Human bone collections force a reckoning with racism p. 148

Ice learns how to flex pp. 158 & 187

Microbial metabolomics wins NOSTER & *Science* Microbiome Prize p. 172

Science

$15
9 JULY 2021
sciencemag.org

AAAS

FROM FLOWERS TO FRACTALS

Perturbed development produces a multitude of spirals p. 192

Advance Your Career at the Frontiers of Science!
October-November 2021

> *"I remember my first GRC experience because the quality of science and depth of discussion were inspirational to me as a young graduate student."*
>
> GEOFFREY VARGISH, Past Chair of the Inhibition in the CNS GRS

Alcohol-Induced End Organ Diseases + GRS
OCTOBER 24 – 29, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Gyongyi Szabo and Bernd Schnabl
VICE CHAIRS: Elizabeth Kovacs and Gavin Arteel

Biotherapeutics and Vaccines Development + GRS
NOVEMBER 14 – 19, 2021
VENTURA BEACH MARRIOTT, VENTURA, CA, US
CHAIRS: Michael Tarlov and Linda Narhi
VICE CHAIRS: Peter Tessier and Jessica Molek

Cannabinoid Function in the CNS + GRS
NOVEMBER 7 – 12, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Matthew Hill and Sachin Patel
VICE CHAIRS: Miriam Melis and Mario Stelt

Carbon Capture, Utilization and Storage + GRS
OCTOBER 24 – 29, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIR: Roger Aines
VICE CHAIR: Andre Bardow

Cardiac Arrhythmia Mechanisms + GRS
OCTOBER 17 – 22, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: David Christini and Ursula Ravens
VICE CHAIRS: Crystal Ripplinger and Edward Vigmond

Cell Biology of Metals + GRS
OCTOBER 17 – 22, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIRS: Caryn Outten and Mitchell Knutson
VICE CHAIRS: Angela Wilks and Kerry Kornfeld

Central Nervous System Injury and Repair + GRS
OCTOBER 17 – 22, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIRS: Yimin Zou and Simone Di Giovanni
VICE CHAIRS: Elizabeth Bradbury and Yishi Jin

Chronobiology + GRS
OCTOBER 10 – 15, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Samer Hattar
VICE CHAIR: Charlotte Helfrich-Forster

Computer Aided Drug Design + GRS
NOVEMBER 7 – 12, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Georgia McGaughey
VICE CHAIR: Matthias Rarey

Lung Development, Injury and Repair + GRS
NOVEMBER 7 – 12, 2021
WATERVILLE VALLEY, WATERVILLE VALLEY, NH, US
CHAIRS: Anne-Karina Perl and Daniel Tschumperlin
VICE CHAIRS: Rory Morty and Xin

Mammalian DNA Repair + GRS
OCTOBER 31 – NOVEMBER 5, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIR: Lee Zou
VICE CHAIR: Patricia Opresko

Medicinal Chemistry + GRS
OCTOBER 24 – 29, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Janeta Popovici-Muller
VICE CHAIR: Cynthia Parrish

Phagocytes + GRS
OCTOBER 24 – 29, 2021
WATERVILLE VALLEY, WATERVILLE VALLEY, NH, US
CHAIR: Michelle Lennartz
VICE CHAIR: Sergio Catz

Physical Science of Cancer + GRS
NOVEMBER 14 – 19, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Josef Käs and Sharon Gerecht
VICE CHAIRS: David Odde and Jennifer Munson

Stochastic Physics in Biology + GRS
OCTOBER 10 – 15, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIR: Yuhai Tu
VICE CHAIR: Jie Xiao

Stress Proteins in Growth, Development and Disease + GRS
NOVEMBER 7 – 12, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIR: Brian Freeman
VICE CHAIR: Harm Kampinga

Three Dimensional Electron Microscopy + GRS
OCTOBER 31 – NOVEMBER 5, 2021
JORDAN HOTEL AT SUNDAY RIVER• NEWRY, ME, US
CHAIR: Sharon Wolf
VICE CHAIR: Daniela Nicastro

Apply now at www.grc.org!

Using physics to better understand and use materials

Studying the characteristics of materials has captured the attention of many scientists, especially physicists. The knowledge that researchers gain from dissecting the properties of materials impacts the lives of people around the world. In three virtual physics presentations organized in April and May 2021 by the Hong Kong Institute for Advanced Study (HKIAS) at the City University of Hong Kong (CityU), expert scientists discussed the history and latest advances in understanding atomic-level interactions in glass, tuning quantum states, and raising the temperature of superconductivity, all part of the HKIAS Distinguished Lecture Series on Physics. This series is supported in part by the Kwang Hua Educational Foundation.

Exploring the grand challenge of glass

On April 14, Xun-Li Wang–Head and Chair Professor of the Department of Physics at CityU–kicked off this series with a lecture titled "Structure and Dynamics of Metallic Glass–Atomistic Insights from Neutron and Synchrotron Scattering Experiments."

Wang emphasized that understanding materials made of glass is one of the biggest challenges in science, leading the audience through his atomic-level findings about the structure of metallic glass and the dynamics of phonons (elementary heat particles) within it.

Scattering is a powerful technique for studying the structure and dynamics of glass, Wang noted. "This is an important topic and there are plenty of opportunities in this field," he said, while emphasizing the scientific prospects available in Dongguan, China.

Tuning the efficiency of quantum states

Basing his lecture on the demand for information about low-dimensional physics theory, Ruiqin Zhang–Chair Professor of Physics and Materials Science and Engineering at CityU–discussed "Tuning of Confined Quantum States" on April 28.

"Confined quantum states open up many possibilities," Zhang said. "We have been working on engineering these confined quantum states using different strategies," he noted–strategies that include surface engineering, stress or strain engineering, and excited-state engineering.

Research in this area facilitates the development of miniaturized and efficient equipment that uses solar energy to produce hydrogen. Zhang added that this research can promote the application of low-dimensional materials in the fields of optoelectronics, nanoelectronics, environmental science, green energy, biology, and medicine.

Warming up superconductors

In the last lecture of this series on May 12, Wei Bao–Chair Professor in the Department of Physics at CityU–delivered a talk entitled "Lattice, Charge, Spin and Orbital Aspects of the Iron Chalcogenide Superconductors."

Discovered in 1911, superconductivity–the ability of a charge to move through a material with no electrical resistance–appears in certain materials at extremely low temperatures. Bao explained how room-temperature superconductivity "has been a dream for us for many decades." To find superconductivity at even higher temperatures, scientists like Bao are studying materials such as iron chalcogenides; these insulating cuprate ceramics created an Iron Age of superconductors and offer hope for realizing Bao's dream.

A commitment to a better world

HKIAS is committed to advancing CityU and the world's academic community by delivering groundbreaking research and encouraging talented young researchers. In recent years, the institute has provided a platform for preeminent visiting scholars to engage in pioneering research and contribute to postdoctoral and postgraduate training.

To advance the frontiers of interdisciplinary research, HKIAS is creating research clusters. Three recent clusters–the HKIAS Materials Science Cluster, the HKIAS Bioscience Cluster, and the HKIAS Mathematics Cluster–encourage collaboration and deliver high-quality research. Led by HKIAS senior fellows, each cluster addresses a significant problem or set of interrelated questions through individual and collective research. Serving as incubators for larger and longer-term collaborations, the clusters aim to attract external funding and generate significant results. They also enrich the institute's distinctive intellectual community through partnerships that spark new approaches and findings, as well as facilitating the mentoring of junior scholars.

To make the world a better place, HKIAS will keep delivering pioneering research and nurturing young, talented scholars across disciplines–driving the development of CityU and of science education across the globe.

Hong Kong Institute for Advanced Study
Phone: +852-3442-6611
Email: hkias@cityu.edu.hk

PHOTOS: COURTESY OF HKIAS

CONTENTS

9 JULY 2021 • VOLUME 373 • ISSUE 6551

NEWS

IN BRIEF

140 News at a glance

IN DEPTH

142 Can immune responses predict which vaccines work best?
Elusive "correlates of protection" could lead to approvals of boosters or new vaccines without big clinical trials *By J. Cohen*

143 Smell proves powerful sense for birds
New studies highlight underappreciated role of avian olfaction *By E. Pennisi*

145 Sex and gender missing in COVID-19 data
Despite suggestions of differential effects, most clinical trials don't report results by sex
By C. O'Grady

146 Something is killing U.S. birds. It's not cicadas
Birds with crusty eyes and neurological damage found in nine states and Washington, D.C. *By D. Malakoff and E. Stokstad*

147 Journal retracts paper claiming COVID-19 vaccines kill
Editors at *Vaccines* quit, protesting "irresponsible" study
By M. Wadman

FEATURES

148 The ghosts in the museum
Anthropologists are reckoning with collections of human remains—and the racism that built them *By L. Wade*

153 Charleston honors Black ancestors, with both science and ceremony *By L. Wade*

INSIGHTS

PERSPECTIVES

154 A shift in taste
The evolution of sugar perception in songbirds began with a savory receptor
By F. K. Barker
REPORT p. 226

155 Sowing the seeds of leukemia before birth
Distinct cooperation between a mutated gene and trisomy 21 triggers leukemia in utero
By I. Roberts and P. Vyas
RESEARCH ARTICLE p. 179

157 A touch more unconventional
Strong antiferromagnetic interactions are revealed in superconducting nickelates *By E. Benckiser*
REPORT p. 213

158 A flexible and springy form of ice
Single-crystal ice microfibers recover their shape after bending to near their breaking limit
By E. M. Schulson
RESEARCH ARTICLE p. 187

159 Host genetics influence the gut microbe
Longitudinal data from nonhuman primates reveal widespread gut microbe heritability
By L. Cortes-Ortiz and K. R. Amato
RESEARCH ARTICLE p. 181

160 Boosting stem cell immunity to viruses
A newly discovered isoform of Dicer protects stem cells by enhancing antiviral RNA interference
By S. Shahrudin and S.-W. Ding
REPORT p. 231

162 Networks of SARS-CoV-2 transmission
Individual and network heterogeneity should inform respiratory pandemic responses *By M. Cevik and S. D. Baral*

164 Retrospective: Richard Ernst (1933–2021)
Chemist who made magnetic resonance a powerful tool
By G. Bodenhausen

CREDITS: (ILLUSTRATION) JOHNALYNN HOLLAND; (PHOTO) MICHAEL QUINTON/MINDEN PICTURES

POLICY FORUM

165 ARPA-H: Accelerating biomedical breakthroughs
A DARPA-like culture at NIH can drive biomedical and health advances *By F. S. Collins* et al.

BOOKS *ET AL.*

168 Medicinal plants, in context
A journalist's meandering meditation probes a trio of mind-altering natural compounds
By K. Sanford

169 Stepping out of the comfort zone
A writer grapples with the environmental and social costs of mechanical cooling
By G. Dreyfus

LETTERS

170 China's ambitious energy transition plans
By X. Shi et al.

170 Institutions key to inclusion and equity
By T. L. Goulet

171 Saving China's blue-crowned laughingthrush
By N. Li et al.

PRIZE ESSAY

172 Can microbes combat neurodegeneration?
Identifying a new link between microbiome and metabolites in amyotrophic lateral sclerosis *By E. Blacher*

RESEARCH

IN BRIEF

175 From *Science* and other journals

REVIEW

178 Light-matter coupling
Manipulating matter by strong coupling to vacuum fields *F. J. Garcia-Vidal* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD0336

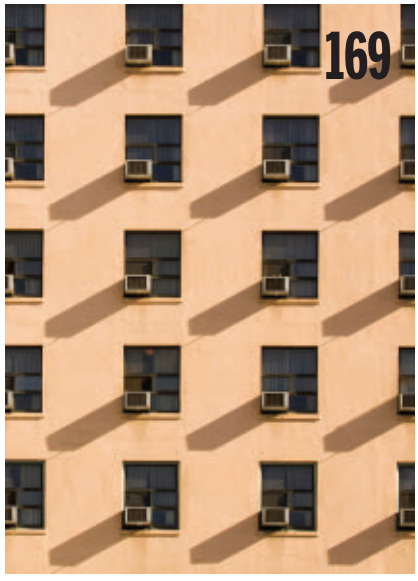
RESEARCH ARTICLES

179 Cancer
Mapping the cellular origin and early evolution of leukemia in Down syndrome
E. Wagenblast et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF6202
PERSPECTIVE p. 155

180 Coronavirus
Estimating infectiousness throughout SARS-CoV-2 infection course *T. C. Jones* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABI5273

181 Microbiomes
Gut microbiome heritability is nearly universal but environmentally contingent
L. Grieneisen et al.
PERSPECTIVE p. 159

187 Ice physics
Elastic ice microfibers *P. Xu* et al.
PERSPECTIVE p. 158



192 Plant science
Cauliflower fractal forms arise from perturbations of floral gene networks
E. Azpeitia et al.

198 Martian geology
Brine-driven destruction of clay minerals in Gale crater, Mars *T. F. Bristow* et al.

REPORTS

204 Geomorphology
The life span of fault-crossing channels
K. Dascher-Cousineau et al.

208 Mechanochemistry
Flyby reaction trajectories: Chemical dynamics under extrinsic force *Y. Liu* et al.

213 Superconductivity
Magnetic excitations in infinite-layer nickelates *H. Lu* et al.
PERSPECTIVE p. 157

217 Catalysis
Stable and selective catalysts for propane dehydrogenation operating at thermodynamic limit *A. H. Motagamwala* et al.

223 Physiology
Skeletal muscle thermogenesis enables aquatic life in the smallest marine mammal
T. Wright et al.

226 Sensory evolution
Early origin of sweet perception in the songbird radiation *Y. Toda* et al.
PERSPECTIVE p. 154

231 Antiviral defense
An isoform of Dicer protects mammalian stem cells against multiple RNA viruses
E. Z. Poirier et al.
PERSPECTIVE p. 160

236 Coronavirus
Fe-S cofactors in the SARS-CoV-2 RNA-dependent RNA polymerase are potential antiviral targets *N. Maio* et al.

242 Neuroscience
Nonlocal spatiotemporal representation in the hippocampus of freely flying bats
N. M. Dotson and M. M. Yartsev

DEPARTMENTS

135 Editorial
Save Earth's global observatories
By Gene E. Likens and David L. Wagner

250 Working Life
Standing my ground *By Manya Ruckhaus*

ON THE COVER

The fractal-like forms of the Romanesco cauliflower are some of the most spectacular in the plant kingdom. These structures arose from the transformation of the cabbage inflorescence during the domestication process. Researchers show how perturbations of the floral gene networks combined with modifications of growth dynamics led to the emergence of such fractal forms. See page 192.
Photo: Laura Chase de Formigny Photography

Science Staff 134
Science Careers 248

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS.

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

PHOTO: LHB PHOTO/ALAMY STOCK PHOTO

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Di Jiang, Marc S. Lavine (Canada), Yevgeniya Nusinovich, Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Kelly Servick, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Menna Ibrahim, Sofia Moutinho, Anil Oza, Alex Viveros **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Tokyo), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Joshua Sokol, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB STRATEGY MANAGER** Kara Estelle-Powers **MULTIMEDIA MANAGING PRODUCER** Joel Goldberg **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Ashley Mastin **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell **INTERN** Claire Hogan

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT AND PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy **PRODUCTION SPECIALIST** Kristin Wowk **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris, Justin Wood **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Anne Mason **SPJ ASSOCIATE MANAGER** Samantha Bruno Fuller **SPJ ASSOCIATE** Casey Buchta

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **MARKETING MANAGER** Kess Knight **BUSINESS DEVELOPMENT MANAGER** Rasmus Andersen **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan, Claudia Paulsen-Young **CUSTOMER SUCCESS MANAGER** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** 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** Yoshimi Toda (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren

---

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services
science_advertising@aaas.org

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

AAAS BOARD OF DIRECTORS
CHAIR Claire M. Fraser
PRESIDENT Susan G. Amara
PRESIDENT-ELECT Gilda A. Barabino
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Rosina M. Bierbaum, Ann Bostrom, Janine Austin Clayton, Laura H. Greene, Kaye Husbands Fealing, Maria M. Klawe, Robert B. Millard, William D. Provine

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

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

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Inst. Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Nandita Basu, *U. of Waterloo*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr.*
Annmarie Carlton, *UC Irvine*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
Amander Clark, *UCLA*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Christian Davenport, *U. of Michigan*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
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 Ehlers, *U. of Tübingen*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *GWU*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Paul Kubes, *U. of Calgary*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
Wendell Lim, *UCSF*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Wash. U. in St. Louis*
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marín, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
Jason Matheny, *Georgetown U.*
Heidi McBride, *McGill U.*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
Jane Memmott, *U. of Bristol*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Suresh Naidu, *Columbia U.*
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Daniel Pauly, *U. of British Columbia*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Lei Stanley Qi, *Stanford U.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UCSF*
Robert Siliciano, *JHU School of Med.*
Lucia Sivilotti, *U. Coll. London*
Alison Smith, *John Innes Ctr.*
Richard Smith, *UNC* **(S)**
Mark Smyth, *QIMR Berghofer*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Dartmouth*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Wim van der Putten, *Netherlands Inst. of Ecology*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
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.*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* **(S)**
Hao Wu, *Harvard U.*
Wei Xie, *Tsinghua U.*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

Save Earth's global observatories

Sitting at the interface of human societies and the natural environment are sentinels tracking environmental change. Across the globe, field stations and marine laboratories (FSMLs) amass crucial information about climate, biodiversity, environmental health, and emerging diseases, anchoring multidecadal data sets needed to solve environmental challenges of the Anthropocene. These observatories are now in danger of being shut down—part of the collateral damage of the COVID-19 pandemic.

On every continent, facilities have been shuttered and field courses canceled because of restricted travel. This has reduced the flow of financial support to these stations, debilitating their capacity to collect essential information and train the next generation of scientists. Two-thirds of university support for the Tiputini Biodiversity Station in eastern Ecuador—situated in Earth's most biologically diverse region, at the confluence of the Amazon Basin and Andes—came from international universities, nearly all of which was permanently terminated during the pandemic. The renowned Asa Wright Centre in Trinidad and Tobago closed in April. Further, FSML budgets are menaced by pandemic-related deficits suffered by their parent institutions—which are generally universities (a reflection of their importance to education and training) but also include museums, government agencies, and nongovernmental organizations—potentially compromising every facet of their operations.

"Nature's struggles are humanity's struggles..."

As Earth's population swells to 8 billion, understanding and predicting human impacts on the planet become ever more urgent. Both long-term and real-time data are needed to quantify the repercussions of deforestation, agricultural intensification, desertification, climate change, ocean acidification, and other stresses if we are to mitigate their effects, plan adaptive responses, and develop national and international policies. Nature's struggles are humanity's struggles: As biodiversity is lost and ecosystems erode, so will the quality of our air, waters, and soils. This degradation will also affect the essential ecosystem services that nature consistently provides. Crop pollination services alone are estimated as a $500 billion annual benefit for society. And emerging pathogens will continue to be a threat across all borders. Environmental data to guide sound, science-based solutions, and broader public understanding and engagement, are necessary to overcome these mounting environmental challenges.

FSMLs are essential for educating and training the next generation of scientists. Immersive in situ experiences are foundational to those seeking careers in biology and ecology, geology and soil science, oceanography, hydrology and limnology, meteorology, conservation, and resource management. Evidence shows that field courses close demographic gaps in science participation and persistence and improve diversity across disciplines. Virtual materials and live-stream research-based field experiences simply cannot supplant place-based learning, curiosity-driven exploration, the life-changing value of discovery, and the realization that Earth is still a little-known planet. Furthermore, FSMLs play a broader role in education. Field course alumni become educators and school administrators, or pursue careers in medicine, law, social services, and business, among other professions. Scientists and nonscientists alike take away a deeper understanding and appreciation for nature and a propensity to embrace an ethic of planetary stewardship.

In January 2021, an international call to protect funding for field stations, marine labs, and field courses, imperiled by the pandemic, was circulated to professional society mailing lists. The signature petition has since been endorsed by 21 past presidents of the Ecological Society of America; active and past leadership of multiple international scientific and educational organizations; and more than 2200 scientists, station directors, educators, and concerned citizens worldwide. And last month, the US Congress passed the National Science Foundation (NSF) for the Future Act (House bill H.R. 2225) with bipartisan support, which includes provisions for national labs, field stations, and marine labs. Despite NSF's pivotal role in supporting FSMLs for decades, its funding is generally limited to new initiatives and infrastructure. NSF dollars are rarely sufficient to support staff, maintenance, courses, and other day-to-day station needs. The issue will hopefully receive bipartisan support from the Senate as well.

The pandemic has cut revenue streams to FSMLs for a second year. At a time when environmental issues demand even greater attention, the world cannot risk undermining their contributions to scientific literacy, environmental research, and student training—all of which are essential to protect Earth's bountiful natural heritage and life-sustaining ecosystems. Universities, governments, and other organizations must find ways to save these global sentinels—all life depends on them.

–Gene E. Likens and David L. Wagner

Gene E. Likens is a distinguished research professor in the Department of Ecology and Evolutionary Biology at the University of Connecticut, Storrs, CT, USA, and founding director and president emeritus at the Cary Institute of Ecosystem Studies, Millbrook, NY, USA. likensg@caryinstitute.org

David L. Wagner is a professor in the Department of Ecology and Evolutionary Biology at the University of Connecticut, Storrs, CT, USA. david.wagner@uconn.edu

10.1126/science.abk2615

Translating the benefits of gut microorganisms into disease treatments and better health care

Noster has launched its first postbiotics product based on decades of research on gut microorganisms including lactic acid bacteria and related lipid metabolites.

"Mankind has come a long way in understanding the causes and treatment of human ailments since Hippocrates proposed over 2,000 years ago that all disease begins in the gut," says Kohey Kitao, CEO of Noster Inc., an innovative Kyoto-based company specializing in the development of therapeutic drugs to treat diseases by controlling and manipulating gut microorganisms and their metabolites. "We established Noster in May 2020 with the goal of 'connecting life and the gut microbiome.' Our in-house expertise has evolved tremendously since my first chance encounter more than 10 years ago with gut microbes and fatty acid metabolites in the human gut. Now, in 2021, we have successfully launched HYA-50, our first postbiotic product under our CUMEC brand, which stands for 'cutting-edge microbiome care.' It is a functional supplement that is the direct result of Noster's knowledge of the gut microbial lipid metabolite 10-hydroxy-*cis*-12-octadecenoic acid (10-HOE, also called HYA) and is so named because it contains 50% HYA. Our research shows that taking three capsules of HYA-50 before meals improves insulin resistance and controls blood glucose levels. This treatment heralds the beginning of an exciting new chapter in our story of the benefits of gut microorganism for human health."

Embracing serendipity, realizing potential, and overcoming challenges

The scientific know-how accumulated by Noster is based on long-term, multidisciplinary collaborations with experts in academia. An interest in the health benefits of HYA is one thing, but turning that interest into a viable product is something quite different. The company's first major challenge was establishing a process for mass-producing HYA.

"I first became aware of the potential health benefits of HYA during serendipitous meetings with two academics," recalls Kitao. "The first was with Jun Ogawa, a professor at Kyoto University, who was excited by his group's discovery that HYA is produced by gut microbes in the intestinal tract and transferred to the host (*1*). Next, I remember long conversations about HYA with Ikuo Kimura, now also a professor at Kyoto University, whose group made important fundamental contributions to our understanding of the potential role of HYA for the treatment of chronic type 2 diabetes, inflammatory bowel disease, and insulin resistance in mouse models (*2*, *3*). I wanted to explore potential pharmaceutical applications of these scientific findings but realized that the first hurdle to overcome for drug discovery applications was producing HYA in large volumes. This was around 10 years ago. It was a daunting task, but we took up the challenge!"

PHOTOS: PROVIDED BY NOSTER

Large-scale production of HYA-50 and the launch of CUMEC

Producing HYA for functional supplement applications necessitates that all the materials used must themselves be composed of foods. "Given the strict regulations for such products, our goal was to produce at least 50%-pure HYA in large quantities," explains Kitao. "We used high-purity sources of linoleic acid, a form of seed oil, and cultivated lactic acid bacteria in-house for high-efficiency conversion of this source oil into HYA, which is produced by gut microbial metabolism in the human digestive tract. It was a difficult task, but we managed to complete it successfully through a collaboration with Kyoto University."

For the mass production of HYA, the Noster R&D team decided to use highly pure (>70%) safflower oil as the source for the linoleic acid. Their next task was devising a procedure that would yield a high conversion rate of the linoleic acid into HYA. Their experiments revealed that it would be critical to protect the safflower oil from oxidation and optimize the emulsion of ingredients to enhance contact between the lactic acid bacteria and linoleic acid.

"Our process was anaerobic, with a highly efficient reaction in the emulsified state," says Kitao. "The know-how for cultivating anaerobic bacteria in the laboratory is one of Noster's key technologies. It is not a straightforward process, necessitating a deep understanding of the behavior of gut microorganisms. We can now extract many kinds of metabolites directly from gut microorganisms cultivated in-house."

The final step was to recover HYA from the reaction solution with a purity of more than 50%. Since it had emulsified in the reaction process, it could not be separated cleanly using normal centrifugation. The Noster team careful analyzed the effect of temperature on the physical characteristics of the fatty acid–water emulsion, and determined an optimal temperature that led to high-efficiency separation of 50% pure HYA from the emulsion.

"Using this process, we went from being able to produce several grams of HYA-50 with a purity of 50% to synthesizing kilogram amounts," says Kitao. "This production technology is at the heart of our HYA-50 CUMEC postbiotic product."

Boosting purity for clinical research

The purity of HYA-50 was adequate in a consumer product, but much higher purity—closer to 99%—was needed to perform advanced clinical studies. The most difficult step in producing ultrapure HYA was developing a highly efficient refining process. The conversion of linoleic acid to HYA is a reversible enzymatic reaction and is therefore never 100% complete. To address this problem, the Noster team attempted to remove residual linoleic acid using a fractionation column after the reaction had run its course. However, this conventional approach did not yield complete separation, leaving too much linoleic acid residue in the refinement product.

Noster's solution was to find lactic acid bacteria strains that increased the conversion rate, thereby minimizing the amount of linoleic acid remaining after the reaction. Notably, even the best lactic acid bacteria strains selected from Noster's library of about 400 strains for mass production of high-purity HYA yielded a maximum conversion rate of only 80%, leaving 20% linoleic acid remaining. To solve this problem, the project team turned to *Escherichia coli*, creating a strain of this bacteria that carried the conjugated linoleic acid hydrase (CLA-HY) enzyme, which converts linoleic acid to HYA. After testing numerous *E. coli* strains, transfection vectors, and culture conditions, the research team succeeded in generating a strain capable of converting close to 100% of linoleic acid to HYA.

Another challenge was that conventional purification processes rely on diethyl ether. Production of large volumes of HYA would require large amounts of this ether, which is extremely flammable and can only be used safely in small amounts, severely limiting the production potential of pure HYA.

Kitao explains the Noster team's approach to this problem: "We experimentally found a solvent with low flammability that could efficiently separate HYA and linoleic acid. But then we hit another wall. This solvent led to the rapid deterioration of the fractionation column, and it was not a practical solution for a continuous purification process. So we still had work to do!"

Eventually, the Noster team developed a customized fractionation column to efficiently separate HYA and linoleic acid using the new solvent, but with significantly less deterioration. "The combination of the new solvent and our new column enabled us to scale up the production of our ultrapure HYA product."

Production of other lipid metabolites

Over the last decade, Noster researchers have established procedures for producing other types of lipid metabolites in

PHOTOS: PROVIDED BY NOSTER

recombinant bacteria. But it was not easy. For example, in the case of linoleic and oleic acids, the fatty acid conversion reaction occurs in multiple steps, and each step of the reaction is catalyzed by a different enzyme.

Furthermore, these are reversible reactions exhibiting conversion rates of less than 100%, with the amount of the desired final product decreasing with each step. Noster researchers experimented with using multiple strains of recombinant bacteria to improve efficiency, but the reactions were difficult to control, and it was not possible to efficiently isolate target compounds. For instance, it was difficult to simultaneously separate and purify individual fatty acids from a solution containing a mixture of multiple fatty acids, because the structures and physical properties of each fatty acid were similar. "After much experimentation we were able to improve our yield of a wide range of lipid metabolites from several milligrams to tens of milligrams," says Kitao. "This was enough for cell-based experiments and animal tests, accelerating the drug discovery process."

Libraries for drug discovery

Noster's technology for cultivating gut microorganisms and related metabolites has enabled the company to establish two unique libraires: a microbial library of more than 1,600 strains and a lipid metabolite collection of more than 300 compounds. "Our libraries offer a wealth of information on gut microorganisms and metabolites that can be candidates for drug discovery. They have attracted attention from researchers working on therapeutics and drug discovery based on the gut microbiome. We continue to welcome collaborations."

