TEM images were well defined, the strain imprint must have been the same every switching cycle (otherwise the images would be blurry). The increasing contrast with increasing pulse width in the differential TEM images also suggested the absence of structural fatigue in Al_2O_3 with repeated electrical cycling (otherwise the contrast would decrease over time). Overall, we found that the substrate strain imprint effect was highly robust, which may enable new practical approaches to couple multiple IMT devices through the substrate-mediated mechanical interactions.

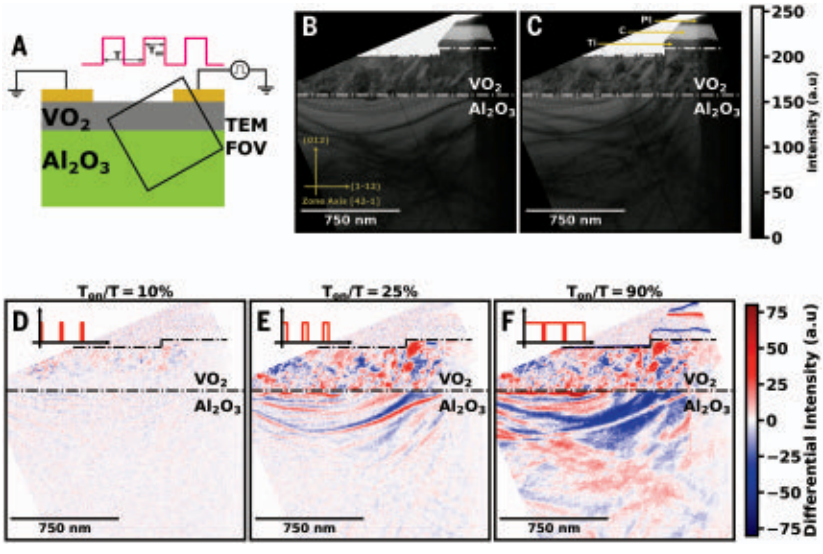


Fig. 5. TEM imaging of the substrate strain imprint. (**A**) Schematic showing the electrical pulse application setup and TEM FOV. (**B** and **C**) TEM images acquired using 1.2 V pulses (above threshold voltage for this sample) with T_{on}/T = 5% (B) and T_{on}/T = 90% (C). Inducing switching in VO_2 (C) changes the TEM image both in VO_2 and Al_2O_3 regions. (**D** to **F**) Differential intensity maps at T_{on}/T = 10% (D), T_{on}/T = 25% (E), and T_{on}/T = 90% (F) with respect to the T_{on}/T = 5% images. The respective pulse trains are schematically shown in the insets. These images show that the structural changes in Al_2O_3 induced by the switching in VO_2 can propagate micrometers deep into the substrate as the switching activation period increases. Black dot-dash lines indicate the boundaries between film and substrate (bottom) and film and electrode/vacuum (top).

Discussion

We demonstrated that the film-substrate interactions could produce structural modulations of the substrate. Without an applied voltage, subtle structural inhomogeneities in the Al_2O_3 substrate regions formed beneath the VO_2 film, and inhomogeneities were absent in the regions where the film was etched away. When an applied voltage triggered the filament formation in VO_2, through a local coupled structural and electronic phase transition, a large compressive-to-tensile asymmetric strain was locally imprinted in Al_2O_3. The strain imprint in Al_2O_3 was likely the consequence of accommodating the VO_2 lattice volume change associated with its phase transition. Furthermore, asymmetric substrate strain imprint fed back into the film and

defined the filament growth direction. The strain imprint asymmetry—the presence of both tensile and compressive strain regions—prevented structural breakage in the substrate or film, allowing for repeatable strain imprint over many switching cycles. The asymmetric strain imprint far exceeded thermal expansion, ±0.12% versus +0.04%, respectively. No strain asymmetry was observed during a uniform heating without an applied voltage, indicating a stark contrast between the electrically driven and the thermally driven transitions. A multiphysics model that includes coupled electrical, thermal, and elastic properties of the film-substrate system is needed to gain further insight into the basic phenomena governing the interactions between IMT films and supporting substrates.

Coincident electronic and structural transitions occur in many IMT systems (*19*, *39*–*42*). We observed the strain imprint effects in seven different samples on 14 different devices on two different substrates (Al_2O_3, Si), using three independent techniques (microdiffraction, reflection and transmission DFXM, and TEM). We expect the strain imprint effect to be an integral part of functional behaviors of IMT film-substrate systems and materials that undergo voltage-induced local structural transitions, such as ferroelectrics (*43*–*48*). Our observations allowed us to envision a path to advance x-ray imaging of functional devices by using the intense Bragg peaks of substrates as a proxy for probing dynamic structural changes in thin films.

The tens-of-micrometers propagation depth of the substrate strain imprint can open promising opportunities to establish electrically tunable elastic coupling between IMT switching devices in three-dimensional integrated architectures. For example, achieving efficient synchronization of spiking IMT-based oscillators is a highly active research area (*49*). Using the substrate strain imprint effect may allow for multilayer oscillator networks in which thin IMT layers alternate with thicker, substrate-like spacing layers that provide a mechanical coupling media to synchronize spiking in the adjacent IMT devices without direct electrical contact. This presents a possible model for an architecture that is closer to the human brain, potentially leading to the development of three-dimensional neuromorphic computing hardware (*50*).

REFERENCES AND NOTES

1. S. J. Callori *et al.*, *Phys. Rev. B* **91**, 140405 (2015).
2. G. Campillo *et al.*, *J. Magn. Magn. Mater.* **237**, 61–68 (2001).
3. R. Frisenda *et al.*, *NPJ 2D Mater. Appl.* **1**, 10 (2017).
4. R. Ghosh, D. Basak, S. Fujihara, *J. Appl. Phys.* **96**, 2689–2692 (2004).
5. Y. Bleu *et al.*, *Mater. Today Commun.* **35**, 105564 (2023).
6. S. Lee, I. N. Ivanov, J. K. Keum, H. N. Lee, *Sci. Rep.* **6**, 19621 (2016).
7. N. F. Quackenbush *et al.*, *Phys. Rev. B* **94**, 085105 (2016).
8. Y. Chen *et al.*, *Nature* **577**, 209–215 (2020).
9. S. Deng, A. V. Sumant, V. Berry, *Nano Today* **22**, 14–35 (2018).
10. W.-C. Tan, W.-H. Shih, Y. F. Chen, *Adv. Funct. Mater.* **24**, 6818–6825 (2014).
11. K. K. Sappati, S. Bhadra, *Sensors* **18**, 3605 (2018).
12. M. Liu *et al.*, *Appl. Phys. Lett.* **104**, 121905 (2014).
13. M. K. Liu *et al.*, *Phys. Rev. Lett.* **111**, 096602 (2013).
14. Y. Kalcheim *et al.*, *Adv. Funct. Mater.* **30**, 2005939 (2020).
15. A. Pofelski *et al.*, *Phys. Rev. Mater.* **8**, 035003 (2024).
16. J. Del Valle *et al.*, *Nature* **569**, 388–392 (2019).
17. J. Del Valle *et al.*, *Science* **373**, 907–911 (2021).
18. M. M. Qazilbash *et al.*, *Science* **318**, 1750–1753 (2007).
19. T. Kawakubo, T. Nakagawa, *J. Phys. Soc. Jpn.* **19**, 517–519 (1964).
20. G. Stefanovich, A. Pergament, D. Stefanovich, *J. Phys. Condens. Matter* **12**, 8837–8845 (2000).
21. A. Zimmers *et al.*, *Phys. Rev. Lett.* **110**, 056601 (2013).
22. P. Salev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2317944121 (2024).
23. S. Kumar *et al.*, *Adv. Mater.* **26**, 7505–7509 (2014).
24. X. Liu, S. Li, S. K. Nandi, D. K. Venkatachalam, R. G. Elliman, *J. Appl. Phys.* **120**, 124102 (2016).
25. P.-H. Huang, R. Coffee, L. Dresselhaus-Marais, *Integr. Mater. Manuf. Innov.* **12**, 83–91 (2023).
26. H. Simons *et al.*, *Nat. Commun.* **6**, 6098 (2015).
27. E. Kisiel *et al.*, *ACS Nano* **19**, 15385–15394 (2025).
28. Z. Qiao *et al.*, *Rev. Sci. Instrum.* **91**, 113703 (2020).
29. A. C. Jakobsen *et al.*, *J. Appl. Cryst.* **52**, 122–132 (2019).
30. J. Plumb *et al.*, *Phys. Rev. Mater.* **8**, 093601 (2024).
31. S. Borgi *et al.*, *J. Appl. Cryst.* **57**, 358–368 (2024).
32. M. P. Smylie *et al.*, *Phys. Rev. B* **110**, 024509 (2024).
33. G. Stone *et al.*, *Adv. Mater.* **36**, e2312673 (2024).
34. D. Lee *et al.*, *Science* **362**, 1037–1040 (2018).
35. J. H. Park *et al.*, *Nature* **500**, 431–434 (2013).
36. D. Lee *et al.*, *Nano Lett.* **17**, 5614–5619 (2017).
37. M. K. Sohn *et al.*, *Appl. Phys. Lett.* **120**, 173503 (2022).
38. Y. Arata, H. Nishinaka, M. Takeda, K. Kanegae, M. Yoshimoto, *ACS Omega* **7**, 41768–41774 (2022).
39. J. A. Alonso *et al.*, *Phys. Rev. Lett.* **82**, 3871–3874 (1999).
40. T. Shao *et al.*, *Appl. Surf. Sci.* **399**, 346–350 (2017).
41. J. A. Alonso, M. J. Martínez-Lope, M. T. Casais, M. A. G. Aranda, M. T. Fernández-Díaz, *J. Am. Chem. Soc.* **121**, 4754–4762 (1999).
42. M. M. Gomes *et al.*, *Phys. Rev. B* **109**, 224109 (2024).
43. M. Lange *et al.*, *Phys. Rev. Appl.* **16**, 054027 (2021).
44. T. Luibrand *et al.*, *Phys. Rev. Res.* **5**, 013108 (2023).
45. H. Madan, M. Jerry, A. Pogrebnyakov, T. Mayer, S. Datta, *ACS Nano* **9**, 2009–2017 (2015).
46. A. Milloch *et al.*, *Nat. Commun.* **15**, 9414 (2024).
47. M.-H. Zhang *et al.*, *Acta Mater.* **200**, 127–135 (2020).
48. I. V. Ciuchi *et al.*, *J. Eur. Ceram. Soc.* **37**, 4631–4636 (2017).
49. E. Qiu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2303765120 (2023).
50. S. J. Kim, H.-J. Lee, C.-H. Lee, H. W. Jang, *NPJ 2D Mater. Appl.* **8**, 70 (2024).
51. E. Kisiel *et al.*, Additional datasets and analysis scripts for: Dynamic asymmetric strain imprinted into the substrate from an oxide thin film, Dryad (2026); https://doi.org/10.5061/dryad.2rbnzs7zq.

ACKNOWLEDGMENTS

We thank the Karlsruhe Nano Micro Facility (KNMFi), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT), for the fabrication of the polymer x-ray optics. We also thank the five reviewers for their critical assessment. **Funding:** This work was supported as part of the "Quantum Materials for Energy Efficient Neuromorphic Computing" (Q-MEEN-C), an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES) under award DESC0019273. This research was performed on Advanced Photon Source (APS) beam time award (DOI: 10.46936/APS-193721/60016636) from the Advanced Photon Source, a US DOE Office of Science user facility, and is based on work supported by Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the US DOE under contract DE-AC02-06CH11357. The electron microscopy facility and the development of frequency-tunable radio frequency imaging capabilities at the Brookhaven National Laboratory were supported by DOE-BES, Materials Science and Engineering Division, under contract DE-SC0012704. The research by A.G. was supported by the US DOE, Office of Science, BES, Materials Sciences and Engineering Division. K.A.Y. was supported by the US DOE, Office of Science, BES, under contract DE-AC02-76SF00515. **Author contributions:** Conceptualization: E.K., I.K.S., A.F., Z.I., P.S.; Methodology: E.K., P.S., I.P., A.P., Y.Z., Z.I., A.G., A.L.; Investigation: E.K., P.S., I.P., A.P., C.L., S.R., W.H., K.A.Y., J. L., R.B., Z.Z., Z.I., A.G., E.Q., U.P.; Visualization: E.K., P.S., A.F.; Funding acquisition: I.K.S., A.F., Y.Z., A.G., Z.I.; Project administration: I.K.S., A.F., Y.Z., Z.I.; Supervision: A.F., I.K.S., Y.Z., Z.I.; Writing – original draft: E.K., A.F., P.S., I.K.S., Z.I.; Writing – review & editing: E.K., A.F., P.S., Z.I., A.P., C.L., S.R., A.L., A.G. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data are available in the main text or the supplementary materials. Additional datasets and analysis scripts are available in Dryad (*51*).

License information: Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt9347
Materials and Methods; Figs. S1 to S18; Table S1; References (*52*–*58*); Movies S1 and S2

Submitted 22 October 2024; resubmitted 5 May 2026; accepted 3 June 2026;
published online 18 June 2026

10.1126/science.adt9347

COMPARATIVE PHYSIOLOGY

Crush-resistant air sacs allow insect larvae to exploit aquatic habitats at extreme depth

Evan K. G. McKenzie[1], Maxon J. R. Ngochera[2], Joseph Chombo[2], Hayley McLennan[3], Roland Proud[3,4], Tahnee Ames[1], Philip G. D. Matthews[1]*

The absence of aquatic insects from pelagic marine habitats has been attributed to their air-filled tracheal respiratory system, which could implode at depth while undertaking diel vertical migrations to escape predatory fish. However, we found that the aquatic larvae of the lake fly *Chaoborus edulis* in Lake Malawi have modified their tracheal system into reinforced buoyancy-regulating air-filled sacs, allowing them to escape predatory fish by undertaking these migrations well over 200 meters deep into the lake's anoxic hypolimnion during the day. The crush depth of their air sacs increases with each instar, with final instars resisting implosion at depths greater than half a kilometer. Contrary to expectations, these insects adapt to the extreme hydrostatic pressures of deep-water habitats, allowing them to coexist with pelagic fish.

Insects represent the most successful expansion of a terrestrial animal lineage into the aquatic environment, constituting more than 60% of all freshwater aquatic animal species (*1*). However, despite their success in both freshwater and land-locked hypersaline habitats, no known aquatic insects live underwater in the open ocean. In this environment insects must contend with the presence of fish—voracious predators of pelagic invertebrates. Zooplankton survive in pelagic environments by performing diel vertical migrations (DVMs): They hide in deep, dark water during the day and only swim up from the depths at night to feed under cover of darkness (*2*). However, because aquatic insects retain the same air-filled tracheal respiratory system as their terrestrial ancestors, it has been hypothesized that they cannot dive deeply enough to escape predatory fish during the day, as they would die from tracheal collapse before reaching the safety of twilight zone depths (*3*). The aquatic larvae of the lake fly *Chaoborus edulis*, which thrive in one of the deepest lakes in the world (Lake Malawi/Nyasa/Niassa), challenge this assumption.

The larvae of *C. edulis*, like those of all *Chaoborus* species, exploit the open water of lakes by modifying their tracheal system into two pairs of air-filled sacs that function as hydrostatic organs (*4*). Although derived from the insect's respiratory system, the closed air sacs play no respiratory role and *Chaoborus* larvae rely entirely on cutaneous respiration to breathe underwater (*4*). The air sacs, however, employ a distinct mechanochemical system for buoyancy control whereby bands of the protein resilin, laminated between bands of hard cuticle in the air sac wall, are driven to expand and contract in response to changes in pH generated by the air sac's epithelium (*5*). Each air sac transduces pH change into mechanical work to regulate its volume against hydrostatic pressure. As the wall of the air sac is gas permeable (*4*, *6*), gases within it exist in equilibrium with gases dissolved in the surrounding water. In air-equilibrated water, this pressure equals the atmospheric pressure at the water's surface, increasing slightly with depth in the same manner as pressure varies within a vertical column of gas (*7*). The air sacs are neither supported nor inflated by internal gas pressure. Rather, they must be sufficiently rigid to resist the total hydrostatic pressure at depth without collapsing and be able to expand their resilin bands against this pressure.

C. edulis larvae are extraordinarily abundant in Lake Malawi (*8*), despite it being deep, clear, and inhabited by a diverse pelagic fish fauna (*9*, *10*). They survive in this challenging aquatic environment by undertaking deep migrations daily (*11*). Plankton trawls revealed that the larvae dive progressively deeper as they develop through four larval instars. They descend more than 200 m into Lake Malawi's permanently anoxic hypolimnion each morning to escape predatory fish before returning to shallower water at night to hunt zooplankton (*12*). However, the depth of their daily descent and the maximum depth they can withstand were previously unknown. Our objectives were to (i) determine the depth of *C. edulis*' DVMs in Lake Malawi and relate their daily migration to the distribution of fish and the depth of the hypolimnion; (ii) determine the capacity of *C. edulis* air sacs to withstand failure and regulate buoyancy at depth; and (iii) compare these mechanical parameters with those measured from the air sacs of four additional *Chaoborus* species: *C. anomalus* and *C. pallidipes* from Lake Victoria, Kenya, an 80-m-deep, nonstratified lake; and *C. americanus* and *C. trivittatus* from Shirley Lake, BC, Canada, a fish-free lake 8 m deep.

Hydroacoustic monitoring of *C. edulis* DVM and pelagic fish

Chaoborus larvae can be monitored using hydroacoustics because their air sacs scatter sound waves (*13*, *14*) (fig. S2). To quantify the pattern of daily migrations of *C. edulis* in Lake Malawi, we deployed a multifrequency sonar system with an upward-facing transducer suspended 20 m above the lake floor in water 295 m deep, approximately 1.6 km offshore from Nkhata Bay, Malawi (Fig. 1, A and B). A 250-m vertical transect of water quality parameters was recorded at the same location, revealing a sharp decline in dissolved oxygen (DO) levels from 6.17 to 0.18 mg/L between 150 and 200 m, coincident with a decrease in temperature of 0.8°C and pH, from 8.15 to 7.37 (Fig. 1D).

Layers of *C. edulis* descending into Lake Malawi's hypolimnion became detectable (i.e., above background noise levels) between 02:22 and 11:13, sinking at a mean rate ± SD of 7.56 ± 6.48 m h^{-1}. The maximum depth recorded for any *C. edulis* layer was 258.6 m, and the overall weighted mean depth (depth weighted by echo intensity, a proxy for *C. edulis* density) of vertically static layers, which all occurred during daytime, was 213.7 ± 13.6 m. The mean shallowest depth of any static layer was 196.8 m. The ascent started between 13:35 and 16:20, and layers became undetectable between 17:09 and 18:29 at the end of their dusk ascent. The mean ascent rate was 19.8 ± 15.2 m h^{-1}. The upward migration started slowly with a mean ascent rate of 15.48 ± 6.48 m h^{-1}, increasing to 23.76 ± 19.8 m h^{-1} (Fig. 1D). These measured rates exceeded the maximum rates of vertical movement recorded from passively sinking and floating larvae in a pressure chamber exposed to hydrostatic pressures between 0 to 282 m depth (−5.54 and 7.23 m h^{-1}, respectively Fig. 1C), suggesting that the migration involved some active swimming. Larvae were observed swimming in short bursts lasting 0.75 ± 0.8 s at an average speed of 105.8 ± 51 m h^{-1} (±SD). Thus, intermittent swimming with passive floating/sinking can adequately account for vertical movement rates measured hydroacoustically. However, as fish easily swim an order of magnitude faster than this (*15*), swimming is clearly no defense against predation.

A total of 394,644 individual fish tracks were detected over the survey area, with only 1605 (0.4%) having a mean depth >196.8 m, the shallowest mean depth of static *C. edulis* layers observed during the day. Fish detections had a mean depth of 91.5 ± 51.1 m, with 99% of all fish detections occurring <189.9 m, where the water had a DO of 0.41 mg L^{-1} (Fig. 1D).

[1]Department of Zoology, University of British Columbia, Vancouver, BC, Canada. [2]Department of Fisheries, Ministry of Natural Resources and Climate Change, Lilongwe, Malawi. [3]Pelagic Ecology Research Group, School of Biology, Scottish Oceans Institute, Gatty Marine Laboratory, University of St Andrews, Fife, UK. [4]Cupar Analytics Ltd, Cupar, UK. ***Corresponding author. Email: pmatthews@zoology.ubc.ca**

Air sac morphology

Chaoborus larval air sacs are curved fusiform tubes that vary substantially in geometry between species and across development (*16*) (Fig. 2). The dimensions of fourth-instar air sacs were measured from profile images of excised air sacs placed in pH 6 buffer (the in vivo pH of neutrally buoyant *C. americanus* air sacs at zero depth) (*5*). Maximum *C. edulis* air sac diameter was 0.223 mm, and the maximum air sac length was 1.099 mm (both measurements from anterior air sacs). Volumes of *C. edulis* air sacs under the same conditions as above were estimated by modeling them as truncated cones capped on either end by prolate ellipsoid hemispheres (fig. S5). The mean volume of an anterior air sac ± SD was 19.5 ± 1.7 nL (n = 7) and the mean volume of a posterior air sac was 11.6 ± 2.3 nL (n = 21). The estimated total air sac volume of 62.2 nL displaced sufficient water to bring a fourth-instar *C. edulis* larva to, or very near, neutral buoyancy.

Diameters of fourth-instar posterior air sacs varied substantially between species. The mean diameter ± SD of *C. edulis* air sacs (n = 52) was 0.163 ± 0.014 mm, which was not significantly different from the 0.165 ± 0.005 mm diameter of *C. anomalus* air sacs (n = 13), whereas all other differences were significant (Fig. 2). *C. pallidipes* air sacs, at 0.195 ± 0.004 mm (n = 4), were slightly wider. The air sacs of *C. americanus* (n = 26) and *C. trivittatus* (n = 17) were significantly wider than all others, with means of 0.271 ± 0.022 mm and 0.307 ± 0.017 mm (Fig. 2).

Crush depth of *Chaoborus* larvae and pupae

Intact larvae (second to fourth instar) and pupae (*C. edulis* and *C. pallidipes* only) were placed in a water-filled pressure chamber (fig. S6) and exposed to stepwise increases in pressure. The pressure increment at which the first air sac or pupal respiratory horn (a pair of static air-filled structures used for buoyancy and respiration) imploded indicated depth of failure. For *C. edulis*, mean crush depth ± SD increased across development, from 242 ± 24 m in the second instar, to 324 ± 39 m in the third instar, and 463 ± 75 m in the fourth instar (Fig. 2). There was no significant difference between the crush depths of fourth-instar larvae and pupae (Fig. 2). In comparison, the crush depths of fourth instar *C. anomalus* and *C. pallidipes* from Lake Victoria (~80 m deep) were far shallower (97 ± 27 and 114 ± 39 m, respectively), and those of *C. americanus* and *C. trivittatus* from Shirley Lake (~8 m deep) were shallower still (16 ± 2 and 22 ± 2 m, respectively).

From fourth-instar data, we calculated air sac safety factors as the ratio of mean crush depth to maximum depth occupied, a metric which indicates the air sac's capacity to withstand hydrostatic pressures beyond those routinely encountered. Fourth-instar *C. americanus* and *C. trivittatus* have been recorded at maximum depths of 10 and 20 m, respectively, in a 42-m-deep, fish-free lake adjacent to Shirley Lake (*17*), giving a safety factor of 1.6 for *C. americanus* and 1.1 for *C. trivittatus*. The maximum depth of Lake Victoria was taken to be the maximum depth occupied for both *C. anomalus* and *C. pallidipes*, giving safety factors of 1.2 and 1.4, respectively. Assuming the maximum detected *C. edulis* layer depth of 258.6 m represents the voluntary dive limit of this species, their air sac safety factor was 1.8.

Air sac compression at depth

Although air sacs vary in shape, the structure of the air sac wall is consistent, comprising alternating bands of resilin and stiff cuticle arranged transverse to the air sac's long axis (Fig. 3, S8) (*5*). This arrangement means that air sacs change volume uniaxially, varying in length but not diameter (*5*). To quantify their mechanical properties at depth, freshly excised air sacs from fourth-instar larvae of *C. edulis*, *C. pallidipes*, *C. anomalus*, and *C. americanus* were exposed to stepwise increases in hydrostatic pressure in the same chamber used above, but filled with insect physiological saline, again buffered to pH 6. Air sacs were contained within shallow wells attached to the inside surface of the pressure chamber's clear acrylic window, allowing imaging from above. The profile area (mm^2) of each air sac was measured from images taken following each pressure increase until failure (movies S1 to S3). Because the air sac's wall constrains profile area to change in direct proportion with long-axis length (*5*), percent area change was plotted against pressure as percent length change to describe the air sacs' lengthwise stiffness in units of MPa. However, as air sacs are hollow and not uniform in cross

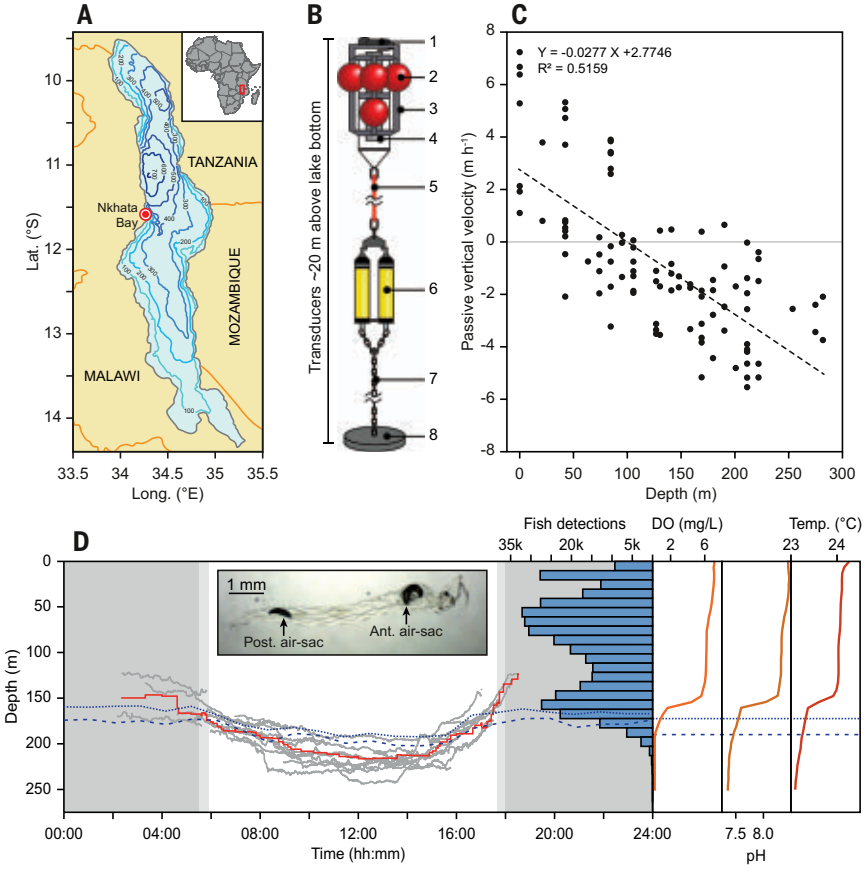


Fig. 1. *Chaoborus edulis* DVM in Lake Malawi. (**A**) Bathymetric map of Lake Malawi showing study location; (inset) location in East Africa. (**B**) Quad-frequency Acoustic Zooplankton Fish Profiler (AZFP): 1: 67, 120, 200, and 455 kHz transducers; 2: 6× ABS floats; 3: Cage; 4: AZFP; 5: 12.3 m rope; 6: Acoustic releases; 7: 3-m drop chain; 8: 120 kg iron ballast. (**C**) Passive vertical velocities of larvae as a function of depth. (**D**) The deepest portion of the DVM of *C. edulis* and its relationship to the vertical distribution of fish and water parameters. Night, twilight, and day indicated by dark gray, light gray, and white panels, respectively. Dark gray lines show mean depth of all detected *C. edulis* layers and the red line shows the overall mean. Blue lines show depth of 95th (dotted) and 99th (dashed) percentiles of fish detections. Histogram shows the distribution of all fish detections in 10-m depth bins. Panels to the right show dissolved oxygen (DO), pH, and water temperature measured at the sonar deployment site. Blue lines show depth of 95th and 99th percentiles of all fish detections over deployment period. (Inset) A fourth-instar *C. edulis* larva, with posterior and anterior air sacs indicated.

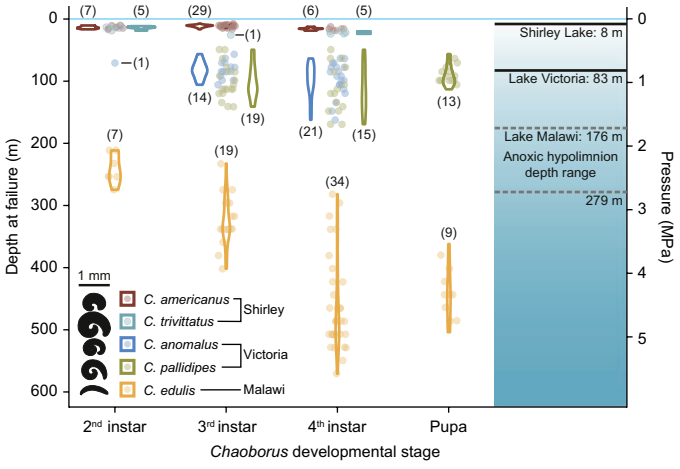


Fig. 2. Implosion depths of larval air sacs and pupal respiratory horns in vivo. Five species of *Chaoborus* were exposed to stepwise increases in hydrostatic pressure until the first air sac or respiratory horn imploded. Pupae were only available for *C. edulis* and *C. pallidipes*. Their habitat depth (from shallowest to deepest) was as follows: *C. americanus* (red) and *C. trivittatus* (teal), both from Shirley Lake; *C. anomalus* (blue) and *C. pallidipes* (green), both from Lake Victoria; and *C. edulis* from Lake Malawi (yellow). Sample size is indicated in brackets. The depth scale on the right indicates maximum lake depth or the depth range of the anoxic hypolimnion in Lake Malawi (*22*). *C. edulis*, *C. anomalus*, and *C. trivittatus* all showed significant positive trends of depth tolerance with instar [analysis of variance (ANOVA): $F = 106.0$, $P < 0.0001$; $F = 4.8$, $P = 0.036$; and $F = 12.4$, $P = 0.006$, respectively]. Through larval development, *C. edulis* gained 119.02 ± 11.56 m depth tolerance instar^{-1} ($F = 106$, $P < 0.0001$). Depth of failure for *C. edulis* fourth-instar larvae and pupae were not significantly different (Welch Two Sample *t* test: $t = 1.2$, $df = 28.0$, $p = 0.2$) but in *C. pallidipes* pupae failed at shallower depths than fourth-instar larvae (Welch Two Sample *t* test: $t = 2.3$, $df = 16.9$, $p = 0.035$). Air sac diameters were compared using a Tukey adjusted pairwise comparison of estimated marginal means, with the only nonsignificant difference being between *C. edulis* and *C. anomalus* ($t = 0.7$, $P > 0.9$). Air sac silhouettes show a fourth-instar posterior air sac from each species to scale and in vivo orientation.

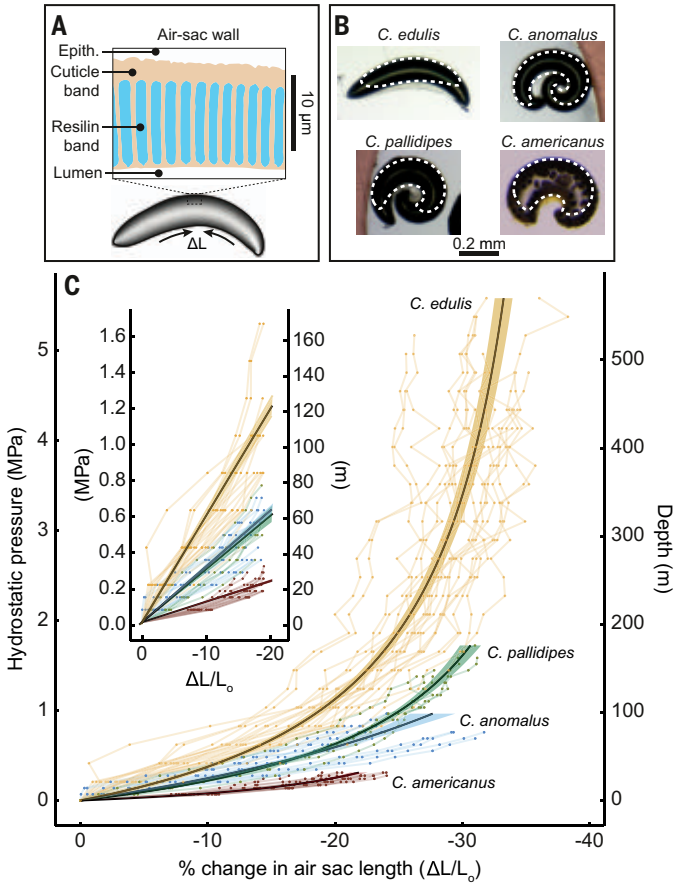


Fig. 3. Air sac compression and stiffness under hydrostatic pressure. (**A**) Transmission electron microscopy of longitudinal section through fourth-instar *C. edulis* air sac wall showing arrangement of alternating resilin and cuticle bands, air sac epithelium (epith.), and direction of air sac compression. (**B**) Photos of *Chaoborus* air sacs at atmospheric pressure with dashed outlines indicating degree of compression before failure. (**C**) The change in excised air sac length in pH 6 buffer with increasing hydrostatic pressure indicates compressive stiffness. Bands show 95% confidence intervals (CIs). Air sac length was normalized to the initial unpressurized length. (Inset) Initial, linear phase of each air sac stiffness curve. A Tukey post-hoc comparison of estimated marginal means was used to compare stiffness between species. Only *C. anomalus* ($n = 13$) and *C. pallidipes* ($n = 6$) were not significantly different ($t = 0.32$, $P = 0.99$); *C. edulis* ($n = 21$) and *C. americanus* ($n = 18$) had stiffnesses significantly different from all other species ($t \geq 4.7$, $P \leq 0.016$).

section, this relationship is not a true compressive modulus, but rather a measure of the whole air sac structure's stiffness (Fig. 3).

The compression curves of the air sacs from these four species are J-shaped, with approximately linear initial decreases in air sac length and increasing stiffness at higher pressures (Fig. 3C). The initial stiffness of the four species' air sacs corresponded with the maximum depth of their environment. Thus, air sacs of *C. edulis* from Lake Malawi had a modeled stiffness and 95% confidence interval (CI) of 5.17 ± 0.53 MPa, followed by the two species from Lake Victoria, *C. anomalus* and *C. pallidipes*, being 2.63 ± 0.59 and 2.80 ± 0.84 MPa, respectively. Finally, *C. americanus*, from Shirley Lake, had the most compliant air sacs (1.17 ± 0.53 MPa). Between species, only the co-occurring *C. anomalus* and *C. pallidipes* had air sacs that did not differ substantially in their stiffness (Fig. 3).

With increased pressure, air sacs either imploded (*C. americanus*) or increased in stiffness, likely due to compression of the cuticular bands against one another before failure (movie S3). This increase was most notable in *C. edulis* air sacs, which had a stiffness ± SD before failure of 76.36 ± 22.82 MPa, nearly a 15-fold increase over their initial stiffness. In comparison with the other species, the maximum stiffness of *C. edulis* air sacs was five times greater than those of *C. pallidipes* (14.69 ± 9.27 MPa), 18 times greater than those of *C. anomalus* (3.86 ± 1.03 MPa), and 24 times greater than those of *C. americanus* (3.21 ± 0.78 MPa), following the trend of dive depth between species. Despite these large differences in stiffness, the degree of linear compression at failure was similar for all species: between 0.21 ± 0.03 (*C. americanus*) to 0.33 ± 0.03 (*C. edulis*) (Fig. 3).

Air sacs and mechanical work

The osmotic work done by resilin expanding the air sac wall following a pH increase is approximately equal and opposite to the mechanical work performed by an increase in hydrostatic pressure compressing the air sac wall as, neglecting losses, the free energy change required to chemically induce an expansion of the air sac wall should be equivalent to the mechanical energy associated with its recompression (*18*). Manipulating pH and pressure while measuring change in air sac length indicated the air sac's capacity to convert pH changes into mechanical work. Air sacs from fourth-instar *C. edulis* larvae were placed in pH 6 buffer and secured inside the pressure chamber. They were then exposed to two 1-unit increases in pH, up to pH 8. Stepwise pressure increases were then applied while the air sac profile area was recorded. We found that the hydrostatic pressure at a depth of 52.1 ± 4.74 m (±95% CI) was sufficient to compress the air sacs back to their unpressurized length at pH 6 (Fig. 4A). Thus, increasing the hydrostatic pressure by 509.42 kPa did the same amount of work

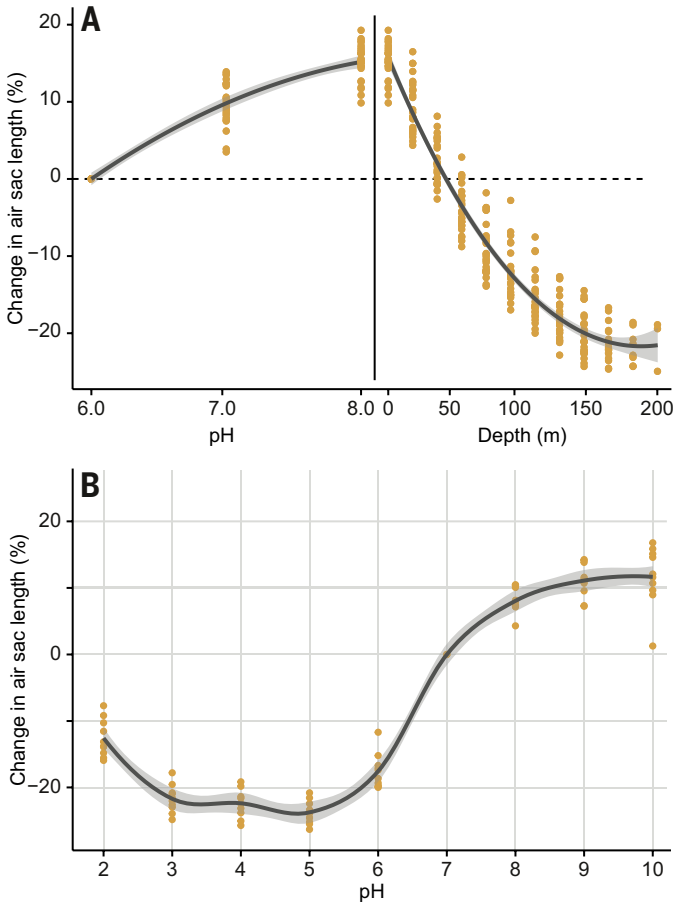


Fig. 4. Relationship between *C. edulis* air sac length, pH, and hydrostatic pressure. (**A**) Expansion of unpressurized air sacs following an increase in pH from 6 to 8 (LHS), and the mechanical work done in compressing the air sacs at pH 8 by increasing hydrostatic pressure (RHS) ($n = 11$). (**B**) Percent change in length of excised *C. edulis* air sacs in relation to pH of the bathing medium ($n = 11$) (measured from change in profile area and normalized to pH 7). The greatest change occurs between pH 6 and 8. Bands show 95% CI.

compressing the air sacs as they could perform expanding against this pressure following a 6 to 8 pH increase. Factoring in the average cross sectional area of the air sac in the direction of expansion, we calculated this work (±95% CI) to be 1.051 ± 0.083 mJ m^{-1}, i.e., work per meter of initial air sac length. The results of this experiment suggest that *C. edulis* can only regulate their air sac volume and buoyancy at depths less than ~50 m. However, this species may have the capacity to vary the pH of their air sac wall over a range greater than 6 to 8 to do work at greater depths. But as air sac volume change was greatest within this pH range (Fig. 4B), only a minor gain in volume control would be achieved by going outside of these limits.

Conclusions

Our research suggests that the tracheal system of *C. edulis* has evolved into a robust hydrostatic organ that enables exploitation of a deep-water pelagic environment. This adaptation, coupled with the unusual ability of *Chaoborus* larvae to respire anaerobically by converting stores of malate to succinate (*19*), are key to their survival in Lake Malawi, allowing them to feed on zooplankton in surface waters at night before diving into the persistently anoxic hypolimnion to escape predatory fish during the day. This predation pressure explains why *C. edulis* larvae undertake a DVM that is likely the deepest of any freshwater invertebrate, more than three times deeper than that performed by the lake's crustacean zooplankton (*12*), and tolerate ~10 hours each day in near total anoxia. Although unable to pursue the larvae into the anoxic hypolimnion, fish congregate between 140 to 180 m, within the hypoxic thermocline just above this region (Fig. 1D). Several deep-water cichlid genera in Lake Malawi possess hemoglobin mutations associated with enhanced O_2 binding (*10*, *20*), suggesting strong selection pressure to enable these fish to pursue and feed upon the high concentrations of *Chaoborus* larvae as they enter and leave their anoxic refuge. Stomach contents of pelagic cichlids confirm that they feed on *C. edulis* at dawn and dusk, while also revealing that fourth-instar larvae and pupae are their principal prey (*21*). This explains why daytime plankton trawls within the lake's upper 150 m collected second and third instars, whereas fourth instars and pupae were rarely encountered at these depths (*11*). Thus, although pre–fourth-instar larvae remain above the hypolimnion, they still increase their air sac crush depth as they develop (Fig. 2), ensuring that when they become most vulnerable to predation as fourth instars, they are equipped to escape into the anoxic depths.

The ability of *Chaoborus* larvae to undertake a DVM requires that their air sacs retain some capacity to regulate their volume, and thus the larvae's buoyancy, while also resisting implosion at depth. As we show here, the ability of *C. edulis* resilin to convert a change in pH into mechanical work is limited: The hydrostatic pressure at 50 m depth would prevent it from expanding when exposed to a biologically reasonable change in pH from 6 to 8. Any deeper, and the air sac volume changes little as the resilin bands become fully compressed and the cuticular bands abut one another, increasing the stiffness of the air sac wall and allowing maintenance of approximately constant buoyancy down to ~200 m (Figs. 3 and 4). Even when compressed to minimum volume, air sacs still substantially reduce the larvae's density, slowing their sinking rate and reducing the energetic cost of holding position at depth. We propose that *C. edulis* larvae use a hybrid buoyancy control strategy across their DVM: regulating buoyancy in shallow water where pH-driven control of air sac volume is possible, then relying on air sac rigidity to reduce body density at depths beyond 50 m.

The mean air sac and respiratory horn crush depths of the four *Chaoborus* species from lakes without permanent anoxic layers closely correspond to the maximum depth of the lake where they live, indicating that these species can find refuge from fish in the lake's sediment. However, both the fourth-instar larvae and pupae of *C. edulis* have crush depths substantially below the maximum recorded depth of the anoxic hypolimnion (Fig. 2; ~279 m) (*22*), indicating that this species could dive far deeper than it does. The safety factor of 1.8 calculated for this species is also the highest measured here, while the other four species have a mean safety factor of 1.3 ± 0.3. The only animals with rigid air-filled hydrostatic organs comparable to *Chaoborus* air sacs are marine cephalopods. Curiously, *Sepia*, *Nautilus*, and *Spirula* also have safety factors of ~1.3 despite diving to very different maximum depths (150, 700, and 1750 m, respectively) (*23*–*25*).

The extreme hydrostatic pressures *C. edulis* experience have driven adaptive changes to their hydrostatic organ morphology. Compared with all other species, their air sacs are narrow and show the least axial curvature (Fig. 2), while the respiratory horns of the pupa are almost spherical, rather than spindle-shaped as in other *Chaoborus* species (*26*). These geometric changes reduce air sac wall stress (*27*) and appear to have evolved comparatively recently, as Lake Malawi filled to its current depth less than 75,000 years ago (*28*).

Our research offers new insights into the depth limits of free-living aquatic insects. Although seal lice retain the record as the deepest-diving insects, they are essentially air-breathing insects with typical tracheal systems that endure intermittent deep submergence while immobile and clinging to their mammal hosts (*29*, *30*). By contrast, *C. edulis* larvae swim, feed, grow, and pupate in open water before surfacing to eclose as adult midges. By completing their entire life cycle

in and over deep water, they prove that depth per se does not exclude insects from pelagic habitats. Their success further indicates that a key requirement for insect survival in open water is a fish-free refuge. Freshwater lakes are not the only bodies of water with suitable conditions. For example, if *Chaoborus* could tolerate the osmotic challenges of seawater, they could conceivably survive in the Black Sea, which is persistently anoxic below 125 to 200 m over most of the basin (*31*). The ability of *Chaoborus* larvae and pupae to live in deep freshwater lakes indicates that physiological and ecological factors other than depth must be considered to explain the absence of insects from open-water marine environments.

REFERENCES AND NOTES

1. K.-D. B. Dijkstra, M. T. Monaghan, S. U. Pauls, *Annu. Rev. Entomol.* **59**, 143–163 (2014).
2. G. C. Hays, *Hydrobiologia* **503**, 163–170 (2003).
3. S. H. Maddrell, *J. Exp. Biol.* **201**, 2461–2464 (1998).
4. A. Krogh, *Skand. Arch. Physiol.* **25**, 183–203 (1911).
5. E. K. G. McKenzie, G. T. Kwan, M. Tresguerres, P. G. D. Matthews, *Curr. Biol.* **32**, 927–933.e5 (2022).
6. S. Teraguchi, *J. Insect Physiol.* **21**, 1659–1670 (1975).
7. W. O. Fenn, *Science* **176**, 1011–1012 (1972).
8. K. Irvine, *The Fishery Potential and Productivity of the Pelagic Zone of Lake Malawi/Niassa*, A. Menz, Ed. (Natural Resources Institute, 1995), chap. 4, pp. 109–140.
9. G. F. Turner, R. L. Robinson, B. P. Ngatunga, P. W. Shaw, G. R. Carvalho*Management and Ecology of Lake and Reservoir Fisheries*, I. G. Cowx, Ed., 2002), chap. 29, pp. 353–366.
10. C. Hahn, M. J. Genner, G. F. Turner, D. A. Joyce, *Evol. Lett.* **1**, 184–198 (2017).
11. K. Irvine, *Freshw. Biol.* **37**, 605–620 (1997).
12. K. Irvine, *The Fishery Potential and Productivity of the Pelagic Zone of Lake Malawi/Niassa*, A. Menz, Ed. (Natural Resources Institute, 1995), chap. 3, pp. 85–108.
13. T. G. Northcote, *Limnol. Oceanogr.* **9**, 87–91 (1964).
14. F. R. Knudsen, P. Larsson, P. J. Jakobsen, *Fish. Res.* **79**, 84–89 (2006).
15. R. Bainbridge, *J. Exp. Biol.* **35**, 109–133 (1958).
16. K. S. Bardenfleth, R. Ege, *Videnskabelige Meddelelser fra Dansk Naturhistorisk Forehing* **67**, 25–42 (1916).
17. A. Y. Fedorenko, M. C. Swift, *Limnol. Oceanogr.* **17**, 721–730 (1972).
18. W. Kuhn, A. Ramel, D. H. Walters, G. Ebner, H. J. Kuhn, *Fortschritte Der Hochpolymeren-Forschung* (Springer, 1960), pp. 540–592.
19. F. Scholz, I. Zerbst-Boroffka, *J. Insect Physiol.* **44**, 427–436 (1998).
20. J. I. Camacho García *et al.*, *Mol. Biol. Evol.* **42**, msaf147 (2025).
21. E. H. Allison, K. Irvine, A. B. Thompson, B. P. Ngatunga, *Freshw. Biol.* **35**, 489–515 (1996).
22. M. K. Vollmer, R. F. Weiss, H. A. Bootsma, *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity*, **vol. 12**, E. O. Odada, D. O. Olago, Eds. (Springer, 2002), pp. 209–233.
23. P. Ward, L. Greenwald, F. Rougeriye, *Lethaia* **13**, 182–182 (1980).
24. E. J. Denton, *Proc. R. Soc. London Ser. B* **185**, 273–299 (1974).
25. A. J. Dunstan, P. D. Ward, N. J. Marshall, *PLOS ONE* **6**, e16311 (2011).
26. O. A. Saether, *Nearctic and Palaearctic Chaoborus (Diptera: Chaoboridae)*, Bulletins of the Fisheries Research Board of Canada (Fisheries Research Board of Canada, 1970), vol. 174, pp. 57.
27. C. T. F. Ross, *Pressure Vessels Under External Pressure* (Elsevier, 1990).
28. C. A. Scholz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16416–16421 (2007).
29. M. S. Leonardi *et al.*, *J. Exp. Biol.* **223**, jeb226811 (2020).
30. A. Preuss *et al.*, *Commun. Biol.* **8**, 852 (2025).
31. J. W. Murray *et al.*, *Nature* **338**, 411–413 (1989).
32. E. McKenzie *et al.*, Data from: Crush-resistant air-sacs allow insect larvae to exploit aquatic habitat at extreme depth, Dryad (2026); https://datadryad.org/dataset/doi:10.5061/dryad.4tmpg4fqk

ACKNOWLEDGMENTS

Our thanks to Malawi's National Commission for Science and Technology (NCST) for approving the research and E. Mataka, J. Chipalamoto, and Y. Muhammad of the Nkhata Bay District Fisheries Office for invaluable fieldwork assistance. We also thank B. Gillespie, Department of Zoology workshop, UBC, for construction of the pressure chamber; WSP Wild Water Productions Limited for assisting with travel and logistics in Kenya; and R. Shadwick for feedback on an earlier draft of this paper. **Funding:** NSERC Discovery project and Accelerator grant RGPIN-2020-07089 (to P.G.D.M.); Company of Biologists (COB) Travelling Fellowship JEBTF23051050 (to E.K.G.M..); The Society for Integrative and Comparative Biology Fellowship of Graduate Student Travel (FGST) (to E.K.G.M.); ASL Environmental Sciences, Saanichton, BC, Canada, 2023 Acoustic Zooplankton Fish Profiler (AZFP) award (to M.J.R.N. and P.G.D.M.). **Author contributions:** Conceptualization: P.G.D.M., E.K.G.M.; Methodology: P.G.D.M., E.K.G.M., M.J.R.N., R.P., H.M.; Investigation: P.G.D.M., E.K.G.M., T.A., M.J.R.N., J.C., H.M., R.P.; Visualization: P.G.D.M., E.K.G.M., H.M.; Funding acquisition: P.G.D.M., M.J.R.N., E.K.G.M.; Project administration: P.G.D.M.; Supervision: P.G.D.M., R.P.; Writing – original draft: P.G.D.M., E.K.G.M., R.P., H.M.; Writing – review & editing: P.G.D.M., E.K.G.M., T.A., M.J.R.N., J.C., H.M., R.P. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data are available at Dryad (*32*). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aed0667
Materials and Methods; Figs. S1 to S8; Tables S1 and S2; References (*33*–*52*); Movies S1 to S3; Data S1 to S8

Submitted 19 November 2025; accepted 5 May 2026

10.1126/science.aed0667

EXOPLANETS

Constraining an exoplanet's magnetic field using star-planet interactions

D. Revilla[1]*, P. J. Amado[1]†, R. Luque[1,2]†, P. Schöfer[1]‡, A. F. Lanza[3], A. Binnenfeld[4]§, J. A. Caballero[5], A. P. Hatzes[6], G. W. Henry[7], S. V. Jeffers[6], S. Kaur[8,9], E. Pallé[10,11], L. Peña-Moñino[1], M. Pérez-Torres[1,12], A. Quirrenbach[13], A. Reiners[14], I. Ribas[8,9], D. Viganò[8,9,15], M. R. Zapatero-Osorio[5], S. Zucker[4,16]

Theory predicts that a planet with a sufficiently strong magnetic field orbiting close to its host star could induce star-planet magnetic interactions. This is potentially observable as an optical or radio stellar activity signal synchronized with the planet's orbital period. We analyzed 18 years of high-resolution optical spectroscopy of GJ 436, a low-mass star orbited by a Neptune-sized exoplanet on a polar eccentric orbit. Stellar activity indicators show enhancements at a period corresponding to the exoplanet orbit, modulated by stellar rotation, and the star's 8-year magnetic cycle. We interpret this as a signal of star-planet magnetic interaction. Using a geometric model, we reproduced these periods if GJ 436 b has a magnetic field strength of 6 to 110 gauss.

In a planetary system, the host star and its orbiting planets undergo continuous physical interactions, which can be gravitational, radiative, or magnetic. These interactions are typically strongly asymmetric, with the star dominating the planet. However, theory predicts that this balance can shift in systems where the planet orbits very close to its host star. At small orbital distances, the planet could influence the star, producing measurable changes in stellar rotation, angular momentum evolution, or magnetic activity. These effects could arise from mechanisms including tidal interactions, spin-orbit coupling, or magnetic connections, which are predicted to have distinct observational signatures (*1*–*3*).

Among these mechanisms, magnetic star-planet interaction (SPI) occurs when a planet interacts with a magnetized stellar wind. If the planet has a sufficiently strong magnetic field, this interaction is potentially observable and could constrain the strength of the planetary magnetic field under specific model assumptions (*4*). The magnetic field of a planet can determine whether it retains an atmosphere (and for how long), constrains its interior structure, or influences its long-term evolution (*5*, *6*). Similar magnetic coupling has been observed in the Solar System, such as the effect on Jupiter's magnetic field produced by its moons Io and Ganymede (*7*). Theoretical models predict that SPI could produce periodic signals that are potentially observable at radio and optical wavelengths. Such signals would typically appear at the planet's orbital period but could potentially be shifted due to stellar rotation (*8*). The predicted strength of these signals depends on multiple factors: the magnetic fields of the planet and the star, the planet's size, the orbital distance, and whether the planet is within the star's Alfvén surface (the boundary within which magnetic forces dominate over kinetic forces) (*9*, *10*). Planetary systems with large planets on close orbits are therefore the most likely to exhibit SPI, and there is evidence of SPI in several such systems (*11*–*13*).

The GJ 436 system

The red dwarf star GJ 436 (also cataloged as Ross 905) is at a distance $d = 9.78$ parsecs from the Sun (*14*). It has a stellar radius $R_\star = 0.42$ solar radii, a stellar mass $M = 0.44$ solar masses (*15*), and a rotation period $P_{\rm rot} = 45 \pm 5$ days (*16*). It is orbited by a large exoplanet, GJ 436 b (*17*), which has a mass of 4.170 ± 0.168 Earth masses ($M_\oplus$) (*18*) and an orbital period $P_{\rm orb} = 2.64$ days (*17*). The stellar magnetic environment is predicted to be conducive to SPI (*19*, *20*). The star is magnetically quiet (*21*), with low background variability in its chromosphere, the hot, low-density layer of the stellar atmosphere above the visible surface, where magnetic activity manifests as variable excess emission from the ultraviolet to the near-infrared.

Spectroscopic datasets

Since the identification of GJ 436 b, this system has been regularly observed using high-resolution optical spectroscopy to constrain the exoplanet's mass. We reanalyzed those archival observations to search for periodic signatures of SPI. We focused on chromospheric activity indicators predicted to be sensitive to SPI (*22*, *11*): the Ca ii H&K lines at 396.84 and 393.36 nm; the Ca ii infrared triplet at 849.8 nm (IRT-a), 854.2 nm (IRT-b), and 866.2 nm (IRT-c); and the H α line at 656.28 nm. We used archival high-resolution spectra observed with two instruments: the High Accuracy Radial velocity Planet Searcher (HARPS) (*23*), which covers 380 to 690 nm at a resolving power of 115,000, and the Calar Alto High-Resolution Search for M Dwarfs with Exoearths with Near-Infrared and Optical Echelle Spectrographs (CARMENES) (*24*), which covers 520 to 1710 nm, with a resolving power of 94,600 in the optical channel and 80,400 in the near-infrared channel. We analyzed a total of 371 spectra of GJ 436: 169 HARPS observations between 2006 and 2010, a further 23 HARPS observations taken in 2020, 112 CARMENES observations taken in 2016, and 51 CARMENES observations taken in 2024.

Because SPI signals are expected to be transient (*25*, *26*), we applied time-series analysis techniques optimized to find periodic and non-continuous signals in unevenly sampled data (*26*). We first used the generalized Lomb-Scargle (GLS) periodogram (*27*, *28*) to search for signals across seven different epochs corresponding to yearly observing intervals [five from HARPS and two from CARMENES (figs. S3 to S10)]. To account for the transitory nature of SPI, we additionally applied rolling periodograms (*29*, *30*) to each epoch to test whether the signal changes in intensity over shorter timescales. We refer to these shorter intervals as subepochs, which correspond to continuous observing runs spanning days to weeks within a given year. We then applied the GLS periodogram to each subepoch to characterize the properties of the modulation.

[1]Instituto de Astrofísica de Andalucía - Consejo Superior de Investigaciones Científicas, Granada, Spain. [2]Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA. [3]Osservatorio Astrofisico di Catania, Istituto Nazionale di Astrofisica, Catania, Italy. [4]Porter School of the Environment and Earth Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. [5]Centro de Astrobiología, Instituto Nacional de Técnica Aeroespacial - Consejo Superior de Investigaciones Científicas, Centro Europeo de Astronomía Espacial Campus, Villanueva de la Cañada, Madrid, Spain. [6]Thüringer Landessternwarte Tautenburg, Tautenburg, Germany. [7]Tennessee State University, Nashville, TN, USA. [8]Institut d'Estudis Espacials de Catalunya, Castelldefels, Barcelona, Spain. [9]Institut de Ciències de l'Espai - Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain. [10]Instituto de Astrofísica de Canarias, La Laguna, Tenerife, Spain. [11]Departamento de Astrofísica, Universidad de La Laguna, La Laguna, Tenerife, Spain. [12]School of Sciences, European University Cyprus, Nicosia, Engomi, Cyprus. [13]Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Heidelberg, Germany. [14]Institut für Astrophysik und Geophysik, Georg-August-Universität, Göttingen, Germany. [15]Institute of Applied Computing Community Code, Universitat de les Illes Balears, Palma, Spain. [16]School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. ***Corresponding author. Email: drevilla@iaa.csic.es** †These authors contributed equally to this work. ‡Present address: Department of Physics, Ariel University, Ariel, Israel. §Present address: Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne, Observatoire de Sauverny, Versoix, Switzerland.

Periodic stellar activity

In the computed periodograms, we identified a peak at 2.81 days, along with its 1-day alias at 1.54 days, in the HARPS observations taken in 2008 (abbreviated H-08) and CARMENES observations taken in 2016 (abbreviated C-16) (Fig. 1, A and B). This peak coincides with the system's synodic period $P_{syn} = \left(P_{orb}^{-1} - P_{rot}^{-1}\right)^{-1}$, one of two potential periods at which SPI signatures have been predicted (the other being the orbital period) (*31*, *32*). The local false alarm probability (FAP) of the 2.81-day peaks appearing at that expected period (*33*) are 0.37% in H-08 and 0.77% in C-16. These FAP values were computed at a fixed, predefined frequency, so they represent the probability of a noise peak at that location in a single epoch, not anywhere in the periodogram at any epoch. Figure 1 shows the global FAP values, which consider the entire frequency range.

For the 2024 CARMENES epoch (abbreviated C-24), no peak appears at 2.81 days, with a local FAP of 99.8%, indicating no signal at the expected frequency. However, there is a peak at 2.46 days, which has a global FAP of 0.17% (Fig. 1E). The frequency of this second peak is consistent with an antisynodic period $P_{a\text{-}syn} = \left(P_{orb}^{-1} - P_{rot}^{-1}\right)^{-1}$. The C-16 periodogram additionally shows an isolated peak at 4.19 days (fig. S9), which does not coincide with any expected SPI signal. We were unable to determine the physical origin of this peak.

Each of these signals is detected in either the Ca ii IRT-a or Ca ii H&K lines, which are known to be strongly correlated with each other (*34*) and are tracers of chromospheric activity (*35*, *36*). We quantified these lines using the pseudoequivalent width (pEW) metric for Ca ii RT-a (*30*) in the CARMENES spectra and the $I_{Ca\ II}$ metric (*21*) for the Ca ii H&K lines in the HARPS spectra. No epoch was observed with either instrument, so we cannot compare the contemporaneous behavior of both metrics simultaneously.

We also searched other activity indicators arising from the star's chromosphere or photosphere (the visible surface of the star lying just below the chromosphere), including the He D_3, Na D, H α, TiO, VO, and Ca I lines (*26*), but found no periodicities. This indicates that the modulation originates in the lower chromosphere, where Ca ii is present. For this type of star, the temperature of the lower chromosphere is ~3000 to 6000 K, and the column mass density is 3 to 8×10^{-4} kg m^{-2} (*37*). H α, which does not exhibit the 2.81- or 2.46-day periodicities, forms higher in the stellar atmosphere at temperatures of 8000 to 11,000 K, and has lower column mass densities (~10^{-4} kg m^{-2}).

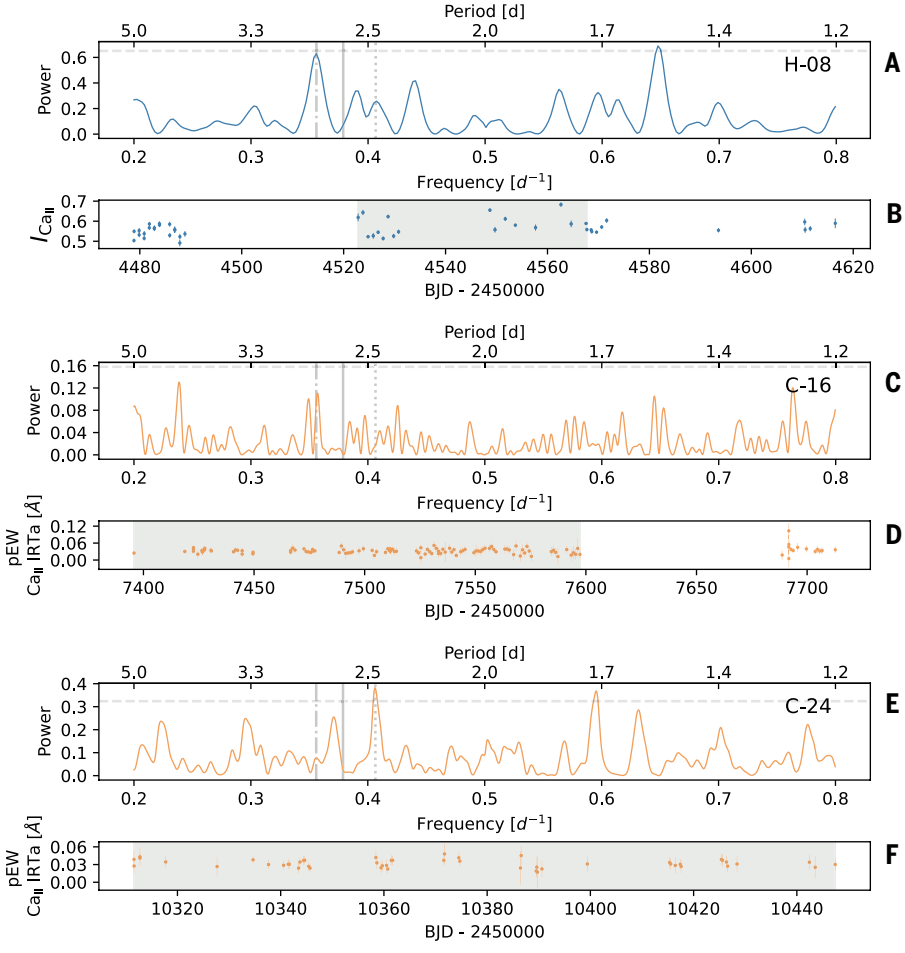


Fig. 1. Periodograms of calcium chromospheric indicators. Each panel is color coded, with blue indicating HARPS data and orange indicating CARMENES data. (**A**) GLS periodogram of the $I_{Ca\ II}$ index from the H-08 subepoch. Gray vertical lines mark the expected SPI periods at $P_{syn} = 2.81$ days (dot-dashed), $P_{orb} = 2.64$ days (solid), and $P_{a\text{-}syn} = 2.46$ days (dotted). The dashed horizontal line shows 1% FAP (*28*). (**B**) The $I_{Ca\ II}$ activity metric as a function of barycentric Julian date (BJD) for the 2008 HARPS epoch. Gray shading indicates the data used to calculate the periodogram in panel (A). Error bars are 1σ uncertainties. (**C** and **D**) Same as panels (A) and (B) but for the pEW (Ca ii IRT-a) metric in the C-16 epoch. (**E** and **F**) Same as panels (C) and (D) but for the C-24 epoch.

Geometric model

To investigate the origin of multiple periodicities in the data, we considered a geometric model containing two distinct activity regions: one associated with intrinsic stellar activity and another induced by magnetic SPI (fig. S19). The first region is fixed on the stellar surface and co-rotates with the star, whereas the second is magnetically linked to the planet and follows its orbital motion. This model accounts for the stellar rotation period, the planet's orbital period, and the obliquity of the planetary orbit (*26*). The relative position between these two regions changes over time, producing a periodically varying separation that modulates the amplitude of the chromospheric emission.