For information about Noster's gut microbiome libraires:
Microbial library: www.noster.inc/products/microbial/.
Lipid metabolite library: www.noster.inc/products/metabolite/.

The importance of standardization and sharing

In 2016, Japan launched the *Society 5.0* initiative as part of its national 5th Science and Technology Basic Plan, which builds on the four historical social infrastructures of hunting, agriculture, industry, and information, and hopes to create a new society where people and technology are increasingly linked, and where artificial intelligence helps us break previous barriers. Notably, molecular-level health care–genetic engineering–based medicine for the personalized treatment of chronic diseases–is an important pillar of the plan. Noster believes that further advances in this field necessitate sharing accurate data on lipidomics as part of an open-access, large-scale data platform.

"Noster is an active member of Japan's Council on Competitiveness-Nippon [COCN] and is providing full access to our library of microorganisms and gut microbial lipid metabolites," says Kitao. "We fully support projects to standardize data on lipids and metabolites for use in lipidomics research."

PHOTOS: PROVIDED BY NOSTER

Notably, Noster has also been conducting its own research on standardization and lipid analysis with Makoto Arita, a professor at Keio University and team leader of the Laboratory for Metabolomics at the RIKEN Center for Integrative Medical Sciences. Arita is using the latest liquid chromatography-mass spectrometry-based lipidomics to standardize the properties of lipids, with the goal of elucidating the structure and function of endogenous lipid mediators that regulate inflammation and tissue homeostasis.

COCN has formulated plans to establish next-generation health care systems centered on providing customized medical care for individuals at the molecular level. This year COCN proposed establishing a national lipids center, which is expected to be completed within the next 3 to 5 years.

Diabetes clinical trials with HYA

One of Noster's major goals is harnessing the power of metabolites produced by gut microorganisms to develop therapeutics to treat the underlying causes of chronic diseases such as diabetes and intestinal inflammation disorders. To achieve this goal, the first important milestone is initiating human clinical trials.

"We have invested considerable time and effort into developing protocols to produce large volumes of high-purity HYA to enable us to initiate these trials," says Kitao. "In November 2020, we started clinical trials with Wataru Ogawa at Kobe University on the efficacy of HYA on 'insulin sensitivity in patients with impaired glucose tolerance or mild diabetes mellitus,' as stated in the Japan Registry of Clinical Trials. This trial is based on oral administration of 99%-pure HYA capsules–a minimally invasive approach compared with conventional highly invasive injection-based treatments."

Noster is also planning to launch clinical trials in 2021 on using HYA to treat individuals with a high body mass index (>30) and type 1 diabetes.

The Noster & *Science* Microbiome Prize

"Global multidisciplinary research partnerships will be critical for further innovation and advances in microbiome research," says Kitao. "In 2020, we launched the Noster & *Science* Microbiome Prize to support and inspire young scientists to find sustainable solutions to chronic diseases. We enjoyed meeting the 2020 recipients during the online awards ceremony held in September 2020 and are pleased to congratulate the 2021 winners."

The 2021 Grand Prize winner is Eran Blacher, postdoctoral research fellow in neurology and neurological sciences at Stanford University in California. The other finalists are Maria Zimmermann-Kogadeeva, a group leader at EMBL Heidelberg in Germany, and Erez Baruch, who is currently doing his internal medicine residency at The University of Texas Health Science Center in Houston, Texas.

Further information about the Noster & *Science* Microbiome Prize can be found at www.sciencemag.org/prizes/noster-science-microbiome-prize.

A summary of the proceedings of the 2020 Prize awards ceremony is available on the Microbiome Research X website: noster-mrx.net.

Microbiome Research Podcast (MRX Podcast)

In April 2020, Noster launched the Microbiome Research X (MRX) website to highlight the latest research results in this field.

"Inspired by the success of the MRX website, we decided to launch the MRX Podcast in the fall of 2020," explains Kitao. "The podcast hosts conversations with scientists working on the gut microbiome. It is a highly effective way of connecting with people on a global scale." The first three podcasts feature interviews with the grand winner and two finalists of the 2020 Noster & *Science* Microbiome Prize. They describe their research as well as how they prepared their essays for the prize.

You can listen to the MRX Podcasts here:

Apple Podcasts

Google Podcasts

A future of postbiotics-inspired pharma

Over the last decade, Noster has made important contributions to the international understanding of the role of gut microorganisms in human health. "Our journey to study the health benefits of gut microorganisms busily working away inside our bodies has only just started," explains Kitao. "I look forward to the day when postbiotic-based therapeutics will be so widespread that we do not even think about it anymore. It will be just like breathing."

References

1. S. Kishino *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17808–17813 (2013).
2. J. Miyamoto *et al.*, *J. Biol. Chem.* **290**, 2902–2918 (2015).
3. J. Miyamoto *et al.*, *Nat. Commun.* **10**, 4007 (2019).

Sponsored by

NOSTER

NOSTER | Science MICROBIOME PRIZE

> “Chucking live lobsters into pots to turn them red is a major issue.”
>
> **Mike Radford**, who studies animal welfare law at the University of Aberdeen, in *The Washington Post*, about a U.K. bill that would protect animals deemed “sentient”—which it doesn’t define.

IN BRIEF

Edited by **Jeffrey Brainard**

People line up to get China’s Sinovac vaccine at a sports stadium in Jakarta, Indonesia, where most hospital beds are full because of a COVID-19 surge.

PHOTO: DITA ALANGKARA/AP IMAGES

COVID-19

Precautions ease as Delta variant spreads, driving up cases

Public health specialists are blaming a troubling mix of lagging immunizations, less mask use, more social gatherings, and the spread of the more contagious Delta variant of SARS-CoV-2 for recent surges in the number of COVID-19 cases in many countries. The World Health Organization says the combination is delaying the pandemic’s end. Last week, cases in Europe increased for the first time in 10 weeks. Growth in Indonesia, Russia, and South Africa was also worrying, with all seeing daily infections double in recent weeks. Some health experts criticized the U.K. government for its 5 July announcement that it would ease almost all restrictions later this month, despite a dramatic increase in daily infections, nearly all of them from the Delta variant. Cases dropped across North America, but in one-third of the United States’s 3000 counties, the share of people who are fully vaccinated remains below 30%. Data from Israel and the United Kingdom suggest that available vaccines still provide strong protection against severe disease from Delta. But protection against mild infections has dropped significantly—suggesting it will be even harder for vaccination campaigns to curb the virus’ toll.

CureVac disappoints again

COVID-19 | The final analysis of a large clinical trial of a candidate vaccine against COVID-19 last week yielded further evidence of its mediocre efficacy—but its manufacturer, German biotech firm CureVac, is still hopeful that the vaccine, which is based on messenger RNA (mRNA), can protect people under age 60. An interim analysis of the trial, which recruited 40,000 participants in 10 European and Latin American countries, had shown the vaccine to have only 47% efficacy (*Science*, 25 June, p. 1381). The final analysis, based on a total of 228 COVID-19 cases in the placebo and vaccinated groups, put overall efficacy at 48%—much lower than other mRNA vaccines. But in participants ages 18 to 60, the rate was 53%, a notch above the 50% deemed acceptable by the World Health Organization. In this group, the vaccine offered 77% protection against moderate to severe disease and 100% against hospitalization and death. Based on

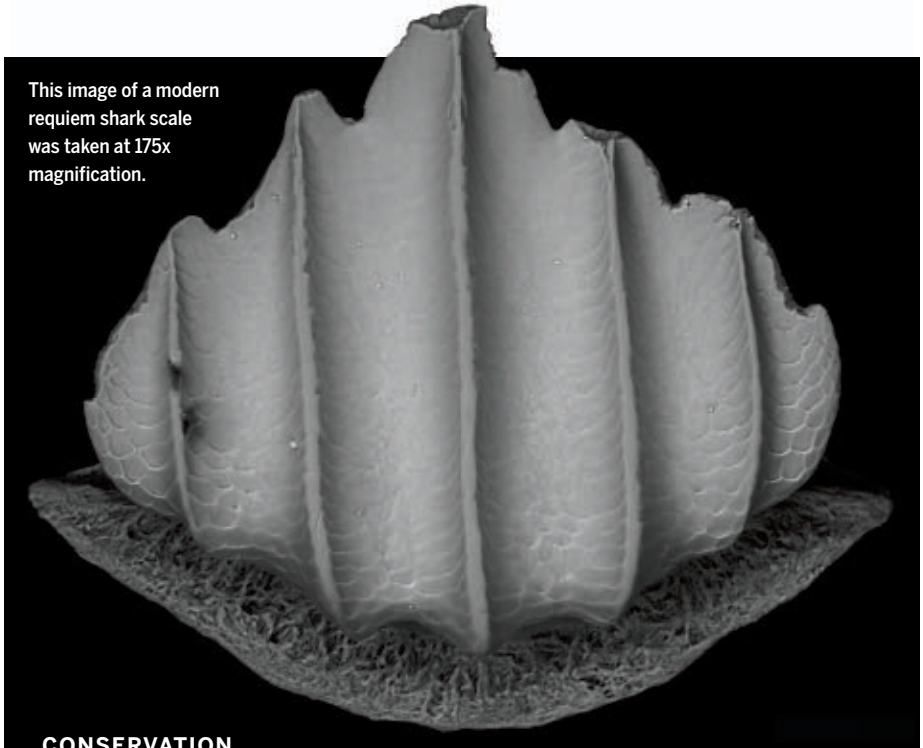
those numbers, the company is in a "continued and constructive dialogue with the European Medicines Agency" to authorize the vaccine, CureVac CEO Franz-Werner Haas said last week.

A genetic 'vaccination' for bats

INFECTIOUS DISEASES | To prevent another pandemic, two researchers have proposed spreading a genetic change through populations of wild bats to protect them from coronaviruses that could spill into humans. The approach, known as a gene drive, is "far from orthodox," say Daniel Douek of the U.S. National Institute of Allergy and Infectious Diseases and Yaniv Erlich of the Interdisciplinary Center Herzliya in their proposal, published last week on Github and first reported by *STAT*. The method would use the gene editing technology CRISPR to introduce a strand of genetic material into the genome of horseshoe bats, a known coronavirus reservoir. The added strand would destroy invading coronavirus and would preferentially pass to an altered bat's offspring, spreading the immunity widely through the population. Gene drive approaches are under development to eradicate populations of other animals, including mosquitos that transmit malaria and other viruses that harm humans. but no animals bearing such modifications have yet been released.

Iran unveils ambitious telescope

ASTRONOMY | Iran last week inaugurated what aspires to be a world-class optical telescope—but the facility is months away from seeing first light. Perched on Mount Gargash in central Iran, the $30 million Iranian National Observatory (INO) will study exoplanets and gamma ray bursts, hunt for dark matter, and probe galaxy formation. Although INO's 3.4-meter mirror is relatively small by today's standards, Gargash's tranquil air rivals that of Hawaii's Mauna Kea, home to some of the world's top telescopes, which could make INO the best general-purpose telescope in the region. Shifting political winds delayed construction for years, and international sanctions on most foreign transactions forced Iranian scientists to find creative ways to import the primary mirror from Germany. The mirror has not yet been installed, and it will take months to calibrate before observations can begin. For that reason, some Iranian astronomers criticized INO's inauguration as premature. It came shortly before Iranian President Hassan Rouhani, an INO supporter, finishes his term next month.



This image of a modern requiem shark scale was taken at 175x magnification.

IMAGE: ERIN DILLON AND JORGE CEBALLOS

CONSERVATION

Shark scales reveal teeming ancient population

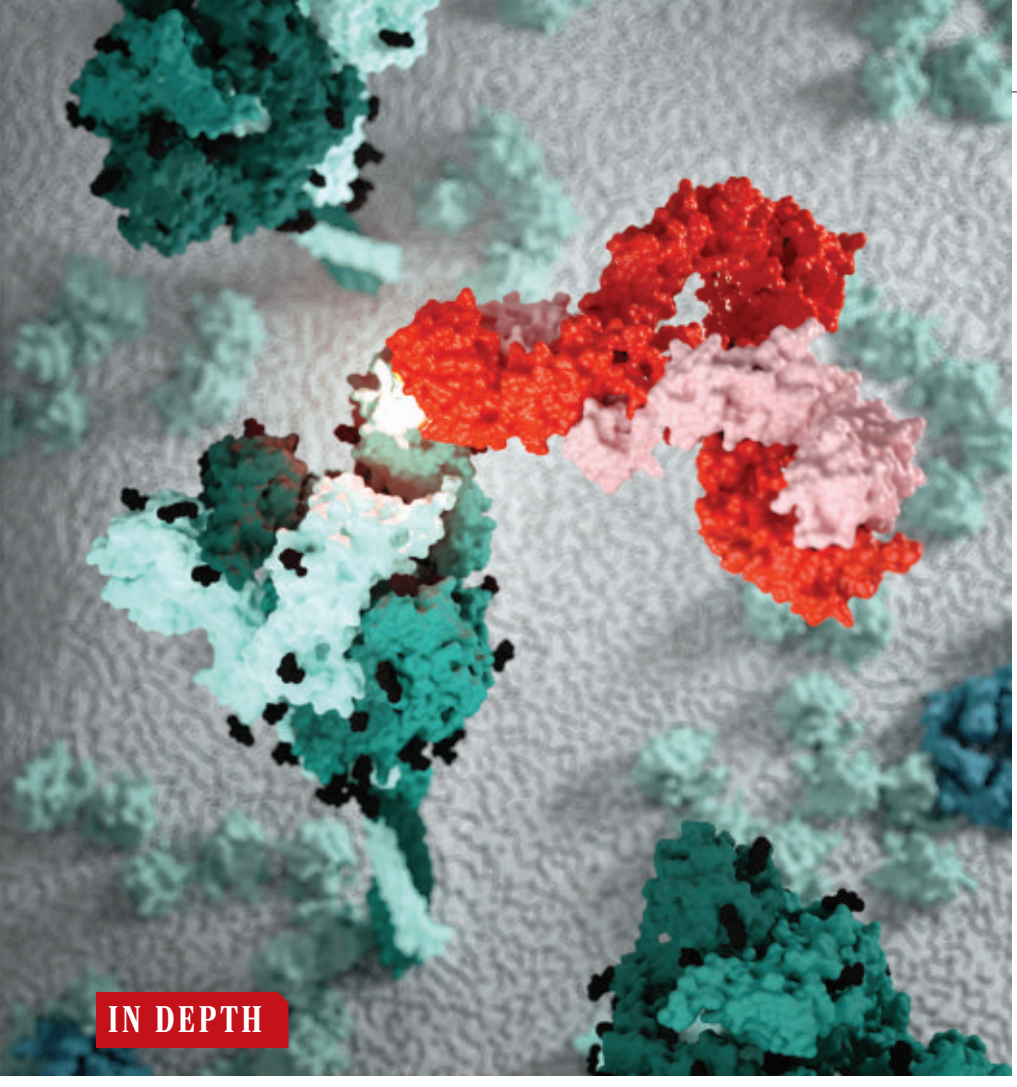
Shark skin may look as smooth as a wetsuit, but it consists of tiny scales that are specialized for speed, toughness, or defense. Now researchers have shown that the scales, called denticles, shed by sharks can reveal changes in population size. Erin Dillon, a Ph.D. candidate at the University of California, Santa Barbara, collected denticles from a 7000-year-old reef that is now on dry ground in Panama, and from nearby living coral reefs. The denticles had accumulated in the ancient sediment three times faster than they do today, Dillion and colleagues report this week in the *Proceedings of the National Academy of Sciences,* suggesting that the shark population was once three times larger. Shark populations in the western Caribbean began to decline precipitously in the 1980s after fishers targeted them.

Europe research body gets chief

LEADERSHIP | The European Commission last week appointed German biologist Maria Leptin as the new president of the European Research Council (ERC), which as Europe's largest basic science funder hands out roughly €2 billion in grants per year. Her 4-year term begins on 1 October. Leptin comes to ERC after 10 years leading the European Molecular Biology Organization, an intergovernmental research institute based in Heidelberg, Germany, that is funded by 30 countries. She has chaired an evaluation panel for ERC grants since 2008. Leptin is seen as a safe pair of hands after the contentious resignation last year of her predecessor, Italian American nanoscientist Mauro Ferrari. Ferrari fell out with ERC's Scientific Council after he proposed soliciting grants for studies of COVID-19, which would have departed from the body's tradition of relying on bottom-up ideas.

Big gravity observatory advances

ASTROPHYSICS | European physicists' plans to build the Einstein Telescope, a radical new gravitational-wave observatory, received a boost last week when an advisory panel added the facility to its road map of future projects. Over the next 3 to 4 years, developers will flesh out the design for the €1.9 billion observatory, which could be built by 2035. It would be nestled in a triangular tunnel, 10 kilometers on a side, containing three pairs of V-shaped detectors called interferometers. With 10 times the sensitivity of existing detectors, it could detect gravitational waves from black-hole mergers throughout the universe. Physicists hope the recommendation by the panel, the European Strategy Forum on Research Infrastructures, will help them expand the collaboration from the current five nations—Belgium, Italy, the Netherlands, Poland, and Spain—to all of Europe.



IN DEPTH

An antibody (red/pink) latches on to the surface protein of SARS-CoV-2 (green).

COVID-19

Can immune responses predict which vaccines work best?

Elusive "correlates of protection" could lead to approvals of boosters or new vaccines without big clinical trials

By **Jon Cohen**

Other than running a placebo-controlled clinical trial lasting many months and involving tens of thousands of people, is there any way to be sure a COVID-19 vaccine will work? Many researchers say studies of existing vaccines point to a shortcut: Simply gauge a new vaccine's ability to elicit so-called neutralizing antibodies, which bind to the virus and prevent it from entering cells. But several recent studies point to other "correlates of protection": "binding" antibodies—which latch on to the virus but don't block entry—and another set of immune warriors called T cells.

Vaccine decisions may soon depend on a better understanding of these supporting actors. Several manufacturers are developing updates of their COVID-19 vaccines that are tailored to protect against new viral variants and could be used as booster shots. The companies hope regulatory agencies won't require evidence of efficacy in big clinical trials, which are time-consuming, expensive, and increasingly ethically fraught because some of the participants receive a placebo even though proven vaccines are now available.

Instead, developers would like to give an updated vaccine to a much smaller group of participants and then check whether they produce the telltale immune responses. (That's how the annual updates of flu vaccines are approved.) Health officials may also turn to correlates when they compare the powers of existing COVID-19 vaccines, authorize new "mix and match" combinations, or even when making decisions about entirely new vaccines.

But finding robust correlates has been challenging. During the megatrials that led to the authorization of COVID-19 vaccines, investigators monitored antibody responses and tried to compare them with the odds of participants getting sick. Different trials, however, used different antibody assays and different definitions of mild COVID-19, the main endpoint in the trials. Many trials lacked the statistical power to measure protection from hospitalization and death, arguably a COVID-19 vaccine's most important task. "It's anarchy because it's always been anarchy," says John Moore, an immunologist at Weill Cornell Medicine. "You're dealing with different academic labs and different companies, and companies tend not to talk to each other." Few trials even looked carefully at T cells, which are cumbersome to measure.

Still, two studies—first published as preprints in March—confirmed the prediction by Moore and many other scientists that neutralizing antibodies ("neuts") play a key role. To "normalize" the different assays used in the trials, they compared levels of antibody elicited by each vaccine with those found in people who naturally became infected in the trial's placebo group. In both analyses, the vaccines that triggered higher levels of neuts than the ones typically seen in recovered people offered the best protection—strong evidence of a correlation, Moore says.

"That's a great relief to me," says Penny Moore (no relation to John Moore), a virologist at the National Health Laboratory Service in South Africa, who helped measure immune responses in different vaccine trials and was "really puzzled" by the results. But she and others say neuts are not the whole story. "I just cannot work out for the life of me how much [other immune responses] are contributing and where they're contributing," she says.

During the efficacy trials of the messenger RNA (mRNA) vaccines made by the Pfizer-BioNTech collaboration and Moderna, for example, the first shot triggered barely measurable levels of neutralizing antibodies, but still offered substantial protection. "It suggests there's more than neutralizing antibodies going on here," says David Montefiori, an immunologist at Duke University who runs a lab that measures neuts for a handful of COVID-19 vaccine trials. Neuts skyrocketed only after the second mRNA shot,

CREDITS: (ILLUSTRATION) V. ALTOUNIAN/*SCIENCE*; (IMAGES) W. SURYA, *BIOCHIM. BIOPHYS. ACTA* (2018); D. WRAPP, *SCIENCE*. (2020); E.O. SAPHIRE, *SCIENCE*. (2001); ORIENTATIONS OF PROTEINS IN MEMBRANES DATABASE

when protection rose to more than 90%.

T cells, which coordinate the B cells that produce antibodies but also clear infected cells when neuts falter, appear to bolster the defense. In a study published in February that included 12 patients whose COVID-19 ranged from mild to fatal, a team led by immunologist Antonio Bertoletti of the Duke–National University of Singapore Medical School reported that patients who early on had the highest levels of immune system messengers that kick T cells into action—an indirect, but relatively simple, way to measure their presence—had milder disease because they cleared the infection faster.

Penny Moore and colleagues also found support for a role for T cells. In an 11 June preprint, they reported that 96% of participants in an efficacy trial of the COVID-19 vaccine produced by Johnson & Johnson (J&J) made antibodies that neutralized a viral strain from early in the pandemic but only 19% had antibodies that neutralized the Beta variant, which is widespread in South Africa and infamous for dodging neuts. Yet the vaccine remained protective against moderate and severe COVID-19. "I think it's entirely plausible … that T cells are doing something really useful here," Moore says. A monkey study with the J&J vaccine published in *Nature* last year also showed that T cells contributed to control of the virus if neut levels weren't high enough to do the job.

Science's COVID-19 reporting is supported by the Heising-Simons Foundation.

Binding antibodies may also be more important than researchers assumed. A 24 June preprint by researchers from the University of Oxford reported that high levels of neuts correlated with the 80% protection seen 28 days after U.K. participants received two shots of the vaccine the team developed with AstraZeneca. But digging more deeply into the data revealed that binding antibodies were as good as a correlate—if not better.

It's not clear exactly why, because binding antibodies don't directly block the infection process. One possibility is that they make the virus more susceptible to being gobbled up by macrophages or other cells that ingest intruders. This mechanism, called phagocytosis, protected children from severe COVID-19, immunologist Galit Alter of the Ragon Institute of MGH, MIT, and Harvard reported in *Nature Medicine* in March. Then again, it may be that binding antibodies are produced in lockstep with neuts, but at higher levels, and are simply a surrogate marker.

A study of the immune reactions of 24 COVID-19 patients whose disease ranged from mild to fatal, by virologist Shane Crotty and Alessandro Sette of the La Jolla Institute for Immunology, showed that people handle SARS-CoV-2 most effectively if they have T cells and antibodies working in sync. "The immune system figures out how to use all the weapons at its disposal," Crotty says.

South Africa, where fewer than 1% of the population is fully vaccinated amid an exploding epidemic, has shown the potential pitfalls of overemphasizing neuts. In February, the country abandoned the AstraZeneca-Oxford vaccine after it had a disappointing 22% efficacy against mild disease in a large trial. Test tube analyses seemed to support the decision: Antibodies triggered by the vaccine had far less neutralizing power against the Beta variant, which then accounted for nearly all infections. But Penny Moore's study of the J&J vaccine has subsequently shown that disappointing levels of neutralizing antibodies don't keep a vaccine from providing good protection against severe disease. "Our obsession with neuts may mean that we missed an opportunity here for AstraZeneca," she says.

Other scientists counter that it makes sense to use neuts as a gauge to rank the relative powers of different vaccines, but acknowledge that this will require standardized assays. "This has not been the most important priority, but it's going to become one if we move away from phase 3 trials," John Moore says.

With the picture still muddy, regulators need to decide whether correlates of protection should offer vaccinemakers a shortcut to bringing improved products to market. Pfizer and Moderna are developing candidates designed to create high levels of neutralizing antibodies against the Beta variant, and the U.S. Food and Drug Administration (FDA) has signaled it will accept this correlate of protection for approval decisions. But Alter worries relying on neuts might lead regulators to approve unnecessary booster shots simply because they outdo existing shots on that measure. "If [regulators] don't adapt, we're going to end up overboosting, and we're going to be making the drug companies really happy," she says.

It's also unclear whether a convincing correlate from a vaccine that uses, say, mRNA, applies to one that uses a different technology. "We're hoping to have more immune correlate of protection information before updates on that," says Peter Marks, who heads FDA's vaccine division.

With more than a dozen vaccines now in use, that information may arrive soon, Sette says. Companies typically control the data from clinical trials, but academic labs can now compare recipients of different vaccines, he says. "In the next few months … a large amount of data will be generated in academic labs," Sette says. "There's going to be a fundamental wealth of information." ■

ANIMAL BEHAVIOR

Smell proves powerful sense for birds

New studies highlight underappreciated role of avian olfaction

By **Elizabeth Pennisi**

Almost 200 years ago, the renowned U.S. naturalist John James Audubon hid a decaying pig carcass under a pile of brush to test vultures' sense of smell. When the birds overlooked the pig—while one flocked to a nearly odorless stuffed deer skin—he took it as proof that they rely on vision, not smell, to find their food. His experiment cemented a commonly held idea. Despite later evidence that vultures and a few specialized avian hunters use odors after all, the dogma that most birds aren't attuned to smell endured.

Now, that dogma is being eroded by findings on birds' behavior and molecular hardware, two of which were published just last month. One showed storks home in on the smell of freshly mowed grass; another documented scores of functional olfactory receptors in multiple bird species. Researchers are realizing, says evolutionary biologist Scott Edwards of Harvard University, that "olfaction has a lot of impact on different aspects of bird biology."

Forty years ago, when ethologist Floriano Papi proposed that homing pigeons find their way back to a roost by sniffing out its chemical signature, his colleagues scoffed at the idea. They pointed out that birds have several other keen senses to guide them, including sight and, in the case of pigeons and some other species, a magnetic sense. "By then, biological textbooks already stated unequivocally that birds have little to no sense of smell, and many people still believe it—even scientists," says Danielle Whittaker, a chemical ecologist at Michigan State University.

Still, contrary evidence was already accumulating. In the 1960s, ornithologist Kenneth Stager found vultures were attracted to boxes with a carcass hidden inside and fans that vented the odors—as long as this bait wasn't too decomposed, as was likely the case in Audubon's experiment. Researchers also found that albatrosses, shearwaters, and some other seabirds find

The smell of cut grass alerts European white storks to exposed prey.

their fish prey by detecting a chemical released by the plankton the fish eat. But these birds, forced to navigate many kilometers across a featureless sea, seemed exceptional. In 2008, "You were part of the dark side if you talked about birds using olfaction," recalls Martin Wikelski, an ecologist at the Max Planck Institute for Ornithology.

That year, though, a graduate student at his institute, molecular ecologist Silke Steiger, analyzed nine bird genomes from across the avian family tree and uncovered many genes for olfactory receptors—proteins in the nasal passages that bind to odors and relay a signal to the brain.

In species that don't rely much on smell (humans are an example), these genes often mutate and become nonfunctional. But the researchers confirmed that many of the birds' olfactory genes were intact. What's more, they found that the number of these genes correlated with the size of a bird species' olfactory bulb, the brain's smell center—further evidence that the receptors were functional.

The genomes in that study were incomplete, however. Last month, Christopher Balakrishnan, an evolutionary biologist at East Carolina University, and graduate student Robert Driver examined some of the best available bird genomes and for some species found many more olfactory genes. Their analysis of genomes from a hummingbird, emu, chicken, zebra finch, and a tropical fruit eater called a manakin revealed scores of new olfactory receptors, they reported on 28 June in the journal *Integrative and Comparative Biology*.

That the emu has so many of these genes excites Whittaker, because this bird sits near the base of the bird family tree. "This result suggests that the ancestor to all birds must have had a very diverse set of olfactory receptor genes as well," she says. Smell must have been important to birds from the beginning, and comparisons of their olfactory receptor genes today confirm it remains so. Balakrishnan and Driver found that one diverse set of receptors unique to birds has split into multiple types specific to different bird lineages. That suggests these genes evolved rapidly as the birds diversified. Natural selection may have honed the genes to perform crucial tasks.

Wikelski and colleagues saw bird smell in action after they were inspired by a question from a curious primary school student. During an outreach program at a school in Radolfzell, Germany, the student asked the scientists how the local population of European white storks found their way to freshly cut meadows, where their insect and rodent prey were most exposed.

To find out, Wikelski piloted his plane in circles to observe a flock of 70 storks on sunny spring and summer days. Even when the storks couldn't see or hear the mowing, he and his colleagues noted, they homed in on mowed fields upwind of them, as if drawn to the smell of the cut grass. To confirm the suspicion, the team sprayed cut-grass smell—a mix of three volatile chemicals—onto fields that hadn't been mowed recently. The storks came flocking, the team reported on 18 June in *Scientific Reports*. The work "shows very clearly that these birds rely exclusively on their sense of smell to make foraging decisions," Whittaker says.

Other bird species may also respond to "calls" from injured plants, recent evidence shows. Two European birds, the great tit and the blue tit, locate insects that are attacking pine trees by detecting the volatile chemicals the stressed trees release, ecologist Elina Mäntylä of the Biology Centre of the Czech Academy of Sciences and colleagues reported in the September 2020 issue of *Ecology and Evolution*.

All these results show bird olfaction "should not be ignored," Mäntylä says. Driver adds that they might also point to a new form of natural pest control, in which farmers or foresters could treat threatened flora with chemicals that entice birds to come and gobble up invasive insects.

Other studies suggest olfaction might guide social interactions between birds. Whittaker's team has focused on preen oil, which birds secrete from a gland at the base of the tail and rub onto their feathers. The oil's chemical composition reveals the bird's species, sex, aggressiveness, and reproductive state. Females produce much more of these odorous chemicals, Whittaker and her colleagues reported in January in the *Journal of Chemical Ecology*, suggesting they depend more on odors to communicate, lacking the flashy feathers and songs that males rely on. Use of these cues is "likely widespread," says Steiger, now at the German chemical company BASF SE, "but simply not yet investigated well enough."

That's changing fast, as studies of bird olfaction expand into new species. Published papers on the topic have doubled every decade since 1992, reaching 80 this past year. The field is, belatedly, putting Audubon's misconception to rest and acknowledging that birds—champions of flight, vision, and song—have another power as well. ■

PHOTO: ARTERRA/UNIVERSAL IMAGES GROUP/GETTY IMAGES

COVID-19

Sex and gender missing in COVID-19 data

Despite suggestions of differential effects, most clinical trials don't report results by sex

By **Cathleen O'Grady**

COVID-19 doesn't strike the genders equally. Globally, for every 10 COVID-19 intensive care unit admissions among women, there are 18 for men; for every 10 women who die of COVID-19, 15 men die. In the United States, a gender gap is emerging in vaccination rates, with women ahead of men by 6 percentage points, according to the Centers for Disease Control and Prevention. And rare adverse effects from the AstraZeneca vaccine appear to strike women more frequently, whereas those from the Pfizer-BioNTech and Moderna vaccines more often affect young men.

But out of 45 COVID-19 randomized controlled trials whose results were published by December 2020, only eight reported the impact of sex or gender, according to a paper published this week in *Nature Communications*. Other recent data show even simple counts of cases and vaccinations are not broken down by sex and gender.

Senior author Sabine Oertelt-Prigione, a gender and health researcher at Radboud University Medical Center, was disheartened by her group's findings. "I would have assumed that [sex] would be picked up in the trials, simply because it's such an evident piece of the puzzle," she says. Skipping that step is potentially dangerous in trials of drugs that may affect men and women differently, given their physiological differences, Oertelt-Prigione says. And it misses an opportunity to learn about the workings of the disease, adds Susan Phillips, an epidemiologist at Queen's University who was not involved in the study.

Martin Landray of the University of Oxford finds the lack of attention to sex effects surprising, too. He led the United Kingdom's Recovery trial, which found the anti-inflammatory drug tocilizumab reduces risk of death from COVID-19, and which did explore results by sex (though it found none worth highlighting). "I just thought that's what everybody did." But Phillips notes that researchers have often skipped gender analyses in published clinical research for more than 30 years. "The problem remains," she says. "And this makes the current paper important."

Oertelt-Prigione's team searched PubMed for all papers on COVID-19 published before December 15 2020, excluding commentaries, observational trials, and other studies to identify 45 randomized controlled trials that tested potential treatments and vaccines. All trials in the study reported numbers of male and female participants. But only eight examined whether results differed by gender.

Even some of the largest COVID-19 trials didn't analyze effects on men and women separately. For example, the giant Pfizer-BioNTech and Moderna vaccine trials explored whether vaccine efficacy differed by sex, finding more than 90% efficacy for both men and women. But neither trial broke out adverse effects by sex, as United Nations University gender and health researcher Lavanya Vijayasingham and colleagues noted in a letter in *The Lancet* in March. Stephen Thomas of the State University of New York Upstate Medical University, a lead investigator of the Pfizer trial, says that those data are still collected and monitored, even if not published in a scientific journal. But low numbers of serious adverse events may make it difficult to detect significant sex differences in side effects, he says.

Men and women line up for COVID-19 vaccination in Ahmedabad, India. Many nations don't report vaccinations by sex.

PHOTO: AJIT SOLANKI/AP IMAGES

The new paper's findings are consistent with other studies. A recent, smaller study of COVID-19 trials, published in *EClinicalMedicine*, found zero out of 30 trials explored whether results were affected by sex. And an April paper in *BMJ Global Health* that examined a broader range of COVID-19 papers found only 14 out of 121 analyzed whether sex affected the results.

Sometimes there may be reasons not to report sex-disaggregated data. Landray's team found one statistically significant sex difference in their study of tocilizumab: In patients who weren't on mechanical ventilation at the start of the trial, the drug overall reduced the risk of dying or needing ventilation—but analyzing by sex suggests the difference was only in men. For other outcomes, such as hospital discharge within 1 month, there was no statistically significant difference between the sexes. The team concluded it didn't have "convincing evidence of there being a sex effect"—and so didn't highlight it, Landray says.

He notes that suggesting a sex difference where one might not exist can be harmful. For example, trials with small numbers of women suggested aspirin doesn't prevent heart attacks and strokes in women. But restricting aspirin's use based on such weak evidence would deprive women of a potentially beneficial drug, Landray argues.

At the moment, it's up to individual investigators to bring sex and gender into their analyses, says Emily Smith, an epidemiologist at George Washington University. But maybe some system-level interventions could help address it," she says. If funding agencies or trial registries required reporting by sex, that could motivate researchers to bake it into their trials.

The lack of data extends beyond clinical trials: Of 198 countries in the most recent monthly report from the Sex, Gender and COVID-19 Project database run by the nonprofit Global Health 50/50, only 37% report sex-disaggregated death data, and only 18% report sex-disaggregated vaccination data. Only Austria and two states in India have reported data for nonbinary people, according to the report, although some U.S. states also record nonbinary or transgender identities.

The COVID-19 pandemic has "shone a light on the importance of sex and gender in a way that very few other conditions have managed to do," says Sarah Hawkes, co-director of Global Health 50/50. She and others say it's time researchers shed their own light on those differences. ■

WILDLIFE BIOLOGY

Something is killing U.S. birds. It's not cicadas

Birds with crusty eyes and neurological damage found in nine states and Washington, D.C.

By **David Malakoff** *and* **Erik Stokstad**

Jennifer Toussaint, chief of animal control in Arlington, Virginia, can't forget the four baby blue jays. In late May, worried residents had delivered the fledglings to her clinic just outside of Washington, D.C., within just a few hours. Each was plump, indicating "their parents had done a great job caring for them," Toussaint says. But the birds were lethargic, unable to keep their balance, and blinded by crusty, oozing patches that had grown over their eyes.

Toussaint and her staff soon reached a gloomy diagnosis: the jays were the latest victims of a mysterious deadly disease that had emerged in their area just a few weeks earlier and had already killed countless wild birds. There was no known treatment, so they euthanized the jays. "It was difficult to feel so helpless," Toussaint recalls.

Others are sharing Toussaint's frustration. Since May, when the illness was first recognized in and around Washington, D.C., researchers have documented hundreds of cases in at least a dozen species of birds in nine eastern and midwestern states. State, federal, and academic scientists are hunting for clues to a cause in bird carcasses and the environment. Last week, they reported some modest progress: Studies have ruled out a number of agents known to cause mass mortality in birds, including *Salmonella* bacteria, several families of viruses, and *Trichomonas* parasites.

"Learning what isn't the cause can be just as helpful as learning what it is," Toussaint says. But it also means "We're still scratching our heads on this one," says wildlife epidemiologist David Stallknecht, director of the Southeastern Cooperative Wildlife Disease Study at the University of Georgia, Athens, which is involved in the effort.

Despite the uncertainty, researchers are beginning to get a clearer picture of the outbreak, thanks in part to thousands of people who have responded to calls from government agencies and scientists to report sick or dead birds. Not all species, for example, appear to be at high risk. "It's been quite species specific," says veterinarian Megan Kirchgessner of the Virginia Department of Wildlife Resources. So far, most cases involve just four species—common grackles, blue jays, American robins, and European starlings—according to a 2 July statement from the U.S. Geological Survey's National Wildlife Health Center. Young birds appear to be especially susceptible.

Those demographics could change as more data come in, especially from rural areas that so far have produced few observations, says Allisyn-Marie Gillet, Indiana's state ornithologist. At this point, the outbreak doesn't appear to pose a serious threat to bird populations, researchers say. Still, they are watching to see whether its geographic scope expands; reports of sick birds now stretch west to Indiana and Kentucky and north to Pennsylvania.

A mysterious disease might have killed this American robin found in Kentucky.

PHOTO: JON CHERRY/STRINGER/GETTY IMAGES

The geography suggested one suspect. In May and June, portions of the outbreak area saw the emergence of billions of periodical cicadas, members of the 17-year Brood X. Birds feast on cicadas, prompting some researchers to wonder whether the outbreak might be linked to the insects. Cicadas spend most of their lives underground, where they may have accumulated pesticides or other contaminants. A type of fungus called *Massospora* that infects cicada broods might also play a role; one *Massospora* species produces compounds that alter the behavior of cicadas, perhaps to increase its spread. But the cicadas appear to be blameless. Birds tend to avoid eating fungus-ridden cicadas, and sick birds have been observed in areas where cicadas were rare. "It does not look like it's a match," says Brian Evans, a migratory bird ecologist with the Smithsonian's National Zoo and Conservation Biology Institute.

Researchers note that mass bird mortalities are not uncommon, especially among species that form dense flocks or gather at feeders. In the mid-1990s, bird watchers in the eastern United States noticed that house finches, a common introduced songbird, were dying in relatively large numbers from an illness characterized by swollen and encrusted eyes. Researchers ultimately determined the cause was a bacterium, *Mycoplasma gallisepticum*, that had likely spread from domestic poultry. Over the past few decades, researchers have also tracked outbreaks of West Nile virus, avian influenza, and *Salmonella* that caused noticeable kills.

Many of those suspects have been ruled out in this case, according to the 2 July statement. But researchers continue to look at other possibilities. They are using electron microscopy to examine tissues for telltale damage, for example, and employing a battery of tests to detect suspect microbes, viruses, parasites, and chemical pollutants.

Many of the dead birds that have been tested were infected with *Mycoplasma* bacteria. That is not uncommon, Evans says, but the bacterium has evolved to become more infectious and deadly, and it might be playing a role in the current outbreak. "In terms of mycoplasma this might be something new," he says. Others, however, are skeptical, noting *Mycoplasma* rarely affects fledglings.

In the meantime, officials in several states told *Science* there are signs that the outbreak might be easing. In Virginia, for example, "the number of birds being brought to rehab centers is starting to decline," Kirchgessner says, and Toussaint's clinic recently had multiple days with no admissions of symptomatic birds. Until the outbreak ends, however, officials are asking bird lovers to take steps that could slow the spread of any disease, including burying dead birds and taking down feeders where birds congregate. ■

A nursing home worker in the Netherlands gets a COVID-19 vaccine. A Dutch database lies at the heart of a controversy around a paper on vaccine safety.

PHOTO: PIROSCHKA VAN DE WOUW/POOL/REUTERS

COVID-19

Journal retracts paper claiming COVID-19 vaccines kill

Editors at *Vaccines* quit, protesting "irresponsible" study

By **Meredith Wadman**

The journal *Vaccines* on 2 July retracted a peer-reviewed article after the angry resignations of at least six editors. They were protesting the publication of a study 1 week earlier that had misused data in a Dutch vaccine adverse events registry to make a startling claim: "For three deaths prevented by [COVID-19] vaccination, we have to accept two inflicted by vaccination."

The retraction, signed by the *Vaccines* Editorial Office, declared: "Serious concerns were brought to the attention of the publisher regarding misinterpretation of data. … The article contained several errors that fundamentally affect the interpretation of the findings."

The editors who resigned also feared the paper would feed antivaccine conspiracy theories. Days after it was published, Katie Ewer, an immunologist at the University of Oxford, wrote in an email to *Science* that the paper "is now being used by antivaxxers and COVID-19-deniers as evidence that COVID-19 vaccines are not safe. [This] is grossly irresponsible, particularly for a journal specialising in vaccines."

The paper had drawn 425,000 readers as of 6 July and has been tweeted by antivaccination activists with hundreds of thousands of followers.

The disaffected editors say they haven't been told what went wrong in the peer-review and editorial processes. To rejoin the board, "I would need a much better description of how this article ever made it through peer and editorial review, given the spectacularly huge flaws listed in the retraction document," says Andrew Pekosz, a respiratory virologist at the Johns Hopkins University Bloomberg School of Public Health who resigned as a section editor.

But another editor who resigned, Diane Harper, an epidemiologist at the University of Michigan, Ann Arbor, and the journal's founding editor-in-chief, rejoined the board after the retraction was published. "The journal management and leadership has acted quickly to retract the article and to change editorial internal processes about review," she wrote in a 2 July email. A third, Florian Krammer, a virologist at the Icahn School of Medicine at Mount Sinai, applauded the retraction but says he will not rejoin.

Damaris Critchlow, head of publication ethics for the journal publisher, MDPI, wrote in an email that the journal's academic editor, Ralph DiClemente, a health psychologist at New York University, made the decision to publish the article. Now, Critchlow wrote, "We are … consulting the Editor-in-Chief and Editorial Board to establish further ways to support our Academic Editors."

The paper's three authors are Harald Walach, a clinical psychologist and science historian who does complementary medicine research at Poznan University of Medical Sciences in Poland; Rainer Klement, a physicist who studies tailored diets in cancer treatment at the Leopoldina Hospital in Schweinfurt, Germany; and Wouter Aukema, an independent data scientist in Hoenderloo, Netherlands. In a 29 June statement, the authors said they stand by their findings.

The authors computed COVID-19 deaths prevented by vaccines by using data from a study of 1.2 million Israelis. They estimated that 16,000 people needed to be vaccinated to prevent one COVID-19 death. To compute deaths "caused" by vaccine side effects, they used EU data on vaccines delivered in the Netherlands and data from the Netherlands Pharmacovigilance Center. That registry, also called Lareb, is a passive surveillance system in which anyone can file a report of an adverse event after vaccination, whatever the cause. Such databases are not used to assess vaccine risks, but to search for early signs of rare vaccine side effects for follow-up studies.

The website of the Dutch registry clearly notes its reports do not imply causality. But the authors used it that way. The day after the paper's publication, Lareb's head of science and research, Eugène van Puijenbroek, sent an email to *Vaccines*'s editors criticizing the paper's use of Lareb's data, and requesting a correction or retraction. He called the assumption that vaccination caused all the reported deaths "far from truth."

The three peer reviewers on the paper offered no substantial criticism of the authors' methodology in brief reviews. One, Anne Ulrich, a chemist who is chair of biochemistry at the Karlsruhe Institute of Technology in Germany, wrote that the authors' analysis "is performed responsibly … and without methodological flaws." Another reviewer, this one anonymous, wrote that the manuscript "is very important and should be published urgently," offering almost no other comment.

"It's very evident from their reviews that [the reviewers] don't have any topic expertise," says another editor who resigned, Helen Petousis-Harris, director of the Vaccine Datalink and Research Group at the University of Auckland. "The authors don't either," she adds. "It's a bit remiss." But once the retraction was published, she said she was "happy to stay on" as an editor. ■

FEATURES

THE GHOSTS IN THE MUSEUM

Anthropologists are reckoning with collections of human remains—and the racism that built them

By **Lizzie Wade**

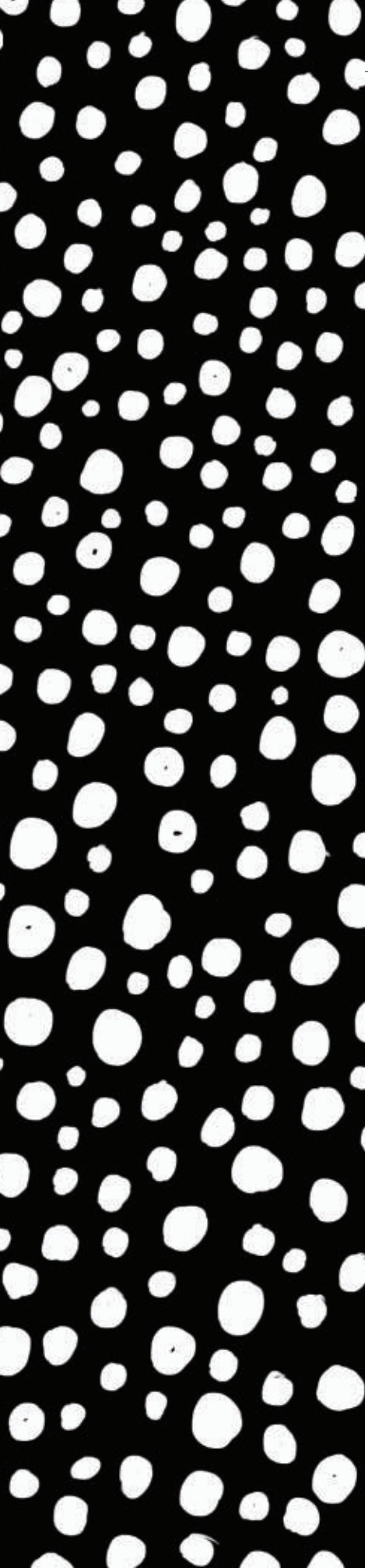


ILLUSTRATION: JOHNALYNN HOLLAND

A 19th century collection of 1300 skulls—symbolized here by white dots—includes some from enslaved people.

They were buried on a plantation just outside Havana. Likely few, if any, thought of the place as home. Most apparently grew up in West Africa, surrounded by family and friends. The exact paths that led to each of them being ripped from those communities and sold into bondage across the sea cannot be retraced. We don't know their names and we don't know their stories because in their new world of enslavement those truths didn't matter to people with the power to write history. All we can tentatively say: They were 51 of nearly 5 million enslaved Africans brought to Caribbean ports and forced to labor in the islands' sugar and coffee fields for the profit of Europeans.

Nor do we know how or when the 51 died. Perhaps they succumbed to disease, or were killed through overwork or by a more explicit act of violence.

What we do know about the 51 begins only with a gruesome postscript: In 1840, a Cuban doctor named José Rodriguez Cisneros dug up their bodies, removed their heads, and shipped their skulls to Philadelphia.

He did so at the request of Samuel Morton, a doctor, anatomist, and the first physical anthropologist in the United States, who was building a collection of crania to study racial differences. And thus the skulls of the 51 were turned into objects to be measured and weighed, filled with lead shot, and measured again.

Morton, who was white, used the skulls of the 51—as he did all of those in his collection—to define the racial categories and hierarchies still etched into our world today. After his death in 1851, his collection continued to be studied, added to, and displayed.

In the 1980s, the skulls, now at the University of Pennsylvania Museum of Archaeology and Anthropology, began to be studied again, this time by anthropologists with ideas very different from Morton's. They knew that society, not biology, defines race. They treated the skulls as representatives of one diverse but united human family, beautiful and fascinating in their variation. They also used the history of the Morton collection to expose the evils of racism and slavery, sometimes using skulls in lectures and exhibits on those topics.

Then, in summer 2020, the history of racial injustice in the United States—built partly on the foundation of science like Morton's—boiled over into protests. The racial awakening extended to the Morton collection: Academics and community activists argued that the collection and its use perpetuate injustice because no one in the collection had wanted to be there, and because scientists, not descendants, control the skulls' fate.

"You don't have consent," says Abdul-Aliy Muhammad, a Black community organizer and writer from Philadelphia. "Black folks deserve to possess and hold the remains of our ancestors. We should be the stewards of those remains." Muhammad and others demanded

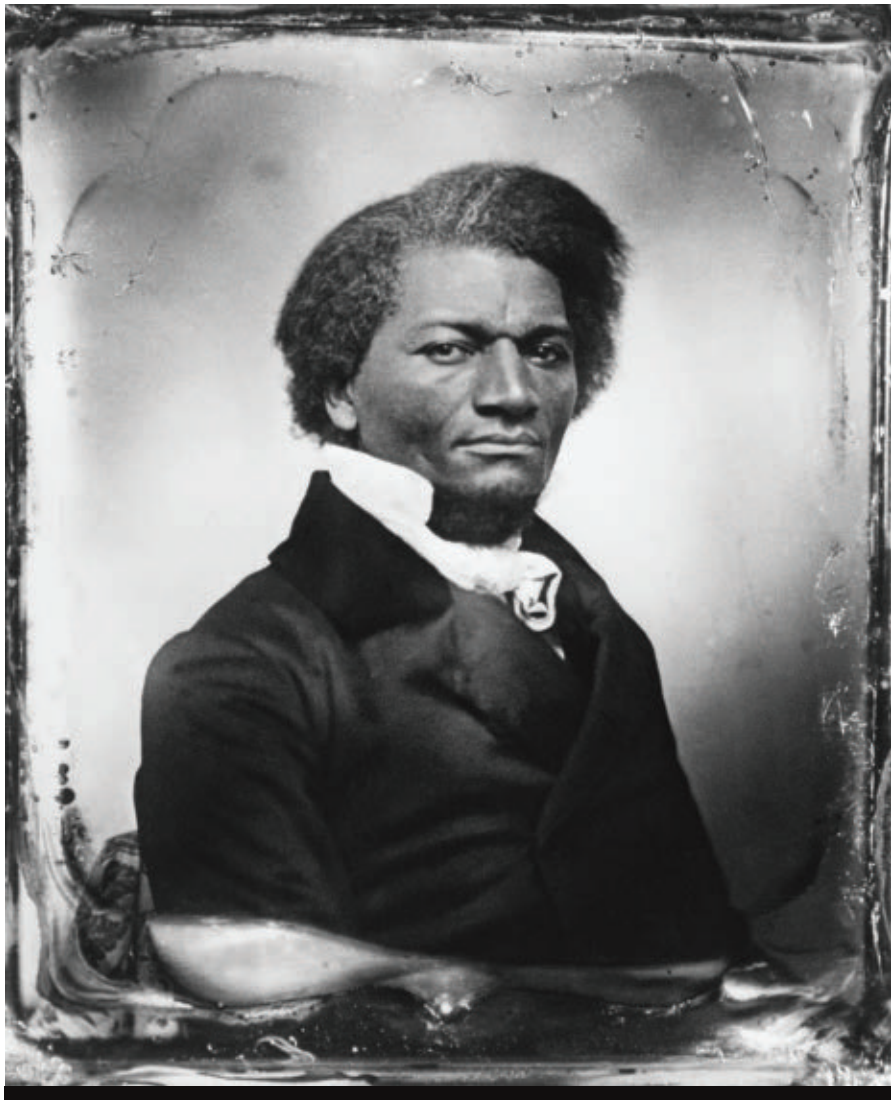
that the Morton collection, now numbering more than 1300 skulls, be abolished.

In July 2020, the Penn Museum put the entire collection in storage and officially halted research.

"One of the things we are having to grapple with now is the idea of possession," says Robin Nelson, a Black biological anthropologist at Santa Clara University. When you study biological material from another person, she says, "your research sample is not, in fact, yours."

That way of thinking could affect many collections in the United States. For example, the Smithsonian Institution's National Museum of Natural History (NMNH) holds the remains of more than 30,000 people, many Indigenous and some likely enslaved. Many remains were taken from their graves without permission, by scientists following in Morton's footsteps through the early 20th century. Other remains were from people who died in institutions, who had no say over the fate of their bodies.

The reckoning over Morton's skulls is also a reckoning for biological anthropology. "The Morton collection has been a barometer for the discipline from the moment of its conception," says Pamela Geller, a white bioarchaeologist at the University of Miami who is working on a book about the collection. Open racism drove its founding, and a new awakening to that legacy is now reshaping its future. "It's always been a gauge for where we are as anthropologists."



"It is strange that there should arise a phalanx of learned men—speaking in the name of *science*—to forbid the magnificent reunion of mankind in one brotherhood."
Frederick Douglass

WHEN THE SKULLS of the 51 were sent to Morton, he was already the world's leading skull collector. Active in the esteemed Academy of Natural Sciences of Philadelphia, Morton had an extensive network of scientifically minded contacts who responded enthusiastically to his requests to send skulls from every corner of the world. Rodriguez Cisneros wrote that he "procure[d] 50 pure rare African skulls" for Morton's collection. The doctor claimed the Africans had recently been brought to Cuba, but some skulls may have belonged to enslaved Africans born on the island, or to Indigenous Taíno people, who were also enslaved in Cuba at the time. (Whether Rodriguez Cisneros sent 53 skulls or 51 is also somewhat unclear.)

As documented in *The Skull Collectors: Race, Science, and America's Unburied Dead*, by Rutgers University historian Ann Fabian, other scientists who sent skulls to Morton included ornithologist John James Audubon, who nabbed five skulls lying unburied on a battlefield during Texas's war with Mexico; John Lloyd Stephens, whose bestselling accounts of expeditions in southern Mexico and Central America jump-started Maya archaeology; and José María Vargas, an anatomist who was briefly president of Venezuela. Military doctors plucked other skulls from the corpses of Native Americans killed in battles against U.S. forces sent to remove them from their own land.

Still other skulls came from the potter's fields of almshouses and public hospitals, where U.S. and European doctors had long sourced bodies for dissection. An 1845 petition to the Philadelphia almshouse board noted that patients, fearing their bodies would be dug up for science, often begged to be buried anywhere but the potter's field "as the last and greatest favor." The Morton collection contains more than 30 skulls from that potter's field—14 from Black people, according to a recent Penn report. "If you were a marginalized or disenfranchised human being, then there's a chance you would end up in Morton's collection," Geller says.

Morton sought a diverse collection of skulls because his life's work was to measure and compare the cranial features of what he considered the human races. Like many scientists of his time, Morton delineated five races: Caucasian, Mongolian, American, Malay, and Ethiopian. Their geographic ori-

PHOTO: BETTMAN/GETTY IMAGES

gins are jumbled to modern eyes, showing how social categories determine race. For example, "Caucasians" lived from Europe to India; the Indigenous people of northern Canada and Greenland were considered "Mongolian," like the people in East Asia; and the "Ethiopian" race included people from sub-Saharan Africa and Australia.

Morton thought skulls could reveal telltale differences among those races. When a skull arrived, he carefully inked a catalog number on its forehead and affixed a label identifying its race; many of the 51 still bear the words "Negro, born in Africa."

Morton meticulously measured each skull's every dimension. He filled them with white peppercorns and, later, lead shot to measure their volumes, a proxy for brain size. The race with the largest brains, he and many scientists thought, would also have the highest intelligence.

Morton found a wide range of cranial volumes within each of his racial categories. But he wrested a hierarchy out of averages: By his accounting, skulls of Caucasians had the largest average volume and skulls of Ethiopians, the smallest. Morton used his findings to argue that each race was a separate species of human.

Even in the 19th century, not everybody agreed. Charles Darwin, whose theory of evolution wasn't published until 8 years after Morton's death, found Morton's understanding of species facile and his arguments unreliable. Frederick Douglass, in a speech 3 years after Morton's death, called research that ranked the humanity of races "scientific moonshine." "It is strange that there should arise a phalanx of learned men—speaking in the name of *science*—to forbid the magnificent reunion of mankind in one brotherhood. A mortifying proof is here given, that the moral growth of a nation, or an age, does not always keep pace with the increase of knowledge," he said.

Despite those critiques, Morton's approach helped lay the foundation for the burgeoning field of physical anthropology. U.S. and European museums vied to build "massive bone collections," exploiting colonial violence to gather bodies from all over the world, says Samuel Redman, a white historian at the University of Massachusetts (UMass), Amherst, and author of *Bone Rooms: From Scientific Racism to Human Prehistory in Museums*. In the early 1900s, Aleš Hrdlička of NMNH, who helped found the American Association of Physical Anthropologists in 1928, continued to use human remains, often stolen from Indigenous communities, to study race and promote eugenics. Hrdlička, who was white and whom Redman describes as "deeply racist," was the driving force behind NMNH's skeletal collection. Last month, the association he founded changed its name to the American Association of Biological Anthropologists to separate itself from the discipline's overtly racist past.