Previous SPI models have predicted periodic signatures at either the synodic or orbital periods (*22*, *31*, *38*). We used simulations to generate synthetic observational datasets using our geometric model (*26*). We found that only the synodic and antisynodic frequencies appeared due to the system's geometry and visibility conditions (equation S7 and fig. S19). The synodic period P_{syn} dominates if the planet's orbit is prograde (with inclination $I \ll 90°$). If the planet is retrograde ($180° > I \gg 90°$), then the antisynodic $P_{a\text{-}syn}$ period dominates. In the intermediate case, polar systems in which the planet orbits perpendicular to the stellar equator ($I \approx 90°$), both periods appear with the same intensity. GJ 436 b has $I = 103 \pm 13°$ (*16*), so we expect the latter regime to apply. The predicted signal exhibits a beat-like modulation composed of two periods (Fig. 2, E and F). The relative probabilities of detecting P_{syn} and $P_{a\text{-}syn}$ depend on the observational cadence and signal-to-noise characteristics (*26*), so either (or both) can appear in a given dataset (Fig. 2, B to D). This geometric model does not predict a detectable signal at P_{orb} under any considered conditions, even in ideal continuous observations.

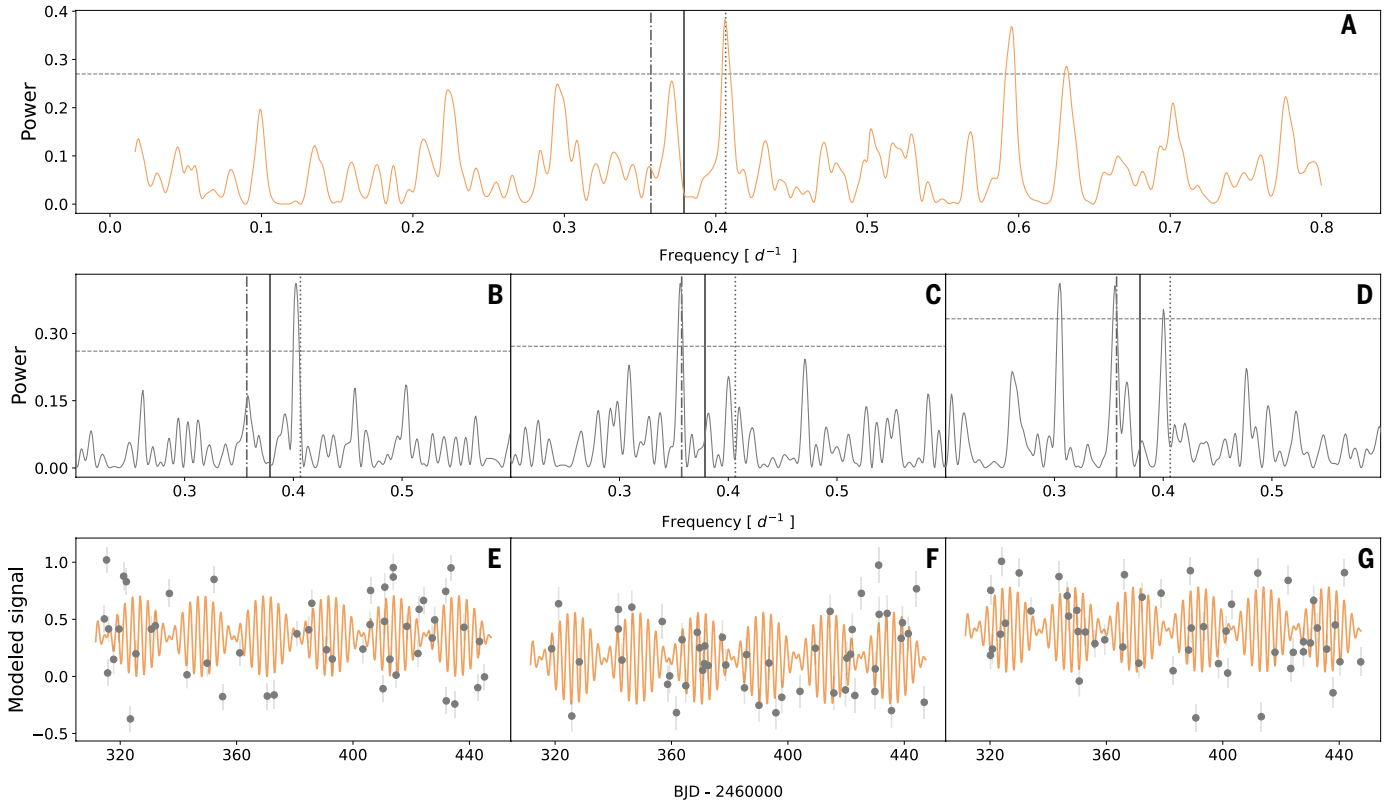


Fig. 2. Observations during C-24 compared with simulations. Line styles are the same as in Fig. 1. (**A**) GLS periodogram of the pEW′(Ca II IRT-a) metric during C-24. (**B** to **D**) GLS periodogram of three synthetic mock observations generated from the geometric model (*26*). Peaks appear at the synodic and antisynodic periods but not the orbital period. Panel (B) reproduces the 2.81-day peak from C-16 and H-08, panel (C) reproduces the 2.46-day peak from C-24, and panel (D) has several peaks simultaneously. (**E** to **G**) The line is the simulated signal from the geometric model, and data points are synthetic observations of that signal with added Gaussian-distributed noise (with SD corresponding to 1σ of the modeled signal). Error bars show synthetic 1σ uncertainties derived from fitting a Gaussian distribution to the C-24 observations. The synthetic observations were taken at different times from the real C-24 observations to illustrate how the sampling affects the resulting periodograms [panels (B) to (D)].

Stellar activity cycle

The geometric model does not explain why SPI signals are only detected during specific epochs (in 2008, 2016, and 2024). Red dwarf stars are known to exhibit magnetic activity cycles on multiyear timescales, which manifest in both photometric variability and chromospheric emission indicators (*39*, *40*). Previous studies have determined that GJ 436's activity cycle is 7.75 ± 0.10 years (*41*, *42*). We analyzed observations of GJ 436 taken with Tennessee State University's 0.8-m Automatic Photoelectric Telescope (APT) from 2003 to 2018 and with the Celestron 14-inches Automated Imaging Telescope (AIT) from 2013 to 2024 (*26*). We measured photometry from those observations and compared it with the chromospheric calcium indices from HARPS and CARMENES (Fig. 3). Although the AIT data show more irregular variability than in previous observations, they are consistent with an ~8-year cycle. The three epochs in which we detected SPI signals are spaced about 8 years apart, coinciding with intermediate activity levels.

We suggest that this temporal coincidence could arise if the detection probability of SPI signals is modulated by the stellar magnetic cycle. The total power involved in magnetic interaction depends on the star's magnetic field strength (*26*). Although stronger magnetic fields at the maximum of the activity cycle could enhance SPI power, the concurrent increase in the star's own activity, with more and larger magnetic active regions, could overwhelm the localized signal produced by SPI (*43*). Conversely, during activity minima, the overall magnetic energy available for SPI is reduced, making it more difficult to detect. We propose that intermediate activity levels, when the field is strong enough to support SPI but surface activity is relatively low, provide more favorable conditions for identifying SPI-induced variability.

The detection of signals at two distinct periods is consistent with our geometric model, and their reappearance at multiple epochs separated by ~8 years indicates an SPI origin. We quantified the statistical significance by computing the FAP of the 2.81-day peaks in the H-08 and C-16 datasets and the 2.46-day peak in the C-24 data. Standard analytical FAPs search over a broad range of frequencies, but our peaks are expected at specific values: the planetary orbital period, the synodic period, or the antisynodic period. We therefore used a bootstrap procedure that incorporates this prior knowledge (*26*). The data points were randomly shuffled to generate multiple realizations simulating pure noise, then the FAP was estimated from how often a peak as strong as the observed one appears near the predicted frequency in the noise realizations. The resulting FAPs are 3.7×10^{-3} for H-08, 7.7×10^{-3} for C-16, and 3.2×10^{-4} for C-24 (table S1), indicating that random noise would produce peaks this strong in 0.4, 0.8, and 0.03% of cases, respectively.

For an overall FAP that also accounts for the four epochs with nondetections, we applied Fisher's test (*44*). This method sums the logarithms of the per-epoch FAPs into a single statistic, $X^2 = -2\sum_i \ln p_i$ (where p_i indicates the individual FAPs) and compares it with the distribution expected if all seven epochs were pure noise. Strong detections produce large contributions to X^2, whereas epochs without signal add only small terms, so the four nondetections dilute the

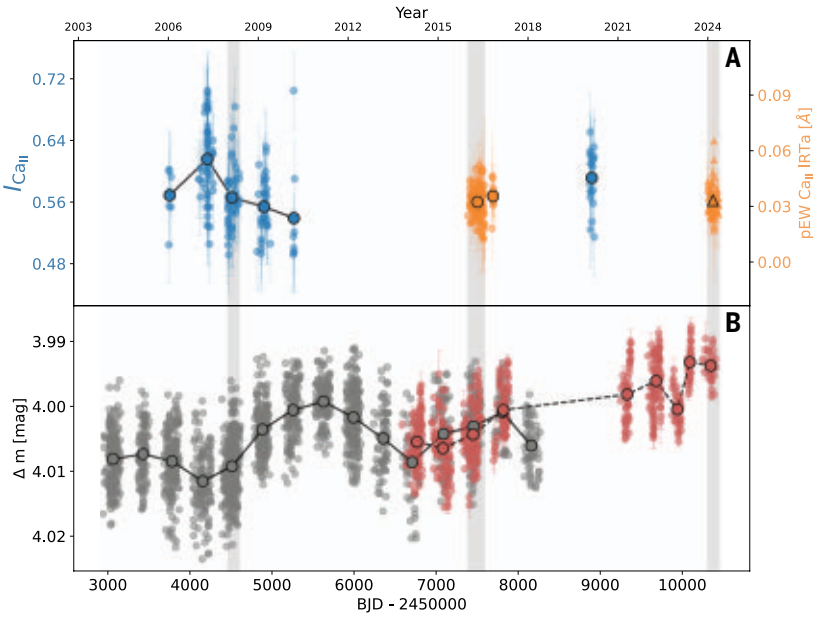


Fig. 3. Chromospheric activity indicators and photometric time series. (**A**) Chromospheric activity indicators $I_{Ca\,II}$ from the HARPS data (blue, left axis), and pEW′(Ca ii IRT-a) from the CARMENES data (orange dots for C-16 and triangles for C-24, right axis). Filled symbols with black outlines are the yearly mean values in each epoch. (**B**) Differential photometry (in the combined Stromgren b- and y-band filters) from the T12 APT (gray) and AIT (red) observations. These data have been corrected for an offset between the two instruments calculated using the four epochs in which both were observing. Filled symbols are as in panel (A). The black lines connect the seasonal means of the T12 APT (solid) and C14 AIT (dashed) datasets. Gray shaded regions show the epochs when SPI was detected (H-08, C-16, and C-24).

Table 1: Planetary magnetic field and magnetospheric radius of GJ 436 b. The calculated planetary magnetic field strength (B_p, in gauss) and magnetospheric radius (r_M, in planetary radii) from the interconnecting loop model calculated under different assumptions for the stellar magnetic field strength and SPI emission efficiency. Scenarios A and B use two extrapolations of the stellar field [(*46*), their models I and II]. Scenario C assumes a purely radial stellar field extrapolation (*49*). The percentages are the assumed fraction of energy converted to flux in the Ca ii IRT-a or Ca ii H&K lines. All scenarios assume a planetary radius of 4.2 Earth radii (*26*).

Fraction of energy in Ca ii	**2%**		**25%**		**100%**		
Scenario	**B_p (G)**	**r_M (R_p)**	**B_p (G)**	**r_M (R_p)**	**B_p (G)**	**r_M (R_p)**	**Reference for r_M**
A	110	20	24	12	10.3	9	Eq. S17 and [(*46*), their model I]
B	95	13	21	8	9	6	Eq. S17 and [(*46*), their model II]
C	68	21	14.7	13	6.3	9.6	Eq. S16 and (*49*)

combined statistic (rather than being ignored). The seven epochs are statistically independent; they are separated by years, and any correlated structure within each epoch is removed by the shuffling procedure. Applying this test to all seven epochs gives $X^2 = 33.78$ (fig. S18), corresponding to an overall probability of 2.22×10^{-3} that the observed pattern of detections and nondetections could have arisen from noise alone. We therefore reject the null hypothesis of no SPI signal at the 99.7% confidence level.

We interpret the signals as arising from a common recurring physical mechanism. To produce magnetic SPI, the planet must be within the stellar Alfvén surface. Spectropolarimetric observations in 2016 (*21*) have previously been used to reconstruct GJ 436's large-scale magnetic field and to determine the location of its Alfvén surface (*45*, *46*). This surface expands from ~1.5 to ~3 times the orbital distance of the planet ($14.6 \times R_\star \simeq 0.028$ astronomical units). This indicates that the planet was located within the Alfvén surface for most of its orbit during the C-16 epoch. Although the Alfvén surface can change during the stellar activity cycle, the other SPI signals were detected during the H-08 and C-24 epochs, which are separated from 2016 by nearly one stellar activity cycle. We therefore consider it likely that GJ 436 b was also within the Alfvén surface during those epochs.

Planetary magnetic field

The power emitted in the Ca ii IRT-a line during the 2016 and 2024 epochs was 10^{19} W (*26*), which we calculated by converting the observed line strengths into absolute fluxes using empirical calibrations that relate the brightness of a line relative to the star's total luminosity (*47*). However, this is only a fraction of the total SPI power because most is likely converted to heat or emitted at other wavelengths. Because the spectral energy distribution of SPI events is unknown, we adopt a flare observed on the red dwarf AD Leo as an empirical proxy for a heated chromosphere. This flare radiated ~2% of its energy in the Ca ii IRT-a line and <50% in the Ca ii H&K lines. We therefore assume that the Ca ii IRT-a line accounts for 2 to 100% of the total SPI power, with the latter value being chosen to provide a conservative limit.

To convert the observed SPI power into constraints on the planetary magnetic field strength (B_p) and planet magnetospheric radius (r_M), we evaluated several theoretical models of magnetic interaction: the Alfvén wing model, the magnetic reconnection model, the helicity variation model, and the interconnecting magnetic loop model (*26*). The magnetospheric radius is defined as the distance at which the planetary field balances the stellar wind pressure. A larger magnetosphere increases the area over which the stellar wind interacts with the planetary field, thereby enhancing the total power that can be extracted from the system. Among the theoretical models considered, only the model of an interconnecting magnetic loop between the planet and the star (*48*) produced sufficient energy output to be consistent with our observations. This model assumes a stable magnetic configuration in which a magnetic loop–like structure connects the star and planet; orbital motion stresses this structure and periodically injects energy, which is rapidly dissipated to maintain equilibrium. The other models that we considered do not reproduce the observed power levels. Table 1 presents the derived values of B_p and r_M using

three assumed stellar magnetic field strengths in gauss (G) at the planet's orbital distance: 0.013 G, 0.026 G [(*46*), their models I and II], and 0.13 G (*49*). Each of these values is consistent with the observed power but different SPI emission efficiencies.

The planetary magnetic field strength required to reproduce the observed Ca ii IRT-a emission depends on both the stellar magnetic field at the orbital distance and the fraction of the total interaction power emitted in this line. Using the interconnecting loop model, we derived a lower bound of 6.3 G by assuming a radial stellar field (0.13 G at the planet distance) and that 100% of the SPI power is radiated in Ca ii IRT-a. At the other extreme, if only 2% of the power is emitted in this line and the stellar field is weaker [0.013 G (*46*)], then the required planetary field is 110 G, which we regard as the upper bound. This range reflects the uncertainty in the total power dissipated by the interaction, the variation in assumed stellar magnetic field strength, and the model's dependence on these parameters.

The corresponding radii of the planet's magnetosphere are 6 to 21 planetary radii (R_p). For comparison, Neptune's magnetosphere is ~26 R_p (*50*). This difference could be due to GJ 436 b's proximity to its host star and the stronger stellar wind pressure that it experiences (*6*). Our transient detections of SPI indicate that long-term variability in stellar magnetic topology and field strength can modulate the detectability of planetary magnetic interactions.

REFERENCES AND NOTES

1. A. F. Lanza, *Astron. Astrophys.* **487**, 1163–1170 (2008).
2. A. Strugarek, A. S. Brun, S. P. Matt, V. Réville, *Astrophys. J.* **795**, 86 (2014).
3. A. Strugarek, E. Shkolnik, Introduction to magnetic star-planet interactions. arXiv:2502.13262 [astro-ph.EP] (2025).
4. P. W. Cauley, E. L. Shkolnik, J. Llama, A. F. Lanza, *Nat. Astron.* **3**, 1128–1134 (2019).
5. A. Strugarek, Physics of star-planet magnetic interactions. arXiv:2104.05968 [astro-ph.SR] (2021).
6. A. A. Vidotto *et al.*, *Astron. Astrophys.* **557**, A67 (2013).
7. R. Prangé *et al.*, *Nature* **379**, 323–325 (1996).
8. E. Shkolnik, G. A. H. Walker, D. A. Bohlender, P.-G. Gu, M. Kurster, *Astrophys. J.* **622**, 1075–1090 (2005).
9. P. Zarka, in *Handbook of Exoplanets*, H. J. Deeg, J. A. Belmonte, Eds. (Springer, 2020), pp. 1–35; https://doi.org/10.1007/978-3-319-30648-3_22-2.
10. A. Strugarek, A. S. Brun, S. P. Matt, V. Réville, *Astrophys. J.* **815**, 111 (2015).
11. E. Shkolnik, G. A. H. Walker, D. A. Bohlender, *Astrophys. J.* **597**, 1092–1096 (2003).
12. P. W. Cauley, E. L. Shkolnik, J. Llama, V. Bourrier, C. Moutou, *Astron. J.* **156**, 262 (2018).
13. E. Ilin *et al.*, *Nature* **643**, 645–648 (2025).
14. A. Vallenari *et al.*, *Astron. Astrophys.* **674**, A1 (2023).
15. L. J. Rosenthal *et al.*, *Astrophys. J. Suppl. Ser.* **255**, 8 (2021).
16. V. Bourrier *et al.*, *Nature* **553**, 477–480 (2018).
17. R. P. Butler *et al.*, *Astrophys. J.* **617**, 580–588 (2004).
18. G. Maciejewski *et al.*, On the GJ 436 planetary system. arXiv:1501.02711 [astro-ph.EP] (2014).
19. V. Bourrier, A. Lecavelier des Etangs, D. Ehrenreich, Y. A. Tanaka, A. A. Vidotto, *Astron. Astrophys.* **591**, A121 (2016).
20. B. Lavie *et al.*, *Astron. Astrophys.* **605**, L7 (2017).
21. M. Kumar, R. Fares, *Mon. Not. R. Astron. Soc.* **518**, 3147–3163 (2022).
22. M. Cuntz, S. H. Saar, Z. E. Musielak, *Astrophys. J.* **533**, L151–L154 (2000).
23. M. Mayor *et al.*, *Messenger* **114**, 20–24 (2003).
24. A. Quirrenbach *et al.*, in *Ground-Based and Airborne Instrumentation for Astronomy V*, S. K. Ramsay, I. S. McLean, H. Takami, Eds. (SPIE, 2014), vol. 9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 91471F; https://doi.org/10.1117/12.2056453.
25. E. Shkolnik, D. A. Bohlender, G. A. H. Walker, A. C. Cameron, *Astrophys. J.* **676**, 628–638 (2008).
26. Materials and methods are available as supplementary materials.
27. S. Ferraz-Mello, *Astron. J.* **86**, 619 (1981).
28. M. Zechmeister, M. Kürster, *Astron. Astrophys.* **496**, 577–584 (2009).
29. O. Herbort, Stellar activity in CN Leo, master's thesis, Institut für Astrophysik Göttingen, Göttingen, Germany (2018).
30. P. Schöfer *et al.*, *Astron. Astrophys.* **623**, A44 (2019).
31. C. Fischer, J. Saur, *Astrophys. J.* **872**, 113 (2019).
32. J. Saur, *Electromagnetic Coupling in Star-Planet Systems* (Springer, 2018); .
33. A. P. Hatzes, *The Doppler Method for the Detection of Exoplanets* (IOP Publishing, 2019); https://doi.org/10.1088/2514-3433/ab46a3.
34. J. Martin *et al.*, *Astron. Astrophys.* **605**, A113 (2017).
35. O. C. Wilson, *Astrophys. J.* **153**, 221 (1968).
36. V. Andretta, I. Busà, M. T. Gomez, L. Terranegra, *Astron. Astrophys.* **430**, 669–677 (2005).
37. D. Hintz *et al.*, *Astrophys. J.* **954**, 73 (2023).
38. A. F. Lanza, *Astron. Astrophys.* **544**, A23 (2012).
39. P. Robertson, M. Endl, W. D. Cochran, S. E. Dodson-Robinson, *Astrophys. J.* **764**, 3 (2013).
40. A. Suárez Mascareño, R. Rebolo, J. I. González Hernández, *Astron. Astrophys.* **595**, A12 (2016).
41. J. D. Lothringer *et al.*, *Astron. J.* **155**, 66 (2018).
42. R. O. P. Loyd *et al.*, *Astron. J.* **165**, 146 (2023).
43. D. H. Hathaway, *Living Rev. Sol. Phys.* **12**, 4 (2015).
44. R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, 1925).
45. S. Bellotti *et al.*, *Astron. Astrophys.* **676**, A139 (2023).
46. A. A. Vidotto *et al.*, *Astron. Astrophys.* **678**, A152 (2023).
47. F. Labarga *et al.*, Chromospheric flux-flux relationships of Cool Dwarfs using VIS and NIR CARMENES spectra. Analysis of different emitters populations, version 2, in *The 21st Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun (CS21)* (Zenodo, 2023); https://doi.org/10.5281/zenodo.7670149.
48. A. F. Lanza, *Astron. Astrophys.* **557**, A31 (2013).
49. A. F. Lanza, *Astron. Astrophys.* **610**, A81 (2018).
50. N. F. Ness *et al.*, *Science* **246**, 1473–1478 (1989).
51. D. Revilla, Planet-induced modulation of stellar activity in GJ 436: Spectroscopic and photometric data, Zenodo (2025); https://doi.org/10.5281/zenodo.20306438.

ACKNOWLEDGMENTS

We thank G. M. Mirouh for discussion of the geometrical model. This publication is partly based on observations collected under the CARMENES guaranteed time observations and 24A-3.5-013 proposal. CARMENES is an instrument at the Centro Astronómico Hispano en Andalucía (CAHA) at Calar Alto (Almería, Spain) operated jointly by the Junta de Andalucía and the Instituto de Astrofísica de Andalucía (CSIC). This instrument was also funded by the Complementary R&D&I Plan in Astrophysics and High-Energy Physics (AST_00001_8) with funding from the European Union – NextGenerationEU, the Spanish Government's Ministry of Science, Innovation and Universities, the Recovery and Resilience Plan, the Spanish National Research Council (CSIC), and the Andalusian Regional Government's Ministry of University, Research and Innovation. **Funding:** D.R. and L.P.-M. were supported by the PRE2021-099654 and PRE2020-095421 grants, funded by MCIU/AEI/10.13039/501100011033. D.R. and P.J.A. were supported by Spanish National grant PID2022-137241NB-C42. P.J.A. was supported by the PID2022-137241NB-C43 project. D.R., P.J.A., L.P.-M., and M.P.-T. were supported by Severo Ochoa grant CEX2021-001131-S funded by MCIN/AEI/ 10.13039/501100011033. R.L. was supported by the European Union (ERC, THIRSTEE, 101164189) and NASA through NASA Hubble Fellowship grant HST-HF2-51559.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. I.R. was supported by the PID2024-158486OB-C31 MCIU CEX2020-001058-M MCIU "SPOTLESS" no. 101140786 Horizon Europe ERC funds. G.W.H. was supported by NASA and NSF. A.B. and S.Z. were supported by the Israel science foundation (grant 1404/22) and the Israel Ministry of Science and Technology (grant 3-18143). S.K and D.V. were supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Starting Grant "IMAGINE" 948582). I.R., S.K., and D.V. were supported by a "María de Maeztu" award to the Institut de Ciències de l'Espai (CEX2020-001058-M). S.K. was supported by the doctoral program in Physics of the Universitat Autònoma de Barcelona. A.F.L. was supported by the Italian Ministry of University and Research through funding to the Italian National Institute for Astrophysics. J.A.C. and M.R.Z.-O. were supported by the PID2022-137241NB-C42 project. I.R. was supported by PID2024-158486OB-C31 MCIU, CEX2020-001058-M MCIU and "SPOTLESS" no. 101140786 Horizon Europe ERC projects. E.P. was supported by the European Union (ERC AdvG SPEAR, GA 101200674), Agencia Estatal de Investigación of the Ministerio de Ciencia e Innovación MCIN/AEI/10.13039/501100011033, and the ERDF "A way of making Europe" through projects PID2021-125627OB-C32 and PID2024-158486OB-C32. **Author contributions:** D.R. and P.J.A. conceived the study. D.R. led the analysis of the data. D.R., P.J.A., and R.L. interpreted the data. D.R., R.L., and P.J.A. wrote the manuscript. P.S. determined the CARMENES activity indicators and helped to interpret the data. A.F.L. developed the geometrical model. A.P.H. contributed to the statistical analysis. G.W.H. provided and analyzed the T12 APT and C14 AIT data. A.B. and S.Z. analyzed the spectra. J.A.C., A.Q., A.R., and I.R. contributed to CARMENES operations, observations, and data extraction. S.V.J., S.K., E.P., L.P., M.P.-T., D.V., and M.R.Z.-O. contributed to the discussion and interpretation of the geometric model. All authors provided comments on the manuscript. **Competing interests:** The authors declare no competing interests. **Data, code, and materials availability:** The HARPS spectroscopic observations are available at http://archive.eso.org/eso/eso_archive_main.html under the target name "GJ 436." The CARMENES 2016 spectroscopic observations are available at http://carmenes.cab.inta-csic.es/gto/jsp/dr1Public.jsp under target names "J11421+267" and "Ross 905." The T12 APT and C14 AIT photometric observations are archived at Zenodo (*51*). The derived photometry and spectroscopic activity indices are also available in the same Zenodo repository (*51*). No physical materials were generated in this work.

License information: Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv3075
Materials and Methods; Figs. S1 to S19; Tables S1 and S2; References (*52*–*85*)

Submitted 16 December 2024; accepted 5 June 2026; published online 25 June 2026

10.1126/science.adv3075

RESEARCH ARTICLE SUMMARY

DRUG DESIGN

Late-stage functionalization with strain-release warheads enables tunable covalent inhibition

Zachary P. Shultz, Ansar Lee-Sam, Yun-Pu Chang, Luxin Sun, Dylan Grassie, Alessio Gabellini, Kyle Pedretty, Thomas Scattolin, Victoria Izumi, Bin Fang, Samer Sansil, Ramu Kakumanu, Lukasz Wojtas, John Koomen, Ernst Schönbrunn, Andrii Monastyrskyi, Derek Duckett, Justin M. Lopchuk*

Full article and list of author affiliations: https://doi.org/10.1126/science.adx7219

INTRODUCTION: Covalent drugs are transforming targeted therapy by forming durable bonds with disease-driving proteins, yet most rely on a narrow set of reactive groups, particularly acrylamides. These conventional approaches, although effective, can lead to off-target interactions and restrict broader application by their limited structural design. Expanding covalent drug design beyond these established chemotypes is essential to improve both selectivity and therapeutic performance.

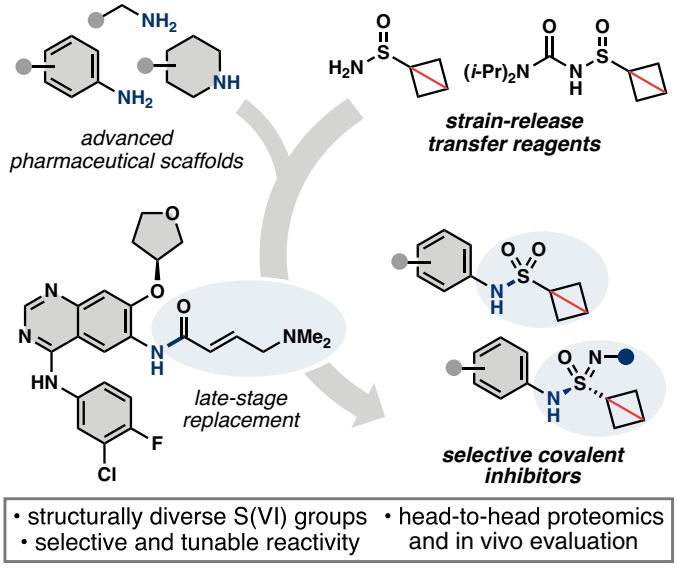
RATIONALE: We sought to establish a general platform for replacing acrylamide-based covalent reactive groups in complex drug molecules with alternative chemotypes that offer improved control over reactivity. Our approach centers on bicyclobutanes that are integrated with sulfur-based functional groups commonly used in medicinal chemistry. To enable broad application, we developed a reagent-based strategy that allows these strain-release elements to be installed at the final stage of a synthesis from widely accessible amine precursors. This modular S(IV)-based platform provides a unified entry to multiple sulfur oxidation states and connectivity patterns, enabling systematic tuning of covalent reactivity and target engagement while preserving the parent-drug architecture. By design, this approach allows direct, head-to-head comparison with established covalent inhibitors.

RESULTS: We developed stable, scalable reagents that enable efficient late-stage installation of strain-release bicyclobutane groups across a wide range of clinically relevant scaffolds, including multiple approved kinase inhibitors. This strategy enables direct bioisosteric replacement of acrylamide warheads without modifying the underlying pharmacophore. These S(VI) strain-release groups are highly chemoselective for thiols, and their intrinsic reactivity can be tuned over a broad range through structural modification.

Notably, compounds with similar intrinsic reactivity displayed markedly different levels of target inhibition, demonstrating that productive covalent engagement depends not only on electrophile reactivity but also on molecular orientation within the protein binding site. In cellular systems, the modified inhibitors retained potent activity and effectively suppressed target signaling. Structural analysis confirmed covalent bond formation at the intended site. Across kinase panels and proteome-wide profiling experiments, these strain-release analogs displayed improved selectivity and reduced off-target interactions relative to their acrylamide counterparts. Importantly, these advances translated beyond in vitro systems, with strain-release analogs demonstrating favorable pharmacokinetic properties and efficacy in preclinical in vivo mice models.

CONCLUSION: This work establishes a reagent-enabled, late-stage functionalization platform for the bioisosteric replacement of acrylamides using strain-release bicyclobutanes, bridging chemical innovation to preclinical validation. More broadly, it further supports that effective covalent inhibition is governed not only by intrinsic electrophile reactivity but also by its integration with molecular recognition, providing a framework for designing more selective and clinically effective covalent therapies. □

*Corresponding author. Email: justin.lopchuk@moffitt.org Cite this article as Z. P. Shultz *et al.*, *Science* **393**, 408 (2026). DOI: 10.1126/science.adx7219



Selective strain-release covalent inhibition. Development of S(IV) strain-release reagents enables late-stage functionalization of complex pharmaceuticals for preclinical evaluation as next-generation covalent inhibitors. *i*-Pr, isopropyl; Me, methyl.

DRUG DESIGN

Late-stage functionalization with strain-release warheads enables tunable covalent inhibition

Zachary P. Shultz[1], Ansar Lee-Sam[1], Yun-Pu Chang[1], Luxin Sun[1], Dylan Grassie[1], Alessio Gabellini[1], Kyle Pedretty[2], Thomas Scattolin[1], Victoria Izumi[3], Bin Fang[3], Samer Sansil[4], Ramu Kakumanu[4], Lukasz Wojtas[2], John Koomen[5,6], Ernst Schönbrunn[1,6], Andrii Monastyrskyi[1,4,6], Derek Duckett[1], Justin M. Lopchuk[1,2,6]*

Covalent inhibition continues to gain momentum as a strategy for selective protein modulation in both therapeutic and chemical biology contexts. Covalent reactive groups (CRGs) typically engage nucleophilic residues such as cysteine, resulting in targeted protein inactivation. However, common electrophiles such as acrylamides often suffer from nonselective reactivity, leading to off-target effects and toxicity. To overcome these limitations, we developed a modular sulfur(IV) reagent platform for the mild, late-stage installation of sulfonyl- and sulfonimidoyl-bicyclobutane motifs with complete cysteine selectivity. This methodology enables access to diverse sulfur(VI) CRGs with tunable strain-release reactivity. Incorporation into US Food and Drug Administration–approved covalent inhibitors demonstrated effective bioisosteric replacement of acrylamides and the potential of strain-release CRGs for selective protein targeting. Preclinical studies in mice have validated this approach, highlighting its promise for next-generation covalent drug design.