"All of us who stand in this field have inherited this history," says Rick Smith, a white biocultural anthropologist at George Mason University. "It's on us to figure out what to do about it."

IN 1982, WHEN JANET MONGE, a white biological anthropologist at the Penn Museum, took charge of the Morton collection, she recognized its potential as a tool to explore anthropology's racist past. She also saw it as a valuable repository of the myriad physical differences among humans, in traits unrelated to the social constructs of race.

"All of us in this field have inherited this history. It's on us to figure out what to do."

Rick Smith, George Mason University

For example, in the late 1990s, a paper claimed that certain skull traits in the nasal cavity were unique to Neanderthals. But the researchers had only used modern human skulls from Europeans for comparison. A University of Pennsylvania student, Melissa Murphy, studied hundreds of skulls in the Morton collection and found some of the "Neanderthal" traits in non-Europeans. "Working with the Morton collection gave me a background in understanding human variation I never would have had otherwise," says Murphy, who is white and now a biological anthropologist at the University of Wyoming.

Between 2004 and 2011, Monge and colleagues expanded scientific access to the Morton collection by using computerized tomography (CT) to scan the skulls and thousands of others held in the Penn Museum. The scans, available online, "really democratized the research process," says Sheela Athreya, a biological anthropologist at Texas A&M University, College Station, who is Indian American and studied with Monge. Monge says more than 70 scientific papers have been published using the Morton scans, on such topics as how tooth alignment has changed over time and how skull growth during childhood affects adult cranial shape. The Penn Museum's website lists more than 100 researchers who used the Morton collection from 2008 to 2018.

Meanwhile, the remains of Native Americans in collections became an ethical and legal flashpoint. In 1990, Congress passed the Native American Graves Protection and Repatriation Act (NAGPRA), requiring federally funded institutions to inventory Native American remains in their collections and to work with tribes to return them to their descendants.

Monge, her students, and colleagues began to dig through historical documents, boosting their efforts to understand where the skulls in the Morton collection came from and contacting tribes about bringing some back home. More than 120 of the 450 or so Native American skulls from the collection have been repatriated.

In researching the skulls' origins, Monge says, "You come to appreciate the people of the collection." Other scholars explored the identities of remains not subject to NAGPRA, often under Monge's guidance. In 2007, one student completed a dissertation on the 51, combining historical analysis with a study of the skulls themselves. Some skulls had filed teeth, then a rite of passage in some West African communities, supporting the idea that the people had grown up in Africa.

The 51 and other skulls were eventually moved to glass-fronted cabinets lining an anthropology classroom at the Penn Museum. There they hovered, year after year, around students learning to study human bones. Monge also used skulls from the collection in classes, public talks, and museum exhibits on how anthropology had helped codify the idea of race and the resulting inhumanity. For example, at the African American Museum in Philadelphia, Monge showed vertebrae fused to the skull of one of the 51, a "major trauma" caused by a painful collar the person was forced to wear. "When you can see what slavery did to the body, it's overwhelmingly powerful," says Monge, who recalls audience members crying.

Such honest, public acknowledgment of the collection's violent past was rare among museums, Athreya says. But in 2020, a renewed reckoning with racism prompted yet another re-evaluation of the collection.

IN 2017, ON HIS SECOND DAY in an archaeology class held at the Penn Museum, Francisco Diaz looked to his right and found himself staring at a skull with the label "Maya from Yucatan" pasted to its forehead. Diaz, an anthropology doctoral student at Penn, is Yucatec Maya, born on Mexico's Yucatán Peninsula. In class, skulls from Black and Indigenous people were "just made part of classroom décor," he recalls. "You have this

institution that has done this type of work on Indigenous people, and then one of you shows up," he says. Seeing that skull in his classroom, "It's kind of like saying, do you really belong here?" This year, he wrote an essay on how study and display of the skulls dehumanized the people they belonged to.

The 51 themselves drew renewed attention in 2019, after a presentation by a group of Penn professors and students investigating the university's connections to slavery and scientific racism. "I was shocked by what I heard," says Muhammad, who attended the presentation. Muhammad wrote op-eds and started a petition to return the 51 and skulls from two other enslaved people to a Black community—either their descendants or a Black spiritual community in Philadelphia. "These people did not ask to be prodded, they did not ask to be dissected, they did not ask for numbers and letters to be imprinted upon their remains. They were brutalized and exploited. They had their lives stolen from them. And they deserve rest," Muhammad says.

After the murder of George Floyd in May 2020 sparked protests for racial justice around the country, more and more people within and outside Penn began to see the Morton collection as a present-day perpetuation of racism and its harms, rather than just a historic example. Until last summer, most researchers thought "the science is justified because we're doing it thoughtfully. And this moment brought to bear, no, that's not enough," says Rachel Watkins, a Black biological anthropologist at American University.

Even with recent research that strove to be respectful, it was almost always scientists who decided how and why to study the skulls, not their descendant communities, Athreya notes. "We were speaking for people without them at the table," she says. To move forward ethically, "Those of us in power are going to have to give up some."

Among anthropologists, Nelson says, "There's a mixture of guilt and fear. Guilt for the ways we have engaged with these kinds of materials and benefited from the data collected in ways that we now may find reprehensible. But there's also fear because we don't know what the field is going to look like [without those practices]."

Yet examples of inclusive, respectful biological anthropology exist. For example, back in 1991, when construction in New York City uncovered the earliest and largest known African burial ground in the United States, Black New Yorkers who identified themselves as a descendant community guided research, and the more than 400 excavated individuals were reburied in 2003. That project has served as a model for others, including for the remains of 36 enslaved people recently found in Charleston, South Carolina (see sidebar, p. 153). But for remains collected a century or two ago, like the Morton collection, applying the same principles can be challenging.

In July 2020, the Penn Museum moved the skulls in the classroom, including the 51, to join the rest of the collection in storage while a committee discussed what to do with it. Protests continued. "Black Ancestors Matter," proclaimed one sign at an 8 April protest.

Four days later, the Penn Museum apologized for "the unethical possession of remains" and announced an expanded repatriation plan for the Morton collection. The museum plans to hire an anthropologist of color to direct repatriation, actively identifying and contacting as many descendant communities as possible and welcoming repatriation requests from them, says Penn Museum Director Christopher Woods, who is Black. The museum has also suspended study of the CT scans while it develops a policy, to be enacted this fall, on the research and display of human remains.

> **"These people did not ask to be prodded, they did not ask to be dissected. ... You don't have consent."**
>
> **Abdul-Aliy Muhammad**, Philadelphia community organizer

Repatriation can be the first step toward building the relationships that make future community-led research possible, says Dorothy Lippert, an archaeologist and tribal liaison at NMNH and a citizen of the Choctaw Nation. "People think about repatriation as something that's going to empty out museum shelves, but in reality, it fills the museum back up with these relationships and connections," she says.

Monge, too, welcomes the new focus on repatriation. "I see a lot of great—honestly, better!—potential research with the collection," she says. "The science person in me says that science can help us a lot" with identifying descendant communities and answering questions they may have about their ancestors. For the 51, Monge thinks analyzing their DNA could answer long-standing questions about their ancestry and descendant communities, which may include both Black and Indigenous people. Once identified, those communities should have decision-making power over the 51, she says.

But some people don't want scientists unilaterally deciding to do more research on the 51. "Healing can't happen at the site of harm," Muhammad says, quoting Black artist Charlyn/Magdaline Griffith/Oro. Muhammad's trust in scientists further eroded beginning 21 April, when news emerged that anthropologists at Princeton University and Penn, including Monge, had kept a sensitive set of remains and used them in teaching: bones presumed to be the remains of Tree and Delisha Africa, who were killed in 1985 when the city of Philadelphia bombed the MOVE community, a Black activist group. (Monge declined to comment because Penn is investigating.)

Muhammad thinks repatriating the skulls of enslaved Black people in the Morton collection to a Black spiritual community in Philadelphia would be more meaningful than launching research to trace their genetic ancestry. "Black people have experienced generational displacement, so there are descendants of these people potentially everywhere and nowhere," Muhammad says. "Ultimately I want them to be in the hands of Black people who love Black people."

Each repatriation case will be unique, says Sabrina Sholts, a white curator of biological anthropology at NMNH. But she and others will be watching Penn's process. "There are many ways [repatriation of the Morton collection] could go that will be really important for all peer institutions and stakeholders to see," she says.

NMNH, like other museums, including the American Museum of Natural History in New York City, is only now beginning to assess how many remains of enslaved African Americans may be in its collection. "What's stunning to me is that we don't even know" how many are held, says Sonya Atalay, a UMass archaeologist who is Anishinaabe-Ojibwe. Ultimately, she and others hope the United States will pass a repatriation law that applies to African American ancestral remains. Many biological anthropologists say institutions should also establish review processes for work with ancestral remains, similar to how institutional review boards evaluate the ethics of research with living people.

On 10 June, the Penn Museum announced it had formed a community advisory group, including Muhammad and other members of Philadelphia community organizations and spiritual leaders, to review the case of the 14 Black people from the Philadelphia potter's field and consider how to respectfully rebury them. Woods says he hopes a decision about their future will be made by year's end. That process could inform future work to repatriate the 51. For now, they are still waiting. ■

Community members from Charleston, South Carolina, photograph a plaque bearing the newly bestowed names of 36 people of African descent, who were reburied with honor in 2019.

PHOTO: BRAD NETTLES/*THE POST AND COURIER*

Charleston honors Black ancestors, with both science and ceremony

Like the 51 enslaved people of African descent whose bodies were dug up in Cuba in 1840 for anthropologist Samuel Morton's collection, the 36 people buried in downtown Charleston, South Carolina, were nameless. No record of the graveyard or those buried in it existed, making it likely they were also enslaved Africans or their descendants.

But the fate of their remains, found during construction in 2013, was different from those in the Morton collection (see main story, p. 148). Instead of being acquired by scientific collectors, the 36—as African American retired teacher La'Sheia Oubré of Charleston calls them—became the responsibility of the city and its Black community, who turned to scientists to help discover their identities and life stories.

Called to investigate, archaeologists noted that the 36, also known as the Anson Street Ancestors after the location of their graves, had been buried with care, in regular rows. Nails and brass pins showed many had been wrapped in shrouds. Buttons, including one made of mother-of-pearl, showed they had been dressed by people who mourned them. Pieces of clay tobacco pipes were buried with two men, and a copper coin—a West African tradition—with another person. One man's incisors had been filed into points, a rite of passage in West Africa. A child had two copper half-pennies placed over their eyes.

The half-pennies, minted in 1773, and other offerings helped date the graveyard to between 1760 to 1790, when enslaved Africans made up nearly half of Charleston's population.

The nonprofit Gullah Society, which protects African and African American burial grounds around Charleston, held consultations about the 36 with the city's Black and African American communities. The community wanted to rebury them with love, honor, and respect. But first they wanted to learn everything they could about the 36, including their genetic ancestry. So Ade Ofunniyin, an African American anthropologist and founder of the Gullah Society, invited anthropological geneticists from the University of Pennsylvania to collaborate.

"Right from the get-go it was set up that we were going to try to levy our resources and expertise to answer the questions and [serve the] needs of the community," says one of the geneticists, Raquel Fleskes, who is white and now at the University of Connecticut, Storrs. When, working alone in a sterile lab, she ground up small pieces of bone from each of the 36 and extracted their DNA, she wore a GoPro camera on her head to share the process with the community. Every bit of sampled bone was saved to be reburied. Other researchers measured strontium isotopes in teeth and bones, which preserve chemical signatures of where a person grew up and lived.

Although 35 of the 36 had types of mitochondrial DNA—genetic material inherited through the mother—common in Central and West Africa, one woman's mtDNA linked her with Native American groups. The finding pointed to the intertwined histories of Black and Indigenous people in Charleston, as people from both communities were enslaved. Most of the 36 had lived in Charleston all their lives. The results were published in October 2020 in biological anthropology's flagship journal, the *American Journal of Physical Anthropology*.

From what was learned about the heritage and sex of the 36, the Gullah Society organized a ceremony, presided over by Yoruba priests, to give each a name. The child with the half-pennies placed on the eyes is Welela. Welela was buried next to Isi, an adult with an identical mitochondrial genome, so at least two of the 36 were buried with family.

On 4 May 2019, a horse-drawn hearse carried some of the remains through Charleston's streets for reburial near their original resting place. A crowd followed, filling the air with drumming and chants. "Every group of people that were identified within the 36 were part of the ceremony," remembers Oubré, who works with the Anson Street African Burial Ground project. "Native American, African, Caribbean, children, adults. We had dancing, we had music. … Never before had Charleston seen such grandeur."

The remains were laid in a burial vault with notes written by the community. "To my beloved ancestors, thank you for life and making your journey to Charleston, SC. You are honored and may God bless your souls," one read. Ofunniyin read out each of the 36 names, and the community echoed them back.

"It is our responsibility to take care of our elders," Oubré says. "Without them, there would be no us." —*L.W.*

INSIGHTS

PERSPECTIVES

EVOLUTIONARY BIOLOGY

A shift in taste

The evolution of sugar perception in songbirds began with a savory receptor

By **F. Keith Barker**

Sensory systems evolve to enable organisms to detect cues pertinent to survival. The diversification of these systems is a critical aspect of the adaptative radiation of animals—that is, how an ancestral species rapidly diversifies into a large number of morphologically diverse descendant species. Birds are the most diverse clade of terrestrial vertebrates, at more than 10,000 currently recognized species. Curiously, the entire avian clade was shaped by the early loss of a gene encoding a sweet receptor. How then, did thousands of bird species that rely on nectar and fruit evolve to perceive sugars? On page 226 of this issue, Toda *et al.* (*1*) report a shift from savory to sweet perception in the early evolution of songbirds. The change may have played a critical role in the radiation of this diverse group.

Painstaking anatomical and physiological studies have yielded important information about sensory systems whose functions depend on cellular- or organ-level adaption, such as echolocation in cetaceans and bats and specialized arthropod visual systems. By contrast, genomic studies of animals have rapidly yielded important insights into the evolution of sensory traits in which function is defined at the molecular level, including hearing (*2*), vision (*3*), olfaction (*4*), and gustation (*5*). For instance, studies on the evolution of opsins—light-sensitive proteins found in photoreceptor cells—demonstrate how gene duplication and loss can either expand or restrict the visual range of animals, affecting their ability to identify resources or to differentiate food resource quality (*6*). Such sensory changes can have profound effects beyond food acquisition, also driving coincident shifts in signaling modality and information content (*7*).

Birds have provided classic examples of adaptive radiation, such as the morphologically diverse Darwin's finches and Hawaiian honeycreepers. Broad comparative studies of avian diversification have also identified other rapidly speciating lineages, where the imprint of adaptive radiation was not so clear cut. One such lineage is the nectarivorous hummingbirds (*8*, *9*). Recent molecular work has shown that the insectivorous ancestor of this lineage retooled its umami receptor (sensitive to amino acids) for the detection of sugars (*5*). This molecular shift—which could

Department of Ecology, Evolution and Behavior and Bell Museum of Natural History, University of Minnesota, St. Paul, MN, USA . Email: barke042@umn.edu

PHOTO: GERALD ALLEN

The early evolution of sweet perception likely played an important role in the diversification of passerines, such as this New Holland honeyeater in Australia.

be described as a key innovation—allowed hummingbirds to exploit nectar, a resource that most early birds, not only the ancestor of hummingbirds, most likely could not taste. This ancestral absence of sweet perception can be inferred because in most nonavian lineages, the sweet receptor is formed by a heterodimer of two taste receptors, T1R2 and T1R3, the former of which is missing in birds whose genomes have been sequenced (now spanning the entire avian tree) (*10*). The umami receptor is formed by pairing T1R3 with another molecular partner, T1R1, which is still present in bird genomes. Toda *et al.* suggest that the evolution of the umami receptor may have played a critical permissive role in some of the classic examples of avian adaptive radiation (e.g., the Darwin's finches and honeycreepers) as well.

Toda *et al.* analyzed taste receptor function and evolution in passerine birds. Passerines are members of the largest of 40 orders of birds, which comprises more than half of all bird species. The authors reviewed the frequency of nectar consumption (the most common source of dietary sugars) across birds, identifying multiple nonpasserine (hummingbirds, parrots, and others) and many passerine lineages that extensively consume sugars. Members of the oscine passerines—commonly known as songbirds—including well-known nectarivore radiations, such as the honeyeaters, sunbirds, and Hawaiian honeycreepers, as well as less-specialized groups, such as wattlebirds, white-eyes, and tanagers, frequently imbibe nectar and similar sugar sources, such as honeydew. Choice tests in a nectar specialist (a honeyeater) and a nonspecialist (the canary) demonstrate that oscine passerines both consume and taste sugars. A comparative evolutionary analysis of the pattern of nectarivory across birds reported by Toda *et al.* suggests that perception of sweetness may be hard to acquire—with possibly a single origin within oscines—but, once present, this trait permits rapid gains and losses of sugar exploitation.

To interrogate the molecular basis of sweet perception, Toda *et al.* cloned the T1R genes from species spanning most of the passerine evolutionary tree. This allowed them to express these native genes (as well as heterospecific combinations and mutants) in cell cultures and detect ligand binding using a cell-based luminescence assay. All tested oscine umami receptors (T1R1-T1R3 dimers) responded strongly to sugars, especially sucrose. By contrast, umami receptors from the two nonoscine passerines retained the ancestral umami sensitivity with no sign of activation by sugar. Expression studies combining honeyeater T1R receptor subunits with the subunits of other oscine species demonstrate conserved sweet perception—and thus presumably a shared molecular mechanism—across most oscines. By contrast, coexpression of these genes with their partners cloned from hummingbirds showed no binding activity for sugars, though they still responded to amino acids, which suggests a different binding mechanism between these deeply divergent lineages.

The shared oscine binding mechanism was further explored by synthesizing inferred ancestral genes for T1R1 and T1R3 and expressing these and chimeras of the two to test for sugar sensitivity. These experiments suggest that the origin of shared sweet perception is nested somewhat within the oscine lineage (excluding two major branches of oscines) and that it involves evolution of residues in both T1R1 and T1R3 subunits, whereas previous work on hummingbirds identified most changes in T1R3. Functional convergence in these lineages is therefore likely based on complex changes in tertiary structure evolving from different starting points, rather than parallel residue-for-residue replacements, which might explain the relatively few origins of this trait.

The diversity of passerines has long intrigued biologists (*11*), and some oscine lineages containing nectarivores have unusually high species diversity given their age (*8*, *12*, *13*). Biogeographic analyses of oscines indicate that they originated in Australasia (*13*, *14*, *15*), and thus sweet perception probably evolved there in the Oligocene (34 to 23 million years ago) or possibly earlier. The early evolution of sweet perception likely played an important role in diversification of this lineage, which is now a numerically and ecologically dominant component of terrestrial avifaunas the world over. ■

REFERENCES AND NOTES

1. Y. Toda *et al.*, *Science* **373**, 226 (2021).
2. Z. Liu, F.-Y. Qi, D.-M. Xu, X. Zhou, P. Shi, *Sci. Adv.* **4**, eaat8821 (2018).
3. R. Borges *et al.*, *BMC Genomics* **16**, 751 (2015).
4. L. R. Yohe *et al.*, *Evolution* **71**, 923 (2017).
5. M. W. Baldwin *et al.*, *Science* **345**, 929 (2014).
6. L. S. Carvalho, D. M. A. Pessoa, J. K. Mountford, W. I. L. Davies, D. M. Hunt, *Front. Ecol. Evol.* **5**, 34 (2017).
7. A. A. Fernandez, M. R. Morris, *Am. Nat.* **170**, 10 (2007).
8. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, *Nature* **491**, 444 (2012).
9. J. A. McGuire *et al.*, *Curr. Biol.* **24**, 910 (2014).
10. S. Feng *et al.*, *Nature* **587**, 252 (2020).
11. R. J. Raikow, *Syst. Zool.* **35**, 255 (1986).
12. R. G. Moyle, C. E. Filardi, C. E. Smith, J. Diamond, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1863 (2009).
13. C. H. Oliveros *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7916 (2019).
14. R. G. Moyle *et al.*, *Nat. Commun.* **7**, 12709 (2016).
15. F. K. Barker, A. Cibois, P. Schikler, J. Feinstein, J. Cracraft, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11040 (2004).

10.1126/science.abj6746

CANCER

Sowing the seeds of leukemia before birth

Distinct cooperation between a mutated gene and trisomy 21 triggers leukemia in utero

By **Irene Roberts**[1,2] *and* **Paresh Vyas**[1]

Each year, ~200,000 babies worldwide are born with Down syndrome (DS), owing to constitutional trisomy of chromosome 21 (T21) (*1*). Children with DS have a markedly increased risk of leukemia, particularly in their first 4 years. Almost 60,000 (30%) will harbor within their blood cells damaging, fetally acquired mutations in the transcription factor gene GATA binding protein 1 (*GATA1*), which encodes a short GATA1 protein (GATA1s) and triggers the first step in the development of leukemia (*2*). *GATA1* mutations are rare in disomic individuals and virtually never cause leukemia in the absence of T21. Why *GATA1* mutations are so frequent in T21 babies and the mechanisms by which a supernumerary chromosome 21 (Hsa21) predisposes to, and cooperates with, genetic events in DS leukemogenesis are not known. On page 179 of this issue, Wagenblast *et al.* (*3*) identify Hsa21 microRNAs (miRNAs) that cooperate with GATA1s and map the cellular origin of the leukemia.

Compared with children of the same age without DS, the risk of myeloid leukemia (modeled by Wagenblast *et al.*) is 150-fold greater, whereas acute lymphoblastic leukemia is more than 20-fold higher in people with DS. DS also causes functional and developmental abnormalities in virtually all organs and tissues. Progressive cognitive impairment, dementia, and cardiac disease are major causes of ill health, with profound effects on the lives of patients and their families. How the supernumerary Hsa21 causes the clinical features of DS remains unclear.

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [2]Department of Paediatrics, University of Oxford, Oxford, UK. Email: irene.roberts@paediatrics.ox.ac.uk; paresh.vyas@imm.ox.ac.uk

Wagenblast *et al.* use CRISPR-Cas9 editing of human fetal hematopoietic stem and progenitor cells (HSPCs) trisomic for Hsa21 to induce DS-specific leukemogenic deletions in *GATA1* and stromal antigen 2 (*STAG2*). These edited cells were then transplanted into immunocompromised mice, creating xenograft mouse models that faithfully recapitulate DS leukemia. In contrast to previous models, it seems likely that targeting primary human fetal HSPCs to induce the pathognomonic *GATA1* mutations provided the permissive cellular substrate for transformation, as predicted by clinical and biological studies in DS (*2*, *4*, *5*). Whereas in disomic fetal cells, expression of GATA1s caused severe impairment of erythropoiesis (corresponding to anemia in non-DS individuals with germline *GATA1* mutations), in T21 fetal cells, GATA1s expression increased megakaryocytes and leukemia-like blast cells similar to those in the preleukemia in DS babies (*2*, *4*).

Studies to decipher the molecular basis for the effects of T21 have generated a wealth of data from a range of cell types, most focusing on gene dosage of Hsa21-located genes. Hsa21, the smallest human chromosome, has ~230 protein-coding genes and almost twice as many non–protein coding genes, including five miRNAs (*1*, *6*). Although many Hsa21 genes are expressed at 1.5-fold higher levels than in matching disomic populations, this is highly tissue and cell population specific, and the overall effects of the supernumerary Hsa21 extend far beyond this. Almost all datasets show genome-wide perturbation of gene expression by T21. This likely explains, at least in part, why causative links between altered gene expression and phenotypes in DS have been so difficult to establish and why none of these phenotypes has so far been explained by a single gene acting alone. Studies aimed at narrowing the region(s) of Hsa21 responsible for specific phenotypes are often inconsistent and hampered by the rarity of individuals with DS owing to partial T21 (<1%) and by cellular systems that imperfectly capture the genetic and epigenetic complexity of DS.

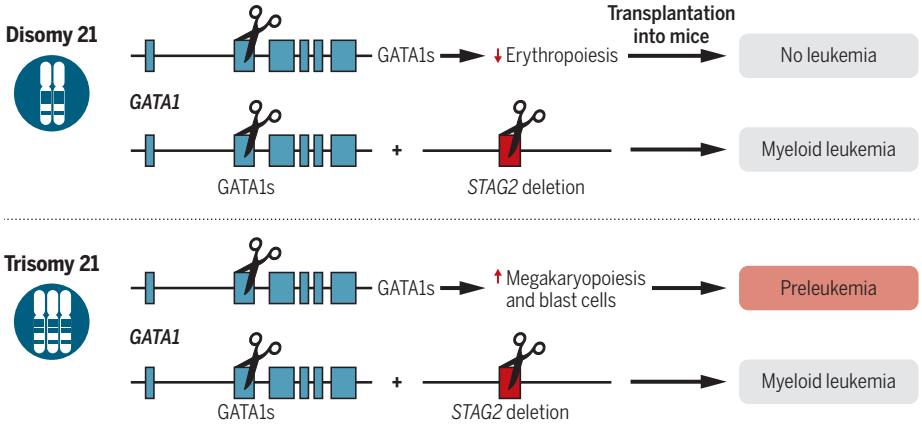
Given the importance of cell context, the development of a biologically accurate model, recreating the precise initiating step necessary for all DS myeloid leukemias (GATA1s-encoding mutation) in the precise cell type where this event occurs (fetal liver HSPCs), should allow more specific and tractable questions about the role of T21 to be addressed. Focusing on the mechanism of cooperation between T21 and GATA1s, Wagenblast *et al.* asked how binding of mutant GATA1s protein differed from full-length, wild-type GATA1 in fetal HSPCs. They found that GATA1s specifically bound to thousands of promoters, including those that regulate genes controlling miRNA production. They focused on three Hsa21 miRNA genes—*MIR-125b-2*, *MIR-155*, and *LET-7C*—whose expression was up-regulated specifically in T21 long-term hematopoietic stem cells (LT-HSCs). Enforced overexpression of these miRNAs in normal fetal LT-HSCs mimicked some features of T21 cells, whereas ablating them partially reduced the preleukemic phenotype of GATA1s, suggesting that GATA1s-mediated up-regulation of these Hsa21 miRNAs' expression could partially explain the mechanism of GATA1s-induced preleukemia in DS. This adds DS myeloid leukemia to the list of human diseases, including other phenotypic abnormalities in DS, that may be caused by altered miRNA expression (*1*). The precise mechanisms underlying the distinct cooperation between T21 and GATA1s in human fetal cells remain enigmatic but are likely to include several Hsa21 genes with known roles in embryonic and fetal hematopoiesis, acting together in a cell context–dependent manner (see the figure).

The reasons for the high frequency of somatic *GATA1* mutations in DS fetal blood cells (most likely owing to selection) and why most *GATA1*-mutant cells are rapidly cleared after birth—a feature not captured in Wagenblast *et al.*'s model, which may involve the T21 hematopoietic cell niche—are not yet known. Evidence of a mutagenic phenotype or generalized defect in DNA repair in DS tissues, including blood cells, is sparse. Children with DS have a lower than expected incidence of solid tumors, pointing to specific properties of T21 in hematopoietic cells. This might also explain why T21 appeared to be less relevant for postnatal progression of myeloid leukemia in Wagenblast *et al.*'s model.

T21 is the most frequent constitutional aneuploidy compatible with survival into adulthood, although fetal loss is high (*1*). Yet, despite the importance of genome organization on function (*7*), the effects of aneuploidy on nuclear architecture of human fetal cells and subsequent consequences for gene expression and protein production are unknown. Similarly, the development of new model systems is needed to better understand the function of each element of Hsa21, only 48% of which has currently been mapped in detail (*1*). Moreover, it can be argued that the value of such models lies with the opportunities for improving outcomes for individuals with DS and their families. This will likely be realized through small steps in specific complications such as leukemia but could include bolder initiatives such as harnessing X chromosome inactivation in female cells by perhaps integrating an X inactive specific transcript (*XIST*) transgene to reduce Hsa21 transcriptional output back to disomic levels (*8*). ■

Cooperating changes to initiate leukemia in fetal cells

The supernumerary copy of human chromosome 21 in Down syndrome [trisomy 21 (T21)] perturbs fetal gene expression and hematopoietic stem and progenitor cell (HSPC) development. Acquisition of exon 2 mutations in the GATA binding protein 1 (*GATA1*) gene in fetal T21 but not disomic HSPCs causes preleukemia in mice because of expression of a short GATA1 protein (GATA1s). Secondary mutations in GATA1s-expressing cells, such as in stromal antigen 2 (*STAG2*), cause myeloid leukemia in disomic and T21 mouse fetal HSPCs.



REFERENCES AND NOTES

1. S. E. Antonarakis *et al.*, *Nat. Rev. Dis. Primers* **6**, 9 (2020).
2. I. Roberts *et al.*, *Blood* **122**, 3908 (2013).
3. E. Wagenblast *et al.*, *Science* **373**, eabf6202 (2021).
4. A. Roy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17579 (2012).
5. M. Labuhn *et al.*, *Cancer Cell* **36**, 123 (2019).
6. M. Hattori *et al.*, *Nature* **405**, 311 (2000).
7. C. Do, Z. Xing, Y. E. Yu, B. Tycko, *Epigenomics* **9**, 189 (2017).
8. J. C. Chiang, J. Jiang, P. E. Newburger, J. B. Lawrence, *Nat. Commun.* **9**, 5180 (2018).

10.1126/science.abj3957

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

SUPERCONDUCTIVITY

A touch more unconventional

Strong antiferromagnetic interactions are revealed in superconducting nickelates

By **Eva Benckiser**

Superconductivity is one of the most fascinating, macroscopically observable quantum phenomena. A notable aspect of its description is the existence of an attractive interaction, a "glue," between electrons. This can bind them into Cooper pairs, which then condense into a common superconducting ground state. In contrast to conventional superconductors, which are described by the theory of Bardeen, Cooper, and Schrieffer, a primary phonon-mediated pairing mechanism is unlikely in the high-temperature (T_c) superconducting cuprates, and therefore they are referred to as unconventional. One characteristic of all high-T_c cuprates are the strong, quasi–two-dimensional antiferromagnetic correlations, with spin waves or magnons as possible alternative pairing glue. On page 213 of this issue, Lu *et al.* (*1*) report on antiferromagnetic magnon-like spin excitations in superconducting infinite-layer nickelate thin films.

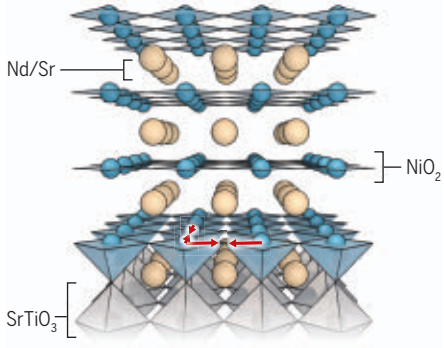
The recent discovery of this new class of potentially unconventional superconductors, $Nd_{1-x}Sr_xNiO_2$ (*2*), has sparked great interest. Although the T_c of 15 K is lower than the $T_c \leq 130$ K in the cuprates, their structural and electronic similarity immediately raised the possibility of magnetic correlations and a related pairing mechanism. To this point, the authors measured the spin wave dispersion of $Nd_{1-x}Sr_xNiO_2$ with resonant inelastic x-ray scattering. A nearest-neighbor superexchange coupling of $J = 64$ meV was determined, which quantifies the coupling of two neighboring magnetic nickel ions through the nonmagnetic oxygen ion in between. This large superexchange interaction is unexpected because it is not much smaller than the J found in cuprates of over 100 meV.

The outstanding question is whether the new nickelate superconductors are just like the high-T_c cuprates. At first glance, the isostructural transition-metal-oxide planes contain monovalent Ni^{1+} ions that are isoelectronic to Cu^{2+}, with a single spin ½ in the planar $3d_{x^2-y^2}$ orbital on a square lattice (see the figure). In both materials, strong electron-electron correlations cause the partially occupied transition-metal d-band to split up into so-called lower and upper Hubbard subbands. One major difference in the electronic structure is the relative energy of the oxygen-2p–derived bands with respect to the Hubbard subbands. Whereas in cuprates, the 2p bands lie in between, in nickelates, they are located below both subbands. In a local picture, this translates into a strongly reduced hybridization of nickel 3d- and oxygen 2p-orbitals with possibly important consequences for the magnetic superexchange interaction. Reduced hybridization reduces the effective hopping t between magnetic nickel ions by way of the oxygen ligand. Because the coupling constant J is proportional to t^2/U, where U is the on-site Coulomb repulsion, values 10 times smaller than those in cuprates have been estimated (*3*, *4*). Quantum chemical calculations (*5*), however, provide an alternative argument that both t and U are reduced in the nickelates. These calculations predict a superexchange coupling $J = 77$ meV for the bulk structure of $NdNiO_2$, which is relatively close to the one reported by Lu *et al.*

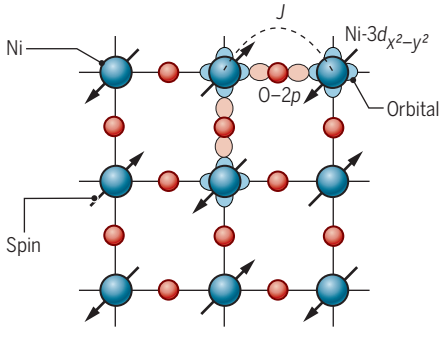
A notable aspect of the authors' infinite-layer nickelate superconductors is that they are synthesized as epitaxial thin films grown on a single-crystalline $SrTiO_3$ substrate. Neither superconductivity nor antiferromagnetic order have been observed in bulk samples (*6*). This calls into consideration the extent to which heteroepitaxial modifications from the substrate play a decisive role in phase stabilization. Effects to be considered are, for example, the biaxial lattice mismatch with the underlying substrate and possible structural and electronic reconstructions at the polar substrate-film interface (*7*, *8*). Because the hopping integral depends inversely on some power of the distance between the ions, a shrinkage of the in-plane lattice as provided by the epitaxy with the underlying $SrTiO_3$ substrate could increase the superexchange coupling. Should heteroepitaxial influences prove to be decisive, extremely interesting perspectives arise from this for a targeted modification of the electronic structure as well as the magnetic interactions and thus the possibility to stabilize superconductivity at higher temperatures.

Addressing the coupling to rare-earth itinerant electrons will also need to be sorted out. This might be accomplished by examining the spin waves in the analogous lanthanum compound. The self-doped Ruddlesden-Popper natural superlattices with varying numbers of infinite-layer nickelate slabs are another possible material system for gaining insight into the electronic structure. A recent x-ray spectroscopy study found sizable antiferromagnetic superexchange but also indicates differences in nickel-oxygen hybridization (*9*). Further information on the symmetry of the superconducting wave function is desired to assess the possibility of a magnon-mediated Copper-pairing mechanism. High-T_c cuprates show a characteristic d-wave symmetry of the superconducting gap function, which can be examined, for example, with scanning tunneling spectroscopy. ■

Superconducting nickelates



A structure on a substrate
The lattice mismatch with the $SrTiO_3$ substrate may affect the antiferromagnetic superexchange coupling (J).



Magnetic interactions in the NiO_2 planes
Antiferromagnetic coupling of neighboring Ni spins arises from exchange interaction involving specific oxygen orbitals.

Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany. Email: e.benckiser@fkf.mpg.de

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. H. Lu *et al.*, *Science* **373**, 213 (2021).
2. D. Li *et al.*, *Nature* **572**, 624 (2019).
3. M. Jiang, M. Berciu, G. A. Sawatzky, *Phys. Rev. Lett.* **124**, 207004 (2020).
4. Z. Liu et al., *NPJ Quantum Mater.* **5**, 31 (2020).
5. V. M. Katukuri, N. A. Bogdanov, O. Weser, J. van den Brink, A. Alavi, *Phys. Rev. B* **102**, 241112 (2020).
6. B.-X. Wang *et al.*, *Phys. Rev. Mater.* **4**, 084409 (2020).
7. F. Bernardini, A. Cano, *J. Phys. Mat.* **3**, 03LT01 (2020).
8. B. Geisler, R. Pentcheva, *Phys. Rev. B* **102**, 020502 (2020).
9. J. Q. Lin *et al.*, *Phys. Rev. Lett.* **126**, 087001 (2021).

10.1126/science.abi6855

ICE PHYSICS

A flexible and springy form of ice

Single-crystal ice microfibers recover their shape after bending to near their breaking limit

By **Erland M. Schulson**

Water ice is ordinarily fragile and breaks if extended by just <0.1 % (*1*). On page 187 of this issue Xu *et al.* (*2*) show that fibers of cold ice a few micrometers or less in diameter can bend without breaking into a near-circular shape tens of micrometers in radius. Upon unloading, the fibers spring back to their original shape. Such strains are near the theoretical limit of ~15% (*3*), so the deformation is completely elastic. The microfibers can transmit visible light as effectively as state-of-the-art on-chip light guides (*4, 5*). The authors also find that extreme bending creates a near-surface layer on the compressive side that transforms relatively quickly from ice I_h (hexagonal crystal structure) to ice II (rhombohedral crystal structure). This pronounced elasticity and transparency reflect the absence of defects within the material, and the structure change implies a low barrier for the ice I_h-to-II transformation.

Ice in nature usually contains pores, microcracks, grain boundaries, crystal dislocations, and other microstructural defects, as well as surface irregularities. Such features originate through the growth and thermal-mechanical history of the material and act both to concentrate stress and to scatter visible light. However, the almost perfect microfibers studied by Xu *et al.* were produced with a method that used electric-field–enhanced growth (see the figure) (*6*). Examination with cryo–transmission electron microscopy revealed single crystals without defects and with very smooth surfaces (surface roughness <1 nm).

Upon cooling to –150°C, a fiber 4.4 μm in diameter could be bent with a micromanipulator to a radius as small as 20 μm. This process created an elastic strain of 10.9% within the near-surface region. Correspondingly, the outer-fiber stress reached ~1.4 GPa (1 GPa is the pressure on Earth at a depth of ~30 km). After manipulator retraction, the fiber had no residual curvature, and multiple fibers exhibited similar mechanical behavior. Rarely have mechanical properties so near theoretical limits been attained in any material.

The optical quality of the microfibers is also attributed to the absence of both interior and exterior defects. The transmissibility was assessed by coupling visible light to one end of a fiber and then by measuring the position-dependent intensity of scattering. As Xu *et al.* suggest, defect-free microfibers of ice have the potential to serve as low-loss, optical waveguides at low temperatures.

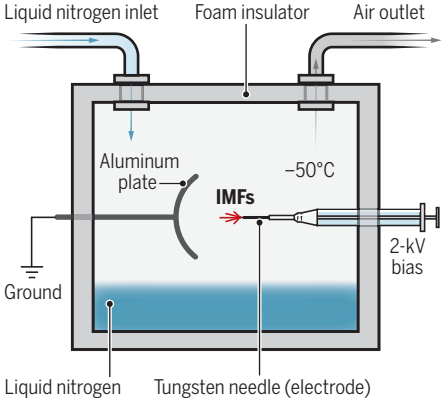
Under terrestrial conditions, ice adopts the I_h crystal structure, and its hexagonal symmetry is reflected in the shape of snowflakes. Under high pressure, however, structures of higher density are favored, of which a dozen or more exist (*7*). The transformation of I_h to ice II is marked by an ~25% increase in density, from 925 to ~1150 kg m^{-3} (at –100°C), accompanied by a proportional decrease in volume. Within the surface region on the compressive side of a highly bent fiber, the stress reached ~0.4 GPa. Through the use of Raman spectroscopy, Xu *et al.* detected evidence of ice II in a cold fiber (–70°C). Under the combination of stress and temperature just noted, thermodynamics dictates that ice II is the more stable phase (*8*). The transformation to ice II within the fibers occurred within ~100 s, indicative of a relatively rapid process compared with sluggish kinetics within bulk ice (*9*). The small sample could contribute to faster kinetics as the fiber dimensions approach the critical nucleus size.

In showing that ice can reach a high level of mechanical integrity, and with high optical quality, Xu *et al.* revealed the potential for similar improvement through appropriate processing in the behavior of other brittle, crystalline materials. Intriguing issues still remain. Although the elastic strain and strength exhibited by the microfibers are extraordinarily high, they are still below the theoretical limit. Dislocations could have been generated at the free surface and nucleated microcracks that then propagated. The mechanism for scattering visible photons in these near-perfect ice fibers is also not apparent. The ice phase transition could occur within micrometer-sized asperities as they slide slowly past each other across opposing surfaces loaded in shear and compression, creating stick-slip friction. Bending microfibers could potentially achieve even greater pressure and detect transformations to other higher-density ices. ■

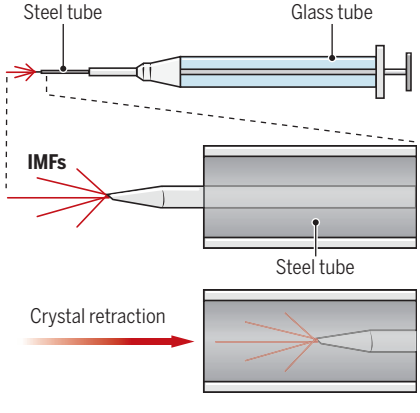
Thayer School of Engineering, Dartmouth College, Hanover NH 03755 USA. Email: Erland.M.Schulson@Dartmouth.edu

Field-induced growth

Xu *et al.* grew ice microfibers (IMFs) using an applied electric bias to enhanced diffusion of water vapor toward the tip of a tungsten needle.



Controlled growth
Schematic of the experimental chamber used to grow IMFs under temperature-controlled conditions is shown.



Microfiber extraction
Retracting the tungsten needle into a cold coaxial steel tube (1-mm diameter) allows for sample transfer.

REFERENCES AND NOTES

1. E. M. Schulson, P. Duval, *Creep and Fracture of Ice*, (Cambridge Univ. Press, 2009).
2. P. Xu *et al.*, *Science* **373**, 187 (2021).
3. A. Kelly, N. H. Macmillan, *Strong Solids* (Clarendon, 1986).
4. T. Horikawa, D. Shimura, T. Mogami, *MRS Commun.* **6**, 9 (2016).
5. S. Wu *et al.*, *Micromachines (Basel)* **11**, 326 (2020).
6. R. Ma *et al.*, *Nature* **577**, 60 (2020).
7. V. Petrenko, R. W. Whitworth, *Physics of Ice* (Oxford Univ. Press, 1999).
8. G. Tammann, *Kristallisieren und Schmelzen* (Barth, 1903), pp. 315–344.
9. W. B. Durham *et al.*, *J. Geophys. Res.* **93** (B9), 10191 (1988).

ACKNOWLEDGMENTS

The author gratefully acknowledges the National Science Foundation (award no. 1947107) for financial support.

10.1126/science.abj4441

GRAPHIC: C. BICKEL/*SCIENCE*

PHOTO: TOM MANGELSEN/MINDEN PICTURES

Sampling 585 yellow baboons across 14 years in the Amboseli National Park, Kenya, demonstrates that most gut microbiota traits exhibit some degree of heritability.

GENOMICS

Host genetics influence the gut microbiome

Longitudinal data from nonhuman primates reveal widespread gut microbiome heritability

By **Liliana Cortes-Ortiz**[1] *and* **Katherine R. Amato**[2]

The influences of the microbiota on host physiology are so pervasive that the microbiota has been hypothesized to play a critical role in host evolution by shaping key host phenotypes (*1*). However, to contribute to host evolution, traits must be transmitted across generations. One way to assess whether some or all microbes are influenced by the genetic composition of the host, and therefore conserved across generations, is by measuring heritability. Host species–specific patterns in the composition of the microbiome—the genetic content of the microbiota—suggest there is some degree of heritability in the microbiota (*2*). However, studies evaluating variation in the microbiota and host genetics within a single host species have generally reported low heritability for a small proportion of microbial taxa (*3*). On page 181 of this issue, Grieneisen *et al.* (*4*) reject this common conclusion by demonstrating that most gut microbiota traits in wild baboons exhibit some degree of heritability.

Grieneisen *et al.* analyzed fecal samples from 585 yellow baboons (*Papio cynocephalus*, some of them admixed with anubis baboons, *P. anubis*) from 10 free-ranging social groups sampled longitudinally across 14 years in Amboseli National Park, Kenya. They determined the microbial composition of each sample and defined 1134 microbiome traits (i.e., relative abundance and presence or absence of microbial taxa, and measures of overall microbiome composition). Then they used kinship data and environmental data, including rainfall, social interactions, and group-level diet, to calculate heritability (the proportion of the variation of a trait attributable to genetic variance as opposed to environmental factors) of each of these microbiome traits with a standard formula used for nonmicrobial traits in livestock and wild animals. The scale of this dataset is currently unmatched in any host-microbe system, which may explain why previous estimates of microbiome heritability have been so low.

Most microbiome studies, including those of humans, analyze a small number of samples or target populations at a single point in time. When Grieneisen *et al.* subsampled their dataset to simulate smaller sample sizes, or sampling at a single time point, most signals of microbiome heritability were not detected. Without data from multiple time points and a large number of individuals to account for stochasticity as well as temporal variation in host environments and the microbiota, it is difficult to accurately estimate microbiome heritability. Therefore, microbiome heritability is likely higher than reported previously in most host species, including humans.

Estimating heritability of the microbiota under natural conditions is essential to advancing knowledge of the extent to which the microbiota affects host ecology and evolution. However, measuring heritability of the microbiota is difficult because there are many environmental variables at play. As Grieneisen *et al.* demonstrate, nonhuman primates provide a strong but currently underutilized system for addressing this challenge. In addition to being genetically closely related to humans, the ecology of nonhuman primate populations is extensively studied and several populations, such as the Amboseli baboons, have decades of relevant data.

Such long-term studies that follow known individuals across their lives (*5*) give primatologists an unmatched ability to collect detailed, longitudinal data describing environment, diet, and individual behavior, as well as noninvasive biological samples. Furthermore, a diversity of physiological and behavioral adaptations to a range of environments exists within the Primates order, allowing targeted testing of specific host-microbe interactions with different primate species and populations. Baboons represent an excellent general model for the human microbiota because previous research has demonstrated that humans share more microbiome traits (e.g., taxonomic and functional profiles) with baboons than chimpanzees (*6*). However, comparative data collected from a variety of primate taxa could reveal underlying mechanisms that maintain specific associations between host genetic traits and particular components of the primate gut microbiota.

Despite reporting evidence of heritability in the majority of microbiome traits they assessed, Grieneisen *et al.* also show that a larger proportion of variation in microbiome data is attributable to environmental factors rather than host genetic factors, as has been shown consistently across studies in other systems (*3*). Furthermore, estimates of microbiome heritability varied between dry and wet seasons, and with diet and host age, as a result of changing environmental contributions to microbi-

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Anthropology, Northwestern University, Evanston, IL, USA. Email: katherine.amato@northwestern.edu

ome variation. These results highlight the plastic nature of the gut microbiota, which could allow it to play a role in facilitating rapid, local adaptation in hosts. However, the mechanisms by which these interactions occur remain unclear. The resolution of data for a single time point or individual precluded Grieneisen *et al*. from empirically identifying which environmental factors or host behaviors were driving temporal patterns in heritability.

A key next step will be to improve host genomic resolution to identify the specific genomic regions and mechanisms through which associations between host genetics and the microbiota occur. Although some of these types of studies are being conducted for humans (*7*), studies of natural populations of other mammalian taxa would help identify generalizable principles that underlie host genetic-microbiota associations. For example, in hybrid zones of genetically and microbially divergent host species (*8, 9*), the study of paired microbiome and host genomic data for individuals with a range of admixed genotypes could help identify specific associations. Knowledge of the mechanisms shaping interactions between host genetics and gut bacterial communities will be critical for generating testable hypotheses for other body sites (e.g. skin, mouth, urogenital tract) and other microbial community members (e.g., microscopic eukaryotes, viruses). Similarly, improving resolution of data describing the microbiota will allow testing of the taxonomic specificity at which these interactions occur as well as the extent to which microbial taxonomy or functions are more strongly associated with host genetics. Technological advances are making it easier and more affordable to generate microbial whole-metagenome data. Together with the development of analytical tools to study the dual genomic composition of hosts and their microbiota in nonmodel organisms, these data will shift explorations of microbial influences on host evolution from correlation and theory to causation and mechanism. ■

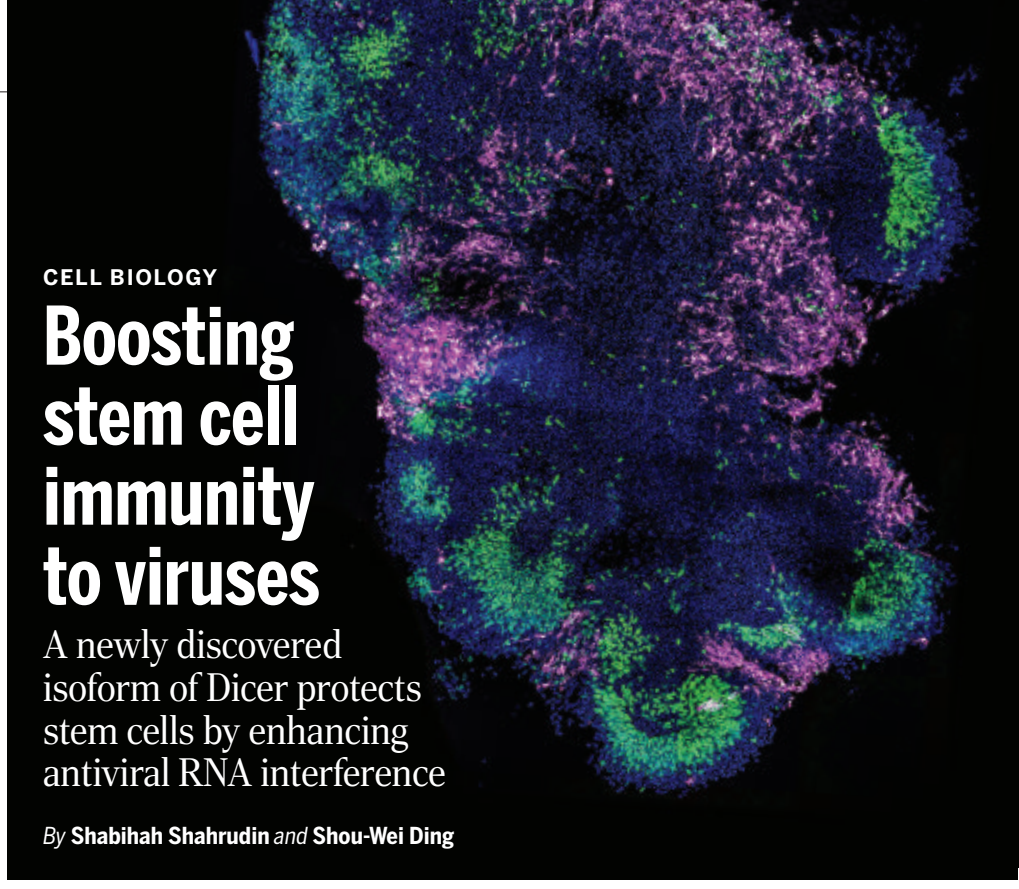
REFERENCES AND NOTES

1. K. R. Theis *et al*., *mSystems* **1**, e00028 (2016).
2. E. K. Mallott, K. R. Amato, *Nat. Rev. Microbiol.*, 1 (2021).
3. D. Rothschild *et al*., *Nature* **555**, 210 (2018).
4. L. Grieneisen *et al*., *Science* **373**, 181 (2021).
5. T. Clutton-Brock, in *Long-Term Field Studies of Primates* (Springer, 2012), pp. 437–449.
6. K. R. Amato *et al*., *Genome Biol.* **20**, e27879v27871 (2019).
7. A. Kurilshikov *et al*., *Nat. Genet.* **53**, 156 (2021).
8. L. Cortés-Ortiz *et al*., *Int. J. Primatol.* **40**, 114 (2019).
9. K. R. Amato *et al*., *Oecologia* **180**, 717 (2016).

ACKNOWLEDGMENTS

K.R.A. is a fellow of the Canadian Institute for Advanced Research "Humans and the Microbiome" program.

10.1126/science.abj5287



CELL BIOLOGY

Boosting stem cell immunity to viruses

A newly discovered isoform of Dicer protects stem cells by enhancing antiviral RNA interference

By **Shabihah Shahrudin** *and* **Shou-Wei Ding**

In the absence of antiviral Dicer, neural stem cells (green) in mouse brain organoids are more susceptible to SARS-CoV-2 infection (magenta).

PHOTO: POIRIER *ET AL*. (*12*)

Mammalian stem cells exhibit deficiencies in innate immunity regulated by interferons (IFNs), so they rely on constitutive expression of some IFN-stimulated genes (ISGs) (*1*) and Argonaute 2 (AGO2)–dependent RNA interference (RNAi) (*2*, *3*) for antiviral protection. Mammalian antiviral RNAi is initiated by Dicer, which processes viral double-stranded RNA (dsRNA) replicative intermediates into small interfering RNAs (siRNAs) that act as specificity determinants for viral RNA cleavage by RNA-induced silencing complex [(RISC) which contains AGO2] (*2*–*10*). However, it remains unclear how stem cells activate antiviral RNAi because deletion of *Dicer* paradoxically enhances virus resistance in mouse embryonic stem cells (*11*). On page 231 of this issue, Poirier *et al*. (*12*) show that mouse and human stem cells have a specialized Dicer isoform for virus-derived siRNA (vsiRNA) production to initiate potent antiviral RNAi. This further indicates that siRNA therapeutic strategies may be viable for RNA viruses such as Zika virus (ZIKV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

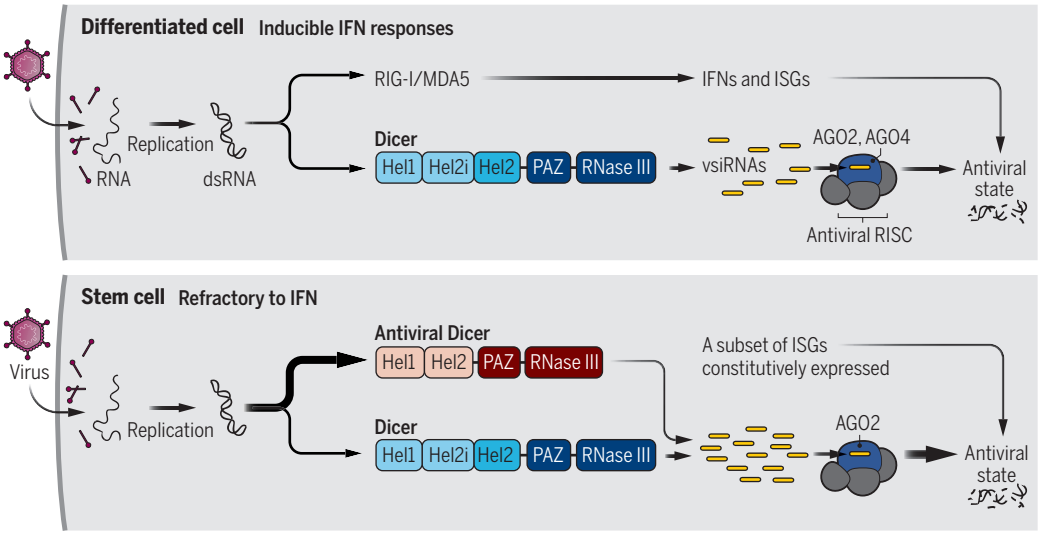
Similar to *Caenorhabditis elegans* nematodes (*13*), mammals encode a single Dicer ribonuclease (RNase) that is responsible for the biogenesis of both microRNAs from their stem-loop precursor transcripts and siRNAs from long dsRNA. Dicers have an amino-terminal helicase domain and tandem RNase III domains as well as additional domains between them, such as the RNA-binding PAZ domain (see the figure). Helicase domains of Dicers and mammalian retinoic acid–inducible gene 1 (RIG-I)–like receptors that trigger IFN-regulated immune responses are highly homologous and contain a distinct helicase-insertion subdomain (Hel2i). The study by Poirier *et al*. began with discovering an alternatively spliced mouse and human *Dicer* messenger RNA (mRNA) from embryonic, neuronal, and tissue stem cells. These transcripts contain an in-frame deletion of exons 7 and 8 so that Hel2i is absent. Poirier *et al*. demonstrate that RNAi initiated by this Dicer isoform, designated antiviral Dicer (aviD), protects mouse stem cells from infections with the RNA viruses ZIKV and SARS-CoV-2.