Over the past decade, targeted covalent inhibition has reemerged as a powerful strategy in pharmaceutical research (*1*–*4*). At the present time, ~130 marketed drugs act through covalent mechanisms (*5*), and both early-stage development and late-phase clinical trials of new covalent inhibitors continue to expand (*6*, *7*). A broad array of covalent reactive groups (CRGs) have been developed to engage proteins, peptides, and nucleobases through reversible or irreversible mechanisms, with acrylamides achieving the greatest clinical success to date (*8*–*10*) (Fig. 1A).

Despite their utility, many CRGs cover a relatively narrow chemical design space that can impose practical and mechanistic constraints and limitations, including metabolic susceptibility, nonselective reactivity, and rapid reversibility (*11*). For example, the acrylamide moieties in the Bruton's tyrosine kinase (BTK) inhibitor ibrutinib and the epidermal growth factor receptor (EGFR) inhibitor afatinib have been associated with widespread off-target proteome engagement at therapeutically relevant concentrations linked to cytotoxicity (*12*). Additionally, common CRGs are often flat sp^2-rich structures that limit steric differentiation at the electrophilic center. These considerations underscore an opportunity to expand the chemical and geometric diversity of CRGs while retaining controllable thiol reactivity under physiological conditions.

In practice, electrophile selection in targeted covalent inhibitor discovery has often been concentrated within a limited subset of warhead chemotypes rather than broadly diversified across mechanistically and structurally distinct classes (*13*). Bioisosteric replacement strategies that move beyond planar Michael acceptors toward alternative cysteine-targeting CRGs, whether in the early development phase or as part of postapproval optimization (Fig. 1A, center), offer a route to improve selectivity, tunability, and therapeutic indices with reduced discovery costs. To this end, we envisioned combining the thiol reactivity and strain-release properties of bicyclobutanes (BCBs) (*14*, *15*) with the modular functionality of sulfonamides and sulfonimidamides. This merger yields stereodefined, tunable CRGs with favorable physicochemical properties, increased hydrogen-bonding potential, and compatibility with established medicinal chemistry workflows (*16*–*18*). For these groups to be broadly useful, they must be available as reagents that are suitable for late-stage functionalization, exhibit chemical stability, and allow structural diversification in a drug discovery context (Fig. 1B).

In 2016, Baran and colleagues reported a de novo synthesis of aryl sulfone BCBs that were stable and selectively reactive toward thiols under physiological conditions, positioning them as promising CRGs (*14*, *15*). Subsequently, Ojida, Shindo, and colleagues developed carboxy-BCB derivatives using a reagent-based strategy (Fig. 1C), generating amide-linked ibrutinib analogs with improved BTK selectivity, albeit requiring changes to the parent scaffold (*19*). Although these advances underscore the potential of BCBs in covalent design, they suffer from limited scaffold diversity and challenging synthetic routes, hampering broader structure-activity relationship development.

More recently, a one-pot de novo synthesis of sulfonyl BCBs via α-metalation and iterative alkylation and cyclization was reported by Lindsay and Jung (*20*) and later extended by Bull and colleagues to N-protected sulfonimidoyl BCBs (*21*) (Fig. 1C). However, these methods afford only tertiary derivatives, require harsh conditions, and offer limited substrate scope—factors that reduce their suitability for late-stage derivatization and pharmaceutical applications.

To address these challenges, we envisioned two pharmacophore-CRG disconnection strategies, S-N–linked and urea-linked designs (Fig. 1D), to enable installation of sulfonamide and sulfonimidamide BCBs at the final stage in a synthetic sequence. Whereas the S-N–linked design mimics classical acrylamide CRGs, the urea-based motif offers enhanced conformational flexibility (e.g., increased rotatable bonds and accessible nucleophile trajectories) for strategic target engagement while recapitulating key electronic, hydrogen-bonding donor-acceptor capacity, and physiochemical features characteristic of sulfonyl ureas (*22*, *23*).

Here, we report the development of a modular, strain-release CRG platform for the late-stage functionalization of complex amines, enabling the bioisosteric replacement of acrylamides and related electrophiles. We describe the synthesis, chemoselectivity, tunable thiol reactivity, and pharmacokinetic (PK) and pharmacodynamic (PD) profiling of sulfonamide and sulfonimidamide BCBs enabled by the development of S(IV) BCB reagents. We used proteomic and biochemical studies to demonstrate targeted cysteine engagement, which was confirmed by cocrystal structures. Finally, we showcase the translational potential of these CRGs in preclinical models of cancer, underscoring the utility of strain-release covalent inhibition for next-generation drug design.

Reagent development

To enable the late-stage incorporation of sulfonyl and sulfonimidoyl BCB motifs, we sought to develop reagents that function analogously to acryloyl (*24*) or sulfonyl chlorides (*25*) yet offer access to all desired S(VI) BCB structural permutations of ***3*** (Fig. 2A). However, the intrinsic

[1]Drug Discovery Department, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [2]Department of Chemistry, University of South Florida, Tampa, FL, USA. [3]Proteomics and Metabolomics Core, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [4]Cancer Pharmacokinetics and Pharmacodynamics Core, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [5]Department of Molecular Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [6]Department of Oncologic Sciences, College of Medicine, University of South Florida, Tampa, FL, USA. ***Corresponding author. Email: justin.lopchuk@moffitt.org**

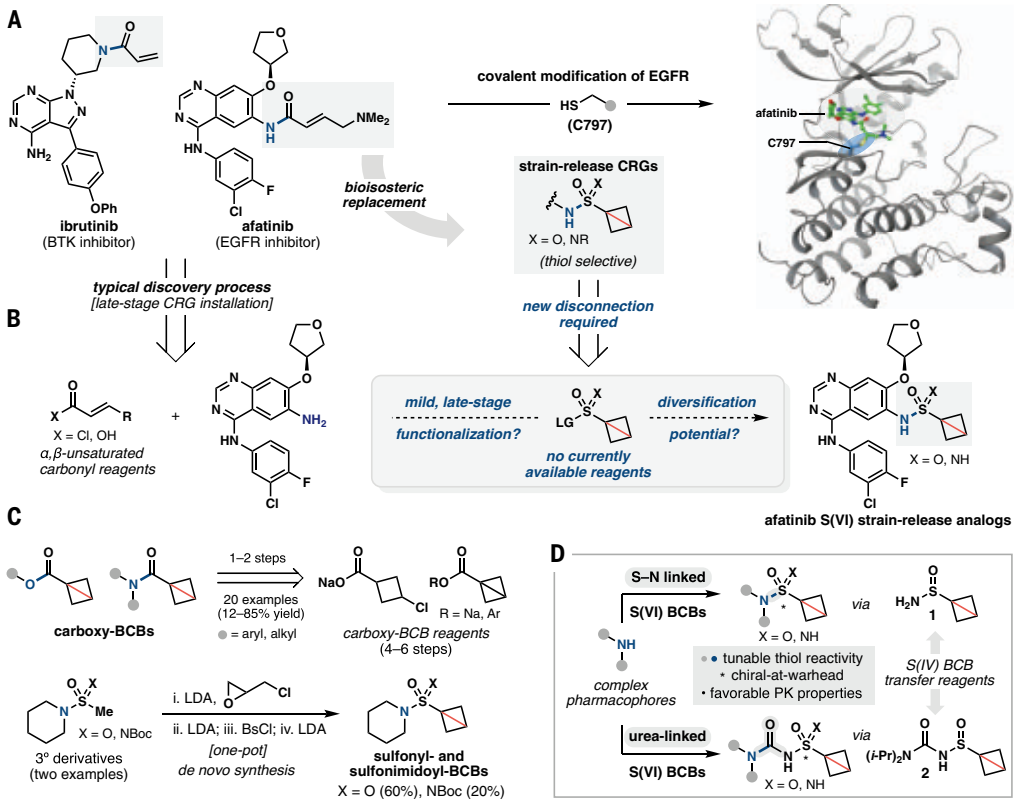


Fig. 1. Covalent inhibition targeting cysteine residues and the context of this work. (**A**) Clinically approved covalent inhibitors containing acrylamide CRGs and proposed replacement with strain-release CRGs. Afatinib bound to EFGR was obtained from the PDB (4G5P) (*10*). (**B**) Methods to incorporate acrylamides and challenges associated with developing analogous methods for strain-release reactive groups. (**C**) Existing methods used to prepare carboxy and S(VI) BCBs. (**D**) Two disconnection approaches are highlighted, S-N–linked and urea-linked, for S(VI) strain-release BCBs to clinical pharmacophores. Boc, *tert*-butoxycarbonyl; LDA, lithium diisopropylamide; LG, leaving group.

electrophilicity of the strained BCB ring allows for alternative reaction pathways, via plausible intermediates **4** and **5**, that pose substantial challenges in achieving site-selective reactivity at the sulfur center using synthetically practical S(VI) reagents.

Initial studies revealed that the preparation of potential reagents was not the primary limitation; rather, poor regioselectivity was the major barrier to viable synthetic access. We investigated a series of sulfonyl and sulfonimidoyl BCB electrophiles bearing varied leaving groups to probe how sulfur polarization influences reactivity with amine nucleophiles (Fig. 2B). After extensive screening of solvents, temperatures, and bases, sulfonyl BCB reagents **6** to **9** yielded the desired monosubstituted product in up to 7% isolated yield, with *N*-methylbenzimidazolium derivative **9** performing best. Bis-addition was frequently observed as the major pathway, likely via intermediate **4**.

Sulfonimidoyl BCB reagents **10** to **13** displayed similar selectivity issues, yielding bis-addition or spirocyclized by-products, particularly when carbonyl-based protecting groups were used. The *N*,*N*-diisopropyl urea-protected sulfonimidoyl BCB fluoride **14** exhibited excellent bench stability (>12 months at ambient temperature) and high regioselectivity at sulfur (>10:1, 77% yield, **15**) in reactions with anilines using NaHMDS in THF [where NaHMDS is sodium bis(trimethylsilyl)amide and THF is tetrahydrofuran; Fig. 2C]. Subsequent deprotection proceeded cleanly to afford NH_2-sulfonimidamide BCB **16** in 84% yield. Unfortunately, attempts to extend BCB S(VI) fluoride exchange (SuFEx) reagent **14** to more complex amines revealed diminished regioselectivity and poor compatibility with aliphatic amines, despite extensive optimization. These limitations highlighted the need for an alternative and more general platform.

We hypothesized that reagents based on the more electron-rich S(IV) oxidation state would display reduced BCB electrophilicity, exhibit improved synthetic utility, and enable orthogonal activation modes. Moreover, S(IV) BCBs could serve as versatile intermediates en route to both S-N–linked and urea-linked S(VI) derivatives (Fig. 2D). The gram-scale synthesis of key intermediates BCB NH_2 **1** and BCB urea **2** was achieved via lithiation of trihalide **17** (*26*) to generate BCB-Li (**18**), followed by reaction with Willis' sulfinyl reagent [triisopropylsilyl sulfinylamine (TIPS-NSO)] (*27*) and the SuFEx reagent developed by our laboratory (*t*-BuSF, where *t*-Bu is *tert*-butyl) (*28*), respectively. Deprotection of TIPS sulfinamide with tetra-*n*-butylammonium fluoride (TBAF) yielded the primary sulfinamide BCB **1** in two steps (67% overall yield), which was further converted to sulfinyl urea BCB **2** in 73% yield with carbamoyl chloride **19**. Alternatively, reagent **2** was accessed directly from *t*-BuSF in 30% yield using a modified protocol with $ZnCl_2$ as an additive. Subsequent transsulfinamidation of primary sulfinamide **1** via release of NH_3 (*29*) and *N*,*N*-diisopropylamine exchange of sulfinyl ureas (*30*) using BCB urea **2** with both aliphatic and aryl amines provided sulfinamide BCB derivatives with complete regioselectivity. These intermediates offer a generalizable route to diversifiable S(IV) species for preparing sulfonamide and sulfonimidamide BCBs.

Reaction scope

With practical S(IV) reagents in hand, we next applied them to the late-stage functionalization of amine pharmacophores found in clinically approved and investigational covalent inhibitors that target cysteine residues. The structural complexity and heterocyclic diversity of these scaffolds presented challenges for S(IV) to S(VI) diversification.

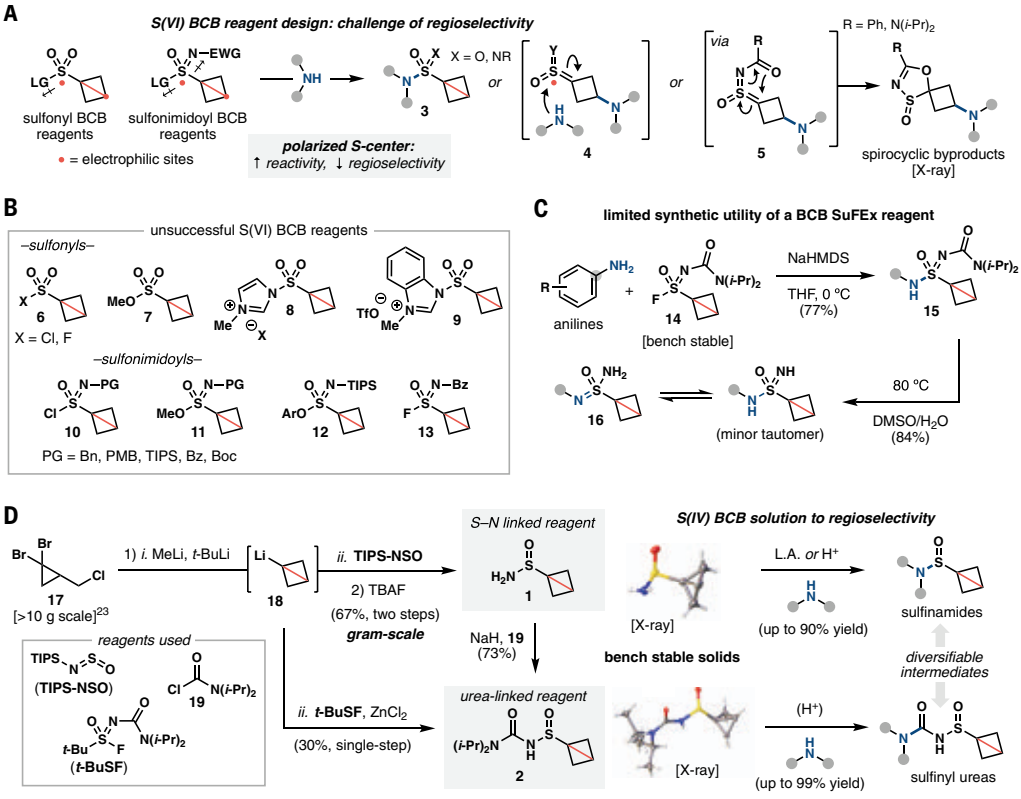


Fig. 2. Reagent synthesis and methods to prepare S(VI) BCBs. (**A**) Challenges and initial reagent and method development strategy. (**B**) Adapted approach using sulfinamide BCBs as versatile intermediates to S(VI) BCB functionality. (**C**) Limited scope of a BCB SuFEx reagent. (**D**) Synthesis of S(IV) BCB reagents used in this work. See SM for additional reaction details and examples. Bz, benzoyl; DMSO, dimethyl sulfoxide; EWG, electron withdrawing group; *i*-Pr, isopropyl; L.A., Lewis acid; PG, protecting group; Tf, trifluoromethanesulfonate.

To address this, we developed multiple S-oxidation and S-oxidative amination strategies to enable the conversion of sulfinamides into sulfonamide and sulfonimidamide BCBs, improving the overall utility of the platform (Fig. 3).

Tertiary acrylamide CRGs are commonly installed in the final synthetic step of covalent inhibitor preparation and are most often derived from secondary aliphatic amines. We began by targeting S-N–linked BCB derivatives through transsulfinamidation. Fully elaborated BTK inhibitor scaffolds (*31*)—ibrutinib, evobrutinib, and acalabrutinib—reacted cleanly with BCB NH_2 **1** and catalytic $Yb(OTf)_3$ to give tertiary sulfinamides **20a**, **21a**, and **22a** in 68 to 90% yield. Oxidation with *meta*-chloroperoxybenzoic (*m*-CPBA) provided sulfonamides **20b**, **21b**, and **22b**, whereas a one-step oxidative amination using hypervalent iodine reagents furnished sulfonimidamides **20c**, **21c**, and **22c** in yields up to 90%.

Analogous methods enabled the late-stage functionalization of $KRAS^{G12C}$ inhibitors adagrasib and ARS-1116 (*32*, *33*), yielding sulfinamides **23a** and **24a**, sulfonamides **23b** and **24b**, and sulfonimidamides **23c** and **24c** using phenyliodine(III)diacetate [$PhI(OAc)_2$] and ammonium carbamate (*34*). The ritlecitinib scaffold (*35*) was similarly converted to sulfinamide **25a** and then oxidized to sulfonamide and sulfonimidamide analogs **25b** and **25c**, respectively. BCB-based analogs of a covalent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) protease inhibitor (*36*) (**26a** to **26c**) were also synthesized efficiently.

Although CRGs derived from primary amines are less common, we investigated the pharamacophores of EN4 (a c-MYC inhibitor) (*37*) and a selective Janus kinase 2 (JAK2) inhibitor (*38*). Secondary sulfinamide BCBs **27a** and **28a** were obtained in high yields (92 and 91%, respectively) and readily oxidized to their corresponding sulfonamides **27b** and **28b**. Direct oxidative amination using $PhI(OAc)_2$ and ammonium carbamate proved inefficient owing to competing S(VI) side products. However, oxidative chlorination of sulfinamides with *t*-BuOCl followed by exposure to an atmosphere of ammonia delivered sulfonimidamide BCBs **27c** and **28c** in 83 and 78% yield, respectively.

Given the prevalence of secondary aryl acrylamides in EGFR inhibitors (*39*), we next explored S(IV) BCB installation on aryl amines (Fig. 4A). Whereas 4-bromoaniline showed excellent conversion to sulfinamide (>90%) via transsulfinamidation, electron-rich or Lewis-basic aryl amines reacted poorly under these conditions. Alternative Brønsted acid–mediated protocols [trifluoroacetic acid (TFA) or acetic acid (AcOH)] enabled sulfinamide formation in these challenging substrates. Using this method, afatinib, dacomitinib, and neratinib were converted to sulfinamides **29a**, **30a**, and **31a**, oxidized to sulfonamides **29b**, **30b**, and **31b** ($RuCl_3/NaIO_4$), and further transformed to sulfonimidamides **29c**, **30c**, and **31c**. A gram-scale transsulfinamidation of the dacomitinib pharmacophore gave **30a** (89% yield), which was oxidized with *m*-CPBA to sulfonamide **30b** for in vivo evaluation (vide infra).

The more complex osimertinib scaffold was converted to sulfinamide **32a** (83% yield, using AcOH), although modified diversification conditions were required. Oxidation with *m*-CPBA afforded sulfonamide **32b** (72%), whereas oxidative amination with phenyliodine(III)bis(trifluoroacetate) [$PhI(CF_3CO_2)_2$] in hexafluoroisopropanol (HFIP) yielded sulfonimidamide **32c** in 55% yield. Additional EGFR inhibitors, olmutinib and rociletinib, underwent sulfinamide formation (**33a** and **34a**, respectively) and subsequent oxidation to sulfonamides (**33b** and **34b**) using *m*-CPBA or a milder method using $PhI(OAc)_2$ with aqueous K_2CO_3. Sulfonimidamides **33c** and **34c** were prepared using modified oxidative amination conditions. Collectively, these tailored oxidation methods enabled access to S(VI) BCB analogs across a broad range of aryl amine substrates.

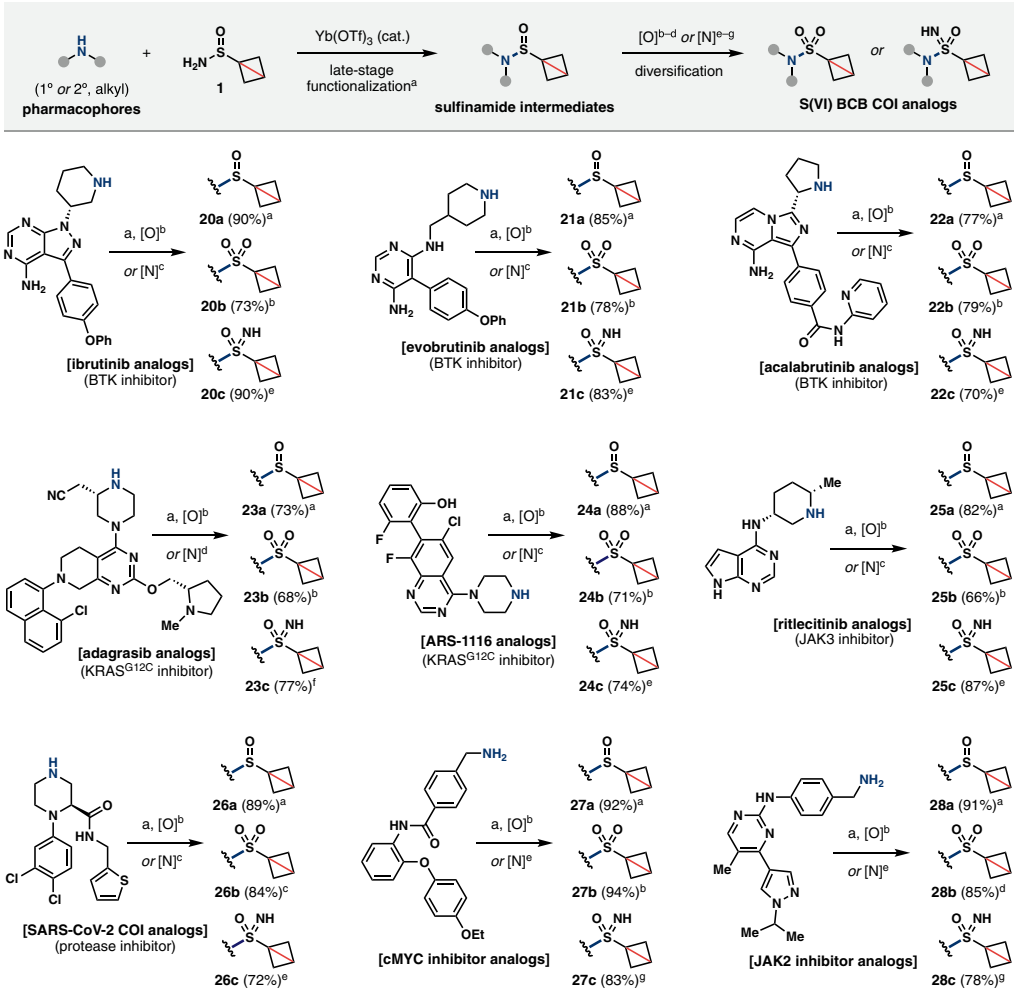


Fig. 3. Substrate scope for the synthesis of S(IV) intermediates and S(VI) strain-release covalent inhibitor derivatives from aliphatic secondary and primary amine pharmacophores. Isolated yields are reported. Reaction conditions are as follows: (a) amine (1 equivalent), **1** (1.2 to 1.5 equiv.), ytterbium(III)trifluoromethanesulfonate [$Yb(OTf)_3$] (10 to 30 mol %), acetonitrile (MeCN) or THF, 40° to 60°C, 4 to 15 hours; (b) sulfinamide intermediate (1 equiv.), *m*-CPBA (1 to 1.2 equiv.), dichloromethane (DCM), 0°C; (c) sulfinamide intermediate (1 equiv.), $RuCl_3$ (1 to 5 mol %), $NaIO_4$ (3 to 6 equiv.), MeCN/THF/H_2O, room temperature (rt); (d) sulfinamide intermediate (1 equiv.), $PhI(OAc)_2$ (1.5 to 2 equiv.), K_2CO_3 (3 equiv.) MeCN/H_2O (4:1), rt, 30 min; (e) amine (1 equiv.), **1** (1.2 to 1.5 equiv.), triethylamine (Et_3N) (3.75 equiv.), $PhI(OAc)_2$ (1.5 to 2 equiv.), THF, rt, 5 to 12 hours; (f) sulfinamide intermediate (1 equiv.), $PhI(OAc)_2$ (2.5 equiv.), ammonium carbamate (4 equiv.), methanol (MeOH), rt, 1 to 6 hours; and (g) sulfinamide intermediate (1 equiv.), *t*-BuOCl (1 to 1.2 equiv.), –40° or 0°C then NH_3 (balloon), 5 to 30 min. See SM section 4 for additional information and examples. COI, covalent inhibitor.

For the urea-linked S(VI) BCB series, we leveraged the reactivity of *N*,*N*-diisopropyl sulfinyl ureas toward amine exchange (*30*) (Fig. 4B). BCB urea reagent **2** underwent efficient substitution with diverse aliphatic and aryl amines, enabling mild CRG installation. Ibrutinib, afatinib, and osimertinib scaffolds were successfully derivatized at room temperature to give intermediates **35a**, **36a**, and **37a**, respectively, the latter two requiring TFA as an additive. Subsequent S(IV) oxidations yielded urea-linked sulfonamides (**35b**, **36b**, and **37b**) and sulfonimidamides (**35c**, **36c**, and **37c**) in good overall yield. This complementary disconnection enhances modularity in CRG design and builds upon the long-standing utility of sulfonyl ureas in drug and agrochemical discovery (*22*, *40*) by introducing a covalent engagement element within the S(VI) framework.

We next prepared an ^{15}N-labeled sulfinamide BCB reagent **^{15}N-1** by transsulfinamidation of **1** with $^{15}NH_4Br$ and in situ generation of sulfinyl reagent **^{15}N-2** (Fig. 4C). Oxidation of ^{15}N-sulfinamides and direct amination of **^{15}N-1** generated isotopically labeled sulfonamide and sulfonimidamide CRGs. For example, reaction with the evobrutinib intermediate afforded sulfonimidamide **^{15}N-21c** with >80% ^{15}N enrichment. Additionally, a one-pot, three-transformation synthesis of a ^{15}N-enriched urea-linked analog of ibrutinib (**^{15}N-35b**) from **^{15}N-1** was achieved in 54% yield while further improving target accessibility. These labeled CRGs expand the utility of the platform into metabolomics, proteomics, PK and PD analysis, and structural biology.

To improve synthetic efficiency in the targeted preparation of S(VI) BCBs, one-pot transsulfinamidation-oxidation protocols were developed and are summarized in the supplementary materials (SM; fig. S22). These conditions enabled direct conversion of primary and secondary amine scaffolds to sulfonamide BCBs **20b**, **26b**, **27b**, **28b**, **30b**, **33b**, **35b**, and **36b** in a single step, providing the corresponding products in 57 to 80% yield. A complementary sequence also furnished sulfonimidamide BCBs **27c** and **35c**.

Strain-release chemoselectivity and reactivity

The chemoselectivity and tunable reactivity of strain-release CRGs were evaluated using a series of S-N–linked and urea-linked S(VI) BCB model compounds alongside representative acrylamides (Fig. 5 and SM section 7). Eight nucleophilic amino acids (10 mM) were incubated

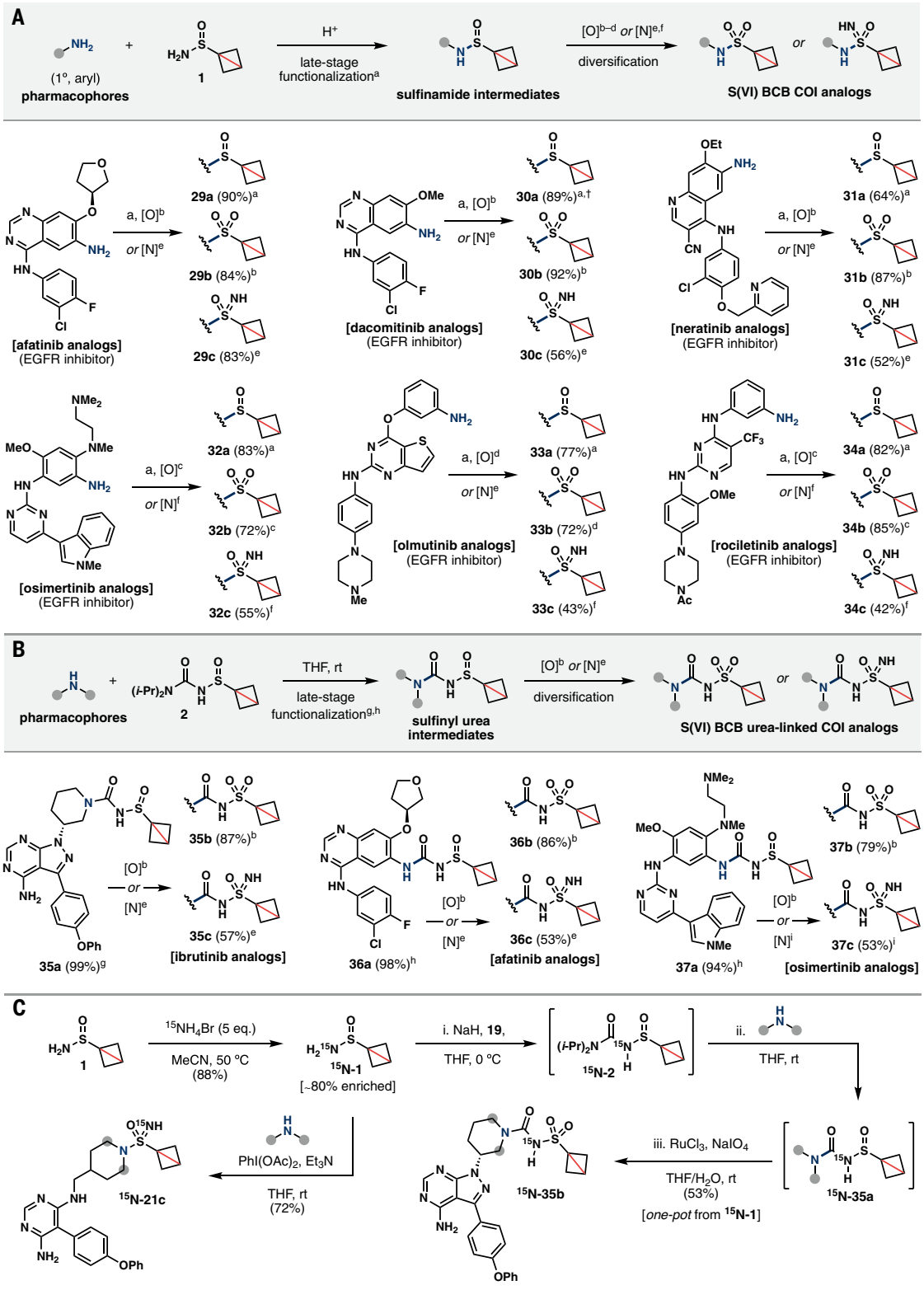


Fig. 4. Substrate scope for the synthesis of S(IV) intermediates and S(VI) strain-release derivatives from aryl primary amine pharmacophores, urea-linked derivatives, and ^{15}N-labeled derivatives. (**A**) S-N–linked S(VI) BCB substrate scope. (**B**) Urea-linked S(VI) BCB substrate scope. (**C**) Preparation of isotopically labeled ^{15}N S(VI) BCB analogs. Isolated yields are reported. ^{15}N enrichment was determined by MS; † indicates gram-scale. Reaction conditions are as follows: (a) amine (1 equiv.), **1** (1.2 to 1.5 equiv.), TFA (3 equiv.), rt, 5 min to 12 hours; (b) sulfinamide intermediate (1 equiv.), $RuCl_3$ (1 to 5 mol %), $NaIO_4$ (3 to 6 equiv.), MeCN/THF/H_2O, rt; (c) sulfinamide intermediate (1 equiv.), *m*-CPBA (1 to 1.2 equiv.), DCM, 0°C; (d) sulfinamide intermediate (1 equiv.), $PhI(OAc)_2$ (1.5 to 2 equiv.), K_2CO_3 (3 equiv.) MeCN/ H_2O (4:1), rt, 30 min; (e) sulfinamide intermediate (1 equiv.), *t*-BuOCl (1.2 equiv.), –40° or 0°C then NH_3 (balloon), 5 to 30 min; (f) amine (1 equiv.), **1** (1.7 equiv.), $PhI(OAc)_2$ or $PhI(CF_3CO_2)_2$ (2 equiv.), HFIP, 0°C, 5 min; (g) amine (1 equiv.), **2** (1.1 equiv.), THF, rt; (h) amine (1 equiv.), **2** (1.1 equiv.), TFA (1 equiv.), THF, rt; and (i) sulfinamide intermediate (1 equiv.), bis(trimethylsilyl)amine [$HN(TMS)_2$] (2.0 equiv.), *N*-chlorosuccinimide (1.2 equiv.), rt, 24 hours. See SM section 4 for additional details and examples. Et, ethyl.