The authors found that the loss of Hel2i enhances aviD processing of long dsRNA into siRNAs without impairing its ability to generate microRNAs. Notably, ZIKV and another RNA virus, Sindbis virus, replicated to lower amounts in human cells forced to express aviD. Moreover, the antiviral activity of aviD was abolished by depletion of AGO2 or ectopic expression of the viral suppressor of RNAi (VSR) encoded by Nodamura

Department of Microbiology and Plant Pathology, College of Natural and Agricultural Sciences, University of California, Riverside, CA, USA. Email: shou-wei.ding@ucr.edu

Intrinsic and innate antiviral immunity in mammalian cells

Inducible IFN-regulated immune responses and antiviral RNAi act additively to establish an antiviral state in differentiated cells. By contrast, stem cells are refractory to IFN signaling, so they predominantly respond to virus infection through RNAi activation using an antiviral Dicer. This isoform of Dicer processes viral dsRNA replicative intermediates into vsiRNAs that act as specificity determinants for viral RNA cleavage by AGO2 in the RISC. This prevents RNA virus replication in stem cells.



AGO, Argonaute; dsRNA, double-stranded RNA; Hel, helicase; Hel2i, helicase-insertion subdomain; IFN, interferon; ISGs, IFN-stimulated genes; MDA-5, melanoma differentiation-associated protein 5; RIG-I, retinoic acid–inducible gene 1; RISC, RNA-induced silencing complex; RNAi, RNA interference; RNase, ribonuclease; vsiRNAs, virus-derived small interfering RNAs.

virus (*1*, *2*, *14*), providing further evidence that aviD confers virus resistance by RNAi.

ZIKV may cause devastating microcephaly in infants born from infected mothers because the virus infects human neural progenitor cells (hNPCs) during brain development. Infection of hNPCs by ZIKV induces AGO2-dependent RNAi, and enhancing antiviral RNAi can provide protection in hNPC-derived brain organoids, which recapitulate much of the composition, diversity, and organization of cell types found in the developing human fetal brain (*3*). Thus, Poirier *et al.* generated brain organoids from wild-type mouse embryonic stem cells, which express both Dicer and aviD, and mutant organoids expressing only Dicer or aviD. They detected no significant differences in growth among these brain organoids. However, the absence of *aviD* compromised stem cell resistance to ZIKV infection in brain organoids, resulting in increased numbers of virally infected stem cells and production of infectious virus particles and more potently inhibited stem cell division and organoid growth.

Sequencing of small RNAs in ZIKV-infected wild-type and mutant mouse stem cell–derived organoids revealed production of predominantly 22-nucleotide vsiRNAs, consistent with previous observations in ZIKV-infected hNPCs (*3*). After normalization by the yield of ZIKV particles, however, brain organoids that do not express *aviD* accumulated much less vsiRNAs than those that do, with or without *Dicer* coexpression. This demonstrated significantly enhanced activity of aviD to process the viral dsRNA replicative intermediates into vsiRNAs compared with full-length Dicer. Thus, aviD protects brain organoids from ZIKV infection by triggering antiviral RNAi.

Poirier *et al.* also provided evidence of vsiRNA production targeting SARS-CoV-2 in infected brain organoids derived from mouse embryonic stem cells engineered to express the human viral entry receptor angiotensin-converting enzyme 2 (ACE2). As for ZIKV infection, *aviD* depletion was more effective in promoting SARS-CoV-2 infection of stem cells in brain organoids than *Dicer* depletion. Consistently, SARS-CoV-2 replicated to higher titers in human cells deficient in *aviD* expression. By contrast, both *aviD*- and *Dicer*-deficient cells supported replication of two DNA viruses at similar levels, possibly because only RNA viruses produce sufficiently abundant dsRNA replicative intermediates to trigger antiviral RNAi. Future studies will be necessary to determine whether specific deletion of only *aviD* or *Dicer* produces in vivo phenotypic differences in development or antiviral RNAi.

Poirier *et al.* propose that *aviD* may have evolved to protect stem cells from RNA viruses in part to compensate for stem cell deficiencies in IFN production and signaling, which may be critical to maintain pluripotency (*1*). Consistently, Poirier *et al.* found that expression of *aviD*, but not *Dicer*, is lost upon the differentiation of cultured neural stem cells into astrocytes, and a previous study observed decreased vsiRNA abundance upon the differentiation of mouse embryonic stem cells (*2*). Virus infection of differentiated cells induces production of type I and III IFNs, which induce the transcription of hundreds of ISGs to establish an antiviral state. Blocking type I IFN signaling in RNAi-defective mouse fibroblasts can further increase the accumulation of several RNA viruses, suggesting an RNAi-independent and additive contribution of IFN signaling to the innate antiviral defense in differentiated cells (*5*–*9*).

Identification of SARS-CoV-2 as both an inducer and target of antiviral RNAi during authentic infections (*12*) supports ongoing efforts to develop siRNA drugs for the human disease, and a recent study has demonstrated efficacy of an siRNA therapeutic against SARS-CoV-2 in an animal model (*15*). However, it is important to note that mammalian RNA viruses from distinct families produce nonhomologous proteins that are RNAi suppressors, such as influenza viral nonstructural protein 1 and SARS-CoV-2 nucleocapsid protein (*2*–*10*, *14*). Moreover, mammalian vsiRNAs made in host cells to target SARS-CoV-2 and other RNA viruses are all predominantly 22 nucleotides long, in contrast to the synthetic therapeutic siRNAs that are exclusively 21 nucleotides long. Thus, designing and testing siRNA cocktails containing 22-nucleotide species plus additional targeting specificity to VSR mRNA may lead to further improvements in antiviral siRNA therapeutics. ■

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. X. Wu *et al.*, *Cell* **172**, 423 (2018).
2. P. V. Maillard *et al.*, *Science* **342**, 235 (2013).
3. Y. P. Xu *et al.*, *Cell Res.* **29**, 265 (2019).
4. Y. Li, J. Lu, Y. Han, X. Fan, S.-W. Ding, *Science* **342**, 231 (2013).
5. Y. Li *et al.*, *Nat. Microbiol.* **2**, 16250 (2016).
6. Y. Qiu *et al.*, *Immunity* **46**, 992 (2017).
7. Y. Qiu *et al.*, *Sci. Adv.* **6**, eaax7989 (2020).
8. F. Adiliaghdam *et al.*, *Cell Rep.* **30**, 1690 (2020).
9. S. W. Ding, Q. Han, J. Wang, W.-X. Li, *Curr. Opin. Immunol.* **54**, 109 (2018).
10. Q. Han *et al.*, *mBio* **11**, e03278 (2020).
11. J. Witteveldt, L. I. Knol, S. Macias, *eLife* **8**, e44171 (2019).
12. E. Z. Poirier *et al.*, *Science* **373**, 231 (2021).
13. A. Fire *et al.*, *Nature* **391**, 806 (1998).
14. J. Mu *et al.*, *Sci. China Life Sci.* **63**, 1 (2020).
15. A. Idris *et al.*, *Mol. Ther.* 10.1016/j.ymthe.2021.05.004 (2021).

ACKNOWLEDGMENTS

We thank H. R. Ding for critical reading and revision suggestions. S.S. and S.-W.D. were supported by NIH grant AI141887 (to S.-W.D.).

10.1126/science.abj5673

VIEWPOINT: COVID-19

Networks of SARS-CoV-2 transmission

Individual and network heterogeneity should inform respiratory pandemic responses

By **Muge Cevik**[1] *and* **Stefan D. Baral**[2]

The basic reproduction number, R_0 (the number of infections caused by a case in a homogeneously susceptible population), for a particular infection is dependent on the epidemiological triad of the biological characteristics of the pathogen, the environment, and the characteristics of the population (*1*). Even for diseases with similar transmission characteristics, R_0 varies by population owing to differential opportunities for onward transmission according to the contact patterns and the size of the transmission network of an infected individual (*1*). Although transmission can happen in many settings, some factors facilitate a greater risk of infection because of compounded risks often driven by network dynamics (frequent contacts, close proximity, and prolonged contact) and structural-level determinants (such as poverty, occupation, and household size) (*2*–*4*). Understanding drivers of transmission risks and heterogeneity could be used to improve modeling and guide population- and setting-specific mitigation strategies.

In the context of an epidemic, although each contact carries a risk of acquiring an infection, real-world social networks are complex, often exhibiting extreme heterogeneity in the number of contacts, which have large-scale effects on the spread of infection (*5*). In infectious diseases, the population attributable fraction (PAF) represents the total contribution of a risk that could be averted if that risk were avoided (*6*). Even for lower-risk exposures, the PAF could increase with higher exposure frequency mediated through greater numbers of contacts (*2*, *6*). For example, the risk of infection depends on the likelihood of transmission within a particular environment and the frequency at which people visit that setting. At an individual level, settings that are associated with higher-risk factors and visited frequently are likely to pose a higher risk of infection and contribute substantially to cumulative infections than those that may have a higher risk but are visited infrequently. This could mean that a small relative risk of a high-frequency exposure can drive the PAF, suggesting that public health interventions could prioritize resources to eliminate a small risk among many.

However, in reality, risk factors concentrate among the relatively few who have disproportionately higher exposure and onward transmission risks (*2*, *7*). This individual heterogeneity is evident in data, which consistently indicate higher risks of infection due to higher frequency of exposure and multiple contacts (see the figure). In many countries, those working in low-paid and public-facing jobs had the highest risk of being infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*4*). Long-term–care facilities such as nursing homes, homeless shelters, and prisons, as well as workplaces such as meat-packing plants, have been associated with large-scale outbreaks of COVID-19, which were then linked to sustained widespread community transmission (*2*, *8*). These settings often represent environments where risks for infection are compounded and multiple transmission networks intersect (*7*). There is also a clear intersection of COVID-19 risk and socioeconomic inequities, given the network effects of occupation, crowded housing, job insecurity, and poverty (*2*, *4*).

The disproportionate risks associated with network dynamics have also led to differential disease burden (*4*, *9*). According to an analysis from Scotland, patients living in areas with the greatest socioeconomic deprivation had a higher frequency of intensive-care admission and higher COVID-19–related mortality (*10*). Health care units in the most deprived areas also operated over capacity for a more prolonged period (*10*). In a US study, those working in food and agriculture, transportation or logistics, manufacturing, health services, and retail had significantly increased excess mortality related to COVID-19 (*9*). Moreover, differential living and working conditions often manifest as racial disparities because of structural racism. An analysis by the Office for National Statistics highlights the finding that occupations in the UK with higher COVID-19–related death rates include health and social care workers, security guards, drivers, construction workers, cleaners, and sales and retail assistants, which are occupations that also feature higher proportions of minority ethnic groups (*4*). For most occupational categories, the risk ratios comparing mortality during the pandemic with that during nonpandemic time were higher in nonwhite ethnic groups (*4*, *9*).

In addition to heterogeneity in risk of exposure and disease burden, there are also heterogeneities in risk of onward transmission. Per-contact, direct onward transmission risks are driven by multiple factors, including closeness of social interactions, symptom status, the severity of illness, environment, and time of exposure (*2*, *6*). For example, the average per-contact risk is lowest for community exposures, intermediate for social and extended-family contacts, and highest in the household (*11*). Transmission risk is lower when the index case is asymptomatic, increasing with symptom severity (*12*). Indirect onward transmission risks or the total number of downstream infections that stem from an individual over multiple chains of transmission represent important contributions to the overall PAF driven by the size of the transmission networks associated with living and working conditions (*4*, *7*, *13*).

Although some high-frequency contacts are driven by social gatherings, which are modifiable with education and enforcement, most high-risk exposures represent nonmodifiable risks due to living and working conditions (*2*, *3*, *7*). Therefore, risk factors that are nonmodifiable in the short term are likely to represent a much larger PAF than those modifiable by individual choices about social contact. Specifically, the onward transmission risks from someone who can work from home and has enough space for self-isolation, even if they are infected, may be minimal; but the PAF will be higher for someone with a large network associated with working and living conditions (see the figure).

There is now international consensus that those living in the most economically deprived neighborhoods and largest households have an increased risk of infection and disease burden (*3*, *4*). In addition, inequities further concentrate risk through connections between networks. In Toronto, long-term–care staff diagnosed with COVID-19 were disproportionately more likely to reside in neighborhoods with the highest infection rates, which are also the most economically deprived and ethnically concentrated (*14*). In a COVID-19 outbreak investigation among large industries in Ontario, one-third of cases linked to workplace outbreaks spilled over to households, further increasing the burden of

[1]Infection and Global Health Division, School of Medicine, University of St Andrews, St Andrews, UK. [2]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, USA. Email: mc349@st-andrews.ac.uk

illness (*15*). Therefore, structural conditions that affect an individual's network and exposure risk are likely far more predictive than individual choices in determining whether the infection will be a terminal event or lead to multiple downstream infections. Thus, comprehensively addressing the needs of a few with disproportionate risks can avert more downstream infections than eliminating a small risk among many.

How can public health strategies address individual heterogeneity and differential infection risk? Early in the pandemic, there was an assumption of relative homogeneity in the risks of infection and the potential impact of interventions across the population. This was included in modeling to inform public health approaches. Compartmental models, which divide populations into distinct sections and assume that individuals in these groups have the same characteristics, are mostly used to model COVID-19 cases and the impact of interventions. However, they infrequently integrate the effects of differential population mixing, socioeconomic factors, and networks across compartment effects. It is now clear that individual heterogeneity has large-scale effects on disparities seen in the risk of infection and disease burden, which is confirmed in network-based disease modeling (*1*, *6*, *7*, *11*). Public health policies implemented based on the assumption of equal risk of acquisition and transmission across all socioeconomic groups, ages, and occupations left certain communities exposed to a higher risk of infection, resulting in differential burdens of disease (*1*–*3*, *7*). Leveraging network heterogeneity in infectious disease models may better demonstrate these differential risks observed in real-life epidemiological analyses and the benefits of prioritizing intensive and targeted interventions to those with differential risks, given the potential for larger numbers of averted downstream infections.

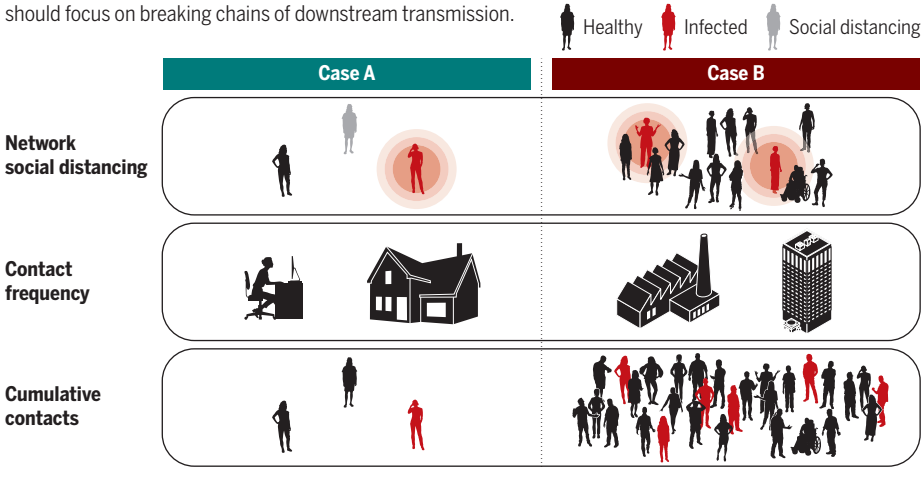
The intersection between direct and indirect onward transmission risks reinforces the need for effective and pragmatic strategies to break chains of transmission, especially in people at high risk of infection. Policy interventions should consider the overall number of contacts that a person has and, subsequently, downstream infections averted based on differential impacts in different communities (*13*). For example, people living in multigenerational households, serving in high-exposure occupations, and residing in densely populated communities could be prioritized for temporary housing support, assurance of employee benefits such as paid leave, and vaccine outreach services. Considering this differential impact, targeted interventions through network-adaptive resource-based interventions could be leveraged according to individual and network-level needs. Such an adaptive approach could inform modeling and prioritize specific resource-based intervention strategies, including testing aligned with lived realities, housing support if insufficient space to isolate, and paid leave from work to support quarantine and isolation, combined with outreach testing and infection prevention and control support in workplaces. In addition, network-adaptive vaccination strategies prioritize those with large networks based on contact heterogeneity. For example, looking at vaccination rates in England by deprivation, vaccination coverage is clearly lower in more deprived areas, where the risk of infection and disease burden is higher. Although this may be due to multiple reasons, including lack of access to care and inability to take time off work, there is a need for areas of high and enduring transmission risk to accelerate vaccination to match the increased risks of infection and onward transmission.

The focus of COVID-19 response strategies has often been on behavior change as a primary means of decreasing contact networks and thus transmission chains. However, contact patterns are driven, in large part, by socioeconomic inequities and structural racism and are nonmodifiable at the individual level in the absence of specific support. Thus, nonadaptive public health interventions fail to address individual heterogeneities and have left socioeconomically marginalized communities at risk of infection, death, and economic hardship. There is a risk that lower vaccine uptake among these communities could perpetuate existing inequalities. Therefore, it is vital that community-led vaccine delivery strategies are strengthened.

The disparities that have defined COVID-19 epidemiology could have been readily predictable given historical data on pandemics. The next respiratory pandemic will also be defined by similar disparities. Using network-driven strategies to inform rapidly emergent epidemic responses represents an evidence-based and equitable path forward where the aim is to invest more to prevent infections in a person with disproportionate risks because disease burden and downstream infection risks vary substantially (*6*, *7*, *13*). ■

Downstream infection risks vary according to network patterns

Case A depicts a person with a small network, who can work from home and self-isolate if needed. Case B represents a person who works in a public-facing job or in an unsafe workplace and lives in a multigenerational or large household. Overall risk of exposure and onward transmission risk differ substantially between these two individuals, representing a disproportionately high transmission chain in case B. Intervention strategies should focus on breaking chains of downstream transmission.



GRAPHIC: K. FRANKLIN/*SCIENCE*; (ILLUSTRATIONS OF PEOPLE) PROPUBLICA'S WEEPEOPLE FONT

REFERENCES AND NOTES

1. T. D. Hollingsworth, J. *Public Health Policy* **30**, 328 (2009).
2. M. Cevik, J. L. Marcus, C. Buckee, T. C. Smith, *Clin. Infect. Dis*. 10.1093/cid/ciaa1442 (2020).
3. M. E. Sundaram *et al*., medRxiv 10.1101/2020.11.09.20223792 (2020).
4. Public Health England (PHE) Transmission Group, "Factors contributing to risk of SARS-CoV2 transmission in various settings" (PHE, 2020); https://bit.ly/3xIbuMM.
5. A. Goyal, D. B. Reeves, E. F. Cardozo-Ojeda, J. T. Schiffer, B. T. Mayer, *eLife* **10**, e63537 (2021).
6. S. Mishra, S. D. Baral, *Lancet Infect. Dis*. **20**, 155 (2020).
7. S. Mishra, J. C. Kwong, A. K. Chan, S. D. Baral, *CMAJ* **192**, E684 (2020).
8. J. E. Lemieux *et al*., *Science* **371**, eabe3261 (2021).
9. Y.-H. Chen et al., medRxiv 10.1101/2021.01.21.21250266 (2021).
10. N. I. Lone *et al*., *Lancet Reg. Health Eur*. **1**, 100005 (2021).
11. K. Sun *et al*., *Science* **371**, eabe2424 (2021).
12. X. Qiu *et al*., *Clin. Microbiol. Infect*. **27**, 511 (2021).
13. B. G. Link, J. Phelan, J. *Health Soc. Behav*. **1995**, 80 (1995).
14. COVID-19 Advisory for Ontario, Update on COVID-19 projections: Ontario COVID-19 science advisory table and modelling consensus table (2020); https://bit.ly/3wRzHjY.
15. M. Murti *et al*., medRxiv 10.1101/2020.11.25.20239038 (2020).

10.1126/science.abg0842

RETROSPECTIVE

Richard Ernst (1933–2021)

Chemist who made magnetic resonance a powerful tool

By **Geoffrey Bodenhausen**

Richard Ernst, professor at the Swiss Federal Institute of Technology in Zurich (ETH Zurich), Nobel laureate, and groundbreaking scientist, died on 4 June. He was 87. By working at the interfaces between chemistry, physics, and biology, Ernst opened the door to magnetic resonance of materials, proteins, and nucleic acids. His work on magnetic resonance imaging (MRI) revolutionized medical diagnosis.

Born on 14 August 1933 and raised in Winterthur near Zurich, Switzerland, Ernst graduated with a diploma (BA) in chemistry from ETH Zurich, followed by a PhD from the same institution under the supervision of Hans Primas and Hans Günthard. Except for a few years with Varian Associates in Palo Alto, California, he spent his entire career as a member of the ETH Zurich faculty. He had a rare talent for picking graduate students and postdoctoral associates, many of whom now hold faculty positions.

In 1945, physicists Felix Bloch at Stanford University and Edward M. Purcell at Harvard University discovered nuclear magnetic resonance in bulk matter by observing very weak signals that emanate from the nuclei of atoms, which possess magnetization like the needle of a compass. Ernst transformed magnetic resonance from an academic curiosity into a universal tool by making it possible to observe all nuclei in a molecule or a human body simultaneously rather than sequentially. The resulting signals can be analyzed by Fourier transformation, a mathematical operation that retrieves with clarity all frequencies in seemingly inextricable signals. Ernst developed methods to unravel signals that overlap in several dimensions, to correlate frequencies of different nuclei in a molecule, and to separate different spatial dimensions in a body. Simultaneous observation of all nuclei drastically improved the sensitivity of magnetic resonance. Information-rich images that are generated allow researchers and radiologists to observe all magnetic nuclei in a molecule or human body.

Ernst's breakthrough applies to work in a variety of fields. In biology, simultaneous observation of magnetic resonance signals allows the structure and dynamics of biomolecules such as proteins and nucleic acids (DNA or RNA) to be determined in solution as well as in microcrystals, fibrils, sediments, and other solids. Magnetic resonance not only circumvents the need for obtaining crystals for diffraction but also opens the way to the study of the flexibility and structural adaptability of biomolecules. In materials science, simultaneous observation of magnetic resonance signals has been exploited with great success in batteries, catalysts, polymers, and porous materials of all sorts, as improved sensitivity makes it possible to determine the arrangement of atoms and ions in space and their ability to move around.

Ernst was honored as the sole recipient of the Nobel Prize in Chemistry in 1991, which coincided with the extravagantly celebrated 90th anniversary of the death of Alfred Nobel and the 700th anniversary of the Swiss Confederation. Upon his return to Switzerland from Stockholm, Ernst was celebrated like a national hero, with spectacular fireworks and many television and radio interviews. He had mixed feelings about his award and the related celebrations, as he was well aware that others had paved the way for his success: Paul Lauterbur and Peter Mansfield, who had made images of humans, and Weston Anderson, Jean Jeener, Ray Freeman, Kurt Wüthrich, and others, who had contributed to obtaining structures of molecules. Three of these scientists would be awarded Nobel Prizes in later years, but Ernst was disappointed that the others did not receive proper recognition.

Ernst also received the Marcel Benoist Prize in Switzerland, the Wolf Prize in Israel, and the first of more than a dozen honorary doctorates from the Swiss Federal Institute of Technology in Lausanne. His autobiography, *Nobelpreisträger aus Winterthur* (Nobel laureate from Winterthur), richly illustrated with works of Tibetan art that he collected, was published in German in 2020. An English translation will be released soon.

Ernst's influence went far beyond the field of physical chemistry. He was for many years one of the most cited Nobel laureates. As a guest speaker at universities around the world, especially after his retirement in 1998, he deftly associated science with other concerns, including Tibetan art and Buddhism. In his talks, he also expressed a biting criticism of university education, which he believed should focus on instilling a sense of responsibility beyond mere knowledge and technical skills, and he sharply denounced inequalities and injustice. His fiery speeches against the evils of US imperialism earned him the enmity of his critics.

The richness of Ernst's career reflects the complexity of the man. Always on the lookout for new methods to probe the secrets of matter, he also sought to unravel the mysteries of music and art. Even though he had given up playing Bach's cello suites because "his fingers did not obey his ear," he continued to study intricate works of 18th-century composers. He also explored the Buddhist painting schools of Tibetan monasteries as well as the pantheon of deities represented on *thangkas* that he ardently collected.

I began working with Ernst as a postdoc in 1980. His support helped me earn the title of privatdozent before departing ETH Zurich in 1985. While there, I helped him complete a book on the principles of nuclear magnetic resonance that took many years to write. Owing in large part to Ernst's demanding sense of breadth and farsighted vision, the book continues to be widely cited as a reference for practitioners of magnetic resonance, even decades after its publication.

Along with his students and other collaborators, I appreciated Ernst's rigor, tinged with a certain stiffness; his incorruptible intransigence; and his oscillations between moods of arrogance and humility, forever unsatisfied and in search of balance. Like the spirits transmitted by reincarnation according to the Tibetan traditions that he admired so much, the spirit of this great scientist will undoubtedly enrich future generations. ■

Department of Chemistry, Ecole Normale Supérieure–Paris Sciences et Lettres, Paris, France.
Email: geoffrey.bodenhausen@ens.psl.eu

10.1126/science.abj9824

PHOTO: ETH-BIBLIOTHEK ZÜRICH, BILDARCHIV/FOTOGRAF: UNBEKANNT/PORTR_13862/CC BY-SA 4.0

POLICY FORUM

RESEARCH POLICY

ARPA-H: Accelerating biomedical breakthroughs

A DARPA-like culture at NIH can drive biomedical and health advances

By **Francis S. Collins[1], Tara A. Schwetz[1,2], Lawrence A. Tabak[1], Eric S. Lander[2]**

The biomedical research ecosystem has delivered advances that not long ago would have been inconceivable, exemplified by highly effective COVID-19 vaccines developed by global partners and approved in less than a year. The United States stands at a moment of unprecedented scientific promise and is challenged to ask: What more can we do to accelerate the pace of breakthroughs to transform medicine and health? Toward that end, President Biden recently proposed to create a new entity, the Advanced Research Projects Agency for Health (ARPA-H), within the National Institutes of Health (NIH) "to develop breakthroughs—to prevent, detect, and treat diseases like Alzheimer's, diabetes, and cancer," requesting $6.5 billion in the fiscal year 2022 budget (*1*). The idea is inspired by the Defense Advanced Research Projects Agency (DARPA), which follows a flexible and nimble strategy, undeterred by the possibility of failure, and has driven breakthrough advances for the Department of Defense (DOD) for more than 60 years. To design ARPA-H, it is critical to understand what is working well within the biomedical ecosystem, where there are crucial gaps, and the key principles of DARPA's success.

WHEN IDEAS DON'T FIT MECHANISMS

Progress in medicine and health in recent decades has been driven by two powerful forces: pathbreaking fundamental research and a vibrant commercial biotechnology sector. Fundamental research is typically performed in university, nonprofit, and government labs. In the United States, it is mostly funded by the federal government, largely through the NIH. By steadily pursuing important fundamental questions in biology and medicine, scientists have made great progress in discovering the molecular and cellular mechanisms underlying health and disease—often suggesting new ideas for clinical treatment. Such fundamental research is what economists term a public good, in that it produces knowledge available to everyone and thus requires public investment. Some have estimated that every dollar of federal investment yields at least $8 in economic growth, and suggested that every new therapeutic approved by the US Food and Drug Administration (FDA) can be traced, in part, to fundamental discoveries supported by NIH (*2, 3*). Given its outsized impact, robust federal investment in fundamental research remains crucial to health and to the economy.

The commercial sector is largely focused on research, development, and marketing of specific products, to bring sophisticated therapies and devices to patients. Biotechnology companies have access to abundant capital to develop products—provided they can protect their intellectual property and recoup the costs by generating sufficient profit in a short enough period of time. Currently, more than 8000 medicines are in development, including 1300 for cancer (*4, 5*).

In many cases, these two components are all that is needed to drive progress toward clinical benefit—though subsequent regulatory approvals, reimbursement, and adoption in health care systems can also be optimized. It's becoming clear, though, that some of the most innovative project ideas, which could yield breakthroughs, don't always fit existing support mechanisms: NIH support for science traditionally favors incremental, hypothesis-driven research, whereas business plans require an expected return on investment in a reasonable time frame that is sufficient to attract investors. As a result, some of the most promising ideas may never mature, representing substantial lost opportunity.

Bold ideas may not fit existing mechanisms because (i) the risk is too high; (ii) the cost is too large; (iii) the time frame is too long; (iv) the focus is too applied for academia; (v) there is a need for complex coordination among multiple parties; (vi) the near-term market opportunity is too small to justify commercial investment, given the expected market size or challenges in adoption by the health care system; or (vii) the scope is so broad that no company can realize the full economic benefit, resulting in underinvestment relative to the potential impact. Evaluations by companies also may not consider the impact of projects on inequities that persist in our health ecosystem. In short, projects with a potentially transformative impact on the ecosystem may not yet be economically compelling or sufficiently feasible for a company to move forward. At the same time, there are no public mechanisms to propel these public goods at rapid speed.

Many such bold ideas involve creating platforms, capabilities, and resources that could be applicable across many diseases. Whereas most NIH proposals are "curiosity-driven," these ideas are largely "use-driven" research—that is, research directed at solving a practical problem.

DARPA AS AN INSPIRATION

DARPA was launched in the wake of Sputnik with a singular mission: to make pivotal investments in breakthrough technologies for national security. DARPA has played a key role in generating bold advances that have shaped the world—such as the internet, Global Positioning Systems, and self-driving cars—and has contributed to the development of many others, including messenger RNA vaccines. However, failure, especially failing early, and learning from that failure are also hallmarks of DARPA.

DARPA has a distinctive organization and culture that contrasts with traditional approaches in biomedical research. It is a flat and nimble organization whose work is driven by approximately 100 program managers (PMs) and office directors. The PMs are often recruited from industry or top research universities, and they come for limited terms of 3 to 5 years. They typically bring bold, risky ideas, and they are given the independence and sufficient resources to pursue them, mitigating risk through metric-driven accountability and by pursuing multiple approaches to achieve a quantifiable goal.

DARPA can support research at three stages (basic research, applied research, and advanced technology development);

[1]National Institutes of Health, Bethesda, MD 20892, USA. [2]Office of Science and Technology Policy, Executive Office of the President, Washington, DC 20502, USA. Email: eric.s.lander@ostp.eop.gov

can fund efforts in multiple sectors (industry, university, national labs, and consortia across these sectors); can provide the critical mass of funding needed to tackle bold goals; and is empowered to promote collaboration and integration across performers. DARPA does not perform its own internal research. Although proposals are reviewed on a competitive basis, PMs have authority to select a portfolio of projects intended to achieve a particular program goal.

DARPA has long encouraged a culture that values a relentless drive for transformative technical results and a willingness to take risks. Notably, it does not focus on merely accelerating ordinary products to the market or making incremental progress, but on creating true breakthroughs. To act in this way, DARPA makes broad use of flexible hiring, procurement, and contracting authorities, provided by law.

Although DARPA is an excellent inspiration for ARPA-H, it is not a perfect model for biomedical and health research. It serves the needs of a single customer, the DOD, and its mission is focused on national security. Its projects typically involve engineered systems. By contrast, health breakthroughs (i) interact with biological systems that are much more complex and more poorly understood than engineered systems, requiring close coupling to a vast body of biomedical knowledge and experience; (ii) interact with a complex world of many customers and users—including patients, hospitals, physicians, biopharma companies, and payers; (iii) interact in complex ways with human behavior and social factors; and (iv) require navigating a complex regulatory landscape. ARPA-H can learn from DARPA but will need to pioneer new approaches.

DARPA-LIKE APPROACHES AT NIH

NIH has some experience with running large, complex programs using DARPA-like approaches to drive highly managed, use-inspired, breakthrough research. A classic example was the Human Genome Project, aimed at reading out the complete 3 billion–nucleotide human genetic code. When the project began in 1990, the technology to accomplish the goal hadn't been invented. By driving innovation, it was completed ahead of schedule and ultimately decreased the cost of sequencing a human genome from $3 billion at the outset to $500 today (*6*). Though estimates vary, it is clear that the overall economic return on investment has been enormous, with notable analyses estimating a nearly 180-fold return (*7, 8*).

A very recent example is the NIH's response to the COVID-19 pandemic. Within weeks, NIH created two programs. The Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) program is an unprecedented partnership with government, industry, nonprofits, and academia to drive preclinical and clinical therapeutics, developing master protocols for testing prioritized compounds in rigorous randomized clinical trials. These efforts accelerated the development and testing of several of the vaccines that are now being widely used. The Rapid Acceleration of Diagnostics (RADx) program used an "innovation funnel" approach to identify promising ideas for COVID-19 tests and support 32 new technology platforms that collectively are contributing 2 million tests per day, mostly at point of care (*9*).

Although these programs have been successful, they required bespoke solutions and herculean efforts to get them off the ground. Because NIH lacks a regular framework for such projects, many bold ideas are hard to realize. That's where ARPA-H can help.

ARPA-H MISSION

ARPA-H should have a clear mission. Building on DARPA's mission statement, an initial mission could be: "To make pivotal investments in breakthrough technologies and broadly applicable platforms, capabilities, resources, and solutions that have the potential to transform important areas of medicine and health for the benefit of all patients and that cannot readily be accomplished through traditional research or commercial activity."

Notably, ARPA-H's focus should be broad—ranging from molecular to societal—because breakthrough technologies are needed and are possible at many levels (see the box). When President Biden challenges researchers to "end cancer as we know it," many basic scientists naturally think about solutions at the laboratory bench: powerful ways to enlist DNA and RNA readouts, genetic regulation, novel chemistry, and the immune system to prevent, detect, and treat cancers. Tech-

Examples of potential projects that ARPA-H could drive

The Advanced Research Projects Agency for Health (ARPA-H) will have a broad focus, and these projects are meant to illustrate the breadth of potential projects that it could support.

Cancer and other chronic diseases

- Vaccines that can prevent most cancers. Use messenger RNA vaccines to teach the immune system to recognize 50 common genetic mutations that drive cancers, so that the body will wipe out cancer cells when they first arise.
- New manufacturing processes to create patient-specific T cells to search and destroy malignant cells, decreasing costs from $100,000s to $1000s to make these therapies widely available.
- Molecular "zip codes" that target a drug or gene therapy vector to any specific tissue and cell type, to make treatments much more effective by treating diseases at their source and eliminating side effects due to impacts on other tissues or cells.
- Small, highly accurate, inexpensive, nonintrusive, wearable 24/7 monitors (e.g., smart watches) for blood pressure and blood sugar.
- New approaches to accelerate discovery of brain imaging and blood biomarkers capable of measuring synaptic loss, neuronal death, and glial inflammatory pathways, as a means of tracking responses to potential Alzheimer's disease therapies.

Infectious diseases

- Ability to design, test, and approve a vaccine against any newly emerging human virus in 100 days.
- Ability to administer vaccines through a skin patch or oral spray to allow rapid, massive vaccination campaigns.

Health care access, equity, and quality

- Platforms to reduce health disparities in maternal morbidity and mortality, which are among the highest in the world, by identifying those at highest risk for pregnancy complications and providing ethically integrated, regular virtual house calls by nurses and midwives, from early in pregnancy through at least 6 months postpartum.
- Platforms to promote better health outcomes through substantially improving how medication is taken, as recommended, on a regular basis or over a standard course (e.g., for hypertension, diabetes, or infections), by engaging community health workers aided by privacy-preserving smart devices and telehealth.

nologists think about new sensors and artificial intelligence–assisted medical decision-making. As importantly, though, there are also opportunities for highly impactful breakthroughs at the macro level to ensure equity in health care access and health outcomes for all patients. Equity considerations (including race, ethnicity, gender/gender identity, sexual orientation, disability, and income level) must be woven throughout the ARPA-H mission—with some projects focused directly on addressing equity and all projects considering equity in their design. Breakthroughs aimed at the most vulnerable groups are not only just and necessary; they will likely improve care for all patients.

ARPA-H's mission will clearly be different from the mission of the existing NIH Institute and Centers (ICs). For example, the name and mission of the National Center for Advancing Translational Sciences (NCATS), an NIH institute created in 2011, might suggest some overlap. However, NCATS' primary focus is to support a national network of clinical research centers and a drug screening hub. These two programs account for nearly 90% of its resources. A modestly sized component within NCATS, the Cures Acceleration Network, is aligned with the general directions of ARPA-H.

Similarly, the NIH Common Fund, a program created by law in 2007, is aimed at a different goal from ARPA-H's use-driven objective: It supports programs to explore new areas of foundational research that cut across multiple ICs—for example, the human microbiome effort. ARPA-H would also be distinct from other existing agencies, such as the Biomedical Advanced Research and Development Authority (BARDA), which focuses on medical countermeasures for public health security threats.

DESIGNING ARPA-H: A DISTINCT DIVISION, CULTURE, AND ORGANIZATION AT NIH

ARPA-H should be housed as a division within NIH, rather than being a stand-alone entity, for two reasons. First, the goals of ARPA-H fall squarely within NIH's mission ("to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability"). Second, ARPA-H will need to draw on the vast range of biomedical and health knowledge, expertise, and activities at NIH. Setting up ARPA-H within NIH will ensure scientific collaboration and productivity and avoid unproductive duplication of scientific and administrative effort.

It is important to acknowledge, however, that a DARPA-like approach is radically different from NIH's standard mechanisms of operation and will require a new way of thinking. The creation of ARPA-H will benefit from transparency, accountability, and a healthy skepticism to ensure that the entity does not become a typical NIH institute.

Taking many features from the DARPA model, ARPA-H needs to have a distinctive culture, organization, authorities, leadership, and autonomy (*10, 11*). ARPA-H's organization should be flat, lean, and nimble. The culture should value bold goals with big potential impact over incremental progress. The organization should lure a diverse cohort of extraordinary PMs from industry or leading universities, for limited terms, with the chance to make a huge impact. They should be empowered to take risks, assemble portfolios of projects, make connections across organizations, help clear roadblocks, establish aggressive milestones, monitor progress closely, and take responsibility for the project's progress and outcomes. Projects should be bounded in time, typically a few years, with longer periods allowed for efforts that are highly complex. ARPA-H should expect that a sizable fraction of its efforts will fail; if not, the organization is being too risk-averse. The best approach is to fail early in the process, by addressing key risks up front. To determine which risks should be taken and to evaluate proposed programs and projects, ARPA-H should adopt an approach similar to DARPA's "Heilmeier Catechism," a set of principles that assesses the challenge, approach, relevance, risk, duration, and metrics of success (*12*).

The ARPA-H director should have substantial authority and independence to act. To keep the entity vibrant, the director should typically serve a single term of 5 years, with the possibility of a single extension in rare cases. For ARPA-H to accomplish its goals, it will need to be provided by Congress with certain authorities parallel to those provided to DARPA, including the authority to recruit, attract with competitive pay, and quickly hire for a set term extraordinary PMs. Unlike DARPA's focus on a single customer, ARPA-H will need to create breakthrough innovations that serve an entire ecosystem and all populations. ARPA-H should have a senior leader responsible for ensuring that issues of equity are considered in all aspects of ARPA-H's work—from scientific program development to staff recruitment and hiring.

Within the Department of Health and Human Services, it will be important for ARPA-H to collaborate with other key agencies such as the FDA, the Centers for Disease Control and Prevention, BARDA, and the Centers for Medicare and Medicaid Services—to identify critical needs and opportunities and to partner on complex projects that interact, for example, with public health infrastructure or medical regulation. DARPA should also play a role in advising ARPA-H on its experiences in driving breakthrough innovation and collaborating on specific projects of shared interest. And it would be valuable to engage science-based agencies and departments, such as the National Science Foundation, the National Institute of Standards and Technology, and the Department of Energy.

It will be critical for ARPA-H to engage with the broader biomedical community, including patients and their caregivers, researchers, industry, and others, to understand the full range of problems and the practical considerations that need to be addressed for all groups and populations.

The potential opportunity is extraordinary. Through bold, ambitious ideas and approaches, ARPA-H can help shape the future of health and medicine by transforming the seemingly impossible into reality. The time to do this is now. ■

REFERENCES AND NOTES

1. *Remarks by President Biden in Address to a Joint Session of Congress* (2021), www.whitehouse.gov/briefing-room/speeches-remarks/2021/04/29/remarks-by-president-biden-in-address-to-a-joint-session-of-congress/.
2. A. A. Toole, *J. Law Econ.* **50**, 81 (2007).
3. E. Galkina Cleary, J. M. Beierlein, N. S. Khanuja, L. M. McNamee, F. D. Ledley, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2329 (2018).
4. G. Long, "The Biopharmaceutical Pipeline: Innovative Therapies in Clinical Development" (The Pharmaceutical Research and Manufacturers of America, 2017).
5. Pharmaceutical Research and Manufacturers of America ,"Medicines in Development for Cancer 2020 Report" (2020).
6. National Human Genome Research Institute, "DNA Sequencing Costs: Data" (2020); www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
7. S. Tripp, M. Grueber, "The Economic Impact and Functional Applications of Human Genetics and Genomics" (American Society of Human Genetics, 2021).
8. "The Impact of Genomics on the U.S. Economy" (Batelle Technology Partnership Practice, for United for Medical Research 2013).
9. National Institute of Biomedical Imaging and Bioengineering, "RADx diversifies COVID-19 test portfolio with four new contracts, including one to detect variants" (2021); www.nibib.nih.gov/news-events/newsroom/radx-diversifies-covid-19-test-portfolio-four-new-contracts-including-one-detect-variants.
10. A. Prabhakar, "How to Unlock the Potential of the Advanced Research Projects Agency Model" (Day One Project 2021).
11. R. E. Dugan, K. J. Gabriel, in *Harvard Business Review* (Harvard Business Publishing, 2013).
12. Defense Advanced Research Projects Agency, "The Heilmeier Catechism" (2021); www.darpa.mil/work-with-us/heilmeier-catechism.

ACKNOWLEDGMENTS

The authors thank R. Fleurence and A. Hallett for helpful input.

Published online 22 June 2021
10.1126/science.abj8547

Although the practice falls into a legal gray area, opium poppies are grown by many gardeners.

HEALTH AND MEDICINE

Medicinal plants, in context

A journalist's meandering meditation probes a trio of mind-altering natural compounds

By **Kiki Sanford**

Michael Pollan's *This Is Your Mind on Plants* combines the author's long-held interest in the natural world with his growing fascination with the workings of the human mind. In the book, he investigates three plant-derived compounds with vastly different stories and effects on human consciousness: the sedative opium, the stimulant caffeine, and the hallucinogen mescaline.

Pollan weaves together three separately engaging stories in a pleasantly meandering style, deftly using his personal experiences with each compound as a jumping-off point for small forays into anthropology, history, politics, psychology, molecular biology, and neuroscience. Even the most distracted reader will come away with an understanding of the physical effects of the spotlighted substances as well as their cultural significance.

Pollan's own garden plays a central role in the book's first section, where readers learn that he (like many other gardeners) has grown opium poppies. For Pollan, growing the gorgeous blooms inspired his current quest to learn more about medicinal plants.

The book's opium chapter is both a telling and a retelling. From 1996 to 1997, Pollan worked on an article about growing the illicit flower that was eventually published in *Harper's Magazine*. Potentially incriminating details, including Pollan's recipe for poppy tea, were left out, however, for legal reasons. *This Is Your Mind on Plants* contains the unredacted article and includes a discussion of America's opiate addiction epidemic, interstitial explanations, and Pollan's present-day reflections on the experience.

Did deleting some of the story's details protect Pollan from litigation? He has no way of knowing, but returning to it now leads him to reflect that "whatever the DEA was thinking in 1996 and '97, the government missed the real story about opium, as in fact did I."

The section of the book that deals with caffeine describes Pollan's struggle to go without the stimulant for several months. Many who drink coffee or tea for its reliable jolt of energy will understand the difficulty inherent in that endeavor. "The day's first cup of tea or coffee acquires most of its power—*its joy!*—not so much from its euphoric and stimulating properties than from the fact that it is suppressing the emerging symptoms of withdrawal," he notes.

The rest of this chapter drifts like an unfocused (uncaffeinated?) mind between Pollan's personal musings and fascinating tales from the colonialist history of England, as he relates his thesis that caffeine—for better or worse—made capitalism what it is today.

The final section of the book details Pollan's interest in mescaline, another compound produced in his garden in the form of a San Pedro cactus. Here, he takes great care with his subject, as he ponders the issue of cultural appropriation with respect to this particular compound, which has long played a vital role in Indigenous American cultures.

Pollan details numerous interactions with individuals from Native American tribes who are working to preserve the slow-growing peyote plant—which he also acquires during the course of his research—from which mescaline is extracted for religious rituals. Such interactions lead him to contemplate his own mission. "This puts the eating of peyote by white people in a long line of nonmetaphorical takings from Native Americans," he writes. "I was beginning to see that, for someone like me, the act of *not* ingesting peyote may be the more important one."

Pollan goes on to describe his experiments with non–peyote-derived mescaline. He reports feeling like "a helpless captive of the present moment." And, although at one point he "felt as though things could easily tip over into terror," the sensation subsided enough to allow him to enjoy the remaining hours of intoxication.

This Is Your Mind on Plants
Michael Pollan
Penguin, 2021. 288 pp.

In the end, the introspective reader is left with more questions than answers—"what exactly is a drug?" for example, a question Pollan poses in the book's first paragraph but fails to answer satisfactorily. Why do we demonize the use of some plant compounds and accept others? And why do we humans seek to alter our consciousness in the first place?

Pollan has crafted a narrative that he hopes will influence the stories we tell about plants, their active molecules, and the relationship we have with them as individuals and as a society. But this tome is not a self-help book. Instead of prescribing a list of potential policy actions, he ultimately leaves the details of how we might move forward with plant-derived compounds open to interpretation. ■

10.1126/science.abi7773

The reviewer is a freelance science communicator and host of the podcast *This Week in Science*.
Email: kirsten@thisweekinscience.com

PHOTO: TTSTUDIO/SHUTTERSTOCK

SCIENCE AND SOCIETY

Stepping out of the comfort zone

A writer grapples with the environmental and social costs of mechanical cooling

After Cooling: On Freon, Global Warming, and the Terrible Cost of Comfort
Eric Dean Wilson
Simon and Schuster, 2021. 480 pp.

By **Gabrielle Dreyfus**

In *After Cooling,* Eric Dean Wilson uses the history of mechanical cooling and the associated creation of comfort as a commodity to tell a tale of control, oppression, and destruction. The book is as much sociopolitical commentary as history of science.

Wilson weaves together the story of the invention of mechanical cooling and Freon—the trademarked name for chlorofluorocarbon (CFC) refrigerants—with recollections from a recent road trip with a friend, Sam, whose job it is to find, purchase, and arrange for the destruction of CFCs. The catch? Almost all the sellers are hostile to climate considerations and the concept of destroying something useful. Sam must win the trust of the sellers without revealing that the ozone-depleting super greenhouse gas chemicals will be eliminated.

Readers should be aware that this history is told from a distinctly American perspective, excluding from discussion, for example, former Singapore prime minister Lee Kuan Yew's assertion that air conditioning played a critical role in improving the productivity of the country's civil service (*1*). Wilson digs deeply into the linkages between Western desire for material comfort and racial oppression, including critiques of capitalism and profit-seeking industry. Like others before him, he blames air conditioning for turning Americans inward, away from porches and community, thus contributing to the nation's alienation and insularity. Wilson's personal experiences and thinking, which manifest, at times, in witty parentheticals, are a frequent reminder that there is no pretense of objectivity in this critique.

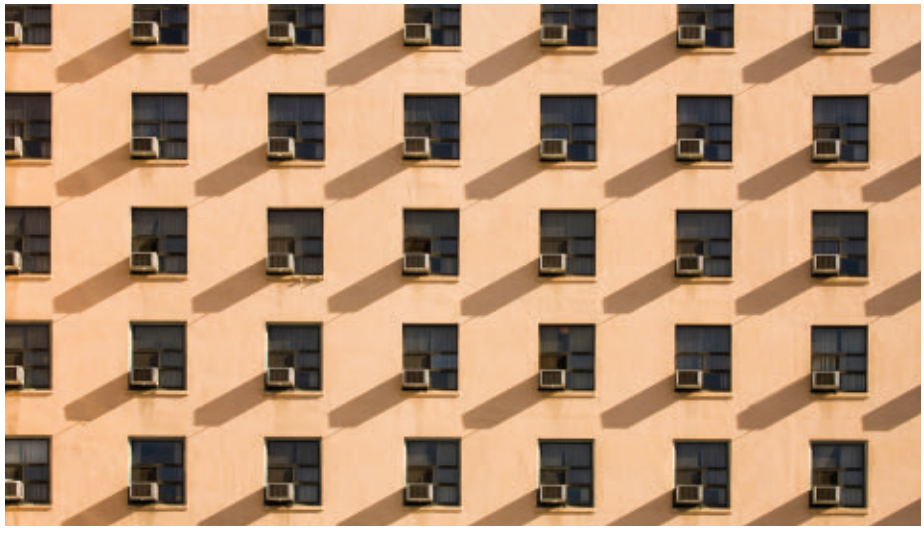
Wilson links air conditioning in cars and homes to Americans' willful ignorance with regard to environmental destruction and the oppression and suffering of others, arguing that the technology impairs our ability to see our interconnectedness. And while he condemns systems of oppression and purveyors of pseudoscientific bigotry, he displays an unanticipated empathy toward General Motors engineer Thomas Midgley Jr., the "two-time environmental loser" who invented both the leaded gasoline that poisoned millions and the CFCs that almost destroyed the ozone layer and destabilized the climate. Rather than dwell on Midgley's denial of the known toxicity of lead, Wilson delves into the engineer's fear of obsolescence, his sudden physical paralysis from polio, and his subsequent suicide.

In doing so, however, Wilson misses an opportunity to describe how close humanity came to calamity. Midgley originally identified eight elements, including fluorine, chlorine, and bromine, as potential candidates for refrigerants (*2*). In his 1995 Nobel acceptance speech, Paul Crutzen noted that if bromine compounds had been chosen, or if chlorine had behaved more like bromine in the stratosphere, "we would have been faced with a catastrophic ozone hole everywhere and at all seasons during the 1970s." "Mankind has been extremely lucky," he concluded (*3*).

Wilson dutifully describes the incredible achievement of the 1987 Montreal Protocol on Substances That Deplete the Ozone Layer, which set the foundation for the phaseout of CFCs and the subsequent recovery of the ozone layer. He criticizes the protocol, however, for limiting its purview to the chemicals and "simple technological fix without addressing any of the underlying, psychosocial errors," specifically "the material conditions for insularity."

One aspect of air conditioning that Wilson fails to discuss is how heat pumps—essentially, air conditioners run in reverse—are now considered by many to be critical to decarbonizing the heating of buildings. Winter heating goes beyond luxury in most of the United States. One is left to wonder whether the same piece of equipment that uses electricity and refrigerants to provide thermal comfort becomes less of a symbol and symptom of society's ills when used for heating.

While making air conditioners more energy efficient and using climate-friendly refrigerants are jointly one of the biggest climate mitigation opportunities available today, such technology fixes do not alone address the need to rethink how we design, build, and live in our cities to achieve greater environmental and social sustainability. *After Cooling* warns of the unintended physical, social, and cultural consequences of seeking to make life more comfortable, easier, and more convenient, urging readers to reflect not just on what might be gained from new technologies but also what might be lost. ■



Air conditioning has amplified America's insularity and willful disregard for the environment, argues Wilson.

PHOTO: LHB PHOTO/ALAMY STOCK PHOTO

The reviewer is at the Institute for Governance and Sustainable Development, Washington, DC 20007, USA. Email: gdreyfus@igsd.org

REFERENCES AND NOTES

1. L. K. Yew, *New Perspect. Q.* **26**, 111 (2009).
2. T. Midgley, *J. Soc. Chem. Industry* **56**, 133 (1937).
3. P. J. Crutzen, "My Life with O_3, NO_x and Other YZO_xs," Nobel Lecture, 8 December 1995.

10.1126/science.abj1417

LETTERS

Edited by **Jennifer Sills**

China's ambitious energy transition plans

China has pledged to achieve peak carbon dioxide emissions by 2030 and carbon neutrality by 2060 (*1*, *2*). Although a national road map has not yet been announced, many cities, companies, and institutions in China have been proposing carbon reduction plans that rely on rapidly transitioning from fossil fuels to other types of energy (*3*, *4*). Such extreme changes could backfire. China's energy plans are most likely to succeed if they are inclusive, gradual, and tailored to the needs of each region.

Radical carbon-neutral targets can lead to unsustainable energy transitions with potentially dangerous unintended consequences. For example, millions of rural residents in China's Hebei province were reportedly left with no heating in the winter of 2017 as a result of a rapid switch from coal to gas in an attempt to reduce pollution (*5*). In mid-December of 2020, factories in China's Zhejiang province were forced to close temporarily to meet unrealistic energy consumption targets (*6*).

China should promote equitable energy transitions that do not leave behind disadvantaged groups or those employed by vulnerable sectors of the economy (*7*, *8*). Energy transition action plans should include strategies to protect the economic and social welfare of the nearly 4 million people who work in fossil fuel mining and many more who work to support fuel supply chains (*9*). Existing policy debates pay insufficient attention to the social and economic challenges involved in transitioning human capital along with energy (*10*).

The energy transition process also needs to be implemented over time rather than in rapid bursts. Energy transition paths should balance human needs and social stability, a lower share of fossil fuels, and energy security. The energy system has strong transition inertia; a radical energy transition path that is not capable of practically and cost-effectively meeting people's basic needs will do more harm than good.

Finally, transition plans need to adapt to regional heterogeneity. Because of the tremendous geographic differences in China, not all regions will be able to achieve energy transitions at the same pace. Regions with more economic development and resources may be able to afford earlier neutrality. Regional heterogeneous targets should be established to account for such differences. Adopting market instruments, such as energy production capacity permits (*11*) and an emissions trading system (*12*), can help to minimize the costs of transitions.

Xunpeng Shi[1], Yongping Sun[2], Yifan Shen[3]*
[1]University of Technology Sydney, Sydney, Australia. [2]Hubei University of Economics, Wuhan, China. [3]Tongji University, Shanghai, China.
*Corresponding author.
Email: shenyifan1989@gmail.com

REFERENCES AND NOTES

1. D. Normile, *Science* **370**, 17 (2020)
2. S. Mallapaty, *Nature* **586**, 482 (2020).
3. M. Walsh, K. Jia, "Chinese provinces home to 250 million plan carbon emission peaks," Caixin (2021).
4. Reuters Staff, "China's top steelmaker Baowu Group vows to achieve carbon neutrality by 2050," *Reuters* (2021).
5. "China does U-turn on coal ban to avert heating crisis," *BBC* (2017).
6. N. Gan, "China turns off the lights in 'Christmas town' as officials race to meet energy targets," *CNN* (2020).
7. F. Green, A. Gambhir, *Clim. Pol.* **20**, 902 (2019).

In 2017, China's Hebei province rapidly shifted from coal to gas, but the resulting energy shortages led to a reversal of the coal ban.

PHOTO: THOMAS PETER/REUTERS

8. S. Carley, D. M. Konisky, *Nat. Energ.* **5**, 569 (2020).
9. National Bureau of Statistics of China, "Communiqué on the Fourth National Economic Census (No. 3)" (2019); www.stats.gov.cn/english/PressRelease/201911/t20191120_1710328.html.
10. G. He *et al.*, *One Earth* **3**, 187 (2020).
11. X. Shi *et al.*, *J. Clean. Prod.* **256**, 120472 (2020)
12. "China's emissions trading scheme," IEA (2020); www.iea.org/reports/chinas-emissions-trading-scheme.

10.1126/science.abj8773

Institutions key to inclusion and equity

In his Letter "Inclusion and equity through STEM training" (28 May, p. 926), M. H. Zaman suggested that better and broader training could increase inclusion and equity in science, technology, engineering, and mathematics (STEM). Training individuals may lead to valuable understanding about workplace inequities, but it is institutional policies, processes, and behaviors—or lack thereof—that allow inequities, harassment, and discrimination to persist. If research and academic institutions are not held accountable, their response to inequities and mistreatment may be inadequate.

Institutions may have offices to deal with allegations of inequity and harassment, but the mere existence of offices is insufficient to ensure fair and effective investigations. In a survey of those who experienced academic bullying, only 8% of individuals who reported abuse thought the process was unbiased and fair (*1*). Even when institutions find wrongdoing, deficient policies (*2*, *3*) and a predisposition toward inaction often result in offenders facing trivial or no sanctions (*4*, *5*). In the case of astronomer Geoff Marcy, allegations first surfaced in 1995, spanning two universities (*6*). Only in 2015, after he was found culpable of sexual harassment, did Marcy resign (*7*). Similarly, it took Harvard University (*5*) and the Salk Institute for Biological Studies (*8*) 40 years to effectively deal with harassment—Harvard acting after the offender retired.