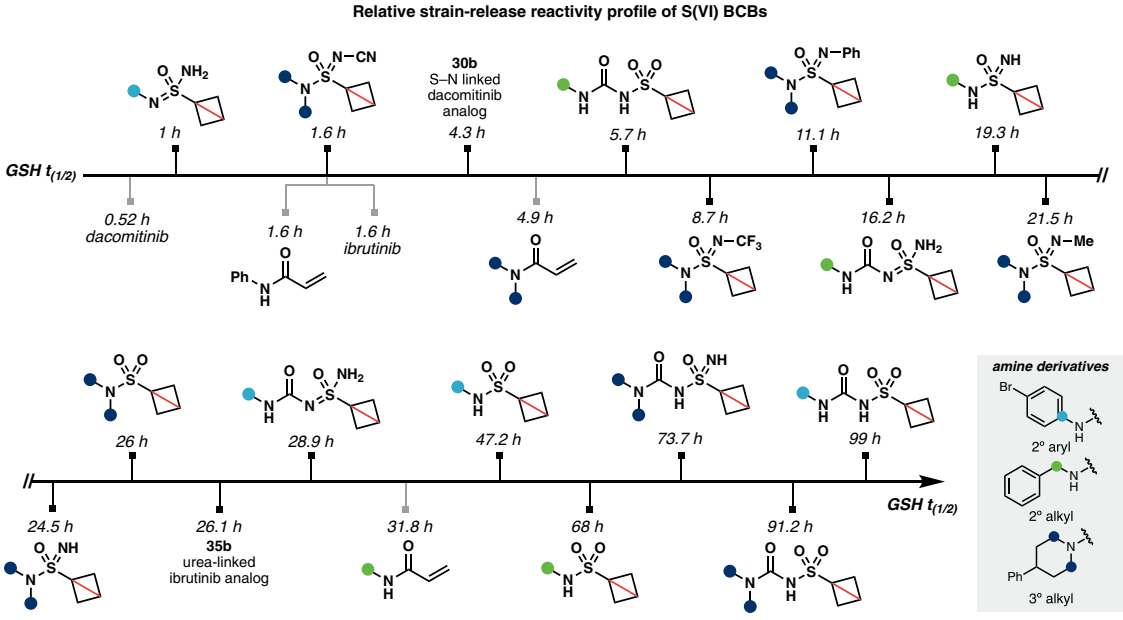


Fig. 5. Relative strain-release reactivity profiling of S(VI) BCBs. Intrinsic reactivity of S(VI) BCBs using a GSH half-life assay. GSH half-life assay conditions are as follows: 90% pH 7.2 100 mM phosphate buffer and 10% *N*,*N*-dimethylacetamide (DMA), 0.5 to 1 mM BCB, 10 mM GSH, at 37°C, monitored over 72 hours. The natural log of electrophile consumption was plotted with respect to time to determine pseudo first-order rate constants and half-life; coefficient of determination ≥0.95. See SM section 7 for additional examples.

with each CRG (1 mM) in 50% aqueous buffer (pH 7.2) over 72 hours to mimic chemical behavior under physiological conditions.

As expected, cysteine displayed the highest reactivity across all CRGs. By contrast, less nucleophilic residues such as histidine, lysine, and serine formed detectable adducts with acrylamides, whereas S(VI) BCBs exhibited exclusive reactivity toward cysteine under the assay conditions. Next, the pH was increased to 10 for approximating the chemoselectivity of lysine and tyrosine in environments where pK_a is lowered. Tyrosine was found to be unreactive, whereas lysine reactivity was observed for the sulfonamide BCB series, albeit substantially lower than that for the acrylamide series (fig. S29). Although designed to capture covalent adducts, reversible reactivity of the acrylamides may preclude detection limits, whereas the irreversible nature of strain-release reactivity in S(VI) BCBs increases confidence in their observed thiol selectivity.

To quantify and compare relative intrinsic reactivities across structurally diverse S(VI) BCBs, a glutathione (GSH) half-life assay was conducted at pH 7.2 (*41*, *42*) (Fig. 5). The resulting GSH half-lives ($T_{1/2}$) ranged from 1 to 99 hours, in contrast to the generally more reactive acrylamides, which showed $T_{1/2}$ values between 0.5 and 32 hours. S(VI) BCB reactivity was strongly influenced by the electronic and structural features of the sulfonyl or sulfonimidoyl group. Incorporation of electron-withdrawing groups (e.g., CN, CF_3, and 4-BrPh, where Ph is phenyl) at the imino nitrogen increased reactivity, whereas electron-donating groups [e.g., H, methyl, alkyl] led to slower adduct formation. Overall, S-N–linked BCBs were more reactive than their urea-linked counterparts, with the exception of a benzyl-substituted sulfonamide ($T_{1/2}$ = 5.7 hours) and sulfonimidoyl ($T_{1/2}$ = 16.2 hours) derivatives.

At physiological pH, the ionizable sulfonamide and sulfonimidamide groups are partially deprotonated (pK_a = 6 to 9) (*43*, *44*), which may increase electron density at sulfur and decrease BCB electrophilicity. However, multiple factors—including steric interactions, solubility, and neighboring group effects—modulate reactivity, and thus model systems serve primarily to establish relative trends. For example, an S-N–linked aryl sulfonamide BCB model compound exhibited a GSH half-life of 47.2 hours, whereas the more functionalized dacomitinib-derived analog **30b** was an order of magnitude more reactive ($T_{1/2}$ = 4.3 hours). A similar comparison was observed with urea-linked derivatives: The ibrutinib-based analog **35b** had a $T_{1/2}$ of 26.1 hours, whereas a tertiary urea model compound reacted more slowly ($T_{1/2}$ = 73.7 hours). These findings underscore the tunability of thiol reactivity within the S(VI) BCB CRGs and demonstrate its potential for selective covalent targeting. Additionally, comparative in vitro PK screening of kinetic aqueous solubility, plasma stability, liver microsomal stability, and permeability revealed no obvious liabilities for the strain-release analogs, with improvements to some properties (tables S2 and S3).

Biochemical and cell-based activity of S(VI) BCB CRGs

To assess whether isosteric replacement of acrylamide groups with strained S(VI) BCB moieties retained enzyme inhibitory activity, we compared the activity of afatinib and dacomitinib with that of their corresponding sulfonamide derivatives **29b** and **30b**. Inhibition of wild-type (WT) EGFR and the $EGFR^{T790M,V948R}$ mutant (which contains mutations Thr^{790}→Met and Val^{948}→Arg) was quantified using a time-resolved fluorescence energy transfer–based kinase assay (*45*). The sulfonamide CRG replacement had little effect on the in vitro inhibitory activity of **29b** and **30b** (Fig. 6A and table S4). To evaluate the compounds further, we calculated the second-order rate constant (k_{inact}/K_I) to gain a measure of the rate of covalent bond formation and overall inactivation efficiency (*46*) (Fig. 6A and fig. S43). As anticipated, the k_{inact}/K_I is high for all four compounds, with comparable maximal covalent inactivation (k_{inact}) despite acrylamide warheads exhibiting higher intrinsic thiol reactivity (GSH half-life). Nevertheless, the k_{inact}/K_I values for afatinib and dacomitinib are ~10-fold higher than those of their respective analogs.

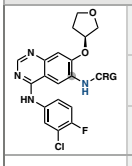
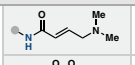
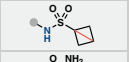
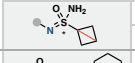
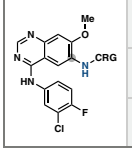
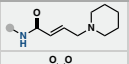
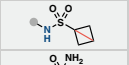
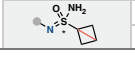
Sulfonimidamide BCBs exhibit higher intrinsic thiol reactivity and introduce stereodefined three-dimensionality to the reactive group. Stereoisomer-resolved biochemical data revealed a pronounced stereochemical dependence on activity, with (–)-**29c** (< 0.9 nM^{WT}, 0.80 $nM^{T790M,V948R}$) and (–)-**30c** (< 0.9 nM^{WT}, 0.55 $nM^{T790M,V948R}$) exhibiting slightly higher activity than their corresponding mixture of stereoisomers at-sulfur (Fig. 6A and table S4). A more substantial change in activity was observed with their (+)-counterparts, (+)-**29c**

(2.3 nM^{WT}, 16.5 nM^{T790M,V948R}) and (+)-**30c** (14 nM^{WT}, 264 nM^{T790M,V948R}). Kinetic parameters could not be determined for the less active enantiomers owing to negligible time-dependent inactivation despite the high GSH reactivity of the aryl sulfonimidamide BCB (Fig. 5), indicating that these species do not form a productive covalent complex under the assay conditions. This disconnect between intrinsic electrophile reactivity and enzymatic covalent engagement suggests that sulfur-centered chirality governs access to a properly aligned prereactive state, emphasizing that binding geometry and warhead presentation are the dominant determinants of covalent inhibition efficiency.

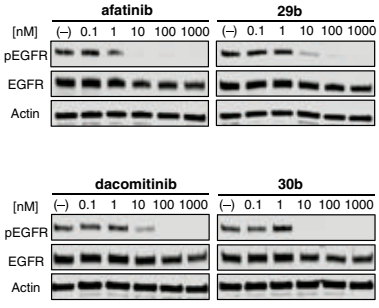
To test whether **29b** and **30b** retained cell-based anticancer activity in EGFR mutant–driven non–small cell lung cancer (NSCLC) cells, we assessed the potency of the compounds in short-term cell-based proliferation assays. Notably, **29b** and **30b** had almost equipotent activity to afatinib and dacomitinib, respectively, with median effective dose (EC_{50}) values in the low single-digit nanomolar range (Fig. 6A and fig. S41). Furthermore, cell killing was consistent with on-target inhibition of EGFR autophosphorylation with total or near total inhibition of Y1068 phosphorylation by administration of 10 nM compound (Fig. 6B) (Y, tyrosine).

The kinase domain of human EGFR harboring the T790M and V948R mutations was cocrystallized with the inhibitor **29b**. The resulting 1.55-Å resolution crystal structure [Protein Data Bank (PDB) ID 9NTP] clearly revealed the inhibitor bound within the adenosine

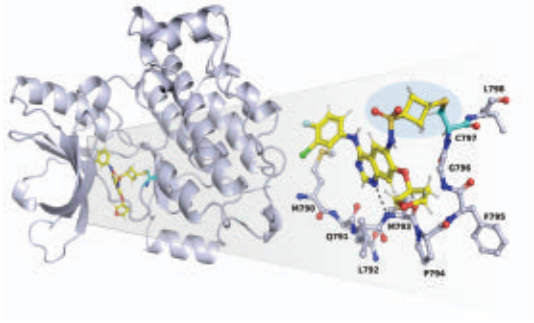
A

Clinical pharmacophore	CRGs	Inhibitor	EGFR^{T790M,V948R} IC_{50} [nM]	EGFR^{WT} IC_{50} [nM]	Cellular EC_{50} [nM]	k_{inact}/K_I ($M^{-1}sec^{-1}$)
		afatinib	0.50 (± 0.05)	< 0.9*	1.33 (± 0.11)	9.97 × 10^6
		**29b**	4.44 (± 0.60)	< 0.9*	2.58 (± 0.36)	1.07 × 10^6
		**(–)-29c**	0.80 (± 0.52)	< 0.9*	5.0 (± 1.3)	0.805 × 10^6
		(+)-29c	16.5 (± 3.0)	2.3 (± 2.6)	59.3 (± 6.9)	ND
		dacomitinib	1.30 (± 0.40)	< 0.9*	1.06 (± 0.29)	9.26 × 10^6
		**30b**	1.60 (± 0.36)	< 0.9*	3.30 (± 0.26)	1.02 × 10^6
		**(–)-30c**	0.55 (± 0.12)	< 0.9*	ND	1.25 × 10^6
		(+)-30c	260 (± 170)	14 (± 12)	ND	ND

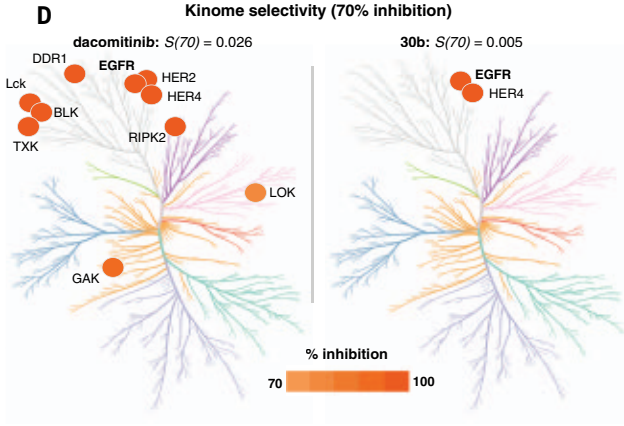
B



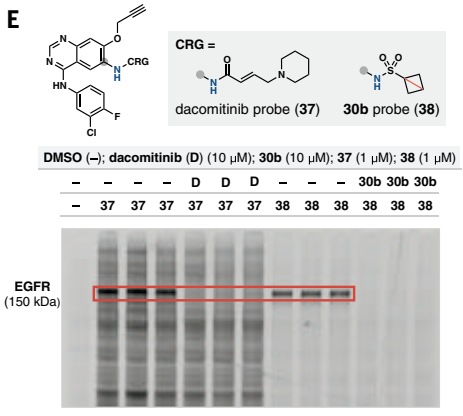
C



D



E



F

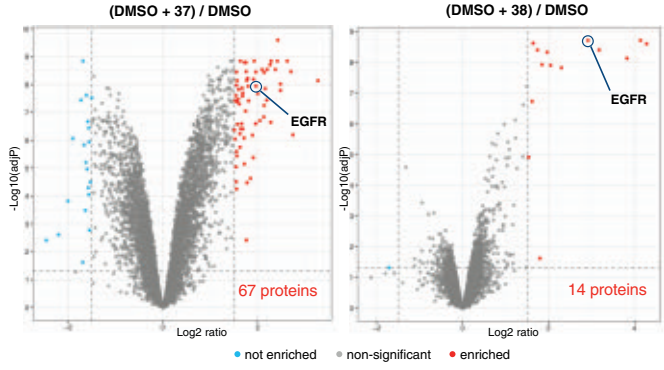


Fig. 6. Biochemistry, crystallography, and in vitro characteristics of S(VI) BCBs. (**A**) Assessment of the biochemical, cellular, and kinetic properties of EGFR tyrosine kinase inhibitors compared with those of S(VI) BCB variants. (**B**) EGFR autophosphorylation analysis by Western blot of the indicated compounds and their BCB analogs in HCC827 cells. (**C**) Crystal structure of BCB variant **29b** bound to C797 in the ATP binding pocket of EGFR. pEGFR, phosphorylated EGFR. (**D**) Kinome selectivity at 70% or greater inhibition for dacomitinib and BCB analog **30b**. (**E**) In situ competitive reactivity profiling of the dacomitinib probe (**37**) and **30b** probe (**38**) in HCC827 cells with in-gel ABPP. (**F**) Global protein enrichment analysis of **37** and **38** in HCC827 cells. All affinity and cell-based activity measurements were performed in triplicate, with error indicated by ±SD. Proteomic experiments were performed in triplicate. Asterisks indicate that measurements less than 50% of the enzyme concentration assayed were reported as <[enzyme]/2; ND indicates not determined. Proteomics data are available through ProteomeXchange (*50*) under identifier PXD075815. F, Phe; G, Gly; L, Leu; P, Pro; Q, Gln.

triphosphate (ATP)–binding site, with the cyclobutane ring forming a covalent bond to the side chain of C797 (Fig. 6C) (C, cysteine). Additionally, a single hydrogen bond was observed between a nitrogen atom in the quinazoline ring and the main-chain amide of M793 in the hinge region. In comparison, a previously resolved crystal structure of EGFR (T790M) bound to afatinib (PDB ID 4G5P) demonstrated a similar binding mode, where the dimethylamino butenamide group covalently interacts with C797 (*10*).

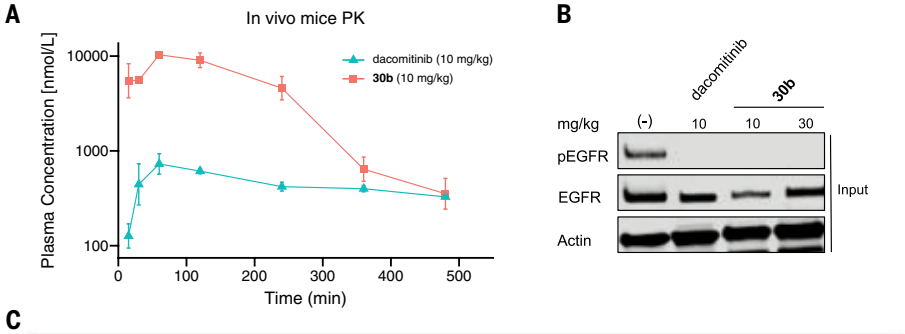
Although irreversible inhibition promotes several distinctive properties that are beneficial for clinical applications, the off-target activity of acrylamide CRGs can be a substantial liability. To determine whether S(VI) BCBs provide enhanced control over cysteine engagement and off-target reactivity under matched experimental conditions, we first qualitatively compared the selectivity of dacomitinib and **30b** across the kinome (Fig. 6D and fig. S44). Compound **30b** is notably more kinome selective than dacomitinib, highlighting that beyond the pharmacophore binding features, the tetrahedral structure and electrostatics of the S(VI) BCB play an important role in kinase target engagement and enhanced selectivity.

Next, in a head-to-head comparison, we compared the global reactivity profiles of dacomitinib and **30b** using in-gel activity-based protein profiling (ABPP) (*12*) and mass spectrometry (MS)–based chemical proteomics (*47*). An NSCLC cell line (HCC827) was exposed to dacomitinib or **30b** followed by treatment with the respective alkynyl probe derivatives of dacomitinib [**37** (*48*)] or **38** and proteins covalently engaged by both probes were isolated and visualized by in-gel fluorescent scanning and liquid chromatography–tandem MS analysis (Fig. 6, E and F). Whereas both probes **37** and **38** showed on-target labeling of EGFR, BCB probe **38** demonstrated significantly fewer off-target interactions (Fig. 6E). Equal protein loading was demonstrated by Coomassie staining (fig. S45). In keeping with these observations, quantitative MS-based analysis revealed that the global reactivity of the dacomitinib probe was markedly higher, with 67 proteins enriched versus 14 by the S(VI) BCB analog (Fig. 6F and figs. S46 and S47). Furthermore, the global reactivity profiles of probes **37** and **38** as a function of time demonstrated that although probe **37** showed a time-dependent accumulation of off-target activity, this was not observed for probe **38**, which also had significantly fewer off-target interactions than **37** (fig. S45).

Whereas acrylamide replacement with the S-N–linked sulfonamide strain-release CRG was accommodated for EGFR targeting, this efficacy was not the case with BTK. Indeed, strain-release derivatives **20b** and **20c** had higher median inhibitory concentrations (IC_{50} than ibrutinib (figs. S39 and S40). Molecular modeling suggested that the distance and orientation of the strained BCB was suboptimal. The more flexible urea-linked sulfonylurea analog **35b** exhibited an almost two orders of magnitude increase in reactivity compared with the S-N–linked derivative. In vitro ADME (absorption, distribution, metabolism, and excretion) profiling of the sulfonylurea (**35b**) and sulfonimidoylurea (**35c**) analogs showed markedly improved solubility (**35b**: 84 μM, **35c**: 83 μM; pH 7.4) and metabolic stability (**35b**: $T_{1/2}$ = 124 min, **35c**: $T_{1/2}$ = 58 min) compared with the parent clinical inhibitor (ibrutinib: 20 μM; pH 7.4; $T_{1/2}$ = 3.4 min; table S2).

Evaluation of S(VI) BCB CRGs in vivo

The PK studies in mice demonstrated substantially improved systemic exposure of **30b** compared with its parent compound dacomitinib, with a ~4-fold increase in area under the curve (AUC) after oral administration and markedly higher maximum serum concentration (C_{max}) values at equivalent doses and formulation (Fig. 7, A and C). The improved exposure is likely related, at least in part, to improved absorption associated with the higher aqueous solubility of the BCB analog, as supported by thermodynamic solubility measurements and visual observations during formulation studies (table S3). Despite exhibiting a shorter half-life and reduced total volume of distribution (V_d), **30b** maintained high oral bioavailability (>100%) and showed lower plasma clearance after intravenous dosing. The relatively



C

Route	Dose	Name	Pharmacokinetic Parameters										
			T_{max}, *min*	C_{max}, *ng/mL*	$AUC_{0-t\ min}$ *ng*min/mL*	$AUC_{0-\infty}$ *ng*min/mL*	$T_{1/2}$ *min*	K_{el} *min^{-1}*	V_d *mL/kg*	MRT *min*	MRT *min*	CL *mL/min/kg*	F, %
IV	1	dacomitinib	-	68.7	11500	20300	397	0.00174	30000	172	173	49	-
PO	10	dacomitinib	60	365	209000	271000	699	0.000992	ND	483	934	-	134
IV	1	**30b**	-	1350	90000	90500	54.6	0.0127	900	82.8	85.0	11	-
PO	10	**30b**	60	4490	1050000	1060000	171	0.00406	ND	172	173	-	117

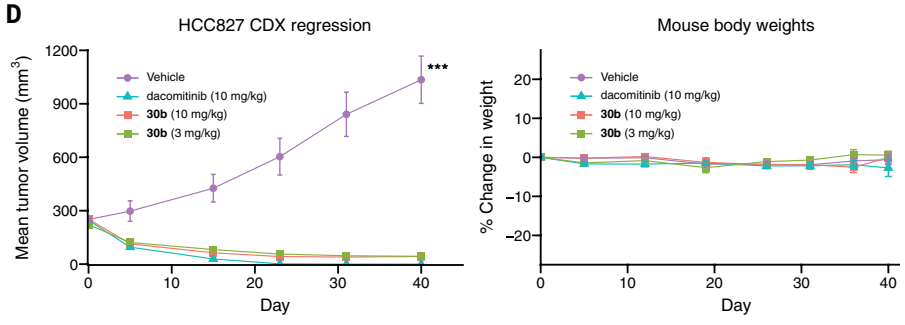


Fig. 7. Pharmacokinetics and in vivo characterization. (**A**) PK properties of the indicated drug and corresponding BCB analog after oral administration at the indicated dose. Plasma concentrations represent geometric mean ± SEM, calculated from log-transformed concentrations (n = 3 mice). (**B**) **30b** is as effective as dacomitinib at suppressing autophosphorylation of EGFR in vivo. HCC827 tumors were harvested from mice 4 hours after oral doses with both compounds (n = 3 mice per group), homogenized, and subjected to SDS-PAGE, and phosphorylation of Y1068 was measured by Western blot analysis. (**C**) Comparison of PK parameters between **30b** and dacomitinib. Dashes indicate parameters not applicable, and ND is not determined. T_{max}, time to maximum plasma concentration; K_{el}, elimination rate constant; MRT, mean residence time; CL, clearance; F, oral bioavailability; IV, intravenous; PO, per os (by mouth). (**D**) As shown on the left, **30b** exhibited similar tumor reduction over time as dacomitinib. HCC827-derived tumors were established in the flank of male NSG mice, followed by randomization into treatment groups once tumors reached 250 to 300 mm^3. Mice were treated with the indicated doses (by mouth, once a day, 5 days per week; n = 9 mice per group) for 40 days, and data are plotted as mean tumor volume ± SEM. As shown on the right, weight changes of mice were <10% for the duration of the study, and data are plotted as mean weight ± SEM. ***P < 0.001 and effect size >0.8 by Wilcoxon rank sum test (see methods).

moderate volume of distribution (~0.9 liters/kg) suggests that **30b** is predominantly retained within the plasma compartment yet remains well within the typical range that is considered optimal for oral small molecules that target systemic tissues (0.5 to 3 liters/kg) (*49*). Taken together, these results indicate that incorporation of the strain-release BCB motif improved the drug-like properties in this series, leading to a more favorable ADME-PK profile and improved systemic exposure relative to dacomitinib.

PD evaluation in HCC827 tumor–bearing mice confirmed robust, dose-dependent target inhibition, with both **30b** and dacomitinib effectively suppressing EGFR phosphorylation in tumor tissues (Fig. 7B and fig. S48). To assess in vivo efficacy, HCC827 cell–derived xenograft (CDX) models were established. Once tumors reached 250 to 300 mm^3, animals were randomized into four groups and treated by oral gavage with vehicle, dacomitinib (10 mg/kg), or **30b** at 3 or 10 mg/kg. Notably, **30b** promoted rapid tumor regression at both treatment doses, with an efficacy indistinguishable from that of dacomitinib treatment (Fig. 7D, left). No palpable tumors were visible in the treated animals, and none of the animals in the treatment groups reached the defined end point over the course of the experiment. All treatment regimens were well tolerated, with no significant changes in body weight observed throughout the study (Fig. 7D, right).

Conclusions

We report the development of a modular reagent platform for the late-stage installation of sulfonamide and sulfonimidamide BCB motifs as tunable strain-release CRGs. This approach enables precise bioisosteric replacement of acrylamides across diverse pharmacophores while preserving or enhancing drug-like properties. Biochemical, proteomic, and in vivo studies confirm that S(VI) strain-release CRGs retain target engagement and inhibitory activity, with reduced off-target reactivity and tractable PK profiles. Preclinical validation highlights the translational potential of this approach, positioning strain-release S(VI) BCBs as valuable assets for next-generation covalent drug discovery.

REFERENCES AND NOTES

1. J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
2. E. De Vita, *Future Med. Chem.* **13**, 193–210 (2021).
3. L. Hillebrand, X. J. Liang, R. A. M. Serafim, M. Gehringer, *J. Med. Chem.* **67**, 7668–7758 (2024).
4. G. Kim, R. J. Grams, K.-L. Hsu, *Chem. Rev.* **125**, 6653–6684 (2025).
5. S. E. Dalton, O. Di Pietro, E. Hennessy, *J. Med. Chem.* **68**, 2307–2313 (2025).
6. F. Sutanto, M. Konstantinidou, A. Dömling, *RSC Med. Chem.* **11**, 876–884 (2020).
7. N. V. Mehta, M. S. Degani, *Drug Discov. Today* **28**, 103799 (2023).
8. R. J. Grams, K. L. Hsu, *Trends Pharmacol. Sci.* **43**, 249–262 (2022).
9. D. Schaefer, X. Cheng, *Pharmaceuticals (Basel)* **16**, 663 (2023).
10. F. Solca *et al.*, *J. Pharmacol. Exp. Ther.* **343**, 342–350 (2012).
11. S. De Cesco, J. Kurian, C. Dufresne, A. K. Mittermaier, N. Moitessier, *Eur. J. Med. Chem.* **138**, 96–114 (2017).
12. B. R. Lanning *et al.*, *Nat. Chem. Biol.* **10**, 760–767 (2014).
13. L. Zheng *et al.*, *MedComm Oncol.* **2**, e56 (2023).
14. R. Gianatassio *et al.*, *Science* **351**, 241–246 (2016).
15. J. M. Lopchuk *et al.*, *J. Am. Chem. Soc.* **139**, 3209–3226 (2017).
16. M. Frings, C. Bolm, A. Blum, C. Gnamm, *Eur. J. Med. Chem.* **126**, 225–245 (2017).
17. U. Lücking, *Org. Chem. Front.* **6**, 1319–1324 (2019).
18. P. Mäder, L. Kattner, *J. Med. Chem.* **63**, 14243–14275 (2020).
19. K. Tokunaga *et al.*, *J. Am. Chem. Soc.* **142**, 18522–18531 (2020).
20. M. Jung, V. N. G. Lindsay, *J. Am. Chem. Soc.* **144**, 4764–4769 (2022).
21. Z. Zhong *et al.*, *Angew. Chem. Int. Ed.* **64**, e202420028 (2025).
22. A. K. Ghosh, M. Brindisi, *J. Med. Chem.* **63**, 2751–2788 (2020).
23. A. Pulgar *et al.*, *J. Phys. Chem. A* **128**, 5941–5953 (2024).
24. C. C. Ayala-Aguilera *et al.*, *J. Med. Chem.* **65**, 1047–1131 (2022).
25. A.-M. D. Schmitt, D. C. Schmitt, in *Synthetic Methods in Drug Discovery*, vol. 2, D. C. Blakemore, P. M. Doyle, Y. M. Fobian, Eds. (Royal Society of Chemistry, 2016), pp. 123–138.
26. B. D. Schwartz, M. Y. Zhang, R. H. Attard, M. G. Gardiner, L. R. Malins, *Chemistry* **26**, 2808–2812 (2020).
27. M. Ding, Z. X. Zhang, T. Q. Davies, M. C. Willis, *Org. Lett.* **24**, 1711–1715 (2022).
28. S. Teng, Z. P. Shultz, C. Shan, L. Wojtas, J. M. Lopchuk, *Nat. Chem.* **16**, 183–192 (2024).
29. D. Wen, Q. Zheng, C. Wang, T. Tu, *Org. Lett.* **23**, 3718–3723 (2021).
30. P. R. Athawale, Z. P. Shultz, A. Saputo, Y. D. Hall, J. M. Lopchuk, *Nat. Commun.* **15**, 7001 (2024).
31. A. Cool, T. Nong, S. Montoya, J. Taylor, *Trends Pharmacol. Sci.* **45**, 691–707 (2024).
32. J. B. Fell *et al.*, *J. Med. Chem.* **63**, 6679–6693 (2020).
33. M. R. Janes *et al.*, *Cell* **172**, 578–589.e17 (2018).
34. F. Izzo, M. Schäfer, R. Stockman, U. Lücking, *Chemistry* **23**, 15189–15193 (2017).
35. A. Thorarensen *et al.*, *J. Med. Chem.* **60**, 1971–1993 (2017).
36. S. Gao *et al.*, *J. Med. Chem.* **65**, 16902–16917 (2022).
37. L. Boike *et al.*, *Cell Chem. Biol.* **28**, 4–13.e17 (2021).
38. Y. Tian *et al.*, *ACS Med. Chem. Lett.* **14**, 1113–1121 (2023).
39. M. Hossam, D. S. Lasheen, K. A. Abouzid, *Arch. Pharm.* **349**, 573–593 (2016).
40. J. V. Hay, *Pestic. Sci.* **29**, 247–261 (1990).
41. V. J. Cee *et al.*, *J. Med. Chem.* **58**, 9171–9178 (2015).
42. M. E. Flanagan *et al.*, *J. Med. Chem.* **57**, 10072–10079 (2014).
43. M. Remko, *J. Mol. Struct. THEOCHEM* **944**, 34–42 (2010).
44. S. R. Borhade *et al.*, *ChemMedChem* **10**, 455–460 (2015).
45. S. Niessen *et al.*, *Cell Chem. Biol.* **24**, 1388–1400.e7 (2017).
46. L. K. Mader, J. E. Borean, J. W. Keillor, *RSC Med. Chem.* **16**, 63–76 (2024).
47. A. J. Rabalski, A. R. Bogdan, A. Baranczak, *ACS Chem. Biol.* **14**, 1940–1950 (2019).
48. C. M. Browne *et al.*, *J. Am. Chem. Soc.* **141**, 191–203 (2019).
49. D. A. Smith, K. Beaumont, T. S. Maurer, L. Di, *J. Med. Chem.* **58**, 5691–5698 (2015).
50. E. W. Deutsch *et al.*, *Nucleic Acids Res.* **51**, D1539–D1548 (2023).
51. A. Lee-Sam, R code for rstatix analysis. Zenodo (2026); https://doi.org/10.5281/zenodo.19240177.

ACKNOWLEDGMENTS

This work was supported in part by the Chemical Biology Core Facility, Proteomics and Metabolomics Core, and Cancer Pharmacokinetics and Pharmacodynamics (PK/PD) Core at the H. Lee Moffitt Cancer Center and Research Institute, a National Cancer Institute (NCI)–designated Comprehensive Cancer Center (P30-CA076292, Moffitt). We thank H. Lawrence (Moffitt), M. Tantak (Moffitt), G. Mahankali (Moffitt), and S. Kallepu (Moffitt) for nuclear magnetic resonance and high-resolution MS support. We also thank M. Teng (Moffitt) for assistance with statistical models and E. Welsh (Moffitt) for assistance with proteomic data analysis. Protein x-ray data collection was performed using the 19-ID/NYX beamline at the National Synchrotron Light Source II, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under contract no. DE-SC0012704. A.G. is affiliated with the University of Florence, and financial support provided by the PhD Program in Drug Discovery Research and Innovative Treatments, University of Florence. **Funding:** This work was funded by National Institutes of Health grant NIGMS R35-GM142577 (J.M.L.) for support of the reagent development and chemical synthesis; the Moffitt Cancer Center Lung Cancer Center of Excellence pilot project (J.M.L., D.D.) for support of the biochemical studies; a Moffitt Team Science Award (J.M.L., E.S.) for support of the cocrystallization studies; and the Moffitt Foundation (J.M.L., D.D.) for support of the in vivo experiments. National Institutes of Health grant NCI P30-CA076292 provides partial support for the Core services and establishes Moffitt's status as an NCI Comprehensive Cancer Center. **Author contributions:** Conceptualization: J.M.L., D.D., A.M., Z.P.S.; Methodology: Z.P.S., A.L.-S., Y.-P.C., L.S., J.K., E.S., A.M., D.D., J.M.L.; Investigation: Z.P.S., A.L.-S., Y.-P.C., L.S., D.G., A.G., K.P., T.S., V.I., B.F., S.S., R.K., L.W.; Visualization: Z.P.S., A.L.-S., L.S., B.F., L.W., E.S., A.M., D.D., J.M.L.; Funding acquisition: J.M.L., D.D.; Project administration: J.M.L., D.D.; Supervision: J.M.L., D.D., A.M., E.S., J.K.; Writing – original draft: Z.P.S., A.L.-S., A.M., D.D., J.M.L.; Writing – review & editing: Z.P.S., A.L.-S., Y.C., T.S., J.K., E.S., A.M., D.D., J.M.L. **Competing interests:** Patent applications naming J.M.L., D.D., A.M., and Z.P.S. as inventors have been filed by the H. Lee Moffitt Cancer Center and Research Institute. J.M.L., D.D., and A.M. are the scientific cofounders of Thyora Therapeutics. All other authors declare that they have no competing interests. **Data, code, and materials availability:** All experimental procedures, graphical procedures for chemical synthesis, compound characterization data, and stability analysis data are available within the article and the supplementary materials. X-ray crystallographic data for the small-molecule structures have been deposited with the Cambridge Crystallographic Data Centre. The data can be obtained free of charge from https://www.ccdc.cam.ac.uk/structures/. Compounds with x-ray structures are as follows: **1** (CCDC 2473999), **2** (CCDC 2474002), **14** (CCDC 2474001), **S1** (CCDC 2474003), and **S94** (CCDC 2474000). Atomic coordinates and structure factors for the reported crystal structures have been deposited in the PDB (www.rcsb.org) under PDB ID pdb_00009ntp. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD075815 and 10.6019/PXD075815. The R code for rstatix analysis of tumor regression is publicly available on Zenodo (*51*). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx7219
Materials and Methods; Figs. S1 to S53; Tables S1 to S28; References (*52*–*76*); Reproducibility Checklist

Submitted 28 July 2025; accepted 20 May 2026

10.1126/science.adx7219

SURFACE CHEMISTRY

Spatially resolving the cone of reaction for a single molecule

Matthew James Timm[1], Adam Matěj[2,3], Ilias Gazizullin[1], Qifan Chen[2,4], Stefan Hecht[5], Pavel Jelínek[2], Leonhard Grill[1]*

Collisions of atoms and molecules are required for bond formation and are thus key to any chemical reaction. Their outcome is determined by the collision energy, relative orientation, and impact parameter of the reactants. Collision studies in the gas phase have shown that energies are easily controlled, whereas it is difficult to steer the molecular orientations and completely impossible to control the impact parameter. Here, we study the collision of individual molecules at a surface where each involved compound is imaged before and after collision by scanning tunneling microscopy, allowing precise control of orientations and impact parameter. We observe that the molecules must collide in a narrow cone of reaction and that surface atom displacements can enable a reaction even for disfavored pathways.

When two atoms or molecules collide in a chemical reaction, typically in solution or the gas phase, the geometries of their collisions are statistically distributed (*1*–*5*). The ability to control the collision, in particular the impact parameter *b* (Fig. 1A), requires precisely confining the paths along which colliding reactants can travel. Early approaches to restrict the possible impact parameters have involved performing photoinduced reactions within van der Waals complexes (*6*) or with molecules aligned at surfaces in "surface-aligned reactions" (*7*–*20*). In the latter, the possible impact parameters are greatly reduced in comparison to the gas phase owing to the underlying crystal affecting the pathways of adsorbed reactants (*7*). By taking advantage of the geometry of the underlying surface, the possible impact parameters can be further reduced—as demonstrated for the photoinduced reaction of O (ads) + CO (ads) → CO_2 (g) on Pt(*339*) and Pt(*779*) single-crystal surfaces, where the step edges channeled "hot" photogenerated O atoms preferentially toward CO molecules adsorbed at the surface steps rather than those at the terraces (*10*).

Besides surface steps, the intrinsic corrugation of the Cu(110) surface can collimate the trajectories of hot dissociation products arising from electron-induced reactions of molecular adsorbates, as shown by Polanyi and co-workers (*17*–*19*). This has been exploited to control the impact parameter of collisions between hot molecular CF_2 "projectiles" with chemisorbed "target" reactants (*18*, *19*). The outcomes of these collisions were shown to depend on their impact parameter, with small values giving rise to bond formation between the projectile and target. However, these studies could not address another important aspect affecting the reactivity, which is the relative orientation of reactants, φ (Fig. 1A). Its importance has been demonstrated by Brook and Jones (*21*) and Beuhler and co-workers (*22*) in gas-phase crossed molecular beam experiments almost 60 years ago, where different reactivities were observed for the reaction M + CH_3I → MI + CH_3 (M is K or Rb), depending on the orientation of the CH_3I species with respect to the incoming metal atom. A more recent study shows that for gas-phase collisions of K atoms with oriented C_6H_5I molecules, the reaction is 28 times more likely if the K atom approaches from the I-atom end as compared to the phenyl end (*23*). Simultaneous control over both parameters *b* and φ, however, has not been achieved so far, yet is highly desirable as it would provide insight into the complete set of steric requirements governing a reaction.

In this work, we demonstrate how simultaneous control over both parameters can be realized. We find that only collisions involving specific geometries and impact parameters led to reaction on a Cu(110) surface, as determined by scanning tunneling microscopy (STM).

How to direct a collision

Figure 1A shows how simultaneous control over the impact parameter, *b*, and the relative alignment of reactants, φ, is achieved. The impact parameter of the collision is determined through the choice of close-packed Cu row along which the CF_2 projectile travels (*18*, *19*). While *b* describes the distance between the projectile and the center of mass of the target, the more important parameter here is *b*′, which we define as the miss distance to the binding site of the target (Fig. 1A and fig. S1). The sign of *b*′, positive or negative, is defined by whether CF_2 travels along a row "above" or "below" the binding site of the target, respectively (fig. S2).

As target molecules, we have chosen dibromoterfluorene (DBTF; $C_{45}H_{36}Br_2$) (*24*, *25*) as it exhibits three characteristics that are important for our collision study: (i) substituent Br atoms at the termini, which can be controllably dissociated by STM manipulation (*26*); (ii) a distinct rod like shape that allows clear identification of the molecular orientation; and (iii) various orientations on the surface upon adsorption. The DBTF molecules were deposited onto the cold Cu(110) surface (see materials and methods), where they adsorb intact, appearing with three lobes (due to the dimethyl groups) and two shoulders (from the terminal Br atoms) in STM images (Fig. 1B and fig. S3A). To improve imaging contrast, the STM tip has been functionalized with an I atom (see materials and methods).

The CF_3 species, which serves as the projectile source, is obtained through deposition of CF_3I onto the Cu(110) surface at a temperature below 9 K, whereupon it decomposes into chemisorbed CF_3 fragments and I atoms (Fig. 1B). An example of CF_2 projectile formation through electron-induced dissociation of a chemisorbed CF_3 "projectile source" is presented in Fig. 1C. Pulsing on CF_3 with 1.5 V (Fig. 1C, initial) causes it to split into a F atom and a CF_2 fragment, which recoiled over 12 Å from the initial CF_3 position (Fig. 1C, final) along the underlying Cu row [an STM image showing such rows of close-packed Cu atoms is given in Fig. 1B (inset)]. The range of projectile kinetic energies can only be estimated. The lower boundary must be above the diffusion barrier for translation along the copper row, which we found to be 0.62 eV (see fig. S4). We estimate the upper boundary to be 2.0 eV from calculations by Anggara *et al* (*18*).

Our calculations suggest that the CF_2 projectile is chemisorbed on the copper row (fig. S5) and retains its orientation, with the two F atoms pointing off the copper row, during its motion (fig. S4). The corrugated Cu(110) plays a key role in this regard as it ensures linear and directed motion of the projectile. On flat Cu(111), for instance, a molecule has additional paths it can follow [three directions as compared to one on Cu(110)]. It is also likely that CF_2 can rotate more easily on Cu(111) owing to the lack of corrugation, reducing its confinement to travel along one surface direction.

The relative alignment of the reactants with respect to the Cu rows of the surface is determined by the orientation, φ, of the target, which is the angle between the long axis of the molecule and the [1$\bar{1}$0] surface direction (Fig. 1A). Thus, the ability to probe the effect of different φ values on the outcome of reaction is defined by the number of orientations the target can adopt at the surface, which can be precisely determined from STM images. Various DBTF adsorption alignments are found (fig. S3). The four major orientations with abundances between

[1]Institute of Chemistry, University of Graz, Graz, Austria. [2]Institute of Physics, Czech Academy of Sciences, Prague, Czech Republic. [3]Department of Physical Chemistry, Faculty of Science, Palacký University, Olomouc, Czech Republic. [4]Faculty of Mathematics and Physics, Charles University, Praha, Czech Republic. [5]Department of Chemistry and Center for the Science of Materials Berlin, Humboldt-Universität zu Berlin, Berlin, Germany. ***Corresponding author. Email: leonhard.grill@uni-graz.at**

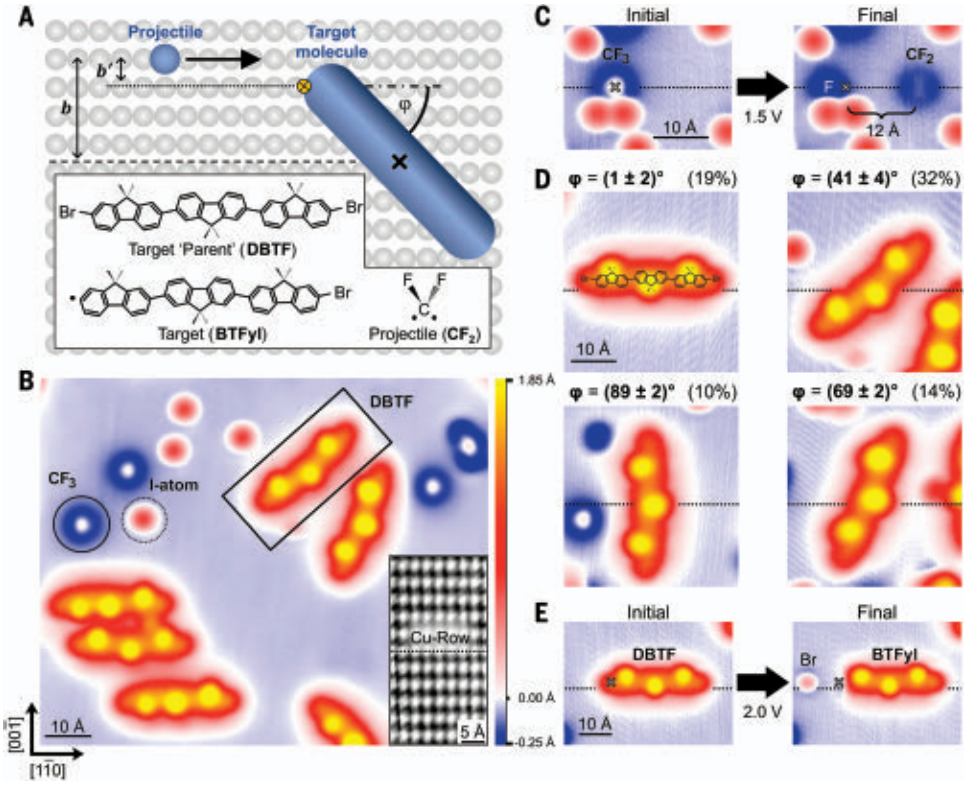


Fig. 1. Directed collision between the CF_2 projectile and BTFyl target. (**A**) Schematic: Projectile traveling along the Cu row, toward chemisorbed target species with alignment angle φ. Double-headed arrows mark the impact parameter (*b*) and distance from the unsaturated C atom (*b′*). Black and orange crosses highlight the center of mass and unsaturated C atom of target species, respectively. (Inset) Chemical structures of the DBTF parent ($C_{45}H_{36}Br_2$), BTFyl target ($\bullet C_{45}H_{36}Br$), and CF_2 projectile. (**B**) Overview STM image showing CF_3 (black circle), DBTF (black rectangle), and chemisorbed I atom (dotted black circle). (Inset) Atomic-resolution STM image of the Cu(110) surface. (**C**) STM images showing CF_3 dissociation to give CF_2, which recoiled along the Cu row (black dashed line). (**D**) Main adsorption alignments of DBTF relative to the Cu row (black dashed line), folded to 90°. (**E**) DBTF converted to partially debrominated target, BTFyl. White crosses in (C) and (E) denote the STM tip position during (C) 1.5-eV or (E) 2.0-eV pulses to dissociate (C) C–F or (E) C–Br bonds. STM images acquired at 6.3 K with imaging conditions: (B and C) $V = -50$ mV, $I = 50$ pA; [(B), inset] $V = -50$ mV, $I = 170$ nA; [(D), top right, and (E)] $V = -20$ mV, $I = 0.2$ nA; [(D) rest] $V = -50$ mV, $I = 50$ pA.

10 and 32% are presented in Fig. 1D, as well as in fig. S6, where the location of the Cu rows, which lie beneath the DBTF, can be seen. Furthermore, DBTF molecules can adopt different configurations upon adsorption due to the fluorene units being able to rotate around their linking C–C single bonds (see fig. S3B). Here, we investigated only the trans–trans DBTF conformation, which serves as the precursor and thus "parent" molecule for the BTFyl target.

To start the collision experiment, the intact DBTF parent molecule is converted into BTFyl ($\bullet C_{45}H_{36}Br$) by applying a 2.0-V pulse above a selected C–Br bond (*26*), with the resulting C–Br dissociation producing the chemisorbed BTFyl fragment (Fig. 1E) that acts as the target in the collision experiment. Notably, it has a similar adsorption alignment to that of its parent, albeit displaced slightly backward (by 2.0 Å) from its parent's initial position. Density functional theory (DFT) calculations aided in identifying the adsorption geometry of DBTF and BTFyl (fig. S7, A and B, respectively). These calculations show that the carbon radical of BTFyl binds to an "atop" Cu atom, which is pulled out by 1.3 Å from its initial position in the copper row underneath the prior C–Br bond (fig. S7B), similar to the displacement of a copper atom arising from the binding of a carbon radical to Cu(111) (*27*).

With this setup, collisions between the CF_2 projectile and the BTFyl target can be induced very precisely in terms of impact parameter and compound orientations. This high control is enabled by the anisotropic Cu(110) surface. The molecule–surface interaction leads to specific molecular adsorption configurations, reducing the parameter space for collisions as compared to the gas phase. The generalizability of collisions on a surface and comparison to the gas phase is therefore not straightforward.

Directed collisions between a projectile and target

To gain detailed insight into the cone of reaction, we have performed collision experiments for various miss distances and orientations of the BTFyl target. Two characteristic examples are given in Fig. 2. The first (Fig. 2A and fig. S8) demonstrates collision with a target aligned at $\varphi = -41°$. BTFyl was created through a 2.0-V pulse on the left-hand C–Br bond of DBTF, resulting in dissociation of that bond (Fig. 2, A and B). The covalent binding of BTFyl to the surface is evidenced by a slight depression near the prior location of the C–Br bond (apparent height, −0.18 Å). This depression, which has also been observed for this molecule on Ag(111) (*26*), is attributed to depletion of electronic density due to the formation of a σ bond between the BTFyl radical and an atop Cu atom (*28*). The lower appearance of the dimethyl group at this end of the molecule, with an apparent height of 1.74 Å (compared with 1.89 Å at the unreacted end), is further evidence for BTFyl surface binding (*26*, *27*). The formation of a chemical bond is supported by DFT-calculated differential electron densities (see fig. S5B) and a calculated C–Cu distance of 1.96 Å between the Cu surface and "unsaturated" carbon atom. Because of this C–Cu bond, the dimethyl group at the bound end of BTFyl is 0.10 Å closer to the surface than the other two dimethyl groups (fig. S7).

Next, the CF_2 projectile was produced and accelerated toward the target through a 1.5-V pulse on the nearby CF_3. The distance traveled by the projectile was 12 Å before collision with the target, resulting in the outcome shown in Fig. 2A ("Collision"). Here, CF_2 appears as a small protrusion with an apparent height of 0.25 Å, located 1.3 Å to the left of the binding point of BTFyl. The central question now becomes what is the chemical composition of the collision outcome, i.e., whether a bond was formed or not. We studied this through lateral manipulation with the STM tip (*29*–*31*) (see section S6.1 in the supplementary materials). Specifically, the STM tip was moved across the CF_2–BTFyl agglomerate (along the black arrow in Fig. 2C). The outcome of this manipulation experiment, which is shown below, reveals that the CF_2 species was pushed away and BTFyl remained unchanged—thus, the two compounds were separated, even under these rather soft manipulation parameters (tunneling resistance at least 33 MΩ). Comparing images and height profiles of BTFyl before collision and after CF_2 was driven away reveals that the BTFyl appearance did not change (fig. S8). Hence, BTFyl was not altered and the collision with CF_2 was unproductive.

A different result is obtained if a target alignment of $\varphi = 1°$ is chosen (Fig. 2B). As before, BTFyl was created by a 2.0-V pulse on the C–Br

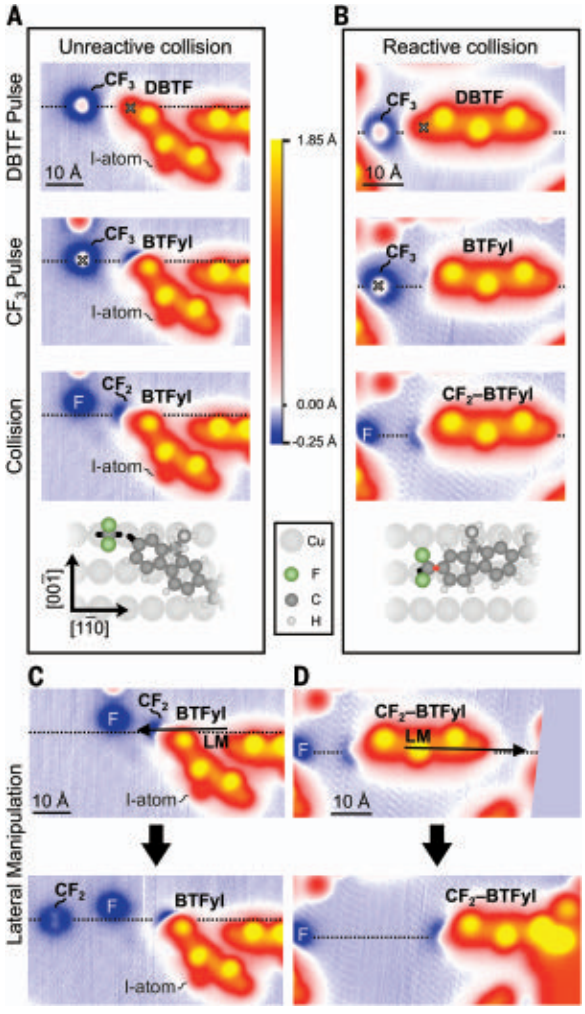


Fig. 2. Directed collisions between the CF_2 projectile and BTFyl target. (**A** and **B**) STM images of directed collisions between the CF_2 projectile and two different BTFyl targets [(A) $b' = 1.2$ Å, $b = 10.2$ Å, $\varphi = -41°$; (B) $b' = -1.3$ Å, $b = -1.6$ Å, $\varphi = 1°$], with schematics of "unreactive" (A) and "reactive" (B) collision outcomes given below (black lines: metal–carbon bonds; red line: newly formed carbon–carbon bond). A 2.0-V pulse on the DBTF parent by the STM tip (at the white cross in the "DBTF pulse") produces BTFyl. A 1.5-V pulse on chemisorbed CF_3 (at the white cross in the "CF_3 pulse") produces, and accelerates, the CF_2 projectile along the Cu row (black dotted line) toward BTFyl, resulting in no reaction (A) or reaction (B) following collision. (**C** and **D**) Collision outcomes demonstrated through lateral manipulations (along black lines marked "LM") of collision "products" (see section S6.1 in the supplementary materials for details). The I atom adjacent to BTFyl in (A) and (C) was a spectator, playing no role in the collision outcome. STM images acquired at 6.3 K with imaging conditions: (A) $V = -20$ mV, $I = 50$ pA; (B) $V = -50$ mV, $I = 50$ pA.

bond, resulting in its dissociation and a slight depression (apparent height, −0.02 Å) at the location of the previous C–Br bond. Then, the CF_2 projectile was accelerated toward BTFyl. For comparability, we use the same pulse voltage as before (1.5 V) and a traveling distance of 11.5 Å, which is close to the previous case. After collision, the product looks rather similar: CF_2 appears as a slight protrusion (apparent height, 0.33 Å) close to the previous surface-binding location of the target. However, the outcome of the collision is very different, as can be discerned through STM manipulation (details are given in section 6.1 in the supplementary materials): When the tip was placed above the center of the BTFyl molecule and then moved in the opposite direction from the CF_2 (i.e., along the black arrow in Fig. 2D), we observe that the CF_2 and BTFyl species move together along the manipulation pathway (fig. S9), in contrast to the easy detachment in Fig. 2C. Hence, CF_2 was moved together with the BTFyl molecule in a pulling motion (*30*). Such a motion of the complete CF_2–BTFyl entity is strong evidence for a covalent bond between CF_2 and BTFyl, as otherwise, in the absence of a chemical bond, BTFyl would be moved with the STM tip and CF_2 would stay in place (*32*). Thus, a reactive collision between the projectile and the target has occurred, forming a new CF_2–BTFyl molecular species. This result is independent of the precise pathway of the STM tip, as the same stability is found for various manipulation directions (fig. S9), supporting our conclusion of a covalent bond formed during the collision. Further evidence comes from high-resolution STM images of such a collision, using a functionalized STM tip (fig. S10), where the CF_2 side group of the reaction product can be identified, and from good agreement of experimental images with the calculated ones (fig. S11).

Reactive and unreactive trajectories

Going beyond comparison of two cases, we present in Fig. 3A a systematic study of the outcomes of many different collision pathways—79 trajectories are plotted for CF_2 projectiles colliding with different BTFyl targets. We have considered only the trajectories of CF_2 projectiles that reached the target (i.e., were found next to the target at the closest possible lattice site). Such statistical analysis allows detailed insight into how the impact parameter and collision geometry affect the collision outcome. Each arrow represents the trajectory of a single projectile, which was induced to collide with a BTFyl target in the same manner as in Fig. 2. The length of the arrows corresponds to the distance CF_2 traveled to reach the target, whereas the color of the arrow, blue or red, shows the unreactive or reactive outcome of the collision, respectively.

The major outcome of these collisions was no reaction ($N = 73$), which is expected for collisions far from the reactive end of the target—highlighting that CF_2 must collide at short distances from the reactive unsaturated C atom of BTFyl for reaction to occur ($N = 6$). As such, we focus on collisions for small values of b' (±3.6 Å; $N = 44$; Fig. 3B and fig. S12)—that is, those trajectories in which CF_2 approached the target along the nearest Cu rows to its attachment point on the surface, i.e., along the row to which BTFyl is attached (Row 0), and the adjacent rows above (Row +1) and below (Row −1) the target (Fig. 3C). We find that reaction is highly sensitive to the precise target alignment—even if CF_2 hits the BTFyl target close to the unsaturated carbon atom [i.e., within ±4 Å of it (fig. S13)], different BTFyl alignments give very different outcomes. Notably, formation of a covalent bond was only found for collisions for targets with alignments of $|\varphi| \leq 3°$—i.e., where the target is oriented with its long axis approximately parallel to the pathway of the CF_2 projectile—whereas collisions for $|\varphi| > 15°$ were found to be unreactive (intermediate angles could not be probed for Row 0; a histogram of probed target alignments is given in fig. S14). This demonstrates that the chemical reaction is not merely proximity driven but requires precise spatial matching of the collision partners to give bond formation, demonstrating a narrow cone of reaction for CF_2 bond insertion.

This cone of reaction is found if CF_2 approaches the BTFyl target on the same copper row to which BTFyl is bound. Intriguingly, an identical product is formed if CF_2 approaches the target along the neighboring copper row (Row +1, Fig. 3C), although this does not correspond to the location of the unsaturated carbon atom within the BTFyl molecule. However, the location of the CF_2 group after the reactive collision is the same in both cases (fig. S15). This finding is unexpected, as one might expect that CF_2 would only react if it approached the target along the same row as that to which the target was bound (Row 0), as observed for smaller reactants (*18*, *19*). It suggests that, over the course of the collision, CF_2 must have transferred from Row +1

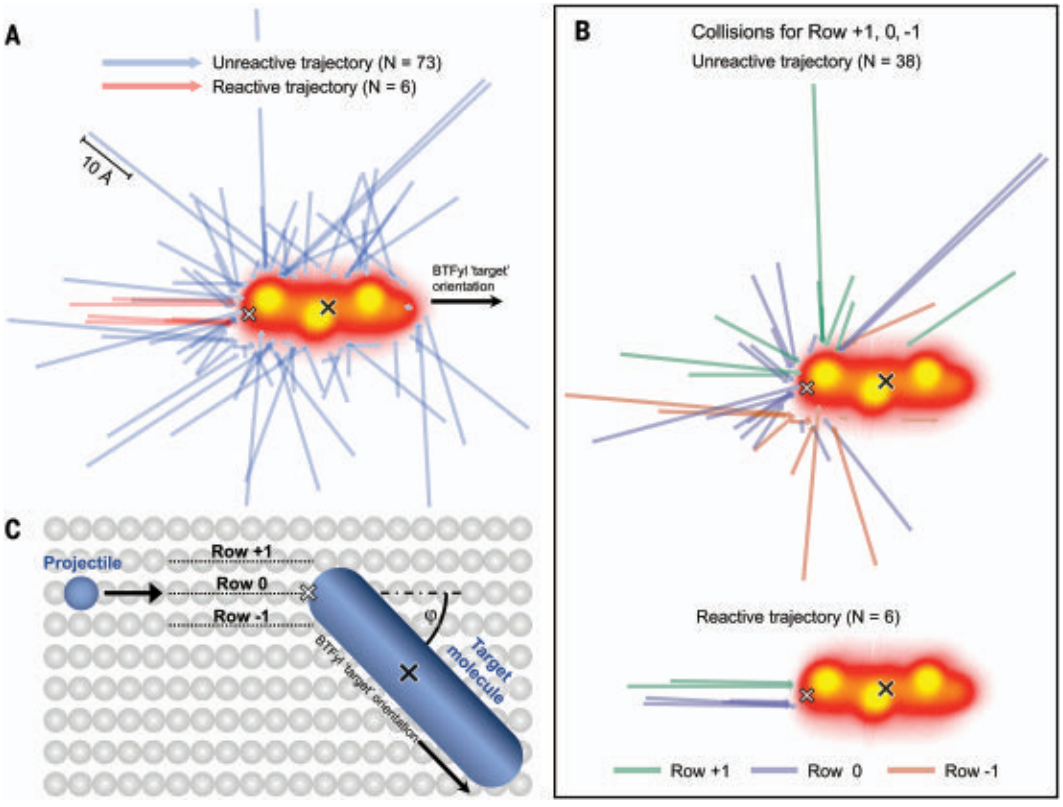


Fig. 3. Trajectories of CF_2 projectiles and outcome of resulting collisions with the BTFyl target. (**A** and **B**) Each arrow represents the trajectory of an individual CF_2 projectile (as shown in Fig. 2), accelerated toward a BTFyl target through a 1.5-V pulse with the STM tip. Each trajectory follows the underlying Cu row; however, here, the trajectories are folded to have the same orientation of the BTFyl target [direction of the black arrow in (A)]. Red (blue) arrows represent trajectories with reactive (unreactive) collisions. The center of mass of the BTFyl target and the unsaturated C atom are marked by black and white crosses, respectively. As can be seen from the arrow lengths in (A), the CF_2 species were generated from up to 48 Å away, to right next to the target. (B) Selection from trajectories shown in (A) where CF_2 approaches the target along Cu Rows +1 (green), 0 (purple), and −1 (orange), as shown by the schematic in (**C**).

to Row 0 to form the product, but this appears unlikely as CF_2 was always observed to stay on the same copper row during diffusion on the Cu(110) surface (after CF_3 dissociation) (*18*). Indeed, the barrier for CF_2 alone to translate across the rows—i.e., to transfer from one row to the next—is calculated to be 2.16 eV (fig. S16), which is a large value, substantially higher than the barrier for CF_2 to diffuse along the Cu row (~0.62 eV; fig. S4). However, the barrier is substantially reduced (from 2.16 to 0.72 eV; fig. S16) if rotation of the CF_2 is included—something that is possible to occur upon collision of CF_2 with the BTFyl target.

As an explanation for the unexpected similarity of reaction products for CF_2 projectiles approaching along Row 0 and Row +1, we propose the 1.3-Å displacement of the Cu atom to which BTFyl is bound (as identified by theory; see fig. S7B). Consequently, this Cu atom is getting close to the neighboring Cu row (i.e., Row +1), only 2.5 Å away from the neighboring copper atom, which is much closer than the normal 3.61-Å separation between copper rows of Cu(110). Indeed, it is comparable to the 2.55-Å distance of close-packed atoms in a copper crystal lattice. Accordingly, we postulate that the proximity of this displaced Cu atom to the next adjacent row may catalyze the transfer of CF_2 from one row to another. This qualitatively fits the difference in reactivities observed in experiment where, even though the total numbers are small, reaction was observed for the projectile approaching along Row 0 with a 100% likelihood ($N = 4$) as compared to 50% ($N = 4$) for the projectile approaching along Row +1. Additionally, no reaction was observed in the cases where CF_2 approached the BTFyl target along the row "below" the binding point of the target, i.e., for Row −1 (fig. S12). Recall that the copper atom is pulled (by BTFyl) to one side of the copper row (fig. S7B), i.e., toward Row +1. This explains why only a projectile moving along Row +1 can be catalyzed by the displaced copper atom, whereas this cannot happen for a projectile on Row −1.

Furthermore, as revealed through nudged-elastic band (NEB) calculations, the very narrow cone of reaction in our experiments can be attributed to the steric effect of the dangling bond's neighboring hydrogen atoms. For a target orientation of $\varphi = 0°$ (Fig. 4A), CF_2 approaches the unsaturated C atom with a calculated barrier for translation of $B_{trans.} = 0.64$ eV ("translation" in Fig. 4A) and $B_{coll.} = 1.29$ eV for CF_2 insertion into the C–Cu bond between BTFyl and the surface ("collision"). As we exclusively find reactive cases (four bond formations for $N = 4$; see the bottom of Fig. 3B) for this configuration ($\varphi = 0°$, Row 0) in our experiments, the projectile apparently overcomes this barrier for reaction. By contrast, for a perpendicular target (orientation of $\varphi = -90°$ in Fig. 4B), it is more difficult for the CF_2 molecule to reach the dangling bond of BTFyl (indicated by a black dot in the calculated configurations at the bottom of Fig. 4B). Our NEB calculations (Fig. 4B) reveal that steric hindrance prevents CF_2 from reaching the unsaturated carbon atom within the BTFyl molecule, which is the site for bond formation. Instead, the system reaches a minimum related to dissociation of the CF_2 molecule rather than bypassing the BTFyl target (Fig. 4B). However, CF_2 dissociation was not observed in our experiments. We attribute this discrepancy to the fact that NEB calculations estimate only the activation barriers but do not account for the momentum transfer of the reactants during the reaction (*33*). Therefore, NEB calculations cannot provide a complete picture of the reaction mechanism. We speculate that upon collision with

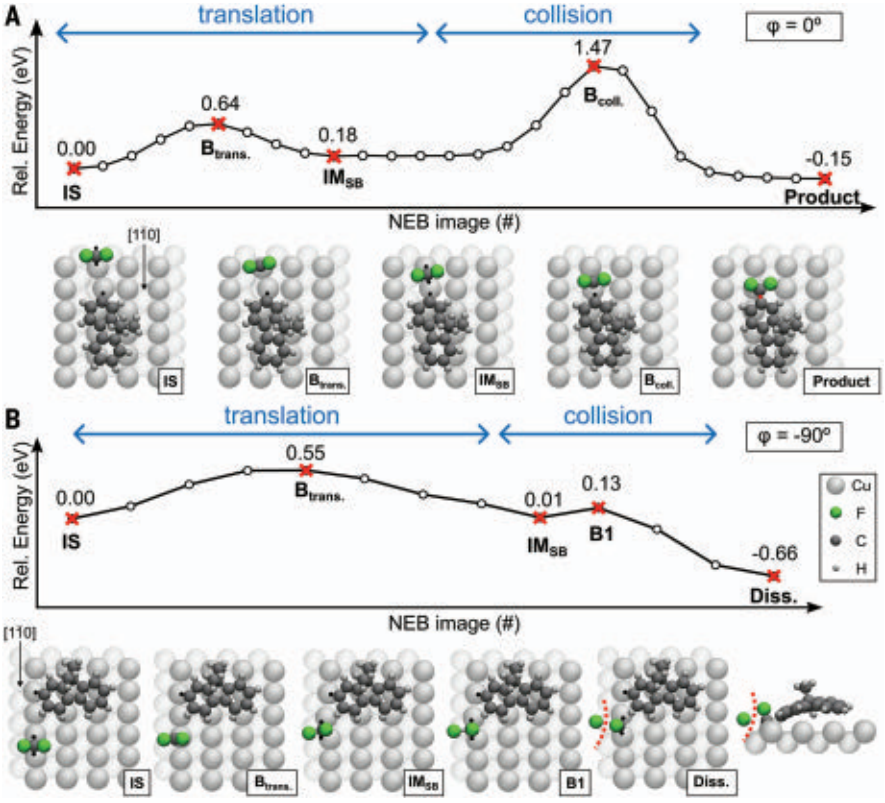


Fig. 4. Minimum energy pathways for different alignments of the target. Calculated relative energies in electronvolts (top) and corresponding on-surface molecular geometries (bottom) for a CF_2 projectile approaching and reacting with a model dimethylfluorenyl ($\bullet C_{15}H_{13}$) target that has an alignment of (**A**) $\varphi = 0°$ or (**B**) $\varphi = -90°$ and an incident parameter (b') of −0.6 Å in (A) and 1.6 Å in (B). Black circles in the geometries denote C–Cu bonding. For both pathways, CF_2 must surmount a barrier ($B_{trans.}$) to translate from one short-bridge (SB) site to the next [from initial state (IS) to intermediate state (IM_{SB})]. This is followed by an encounter with the target and, to form the product, the projectile must overcome a barrier to reaction ($B_{coll.}$). For (A), this involves surmounting a 1.47-eV barrier; whereas for (B), a low-barrier pathway was identified, leading to CF_2 dissociation rather than reaction.

BTFyl, CF_2 is cooled by transferring the energy of its translation and "ratcheting" mode (*18*) into the substrate or target, rather than into C–F bond stretching and, consequently, dissociation. This agrees with the observation that the CF_2 projectile remains on the same copper row and was found intact after collision with a perpendicular target.

Conclusions

Imaging collision cross sections at the level of single molecules highlights the extreme sensitivity of the reaction outcome to the angular orientation of the target and gives a detailed understanding of how surface atoms can aid the process for otherwise disfavored collision geometries. It represents a direct, real-space demonstration of a "cone of reaction" for a bimolecular reaction on a metal surface. Such insight is vital when it comes to tailoring reaction outcomes within the confinement of substrate surfaces. Our results thus advance our ability to control chemical couplings at surfaces toward the design of molecular precursors for on-surface reactions and efficient reactions for the bottom-up assembly of nanostructures.

REFERENCES AND NOTES

1. Y. T. Lee, *Angew. Chem. Int. Ed.* **26**, 939–951 (1987).
2. D. R. Herschbach, *Angew. Chem. Int. Ed.* **26**, 1221–1243 (1987).
3. A. J. Orr-Ewing, *J. Chem. Soc., Faraday Trans.* **92**, 881–900 (1996).
4. R. D. Levine, *Molecular Reaction Dynamics* (Cambridge University Press, 2005).
5. H. Pan, K. Liu, A. Caracciolo, P. Casavecchia, *Chem. Soc. Rev.* **46**, 7517–7547 (2017).
6. S. Buelow, G. Radhakrishnan, J. Catanzarite, C. J. Wittig, *J. Chem. Phys.* **83**, 444–445 (1985).
7. E. B. D. Bourdon *et al.*, *Faraday Discuss.* **82**, 343–358 (1986).
8. J. C. Polanyi, A. H. Zewail, *Acc. Chem. Res.* **28**, 119–132 (1995).
9. B. C. Stipe *et al.*, *Phys. Rev. Lett.* **78**, 4410–4413 (1997).
10. C. E. Tripa, J. T. Yates Jr., *Nature* **398**, 591–593 (1999).
11. L. J. Lauhon, W. Ho, *Faraday Discuss.* **117**, 249–255, discussion 257–275 (2000).
12. S. W. Hla, L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **85**, 2777–2780 (2000).
13. P. Maksymovych, D. C. Sorescu, K. D. Jordan, J. T. Yates Jr., *Science* **322**, 1664–1667 (2008).
14. T. Kumagai *et al.*, *Nat. Mater.* **11**, 167–172 (2011).
15. A. Eisenstein, L. Leung, T. Lim, Z. Ning, J. C. Polanyi, *Faraday Discuss.* **157**, 337–353, discussion 375–398 (2012).
16. Z. Ning, J. C. Polanyi, *J. Chem. Phys.* **137**, 091706 (2012).
17. A. Chatterjee *et al.*, *J. Phys. Chem. C Nanomater. Interfaces* **118**, 25525–25533 (2014).
18. K. Anggara, L. Leung, M. J. Timm, Z. Hu, J. C. Polanyi, *Sci. Adv.* **4**, eaau2821 (2018).
19. K. Anggara, L. Leung, M. J. Timm, Z. Hu, J. C. Polanyi, *Faraday Discuss.* **214**, 89–103 (2019).
20. Q. Zhong *et al.*, *Nat. Chem.* **13**, 1133–1139 (2021).
21. P. R. Brooks, E. M. Jones, *J. Chem. Phys.* **45**, 3449–3450 (1966).
22. R. J. Beuhler Jr., R. B. Bernstein, K. H. Kramer; Crossed Beam Study of the Reaction of Rubidium with Oriented Methyl Iodide Molecules, *J. Am. Chem. Soc.* **88**, 5331–5332 (1966).
23. H. J. Loesch, J. Möller, *J. Phys. Chem. A* **101**, 7534–7543 (1997).
24. L. Lafferentz *et al.*, *Science* **323**, 1193–1197 (2009).
25. D. Civita *et al.*, *Science* **370**, 957–960 (2020).
26. D. Civita, M. Timm, J. Schwarz, S. Hecht, L. Grill, *ACS Nano* **19**, 10255–10262 (2025).
27. Y. An *et al.*, *ACS Nano* **16**, 18592–18600 (2022).
28. L. Vitali *et al.*, *Nat. Mater.* **9**, 320–323 (2010).
29. D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524–526 (1990).
30. L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **79**, 697–700(1997).
31. S. W. Hla, *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct. Process. Meas. Phenom.* **23**, 1351–1360 (2005).
32. L. Grill *et al.*, *Nat. Nanotechnol.* **2**, 687–691 (2007).
33. B. de la Torre *et al.*, *Nat. Commun.* **11**, 4567 (2020).
34. M. J. Timm *et al.*, Spatially resolving the cone of reaction for a single molecule [Data set], Zenodo (2026). https://doi.org/10.5281/zenodo.20267064.

ACKNOWLEDGMENTS

Funding: Financial support from the FWF (project numbers I 5145 and ESP 473), the Federal State of Styria (ABT12-603604), the European ITN Network ULTIMATE, and the European Research Council Advanced Grant AMOS is gratefully acknowledged. **Author contributions:** M.J.T. and L.G. conceived the experiments. M.J.T. performed the experiments with help from I.G. M.J.T. analyzed the data. A.M., Q.C., and P.J. did the calculations. S.H. provided the molecules. M.J.T. and L.G. discussed the results and wrote the manuscript with feedback from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and details needed to evaluate the conclusions in the paper are available in the main text or supplementary materials or are deposited at Zenodo (*34*). **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aec7913
Materials and Methods; Figs. S1 to S16; References (*35–52*)

Submitted 2 October 2025; accepted 5 June 2026

10.1126/science.aec7913


Science Podcast
AAAS
Stay curious, stay informed.
Tune into the *Science* Podcast for the latest from the world of science
LISTEN NOW
NEW EPISODES EVERY THURSDAY

NANOMATERIALS

Step geometry–guided growth of rhombohedral graphene

Mengze Zhao[1]†, Zhibin Zhang[1,2]†*, Xin Sui[3]†, Quanlin Guo[1]†, Cheng Qian[4]†, Zaizhe Zhang[3], Hanbo Xiao[5], Jingyi Zhu[6], Zuo Feng[1,3], Yilong You[1], Tianze Dong[1], Xin-Zheng Li[1,2], Sheng Wang[6], Li Wang[7], Enge Wang[2,3,8], Feng Ding[4]*, Zhongkai Liu[5]*, Xiaobo Lu[3]*, Kaihui Liu[1,3,9]*

Rhombohedral (ABC)–stacked graphene has emerged as a platform for exploring correlated and topological physics. However, its large-scale, pure-phase synthesis has been hindered by intrinsic thermodynamic metastability and kinetic instability. In this work, we introduce a step geometry–guided epitaxial strategy to deterministically control interlayer slip, enabling the synthesis of pure-phase rhombohedral graphene. Using this approach, we obtained pure (phase purity > 99%) rhombohedral graphene with an area reaching 160 micrometers by 80 micrometers and thickness ranging from ~15 layers to ~120 nanometers. The resulting samples establish the first comprehensive reference dataset of Raman fingerprints and intrinsic band structures from few-layer films to ~200-layer bulk. Electronic transport measurements reveal a layer-antiferromagnetic state and the quantum anomalous Hall effect, enabling scalable exploration of next-generation quantum science and technology applications.

Interlayer stacking is a key degree of freedom that dictates the properties of multilayer two-dimensional (2D) materials. For graphene, two prominent stacking orders are broadly observed: the common Bernal (AB) stacking (hexagonal, D_{6h} symmetry) and the rare metastable rhombohedral (ABC) stacking (trigonal, D_{3d} symmetry). With a simple interlayer slip, Bernal-stacked graphene can be transformed into the rhombohedral form, and such a transition would reshape its electronic landscape by generating flat bands and nonzero Berry curvature that are absent in the Bernal phase (*1*–*7*). These features are indicative of strong electronic correlations and nontrivial topology and would make rhombohedral graphene an ideal platform for exploring exotic quantum phenomena, including superconductivity (*8*–*12*), Chern insulating states (*13*–*20*), and the quantum anomalous Hall effect (*21*–*25*).

The exploration of these phenomena requires the stable production of pure-phase rhombohedral graphene. However, most previous studies have relied on mechanical exfoliation to obtain rhombohedral graphene, which is inherently limited by small sample sizes, appreciable phase impurities, and laborious phase identification. For example, five-layer pure-phase rhombohedral graphene is buried in eight possible mixed-phase variants, with a theoretical fabrication probability of <1% (*26*, *27*). Such intrinsic limitations hinder systematic investigations and scalable applications.

Substantial efforts have been devoted to the synthesis of rhombohedral graphene, including methods based on silicon carbide (SiC) surface decomposition (*28*) and strain-assisted stabilization of local ABC domains on curved substrates (*29*, *30*). These studies provided an important foundation for understanding how the substrate can influence stacking order; however, the growth of rhombohedral graphene with both large in-plane size and high out-of-plane phase purity remained elusive. This difficulty arises from two fundamental barriers. First, thermodynamically, the formation energy of rhombohedral graphene is ~0.2 meV per atom greater than that of Bernal stacking (*31*). Second, kinetically, postgrowth perturbations, such as thermal fluctuations and cooling-induced local strains, would readily trigger interlayer sliding that irreversibly converts the rhombohedral phase into the Bernal phase.

Recent advances in 2D material growth highlight both challenges and opportunities (*32*–*34*). In C_3-symmetry binary systems, such as boron nitride (BN) and transition-metal dichalcogenides (TMDs), the rhombohedral phase has already been realized. In these systems, the ABC (3R) phase is the local lowest-energy configuration when layers share the same in-plane orientation, and their strong coupling with the substrate can be used to lock the orientation of each layer and enable stacking control (*35*, *36*). Moreover, once the 3R phase is formed, its conversion to the globally stable AA′ (2H) phase requires a 60° rotation of the entire layer, imposing a kinetic barrier that protects the rhombohedral order. However, graphene has a single-element composition and a higher C_6 lattice symmetry, so the layers intrinsically align in the same orientation, making Bernal stacking the most stable configuration. This higher symmetry also lacks orientational protection that is present in BN and TMDs. A graphene lattice rotated by 60° is identical to its original lattice, so the rhombohedral stacking order can be switched to the Bernal stacking easily by a simple slip of one C–C bond length. Therefore, achieving stable, pure-phase rhombohedral graphene requires a fundamental reconceptualization of the growth strategies that can thermodynamically guide the formation of ABC stacking and kinetically suppress its relaxation to Bernal stacking.

Growth mechanism of rhombohedral graphene

We present a step geometry–guided mechanism for rhombohedral graphene growth by controlling the interlayer displacement. A graphene sheet on a substrate has an edge plane serving as a terminating boundary (Fig. 1A). This edge plane enforces a precise slip magnitude (governed by the slip angle φ, defined between the edge plane and the substrate plane) and slip direction (governed by the orientation angle θ, defined between the confinement direction $\boldsymbol{n}$ and the graphene armchair axis $\boldsymbol{a}$) during the growth process. The slip angle φ determines the degree of geometric constraint on interlayer displacement. When φ = 90° (2φ = 180°, corresponding to an unconfined edge on a flat surface), Bernal stacking is energetically favorable. Therefore, achieving rhombohedral stacking requires φ < 90° to impose appropriate confinement (Fig. 1, B and C).

The orientation angle θ determines the alignment between the confinement direction and the interlayer displacement. When θ = 0°, the confinement aligns precisely with the displacement required for rhombohedral stacking, enabling ideal phase control (Fig. 1, C and D, and fig. S1). When θ = 30°, however, the confinement direction becomes parallel to the graphene zigzag axis, eliminating confinement and resulting in Bernal stacking.

By controlling φ and θ, the overall displacement toward rhombohedral stacking can be achieved. However, stable phase formation also requires that every graphene layer responds coherently to this confinement. During growth, φ enforces discrete bending of each layer at distinct atomic sites (Fig. 1E). If these bending sites are random across layers, the displacement would lose coherence, and rhombohedral stacking could not form. Through our density functional theory calculations,

[1]State Key Laboratory for Mesoscopic Physics, Frontiers Science Centre for Nano-optoelectronics, School of Physics, Peking University, Beijing, China. [2]Interdisciplinary Institute of Light-Element Quantum Materials and Research Centre for Light-Element Advanced Materials, Peking University, Beijing, China. [3]International Centre for Quantum Materials, Collaborative Innovation Centre of Quantum Matter, Peking University, Beijing, China. [4]Suzhou Laboratory, Suzhou, China. [5]School of Physical Science and Technology, ShanghaiTech University, Shanghai, China. [6]School of Physics and Technology, Wuhan University, Wuhan, China. [7]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China. [8]Tsientang Institute for Advanced Study, Hangzhou, China. [9]Songshan Lake Materials Laboratory, Dongguan, China. ***Corresponding author. Email: khliu@pku.edu.cn (K.L.); xiaobolu@pku.edu.cn (X.L.); liuzhk@shanghaitech.edu.cn (Z.L.); dingf@szlab.ac.cn (F.D.); zhibinzhang@pku.edu.cn (Z.Z.)** †These authors contributed equally to this work.

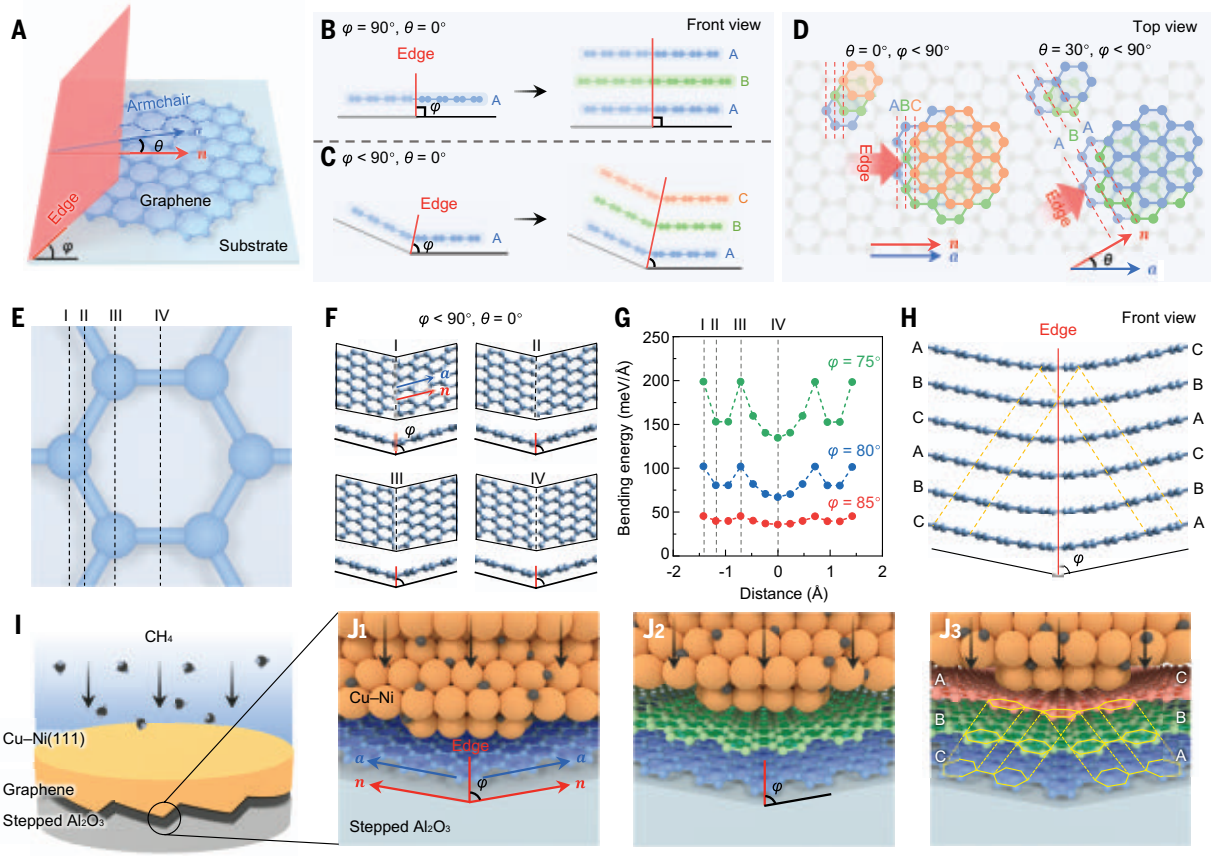


Fig. 1. Step geometry–guided growth mechanism for rhombohedral graphene. (**A**) Schematic of the geometric alignment between the graphene lattice and the edge plane on a substrate. (**B** and **C**) Role of the slip angle φ. (B) At φ = 90°, the geometric constraint is absent, leading to thermodynamically favorable Bernal stacking. (C) At φ < 90°, the geometric constraint induces interlayer slip and leads to rhombohedral stacking. (**D**) Role of orientation angle θ: alignment at θ = 0° favors rhombohedral stacking, whereas misalignment at θ = 30° yields Bernal stacking. (**E**) Atomic structure of graphene with possible bending sites I to IV. (**F**) Relaxed atomic structures after bending at the four sites. (**G**) Bending energy as a function of φ for different bending sites. (**H**) MD-simulated rhombohedral structures at φ = 78° and θ = 0°. (**I**) Schematic of the growth design. (**J_1** to **J_3**) Schematic of the growth process, with the formation of mirror-symmetric ABC and CBA domains.

we reveal that bending preferentially occurs at the highly symmetric midpoints of C–C bonds (Fig. 1, F and G, site IV), and such a preference becomes more dominant as φ decreases (that is, as the confinement increases) (Fig. 1G). Our calculations also show that nucleation is energetically preferred at the step edge, making graphene formation overwhelmingly more likely to initiate at the step-confined region (fig. S2). As a result, the graphene layers nucleate and bend at the same site IV, collectively forming an edge plane that anchors the graphene edge and systematically guides the stacking sequence toward the rhombohedral order (Fig. 1H).

To achieve such coordinated control over both angles, we developed a geometrically confined architecture using stepped sapphire (Al_2O_3) and single-crystal copper-nickel alloy (Cu–Ni) foils for rhombohedral graphene growth (Fig. 1I and supplementary materials, materials and methods). In this work, stepped Al_2O_3 serves as a geometrical confinement to enforce the interlayer displacement of the graphene, and Cu–Ni foil serves as the growth substrate that enables interfacial epitaxy of the graphene. The introduction of Ni increases the carbon solubility in the substrate, facilitating carbon transport. Consequently, carbon atoms from methane (CH_4) decomposition continuously diffuse through the Cu–Ni foil to the buried growth interface, ensuring a sufficient and sustained carbon supply throughout the growth process. For the geometrical confinement, the angle φ is determined by the Al_2O_3 step morphology, and the angle θ is controlled by the alignment of the Al_2O_3 steps with the lattice orientation of the Cu–Ni foil (fig. S3). Because graphene nucleates at the Cu–Ni/Al_2O_3 interface and conforms to both surfaces, each newly grown layer must fit the step geometry (Fig. 1, J_1 to J_3). This step confinement guides each new layer to nucleate and grow with a specific, atomically precise displacement relative to the layer below it, ensuring that the entire stack grows as a single, coherent rhombohedral crystal (fig. S4).

With this strategy, we synthesized rhombohedral graphene at a growth temperature of ~1150°C. However, during subsequent cooling to room temperature, the relaxation of interfacial stress and local temperature fluctuations can kinetically introduce interlayer sliding, converting the metastable rhombohedral stacking into the Bernal phase. Fortunately, our growth strategy incorporated a kinetic stabilization mechanism. Because of the structural equivalence across the step, graphene grew symmetrically and formed mirror-symmetric ABC and CBA domains (Fig. 1J_3). In this regime, any spontaneous transition to Bernal stacking on one side would correspondingly induce AA stacking on the other side (fig. S5), whose energy is ~10 meV per atom higher than that of AB stacking (*37*). Such an energetically unfavorable pathway effectively locks the rhombohedral stacking order, suppressing phase conversion throughout the entire growth process.

Phase control of rhombohedral graphene

On the basis of the designed mechanism, we first manipulated the slip angle φ. At φ ~75° (with θ = 0° if not specified) (Fig. 2A_1), the increased geometric confinement enabled the conformal growth of ultrathin

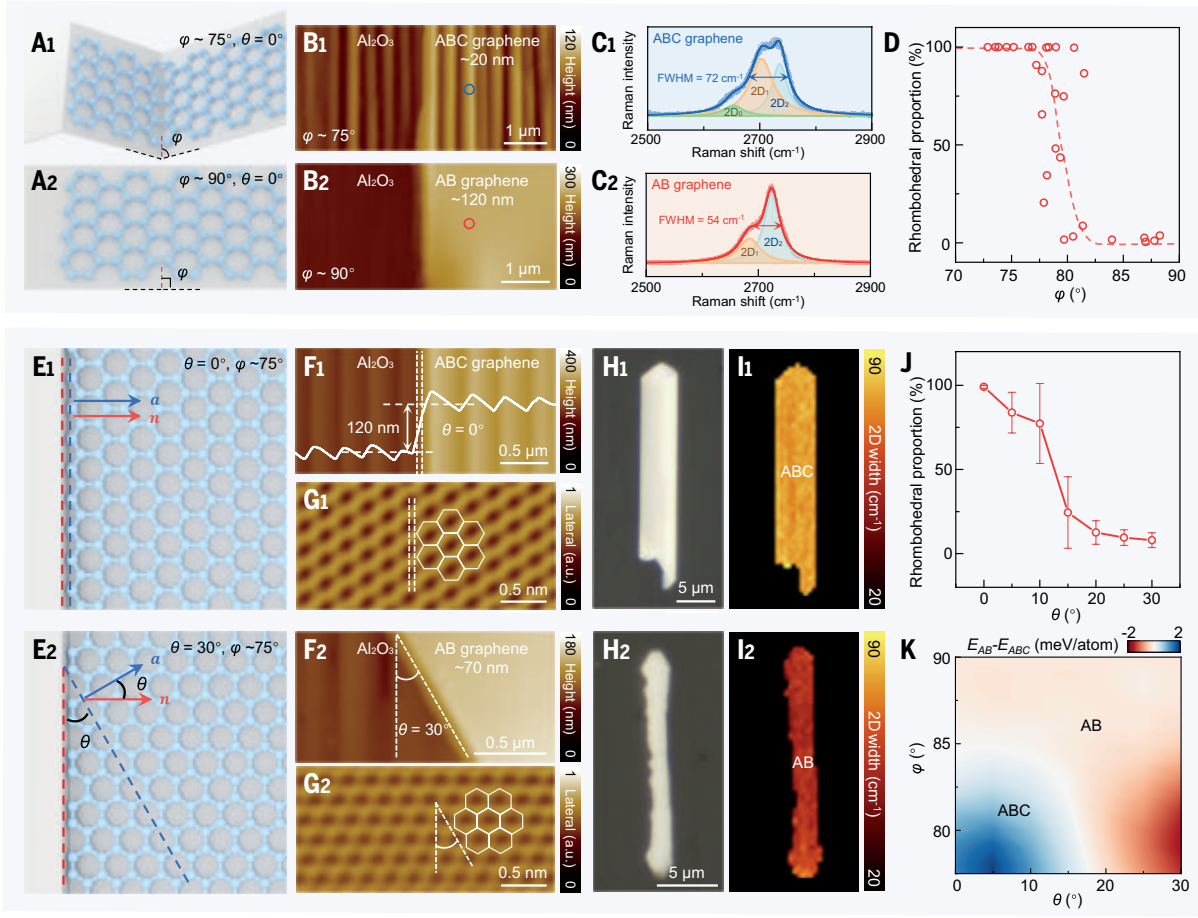


Fig. 2. φ- and θ-dependent phase control of rhombohedral graphene. (**A_1** and **A_2**) Schematics of graphene growth (A_1) with and (A_2) without geometric confinement. (**B_1** and **B_2**) AFM images near graphene edges showing (B_1) a conformal morphology with confinement and (B_2) a flat morphology without confinement. (**C_1** and **C_2**) Raman spectra (circles) with Lorentzian multipeak fits (lines) displaying (C_1) ABC stacking at φ ~ 75° and (C_2) AB stacking at φ ~ 90°, corresponding to the marked circle regions in (B_1) and (B_2). (**D**) Rhombohedral proportion versus φ across samples. (**E_1** and **E_2**) Schematics of the graphene lattice alignment at (E_1) θ = 0° and (E_2) θ = 30°. (**F_1** and **F_2**) AFM images near graphene edges at (F_1) θ = 0° and (F_2) θ = 30°. (**G_1** and **G_2**) High-resolution AFM images identifying lattice directions at (G_1) θ = 0° and (G_2) θ = 30°. (**H_1** and **H_2**) Optical images of the graphene domains at (H_1) θ = 0° and (H_2) θ = 30°. (**I_1** and **I_2**) Raman 2D peak width maps revealing (I_1) the pure ABC stacking at θ = 0° and (I_2) the complete AB stacking at θ = 30°. (**J**) Experimental dependence of the rhombohedral proportion on θ. (**K**) MD-simulated phase diagram showing the relative energy of AB and ABC stacking as a function of θ and φ.

graphite on stepped Al_2O_3, as revealed with atomic force microscopy (AFM) (Fig. 2B_1). Under these growth conditions, rhombohedral stacking was readily obtained and confirmed by characteristic Raman signatures (*38*, *39*): a broad 2D peak with a full-width at half-maximum (FWHM) of ~72 cm^{-1} and three well-resolved subpeaks ($2D_0$, $2D_1$, and $2D_2$) (Fig. 2C_1).

By contrast, when φ approaches 90°, geometric confinement was insufficient to drive interlayer slip (Fig. 2A_2). In this regime, the grown ultrathin graphite exhibited a flat morphology (Fig. 2B_2), and its Raman spectrum featured a narrow 2D peak (~54 cm^{-1}) without the presence of a $2D_0$ subpeak (Fig. 2C_2), revealing a Bernal structure (*38*, *39*). These Raman signatures were also validated by means of infrared scanning near-field optical microscopy (IR-SNOM) (*29*, *40*, *41*), confirming the stacking identification (fig. S6). Through our systematic investigations, we established φ ~ 80° as the critical threshold, only below which the geometric confinement could stabilize the rhombohedral order (Fig. 2D and fig. S7).

Next, we examined the role of the orientation angle θ. At θ = 0° (with φ ~ 75° if not specified) (Fig. 2E_1), the confinement direction aligned perfectly with the rhombohedral slip direction, yielding ultrathin pure-phase rhombohedral graphite (Fig. 2, F_1 to I_1). Atomic-resolution AFM confirmed directional alignment and revealed that the steps could confine graphite growth to thicknesses reaching 120 nm (Fig. 2, F_1 and G_1). This thickness scalability arises from the interfacial-growth characteristics of our process, in which each new layer forms under identical geometrical constraints, enforcing a deterministic interlayer displacement and stacking sequence (fig. S4). Further statistical analysis across multiple samples consistently exhibited a high phase purity of >99% under the optimized growth conditions (Fig. 2J), demonstrating the robustness of this strategy against thermal and mechanical perturbations during growth.

We then explored how variations in θ influence stacking stability (fig. S8). At θ = 5°, although mixed Bernal domains began to appear, rhombohedral stacking remained dominant (fig. S9, A_1 to E_1). At θ = 10°, however, coexisting rhombohedral and Bernal regions were observed (fig. S9, A_2 to E_2), and the average rhombohedral portion substantially decreased (Fig. 2J). When θ ≥ 20°, the geometric manipulation completely lost phase control, and the ultrathin graphite converted entirely to Bernal stacking (Fig. 2, E_2 to I_2, and fig. S9, A_3 to E_3). These results demonstrate that the stacking order of the graphite layers could be precisely regulated by deliberately adjusting the orientation angle θ, enabling reliable production of graphene with the desired stacking configuration for practical applications (fig. S9F).

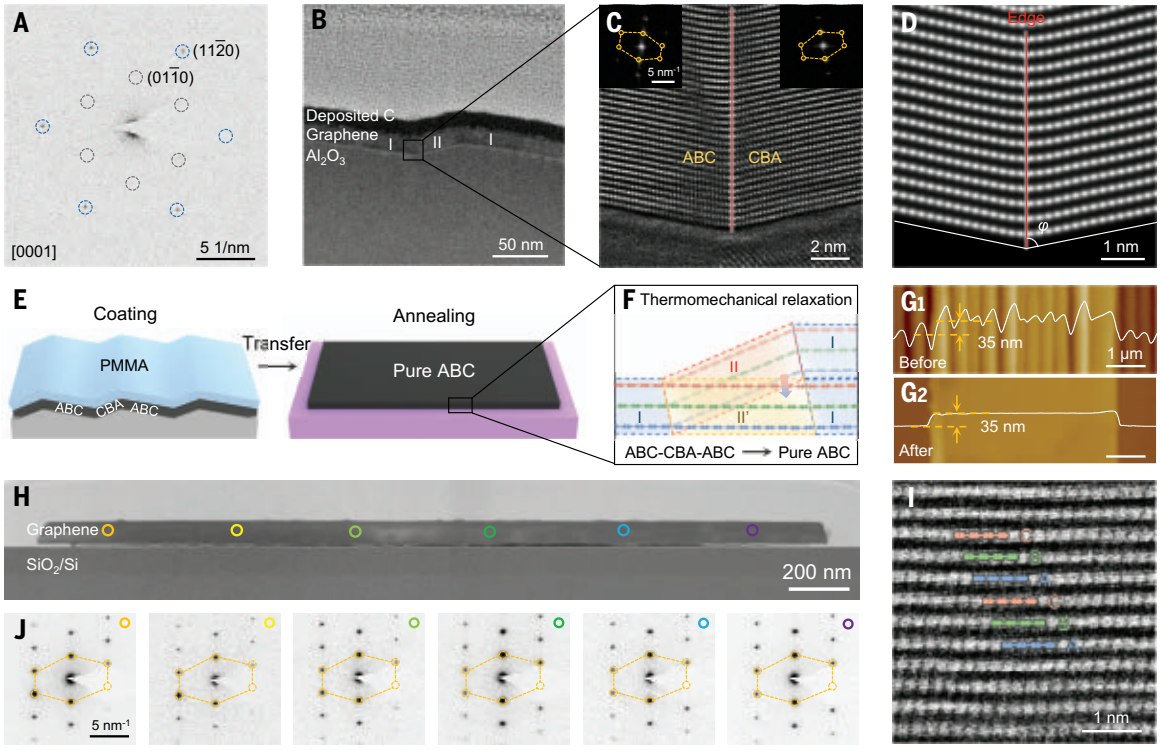


Fig. 3. Structural characterization and thermomechanical relaxation of rhombohedral graphene. (**A**) SAED pattern of rhombohedral graphene along the [0001] axis, showing pure ABC stacking signatures with the absence of Bernal-specific primary diffraction spots (dashed gray circles). (**B**) Low-magnification cross-sectional STEM image revealing conformal graphene growth along the Al_2O_3 step. (**C**) Atomic-resolution STEM image from the boxed region in (B). (Insets) Fast Fourier transform (FFT) patterns from each side. (**D**) Simulated cross-sectional STEM image. (**E**) Schematic of the thermomechanical relaxation process. (**F**) Schematic of interlayer slip (region II → region II′) during relaxation. (**G_1** and **G_2**) AFM morphology of rhombohedral graphene (G_1) before and (G_2) after relaxation. (**H**) Cross-sectional STEM image of atomically flat ABC graphene after relaxation. (**I**) Atomic-resolution STEM image of rhombohedral graphene. (**J**) SAED patterns from multiple positions in (H), indicated with circles with corresponding colors.

These findings, together with molecular dynamics (MD) simulations, outline a phase diagram that is coregulated by the orientation angle θ and slip angle φ (Fig. 2K). The diagram defines the stability windows of different stacking orders and provides guidance for their controlled growth. The optimal rhombohedral growth conditions were identified at $\theta < 10°$ and $\varphi < 80°$, at which coupled angle control reshapes the energy landscape and ensures the precise growth of rhombohedral graphene. In this optimal window, we ultimately achieved the growth of large-area (~160 μm by 80 μm), thin (~24 layers) pure-phase rhombohedral graphene (fig. S10).

Structural characterization of rhombohedral graphene

To validate the phase purity and material quality of the grown rhombohedral graphene, we conducted comprehensive structural analyses. Selected-area electron diffraction (SAED) revealed characteristic out-of-plane rhombohedral signatures, with Bernal-specific primary diffraction spots entirely absent (Fig. 3A). Cross-sectional scanning transmission electron microscopy (STEM) characterization demonstrated conformal growth along the substrate steps, with the graphene layers maintaining intimate contact and structural continuity across the terraces (Fig. 3B). Atomic-resolution STEM imaging exhibited mirror-symmetric ABC and CBA configurations (Fig. 3C), which was consistent with the simulated STEM images that reproduce rhombohedral stacking (Fig. 3D). In addition, rhombohedral graphene samples with thicknesses ranging from ~15 layers to ~120 nm were also characterized, all of which exhibited rhombohedral stacking features and consistent step-geometry–guided growth behavior (figs. S11 and S12). Together, these results provide direct evidence of long-range ordered rhombohedral stacking and validate the step-geometry–guided growth mechanism.

Although the mirror-symmetric ABC-CBA structure kinetically stabilized the rhombohedral phase during the cooling process, it also introduced linear domain boundaries in the samples. To remove these boundaries and achieve a fully uniform structure, we developed a thermomechanical relaxation process (Fig. 3E). After etching away the metal substrate, the rhombohedral graphene was coated with a soft polymethyl methacrylate (PMMA) layer and transferred onto a flat substrate. Afterward, we used a mild annealing procedure (≤400°C) to release the residual strain. Because graphene grown on Al_2O_3 steps consisted of long (region I) and short (region II) planes (Fig. 3B), strain redistribution during annealing was dominated by the long planes and drove the graphene on the short planes to adopt the major stacking order (region II′), ultimately forming a uniform crystal order (Fig. 3F and fig. S13). This behavior is supported by MD simulations, which show that once the step-geometry constraint is removed during transfer, the CBA-to-ABC rearrangement proceeds spontaneously with no obvious barrier, enabling the formation of a uniform ABC sequence (fig. S14).

As a result, our method used thermally and mechanically driven structural reorganization to transform the initial ABC-CBA structure into a pure ABC rhombohedral phase (Fig. 3, G_1 and G_2, and fig. S15), yielding a structurally uniform material essential for reliable characterization and device fabrication. We also found that the annealing temperature was a key parameter: Mild annealing (≤400°C) fully preserved the rhombohedral phase during relaxation, whereas annealing above ~800°C could activate a transition toward Bernal stacking (fig. S16). We further characterized the structure of the transferred rhombohedral samples. Cross-sectional STEM imaging confirmed that the film thoroughly relaxed into an atomically flat morphology with no residual

step structures (Fig. 3H). Atomic-resolution imaging revealed a strict three-layer periodicity with no detectable Bernal configuration (Fig. 3I). Consistent SAED patterns across the sample with identical orientations confirmed the homogeneous rhombohedral phase, demonstrating the successful elimination of the ABC-CBA boundaries (Fig. 3J). This thermomechanical relaxation process was also effective for thinner rhombohedral graphene films, yielding uniform rhombohedral graphene after transfer and relaxation (fig. S17).

Spectroscopic and electronic properties of rhombohedral graphene

The successful large-scale growth of pure-phase rhombohedral samples enabled the exfoliation of high-quality graphene for few-layer characterizations. The as-grown samples were transferred onto SiO_2/Si substrates and then thinned by means of a standard mechanical exfoliation method. These samples established a comprehensive layer-resolved reference dataset of Raman fingerprints and intrinsic band structures, spanning from a few layers to ~200 layers. Stacking-sensitive Raman spectroscopy revealed distinct 2D-band features across this thickness range (Fig. 4, A and B). Compared with Bernal stacking, rhombohedral graphene consistently exhibited higher $2D_1$ peak area to total 2D peak area ratios (I_{2D1}/I_{total}) at all thicknesses (Fig. 4C). In addition, the 2D peak in rhombohedral graphene broadened progressively with increasing thickness, in contrast to the narrowing tendency in Bernal stacking (Fig. 4D). Across the measured range, the rhombohedral 2D peaks were 4 to 18 cm^{-1} wider than their Bernal counterparts, reflecting their distinctive band structures with broader resonances in the phonon momenta (*42*). These systematic spectroscopic fingerprints established a reliable reference for the identification of rhombohedral stacking.

The exceptional phase purity of these samples also enabled direct visualization of the stacking-dependent electronic structures with the use of nanoscale angle-resolved photoemission spectroscopy (nano-ARPES) (Fig. 4, E_1 to E_4). The measured band dispersions exhibited well-defined flat surface bands near the Fermi level ($E_F \pm 50$ meV). The flat bands were most profound in three- to five-layer samples and persisted up to the ~200-layer bulk (*43*, *44*), directly revealing the origin of the reduced quasiparticle velocity and increased electronic correlations of rhombohedral graphene (*23*). In addition, the number of subbands away from

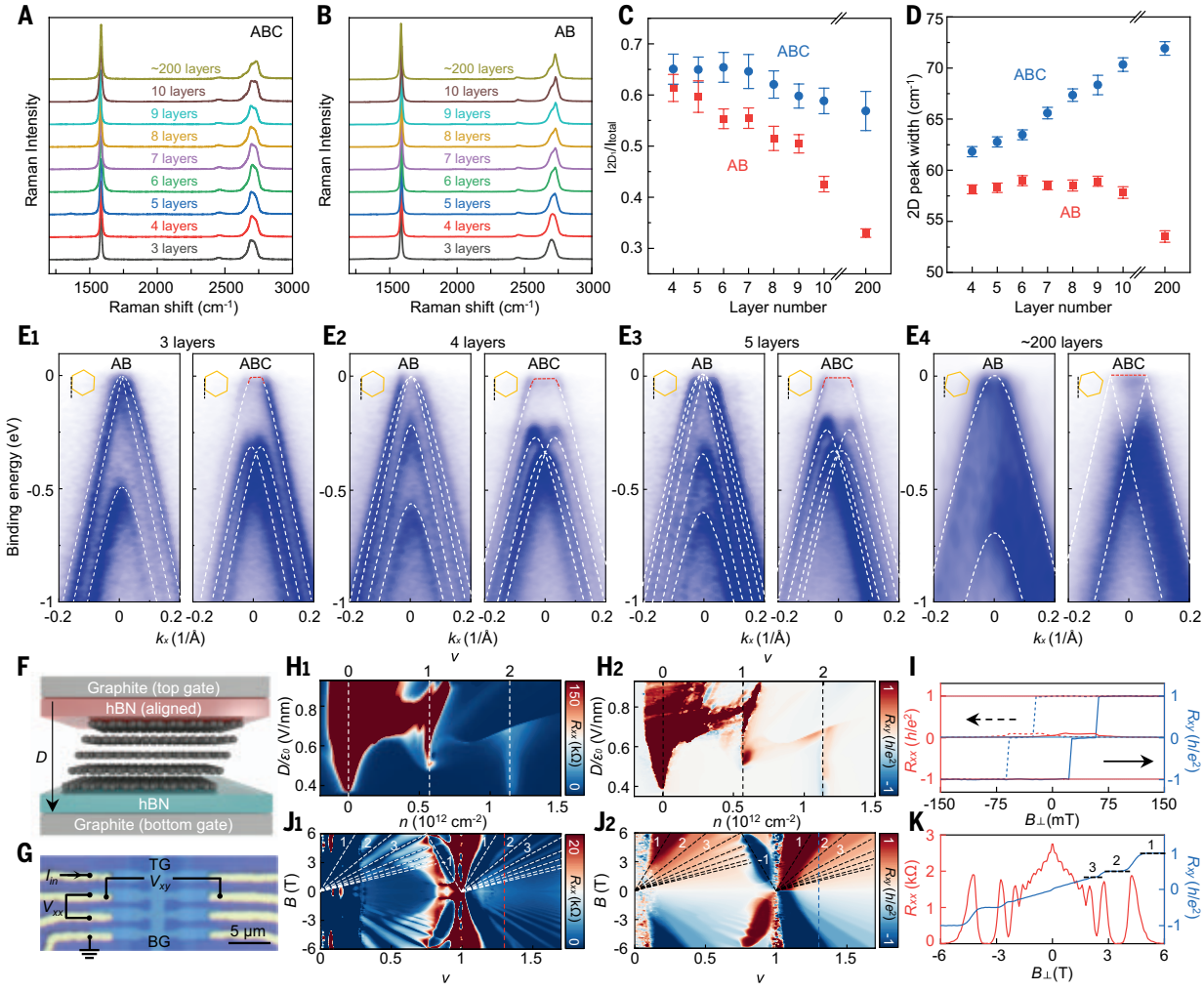


Fig. 4. Spectroscopic and electrical characterization of rhombohedral graphene. (**A** and **B**) Raman spectra of three- to ten-layer and bulk graphene with (A) rhombohedral and (B) Bernal stacking. (**C**) Ratios of the $2D_1$ peak area to the total 2D peak area for rhombohedral and Bernal stackings as a function of layer number. (**D**) 2D peak widths versus layer number for both stackings. (**E_1** to **E_4**) Nano-ARPES spectra with corresponding simulations (dashed curves) for three- to five-layer and bulk (left) Bernal and (right) rhombohedral graphene. Red dashed lines indicate the flat bands. (Insets) Cut directions (black lines) and the graphene Brillouin zone (yellow hexagons). (**F**) Schematic of the five-layer rhombohedral graphene/hBN moiré device. (**G**) Optical image of the device. BG, bottom gate; TG, top gate. (**H_1** and **H_2**) Symmetrized longitudinal resistance (R_{xx}, H_1) and antisymmetrized Hall resistance (R_{xy}, H_2) versus carrier density n and displacement field D under $B = 0.1$ T. (**I**) R_{xx} and R_{xy} as functions of B at $\nu = 1.03$ and $D/\varepsilon_0 = 0.54$ V/nm. (**J_1** and **J_2**) Landau fan diagrams of (J_1) R_{xx} and (J_2) R_{xy} at $D/\varepsilon_0 = 0.54$ V/nm. (**K**) Linecut taken along the dashed lines in (J_1) and (J_2) at $\nu = 1.3$.

E_F increased with increasing layer thickness, and the subbands eventually merged into a bulk Dirac cone in the bulk limit. The measured band structures differ drastically from those of Bernal-stacked graphene with the same layer numbers and show excellent agreement with tight-binding and first-principles calculations (*45*–*47*). The sharp spectra and well-resolved thickness-dependent band structure evolution confirm the quality and uniformity of our samples.

Then, we performed high-resolution electronic transport investigations on devices fabricated from graphene exfoliated from the as-grown samples. In a dual-gated seven-layer device at 1.5 K (fig. S18, A and B), we observed two well-defined insulating states near charge neutrality—that is, a correlation-driven layer-antiferromagnetic insulating (LAF) state at zero displacement field (D) and a band insulating (BI) state at $|D| > 0.3$ V/nm (fig. S18, C and D) (*14*, *15*). Quantum oscillations were also revealed at low magnetic fields (fig. S18E). These results confirmed that such quantum phenomena can be accessed in our grown rhombohedral graphene samples.

We next fabricated a five-layer rhombohedral graphene device aligned with hexagonal boron nitride (hBN) at a twist angle of 0.27° (Fig. 4, F and G). At an ultralow temperature (10 mK), we identified a series of vertical resistance peaks corresponding to correlated states at integer fillings of 0, 1, and 2 electrons per moiré unit cell in the phase diagram (Fig. 4, H_1 and H_2) (*19*–*22*). At a moiré filling of $\nu = 1$, the system exhibited the quantum anomalous Hall effect, in which the Hall resistance (R_{xy}) reached a quantized resistance of h/e^2 (where h is Planck's constant and e is the electron charge), whereas the longitudinal resistance (R_{xx}) nearly disappeared at zero magnetic field (Fig. 4I). Landau fan diagram measurements further resolved the gap at $\nu = 1$, with a Chern number $C = 1$ that could survive at zero magnetic field (Fig. 4, J_1 and J_2). In addition to the $C = 1$ gap, other quantum Hall states were also well developed under finite magnetic fields, exhibiting quantized R_{xy} and vanishing R_{xx} (Fig. 4K). Collectively, the observation of these well-resolved correlated states, pronounced phase transitions, and distinct topological quantum phenomena reflects the exceptional homogeneity and low intrinsic disorder of our grown samples.

Discussion and outlook

We introduce a step geometry–guided epitaxial strategy that overcomes the intrinsic thermodynamic and kinetic barriers of rhombohedral graphene growth. This method yields high-purity (>99%) rhombohedral graphene with both a large in-plane area (160 μm by 80 μm) and high out-of-plane coherence (from ~15 layers to ~120 nm), providing an unprecedented platform for exploring flat-band physics, strong correlations, and topological order. In addition, our strategy opens the possibility for the controlled synthesis of graphene beyond Bernal and rhombohedral stacking orders, such as phases with larger unit cells along the out-of-plane direction (for example, graphene with ABAC and ABCBA stacking orders). Our approach establishes a paradigm for stacking-sequence engineering in emergent quantum materials and paves the way for their scalable applications in future quantum science and technologies.

REFERENCES AND NOTES

1. K. F. Mak, J. Shan, T. F. Heinz, *Phys. Rev. Lett.* **104**, 176404 (2010).
2. W. Bao *et al.*, *Nat. Phys.* **7**, 948–952 (2011).
3. C. H. Lui, Z. Q. Li, K. F. Mak, E. Cappelluti, T. F. Heinz, *Nat. Phys.* **7**, 944–947 (2011).
4. G. R. Chen *et al.*, *Nat. Phys.* **15**, 237–241 (2019).
5. Y. Shi *et al.*, *Nature* **584**, 210–214 (2020).
6. H. Zhou *et al.*, *Nature* **598**, 429–433 (2021).
7. T. Han *et al.*, *Nature* **623**, 41–47 (2023).
8. G. Chen *et al.*, *Nature* **572**, 215–219 (2019).
9. H. Zhou, T. Xie, T. Taniguchi, K. Watanabe, A. F. Young, *Nature* **598**, 434–438 (2021).
10. Y. Choi *et al.*, *Nature* **639**, 342–347 (2025).
11. T. Han *et al.*, *Nature* **643**, 654–661 (2025).
12. J. Yang *et al.*, *Nat. Mater.* **24**, 1058–1065 (2025).
13. G. Chen *et al.*, *Nature* **579**, 56–61 (2020).
14. T. Han *et al.*, *Nat. Nanotechnol.* **19**, 181–187 (2024).
15. K. Liu *et al.*, *Nat. Nanotechnol.* **19**, 188–195 (2024).
16. Y. Sha *et al.*, *Science* **384**, 414–419 (2024).
17. J. Ding *et al.*, *Phys. Rev. X* **15**, 011052 (2025).
18. J. Xie *et al.*, *Nat. Mater.* **24**, 1042–1048 (2025).
19. S. H. Aronson *et al.*, *Phys. Rev. X* **15**, 031026 (2025).
20. D. Waters *et al.*, *Phys. Rev. X* **15**, 011045 (2025).
21. Z. Lu *et al.*, *Nature* **626**, 759–764 (2024).
22. Z. Lu *et al.*, *Nature* **637**, 1090–1095 (2025).
23. T. Han *et al.*, *Science* **384**, 647–651 (2024).
24. B. Zhou, H. Yang, Y. H. Zhang, *Phys. Rev. Lett.* **133**, 206504 (2024).
25. F. Winterer *et al.*, *Nat. Phys.* **20**, 422–427 (2024).
26. Y. Yang *et al.*, *Nano Lett.* **19**, 8526–8532 (2019).
27. C. H. Lui *et al.*, *Nano Lett.* **11**, 164–169 (2011).
28. W. Norimatsu, M. Kusunoki, *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 161410 (2010).
29. Z. Gao *et al.*, *Nat. Commun.* **11**, 546 (2020).
30. C. Bouhafs *et al.*, *Carbon* **177**, 282–290 (2021).
31. J. P. Nery, M. Calandra, F. Mauri, *2D Mater.* **8**, 035006 (2021).
32. B. Qin *et al.*, *Science* **385**, 99–104 (2024).
33. L. Wang *et al.*, *Nature* **629**, 74–79 (2024).
34. L. Liu *et al.*, *Nat. Mater.* **24**, 1195–1202 (2025).
35. C. Cazorla, T. Gould, *Sci. Adv.* **5**, eaau5832 (2019).
36. A. Fan *et al.*, *Sci. Adv.* **11**, eadx8192 (2025).
37. G. Savini *et al.*, *Carbon* **49**, 62–69 (2011).
38. C. Cong *et al.*, *ACS Nano* **5**, 8760–8768 (2011).
39. S. L. L. M. Ramos, M. A. Pimenta, A. Champi, *Carbon* **179**, 683–691 (2021).
40. K. G. Wirth *et al.*, *ACS Nano* **16**, 16617–16623 (2022).
41. D. Beitner *et al.*, *Nano Lett.* **23**, 10758–10764 (2023).
42. A. Torche, F. Mauri, J. C. Charlier, M. Calandra, *Phys. Rev. Mater.* **1**, 041001 (2017).
43. H. Xiao *et al.*, *Adv. Sci. (Weinh.)* **12**, e2412609 (2025).
44. H. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2410714121 (2024).
45. M. Koshino, *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 125304 (2010).
46. H. Henck *et al.*, *Phys. Rev. B* **97**, 245421 (2018).
47. Y. Zhang *et al.*, *Nat. Nanotechnol.* **20**, 222–228 (2025).
48. M. Zhao *et al.*, Dataset for "Step geometry–guided growth of rhombohedral graphene". Zenodo (2026); https://doi.org/10.5281/zenodo.20374368.

ACKNOWLEDGMENTS

Funding: This work was supported by the National Natural Science Foundation of China (92577201, 52025023, 12427806, 52402043, 12488201, T2188101, 12274006, 12141401, 92365204, 22461160283, 22333005, 12550403, and 12274298), the Quantum Science and Technology-National Science and Technology Major Project (2025ZD0300502), the National Key R&D Program of China (2022YFA1403500, 2024YFA1409002, 2025YFE0200800, and 2022YFA1604403), the research program from Suzhou Laboratory (SK-1502-2024-055), the New Generation Artificial Intelligence-National Science and Technology Major Project (2025ZD0121802), the Guangdong Major Project of Basic and Applied Basic Research (2021B0301030002), and the New Cornerstone Science Foundation through the XPLORER PRIZE to K.L. We thank the Peking Nanofab for process support.
Author contributions: K.L., X.L., Z.L., F.D., and Zh.Z. supervised the project. K.L., Zh.Z., and M.Z. conceived the experiments. M.Z., Zh.Z., and X.S. conducted the growth and transfer experiments. X.S. fabricated the devices. Q.G. performed the TEM experiments. C.Q. and F.D. conducted the MD and DFT simulations. Za.Z., Z.F., and X.L. performed the electronic transport experiments. H.X. and Z.L. performed the ARPES experiments. J.Z. and S.W. performed the IR-SNOM experiments. Y.Y. and T.D. conducted the tight-binding calculations. M.Z., Zh.Z., X.S., and K.L. wrote the article. X.-Z.L., L.W., and E.W. revised the manuscript. All the authors discussed the results and commented on the manuscript. **Competing interests:** A Chinese patent application (no. 2026102199718) related to this article was submitted by Peking University (K.L., M.Z., Zh.Z., X.S., and Q.G.). All other authors declare that they have no competing interests. **Data, code, and materials availability:** All data and details of materials synthesis are available in the main text or the supplementary materials. Additional datasets are available in Zenodo (*48*).

License information: Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aed9202
Materials and Methods; Figs. S1 to S18; References (*49*–*59*)

Submitted 15 November 2025; accepted 3 June 2026

10.1126/science.aed9202

LET YOUR KNOWLEDGE BLOSSOM

Visit the *Science* Custom Publishing sites today and grow your knowledge with a wide selection of booklets, podcasts, posters, sponsored features, and webinars!

Booklets | **Podcasts** | **Posters** | **Sponsored Features** | **Webinars**



Brought to you by the *Science*/AAAS Custom Publishing Office.

Science AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing



Let *Science* Careers help advance your career.

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings.
- Sign up to receive job alerts that match your criteria.
- Upload your resume into our database to connect with employers.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS



SCIENCECAREERS.ORG

USF Health
Institute for Translational Virology & Innovation

Faculty Position Search for Human Papillomavirus (HPV), Epstein-Barr Virus (EBV), and Influenza Research

The University of South Florida (USF) Health Morsani College of Medicine and the Institute for Translational Virology and Innovation (ITVI), directed by **Dr. Robert C. Gallo**, invite applications for an open-rank, tenure-track/tenured faculty position focused on Human Papillomavirus (HPV), Epstein-Barr Virus (EBV), and/or Influenza virus research.

We seek an internationally recognized investigator with an established record of scientific excellence and sustained extramural funding. Areas of interest include, but are not limited to, viral pathogenesis, immunology, epidemiology, host-pathogen interactions, viral evolution, diagnostics, therapeutics, vaccine development, and virus-associated cancers and chronic diseases.

The successful candidate will establish and lead an innovative, externally funded research program while joining a highly collaborative scientific community dedicated to advancing translational medicine. Faculty will work closely with ITVI investigators and the Tampa General Hospital Cancer Institute to strengthen initiatives in microbial oncology, infectious diseases, immunology, cancer biology, vaccine research, and global health. Responsibilities include conducting high-impact research, securing competitive funding, publishing in leading scientific journals, mentoring trainees and junior investigators, and contributing to the Institute's strategic growth.

Qualifications

- PhD, MD, MD/PhD, or equivalent doctoral degree in virology, microbiology, immunology, molecular biology, cancer biology, epidemiology, public health, or a related discipline.
- Demonstrated excellence in research, publication, scientific leadership, and competitive extramural funding.
- Experience leading collaborative research programs and mentoring trainees.

Preferred Qualifications

- Continuous NIH funding or equivalent international support, including current R01-level funding or comparable awards.
- Outstanding publication record in high-impact journals.
- Expertise in translational, clinical, epidemiological, molecular, immunological, vaccine, or cancer-focused research related to HPV, EBV, and/or Influenza viruses.

Located in Tampa, Florida, USF is a member of the Association of American Universities (AAU) and one of the nation's fastest-rising research universities. ITVI provides exceptional opportunities to collaborate with world-class investigators in virology, immunology, oncology, and infectious diseases.

To Apply: Submit a cover letter, curriculum vitae, research statement, and contact information for references through the USF Careers (https://jobs.usf.edu) Open Rank – Virology (43523) job posting. Review of applications will begin immediately and continue until the position is filled.



myIDP ScienceCareers
INDIVIDUAL DEVELOPMENT PLAN

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- 20 career paths with a prediction of which ones best fit your skills and interests.

Start planning your future today!
myIDP.sciencecareers.org

In partnership with:

FASEB Federation of American Societies for Experimental Biology
University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

ScienceCareers.org
ScienceCareers



A geologist learns AI

Isaac Pope

The maps and colors of my geology work were blurring together in my mind, and it was time for a break. I opened my email and saw the long-awaited result of my application to the 2025 U.S. National Science Foundation (NSF) Graduate Research Fellowship Program: honorable mention. When I checked the awardee list, I noticed an overwhelming trend: computer science, quantum physics, machine learning. Even the few natural science proposals awarded funding seemed more focused on the application of artificial intelligence (AI) than on the actual research question. At first, the tilt toward AI felt like a betrayal, as if the fundamental science was no longer "good enough." But I have since come to see it as an opportunity.

In my undergraduate studies, I had seen how AI was looming large over our future careers, but it hadn't made much of an appearance in my earth sciences training. I reveled in studying the tangible things around me, quite unlike the layers of abstractions in AI black boxes. I never envisioned myself using those sophisticated algorithms. When I drafted my fellowship proposal, I followed in the strong tradition of observation, mapping, and statistics.

Then, new federal directives prioritized AI research. Those students who had anticipated the shift were successful. Those of us who worked by the old rules found ourselves left behind.

As I began my Ph.D. not long afterward, my adviser—a computational geophysicist who pushed me to think beyond the tangible into the abstract—and I began to collaborate with an aerospace engineering professor who excelled in the use of machine learning. I began to realize those tools were really an application of the math I already knew. Maybe I, too, could learn to use them.

At first I looked for university courses, but most were locked behind heavy computer science prerequisites. My adviser encouraged me to look beyond the classroom; moving from coursework instruction to experiential learning is key to the Ph.D. journey, after all. I started to design projects to "learn by doing," but I still needed a guide to get started. That's when I realized: What better tool to introduce me to AI than AI itself? I began my dive into deep learning by using a large language model to guide me as I installed and trained a machine learning model to tackle an interesting problem in earth sciences.

It was jarring to say the least. Rather than enjoying the beauty of the natural world as I gathered data, I was wading through the harsh computer glare to install deep learning environments. You can walk up to a sinkhole, kick a rock over the edge, and dangle your legs as you enjoy your bagged lunch, but I have yet to do that with any part of an AI model. Every error code and missing dependency added to the frustration.

Slowly but surely, however, the pieces began to come together. In a matter of weeks I had my first working deep learning model, which brought insights I never expected. Before long, I was developing AI tools to identify various geologic landforms from publicly available topographic data. Rather than spending weeks mapping, I delineated the very same structures in a few afternoons. Automation not only eliminated the sampling bias of manual mapping, but it freed me to ask deeper questions about the thousands of natural features I was finding.

It's now been just over a year since I received that email from NSF. That feeling of disappointment has faded, and I feel empowered. There is still much to learn, but I have gained the confidence to learn tools I never imagined I would use. Most important, I have learned that this evolving world doesn't have to be the end for the science I love. I can embrace the blocks of code without losing the rocks that brought me here in the first place. □

Isaac Pope is a Ph.D. candidate at the Missouri University of Science and Technology. Do you have an interesting career story to share? See our author guidelines at https://scim.ag/WorkingLife.

ILLUSTRATION: ROBERT NEUBECKER

Who's the Top Employer for 2025?



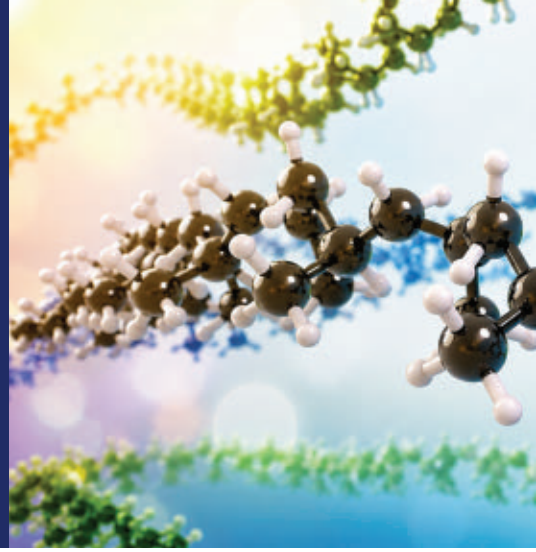
Science 2025 TOP EMPLOYERS

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Explore these highly-rated employers in our interactive experience:
sciencecareers.org/topemployers

2026 Welch Conference

October 26–27, 2026 | The Post Oak Hotel | Houston, Texas



The 2026 Welch Conference on Chemical Research explores the theme, "Frontiers in Macromolecular Science," featuring speakers who are advancing polymer science by redefining how we fabricate, sense, heal, and recycle—laying the groundwork for a cleaner, smarter, more sustainable future. Speakers include a keynote address by Joseph M. DeSimone and the 2026 Welch Awardee Lecture by Robert S. Langer.



GEOFFREY W. COATES
Welch Conference Chair
Tisch University Professor of Chemistry and Chemical Biology, Cornell University
Frontiers in Macromolecular Science

The conference will convene an exceptional group of scientists who have driven foundational advances in polymer synthesis, nature-inspired materials, and sustainable chemistry. The first session explores adaptive polymer networks, skin-inspired sensing materials, and bioinspired systems that range from antimicrobial designs to sugar-responsive constructs. The second session highlights circular monomer-reclamation strategies and precision synthesis. The third session connects molecular design to therapeutic delivery, two-dimensional polymers, and sustainable radical and cationic polymerization. The final session closes the loop with renewable feedstocks, recyclable thermoplastics, and end-of-life chemical strategies.



JOSEPH M. DeSIMONE
Keynote Speaker
Sanjiv Sam Gambhir Professor of Translational Medicine and Chemical Engineering, Stanford University
The Delicate Interplay Between Light, Interfaces, and Design to Scale 3D Printing

I will describe a breakthrough in additive manufacturing—3D printing—referred to as Continuous Liquid Interface Production (CLIP) technology. CLIP, and its cousin injection CLIP (iCLIP), embody a convergence of advances in software, hardware, and materials to bring the digital revolution to polymeric product manufacturing. CLIP uses software-controlled chemistry to produce commercial quality parts rapidly by capitalizing on oxygen-inhibited photopolymerization to generate a continual liquid interface between a forming part and the printer's exposure window, allowing layerless parts to 'grow' from a pool of resin. Compatible with a wide range of polymers, CLIP has already enabled previously unmakeable products at scale. We are pursuing new advances including digital therapeutic devices, multi-material printing, and a high-resolution printer to advance microelectronics, battery electrodes, drug delivery, and minimally-invasive liquid biopsies.

2026 Welch Awardee



ROBERT S. LANGER
David H. Koch Institute Professor, Massachusetts Institute of Technology

Welch Award in Chemistry Lecture

—

A Road Not Taken: My Path from Chemical Engineering to Nanotechnology and Medicine

Recognized for pioneering discoveries at the intersection of chemistry and medicine that led to understanding how to control the release of therapeutic macromolecules through materials and how to create new tissues and organs.

For registration, live-stream information, poster session guidelines and the complete conference program, visit our website at **welch1.org/conference**.

The Welch Foundation
Advancing Chemistry, Improving Life