Institutions may be required to report allegation outcomes to funding agencies that can implement sanctions [e.g., (*9*, *10*)], but funder consequences rely on institutions acknowledging wrongdoing by individuals. Institutions failing to do so perpetuate the lack of consequences for perpetrators. Funding agencies and private philanthropies should instigate anonymous reporting mechanisms, including ombudspersons, for institutional inaction. Funding sanctions and fines imposed on institutions

may incentivize institutional reform. Local and national laws should also be amended to ensure that they help rather than hinder expeditious and fair litigation and consequences. For example, effective laws would extend the reporting window and address detrimental incremental trends rather than focusing on isolated incidences.

Opaque processes, delays, and inadequate institutional responses often adversely affect the target rather than the perpetrator (*2*, *11*). If targets suffer retaliation for stepping forward, cultures of fear arise (*8*), with targets often leaving their institutions (*5*) or academia (*11*). Individuals compose institutions, but institutions set the standards for individuals (*11*). Institutional leadership should take ownership and responsibility, holding every individual within the institution accountable for their actions.

Tamar L. Goulet
Department of Biology, University of Mississippi, University, MS 38677, USA.
Email: tlgoulet@olemiss.edu

REFERENCES AND NOTES

1. S. Moss, M. Mahmoudi, "STEM the bullying: An empirical investigation of abusive supervision in academic science" (2021); https://ssrn.com/abstract=3850784.
2. C. Penrod, M. Fusilier, *J. Workplace Rights* **15**, 151 (2010).
3. M. Fusilier, C. Penrod, *Employee Responsibilities Rights J.* **27**, 47 (2015).
4. C. P. Smith, J. J. Freyd, *Am. Psychol.* **69**, 575 (2014).
5. N. Gluckman, "'Harvard failed her': University apologizes to scholar who endured harassment," *Chronicle of Higher Education* (2021).
6. R. Wilson, "Geoff Marcy's downfall," *Chronicle of Higher Education* (2016).
7. J. Kaiser, "Astronomer Geoff Marcy booted from National Academy of Sciences in wake of sexual harassment," *Science* (2021); https://scim.ag/NASMarcy.
8. M. Wadman, *Science* **360**, 480 (2018).
9. National Science Foundation, Welcome to the Office of Diversity and Inclusion (www.nsf.gov/od/odi/index.jsp).
10. National Institutes of Health, Expectations, Policies, and Requirements (2019); https://grants.nih.gov/grants/policy/harassment/policy-requirement.htm.
11. National Academies of Sciences, Engineering, and Medicine, "Sexual harassment of women: Climate, culture, and consequences in academic sciences, engineering, and medicine" (National Academies Press, 2018).

10.1126/science.abj8196

Save China's blue-crowned laughingthrush

The blue-crowned laughingthrush (*Garrulax courtoisi*), a songbird species endemic to Jiangxi Province in central China (*1*), is designated as Critically Endangered on both the International Union for Conservation of Nature Red List (*1*) and the China Species Red List (*2*, *3*). The laughingthrush, which breeds in low-altitude, broad-leaved forests (*4*), was thought to be extinct until it was rediscovered in Wuyuan in 2000 (*5*). In the 20 years since, the population has not substantially grown (*1*). There are now only about 240 individuals in the wild (*1*). Without effective protection, the species may not survive.

Human activities have exacerbated the decline of the blue-crowned laughingthrush population. Trapping for the bird trade led to the species' extinction in Yunnan, with 400 birds reported to have been trapped from 1987 to 1992 (*5*). In Wuyuan, urban development has caused habitat destruction, which forces the birds to relocate to breed (*1*, *6*). Because Wuyuan is known as "China's most beautiful landscape" (*7*), with rich bird diversity (*8*), human tourism activities have also disrupted the bird's reproductive processes (*6*, *9*).

Protecting the bird and its breeding habitats requires urgent and coordinated action from the government and the public. The species was recently updated to class I, the most critical, in the national key protected species list (*10*), a step in the right direction that should raise public awareness about the need for its conservation. Because the laughingthrush depends on patches of remnant forest and old-growth trees near villages for both breeding and nesting (*6*, *9*), the most effective measure might be managing and restoring this habitat and reducing nearby human activities. Given that most breeding sites of the blue-crowned laughingthrush do not fall within the current system of national natural reserves (*6*), China should also implement policies to protect the gaps before it is too late.

Ning Li[1], Xinglong Huang[1], Qi Yan[1], Weiwei Zhang[2]*, Zheng Wang[3]
[1]Institute of Applied Ecology, College of Food Science, Nanjing Xiaozhuang University, Nanjing, Jiangsu, China. [2]Center for Wildlife Resources Conservation Research, Jiangxi Agricultural University, Nanchang, Jiangxi, China. [3]Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing, Jiangsu, China.
*Corresponding author.
Email: zhangweiwei_nefu@163.com

REFERENCES AND NOTES

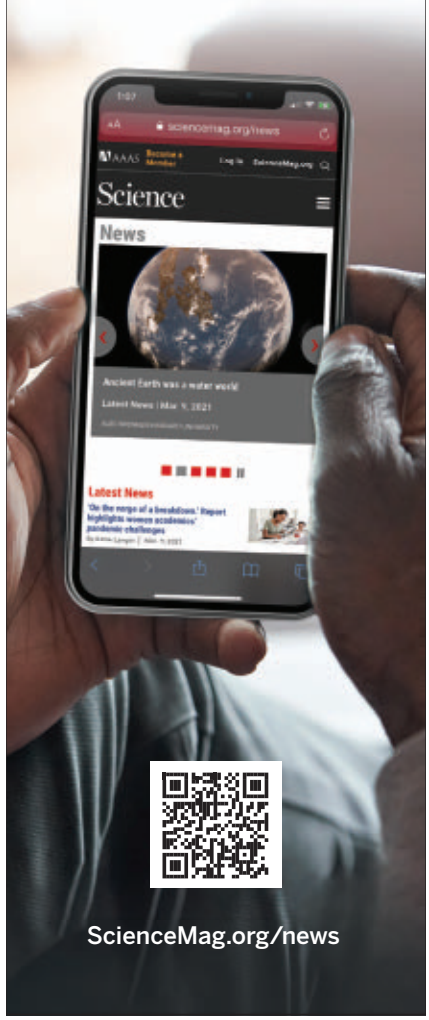
1. BirdLife International, *Garrulax courtoisi* (The IUCN Red List of Threatened Species, 2018).
2. S. Wang *et al.*, *China Species Red List, Vol. 2, Vertebrates* (Science Press, Beijing, 2009) [in Chinese].
3. Y. Y. Zhang *et al.*, *Biodivers. Sci.* **24**, 568 (2016).
4. H. Q. Huang *et al.*, *Acta Ecol Sin.* **38**, 493 (2018).
5. Y. H. Hong, *et al.*, *Oriental Bird Club Bull.* **38**, 35 (2003).
6. W. W. Zhang *et al.*, *Avian Conserv. Ecol.* **12**, 15 (2017).
7. M. Y. Xu, *J. Ser. Sci. Manag.* **13**, 649 (2020).
8. F. Q. He *et al.*, *Chin. J Zool.* **49**, 170 (2014) [in Chinese].
9. T. Liu *et al.*, *Peer J.* **8**, e8785 (2020).
10. National Forestry and Grassland Administration of China, "Official release of the updated list of Wild animals under Special State Protection in China" (2021); www.forestry.gov.cn/main/586/20210208/09540379167571.html [in Chinese].

10.1126/science.abj4535

NEWS FROM Science

Up-to-the-minute research and policy news you won't find in print

Visit us online to read all the news coverage that there just wasn't enough room to print in this issue.



ScienceMag.org/news

PRIZE ESSAY

MICROBIOME

Can microbes combat neurodegeneration?

Identifying a new link between microbiome and metabolites in amyotrophic lateral sclerosis

By **Eran Blacher**

Millions of people worldwide suffer from neurological disorders such as Alzheimer's disease (AD), Parkinson's disease, and amyotrophic lateral sclerosis (ALS). By gradually destroying motor abilities, communication skills, memory, and clear thinking, these devastating diseases rob patients of their independence and take a heavy toll on family members and caregivers.

The exact causes of neurodegeneration remain unclear. Only 10% of ALS cases and 5% of AD cases are familial, whereas the vast majority are of unknown etiology (*1*). In 1993, mutations in the Superoxide dismutase-1 (*Sod-1*) gene were shown to cause ALS and now account for 18.9% of familial ALS cases. Since that discovery, sequencing-based studies revealed additional relevant disease-associated mutations, but limited progress has been made in explaining the molecular mechanisms of neurodegeneration (*2*). With so little causative understanding of neurodegeneration, we must ask: Do environmental factors—such as nutrition, commensal bacteria, and their metabolites—play a role in neurological disorders?

The past decade has witnessed a paradigm shift in brain research. A transition from a dogmatic brain-focused approach toward a holistic conception of health that integrates key signaling hubs of the human body—such as the gut and its microbial populations, the peripheral immune system, and other mucosal barrier surfaces—is increasingly acknowledged as necessary to understand and cure neurodegenerative diseases.

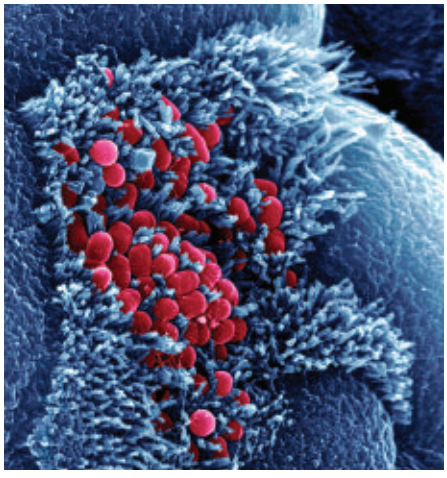
I studied the role of the gut microbiome and its associated molecules in ALS as a postdoctoral fellow in Eran Elinav's laboratory at the Weizmann Institute of Science. The results of this study suggest that gut microbes may secrete small-molecule metabolites that potentially have unexpected regulatory functions in ALS progression both in mouse models and in human patients (*3*).

Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94304, USA.
Email: eranblac@stanford.edu

NERVOUS SYSTEM–MICROBIOME CROSS COMMUNICATION

Recent evidence suggests that the human brain constantly communicates with the gut microbiome—an ecosystem of thousands of bacterial species that inhabit the gastrointestinal tract along a "microbiome-gut-brain axis" (*4, 5*). Cross-talk on this axis can be mediated by small-molecular metabolites secreted by gut bacteria and absorbed into the blood stream. These metabolites can then access the central nervous system through the choroid plexus, where it is believed they reprogram transcriptional responses of brain cells (*6*). The gut microbiome responds quickly to environmental factors and represents a central component in their impact on the host's physiology. Therefore, we hypothesized that gut bacteria influence ALS pathogenesis.

We began our investigation by depleting the microbiome of *Sod1*-transgenic (*Sod1*-Tg) mice through wide-spectrum antibiotic treatment. Microbiome depletion resulted in a substantial exacerbation of ALS symptoms (*3*). We then compared the gut microbiome of antibiotic-treated *Sod1*-Tg mice to that of wild-type littermate controls raised in several specific-pathogen–free facilities. We discovered vivarium-dependent dysbiosis and microbiome-driven alteration in systemic metabolite's configuration that preceded clinical motor symptoms. We then demonstrated that several key bacterial genes that encode for biosynthetic enzymes of nicotinamide and its precursor, tryptophan, were reduced in the microbiome of *Sod1*-Tg mice.



Shown is a colored scanning electron micrograph (SEM) of *Escherichia coli*, one of hundreds of bacterial species residing in the human gut. Research is now revealing cross-talks between those microbes and distant organs, such as the brain, in health and disease.

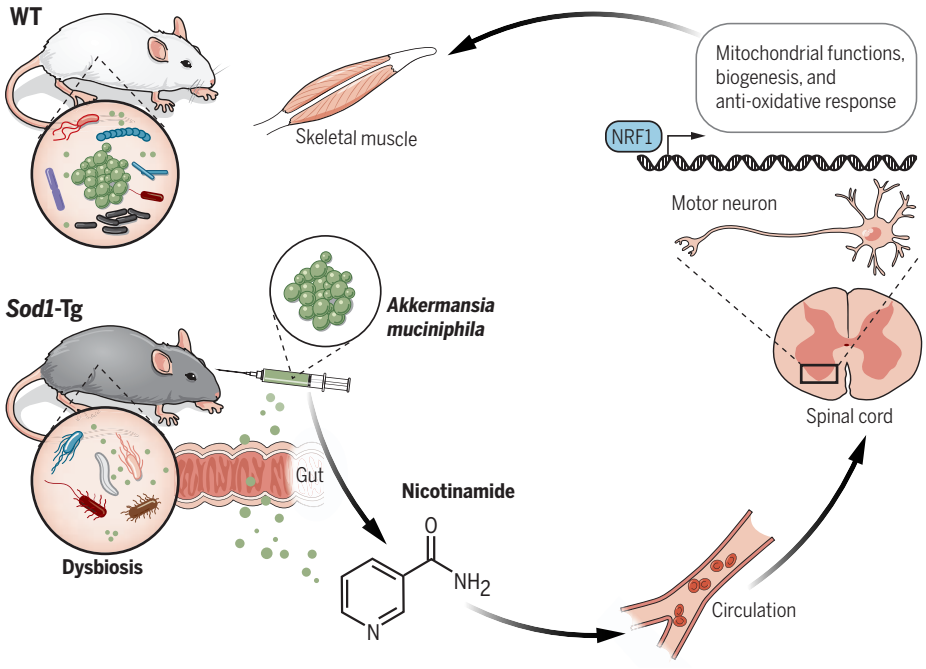
PHOTO: STEPHANIE SCHULLER/SCIENCE SOURCE

DISTINCT MICROBIAL TRANSPLANTATION AMELIORATES MOUSE ALS

Using a comprehensive metagenomic assessment throughout disease progression, we identified 11 distinct microbial strains that were correlated to disease severity. To test their clinical effects on ALS severity, we adopted a "probiotic" approach in which we anaerobically cultured individual strains and administered them to *Sod1*-Tg mice pretreated with antibiotics. Supplementation of these strains demonstrated that *Akkermansia muciniphila* ameliorated whereas *Ruminococcus torques* and *Parabacteroides distasonis* exacerbated ALS symptoms in the mice. We then used metabolomic approaches to characterize bacterial-associated metabolites in the *A. muciniphila*–treated *Sod1*-Tg mice and found that supplementation with the bacterium significantly increased nicotinamide concentrations in the nervous system. Direct administration of nicotinamide, through subcutaneously implanted osmotic pumps, also substantially improved motor abilities and spinal cord gene expression patterns in *Sod1*-Tg mice. These findings highlight nicotinamide as a potential therapeutic agent for ALS. Treating *Sod1*-Tg mice with either the bacterium *A. muciniphila* or with its associated metabolite nicotinamide enriched the expression of neuroprotective genes involved in mitochondrial structure and function, nicotinamide adenine dinucleotide^{+} (NAD^{+}) homeostasis, and removal of superoxide radicals in the spinal cord—functions that are known to be disrupted in ALS (see the figure).

Sod1-Tg ALS mice harbor preclinical dysbiosis

Treatment with *Akkermansia muciniphila*, or with nicotinamide, ameliorates amyotrophic lateral sclerosis symptoms and elicits neuroprotective transcriptional program in the spinal cord. WT, wild type.



CREDITS: (GRAPHIC) ADAPTED FROM E. BLACHER BY N. CARY/*SCIENCE*; (PHOTOS, TOP TO BOTTOM) COURTESY OF NADIA BELKIND; COURTESY OF EREZ BARUCH; MASSIMO DEL PRETE/EMBL

MICROBIOME AND NICOTINAMIDE CHANGE IN ALS PATIENTS

To determine whether our findings could be translated into a potential cure for human ALS, we sequenced the gut microbiome metagenomes of ALS patients and healthy family members that shared the same household environment. This observational study showed that the composition and function of the microbiome of ALS patients substantially differed from that of healthy family members. Moreover, we found a significant reduction in nicotinamide concentrations in both sera and cerebrospinal fluids of ALS patients. We posit that these findings are linked to our previous observations in mice and may lay the foundation for a larger clinical study in the future.

THE FUTURE: MICROBIOME-METABOLOME–BASED THERAPIES?

Harnessing rapidly developing microbiome sequencing, culturing, and computational technologies enabled us to identify a skewed metabolic pathway involved in ALS pathogenesis in mice that is highly affected by the composition and function of the gut microbiome. Similarly, studies performed during my graduate work in the laboratory of Reuven Stein showed that inhibition of CD38, the most efficient NAD^+-consuming enzyme, is a promising strategy to treat brain pathologies (*7*–*10*). Disrupted microbial metabolites profiles may also contribute to neurodegeneration, as we demonstrated in *Sod1*-Tg ALS mice (*3*). These results exemplify how microbiome profiling can be used to identify disease-modifying metabolites.

Further research implementing mass-spectrometry informatics with molecular networking has the potential to reveal the mechanisms behind microbiome-associated phenotypes (*11*, *12*). This approach may pave the way to rationally genetically engineer a transplantable metabolome that would hopefully assist in delaying or even preventing detrimental age-related illnesses. ■

REFERENCES AND NOTES

1. L. Bertram, R. E. Tanzi, *J. Clin. Invest.* **115**, 1449 (2005).
2. G. Kim, O. Gautier, E. Tassoni-Tsuchida, X. R. Ma, A. D. Gitler, *Neuron* **108**, 822 (2020).
3. E. Blacher *et al.*, *Nature* **572**, 474 (2019).
4. K. Berer *et al.*, *Nature* **479**, 538 (2011).
5. T. R. Sampson *et al.*, *Cell* **167**, 1469 (2016).
6. P. Kundu, E. Blacher, E. Elinav, S. Pettersson, *Cell* **171**, 1481 (2017).
7. E. Blacher *et al.*, *Glia* **63**, E460 (2015).
8. E. Blacher *et al.*, *Ann. Neurol.* **78**, 88 (2015).
9. J. Camacho-Pereira *et al.*, *Cell Metab.* **23**, 1127 (2016).
10. A. Levy *et al.*, *Neuro-oncol.* **14**, 1037 (2012).
11. J. M. Gauglitz *et al.*, *mSystems* **5**, e00635-19 (2020).
12. R. A. Quinn *et al.*, *Nature* **579**, 123 (2020).

ACKNOWLEDGMENTS

I wish to thank my supervisor, Eran Elinav, and his wonderful lab members for an inspiring, fruitful, and joyful experience. I would also like to thank my Ph.D. advisor, Reuven Stein, for his incredible support. I am generously supported by the Marie Skłodowska-Curie Global Postdoctoral Fellowships (Grant 888494) and the Stanford School of Medicine Dean's Postdoctoral Fellowship.

10.1126/science.abi9353

GRAND PRIZE WINNER

Eran Blacher

Eran Blacher received undergraduate degrees and a PhD from Tel Aviv University, Israel, and performed a postdoctoral fellowship at the Weizmann Institute of Science, where he studied the role of the microbiome-gut-brain axis in the context of neurodegenerative diseases. He is currently a senior postdoctoral fellow at Stanford School of Medicine studying the immune system–gut–brain axis in aging and neurological disorders.

FINALIST

Erez Baruch

Erez N. Baruch received undergraduate, MD, and PhD degrees from Tel Aviv University, Israel. After completion of his graduate studies, he started an internal medicine residency in a research (Physician-Scientist) track to medical oncology. Internal medicine training is conducted in the McGovern Medical School in Houston, Texas. The research work is conducted in Dr. Jennifer Wargo's lab at the Department of Genomic Medicine, MD Anderson Cancer Center, and is focused on mechanisms of immunotherapy resistance and toxicity, modulation of the gut microbiota, and interaction between innate and adaptive immune cells. **www.sciencemag.org/content/373/6551/173.1**

FINALIST

Maria Zimmermann-Kogadeeva

Maria Zimmermann-Kogadeeva received undergraduate degrees from Lomonosov Moscow State University in Russia and a PhD from ETH Zürich, Switzerland. After completing her postdoctoral fellowships at Yale University in the Goodman group and at European Molecular Biology Laboratory (EMBL) Heidelberg in the Bork group, Maria will start her laboratory in the Genome Biology Unit at EMBL Heidelberg in 2021. Her research combines computational modeling and multi-omics data integration to investigate how microbes adapt to their surroundings and how metabolic adaptations of individual bacteria shape the functional outcome of microbial communities and their interactions with the host and the environment. **www.sciencemag.org/content/373/6551/173.2**

PRIZE ESSAY

FINALIST
Erez Baruch
Erez N. Baruch received undergraduate, MD, and PhD degrees from Tel Aviv University, Israel. After completion of his graduate studies, he started an internal medicine residency in a research (Physician-Scientist) track to medical oncology. Internal medicine training is conducted in the McGovern Medical School in Houston, Texas. The research work is conducted in Dr. Jennifer Wargo's lab at the Department of Genomic Medicine, MD Anderson Cancer Center, and is focused on mechanisms of immunotherapy resistance and toxicity, modulation of the gut microbiota, and interaction between innate and adaptive immune cells.
www.sciencemag.org/content/373/6551/173.1

MICROBIOME

Microbiota modulation to fight cancer

Gut microbiota transplants promote response in immunotherapy-refractory melanoma

By **Erez N. Baruch**[1,2]

Treatment with immune checkpoint blockade therapies that target the programmed death receptor 1 (PD-1) pathway has revolutionized care for patients with melanoma and numerous other cancers. In melanoma, median overall survival rates for patients with metastatic disease have doubled in the anti-PD-1 therapy era (*1*, *2*). Cure of metastatic disease is now a tangible prospect. Complete disappearance of all metastatic lesions is observed in 10 to 20% of patients undergoing anti-PD-1 therapy (*3*), many of whom remain cancer free for years (*4*). Despite this outstanding clinical success, most metastatic patients will experience disease progression, either immediately or after an initial response to treatment (*5*). Today, extensive efforts worldwide are focused on understanding resistance to anti-PD-1 therapy, with the aim of developing new treatments to overcome it.

To better understand resistance mechanisms to anti-PD-1 therapy, we studied tumors from cancer patients who responded to treatment and from those who did not. Prior to treatment, tumor biopsies were collected from melanoma patients at the Sheba Medical Center, Israel. We performed proteomic analyses on these samples. The tumors of anti-PD-1 therapy responders versus nonresponders revealed significant differences in two major functions: metabolism and antigen presentation (*6*). Specifically, we observed that responding tumors exhibited up-regulation of oxidative phosphorylation and lipid metabolism pathways, as well as enhanced antigen presentation machinery pathways.

Emerging data from preclinical models have demonstrated an impact of gut microbes on immunotherapy response (*7*, *8*). We hypothesized that melanoma patients' gut microbiota were affecting both metabolism and antigen presentation in their tumors and turned our efforts to modulating gut microbes to enhance responses to anti-PD-1 therapy in cancer patients.

We decided to study the gut microbes in patients with metastatic melanoma undergoing treatment with anti-PD-1 therapy. We performed 16*S* ribosomal RNA (rRNA) gene sequencing on patients' baseline stool samples and compared the composition of the gut microbes of anti-PD-1 therapy responders and nonresponders. We noted that responders had a higher relative abundance of *Ruminococcus* and *Prevotellaceae* in their gut microbiota, whereas nonresponders had a higher relative abundance of *Betaproteobacteria*.

Concurrent with our studies, a series of independent reports were published in 2018 (*9*–*11*), establishing notable differences in the gut microbiota of responders and nonresponders to anti-PD-1 therapy. These reports also associated *Ruminococcus* and *Prevotellaceae* with a favorable clinical response to immunotherapy. One of the groups, led by my current research mentor, Jennifer Wargo, reported that fecal microbiota transplantation (FMT) leads to enhanced anti-PD-1 effectiveness (*9*). With these findings in mind, we decided to assess the feasibility of conducting FMT in patients with anti-PD-1–resistant metastatic melanoma to improve their clinical response.

We began with the goal of identifying whether the optimal FMT donor was a healthy donor or a melanoma patient who has achieved a durable and profound tumor regression on anti-PD-1 therapy for at least 1 year (a "durable responder"). We studied stool samples from 9 healthy individuals and 10 patients with melanoma who were durable responders to anti-PD-1 therapy. Analysis of the gut microbiota in these cohorts demonstrated that the durable responders had microbiome signatures consistent with those of responders in prior cohorts. Healthy individuals, however, demonstrated substantial variability, with some

[1]Department of Internal Medicine, McGovern Medical School, The University of Texas Health Science Center, Houston, TX 77030, USA. [2]Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA. Email: erez.n.baruch@uth.tmc.edu

PHOTO: COURTESY OF EREZ BARUCH

having gut microbiome signatures similar to those of responders, and others having gut microbiome signatures similar to those of nonresponders. We therefore chose to move forward with durable responder donors and identified two melanoma patients (i.e., Donor 1 and Donor 2) who experienced complete tumor regression following anti-PD-1 therapy to serve as FMT donors on a clinical trial.

We next designed a first-in-human phase 1 clinical trial, aimed at assessing the safety and potential immune effects of FMT in anti-PD-1–resistant metastatic melanoma patients. The patients who enrolled in this study (i.e., FMT recipients) underwent an initial "native microbiota depletion" phase in which they ingested oral antibiotics. FMT was then performed by both colonoscopy and administration of oral stool capsules. Anti-PD-1 therapy was then reintroduced. For each recipient patient, stool, gut, and tumor samples were collected before treatment and at several time points during the treatment protocol.

Our findings demonstrated clinical responses to anti-PD-1 therapy in 3 out of the 10 patients treated, including one patient with complete tumor regression—an exceptional outcome for this immunotherapy-resistant patient population (*12*). These clinical responses may be explained by FMT's profound impact on the recipient patients' gut microbiota and immune activity. Engraftment of the FMT was confirmed by 16*S* rRNA gene and metagenomic sequencing of longitudinal gut microbiota samples, revealing that recipient patients' gut microbiota composition was similar to that of their donor. Treatment with FMT was also associated with an increase in activity of antigen-presenting cells in both the gut and tumors of the recipient patients, as well as an increase in $CD8^+$ T cell activity in their tumors—a critical step in mounting an anti-tumor immune response (*13*).

I have completed my MD-PhD and am currently enrolled in internal medicine residency in the Physician-Scientist track at the University of Texas, McGovern Medical School. As a member of Dr. Wargo's laboratory at the MD Anderson Cancer Center, I am continuing my efforts to uncover the mechanisms behind microbiota effects on anticancer immunity. We are working to understand host-microbiome interactions in cancer using different modulation strategies, such as FMT, consortia, and dietary intervention to develop new treatments that will enable metastatic patients to harness the advantages of microbiota modulation in their fight against cancer. ■

REFERENCES AND NOTES

1. A. Ribas *et al.*, *JAMA* **315**, 1600 (2016).
2. X. Song *et al.*, *Curr. Med. Res. Opin.* **31**, 987 (2015).
3. P. A. Ascierto *et al.*, *JAMA Oncol.* **5**, 187 (2019).
4. A. B. Warner *et al.*, *J. Clin. Oncol.* **38**, 1655 (2020).
5. O. Hamid *et al.*, *Ann. Oncol.* **30**, 582 (2019).
6. M. Harel *et al.*, *Cell* **179**, 236 (2019).
7. A. Sivan *et al.*, *Science* **350**, 1084 (2015).
8. M. Vétizou *et al.*, *Science* **350**, 1079 (2015).
9. V. Gopalakrishnan *et al.*, *Science* **359**, 97 (2018).
10. V. Matson *et al.*, *Science* **359**, 104 (2018).
11. B. Routy *et al.*, *Science* **359**, 91 (2018).
12. E. N. Baruch *et al.*, *Science* **371**, 602 (2021).
13. P. Sharma, S. Hu-Lieskovan, J. A. Wargo, A. Ribas, *Cell* **168**, 707 (2017).

10.1126/science.abi9360

PRIZE ESSAY

FINALIST
Maria Zimmermann-Kogadeeva
Maria Zimmermann-Kogadeeva received undergraduate degrees from Lomonosov Moscow State University in Russia and a PhD from ETH Zürich, Switzerland. After completing her postdoctoral fellowships at Yale University in the Goodman group and at European Molecular Biology Laboratory (EMBL) Heidelberg in the Bork group, Maria will start her laboratory in the Genome Biology Unit at EMBL Heidelberg in 2021. Her research combines computational modeling and multi-omics data integration to investigate how microbes adapt to their surroundings and how metabolic adaptations of individual bacteria shape the functional outcome of microbial communities and their interactions with the host and the environment.
www.sciencemag.org/content/373/6551/173.2

MICROBIOME

Quantifying host-microbiota interactions

Modeling the microbiome increases understanding of its role in drug metabolism

By **Maria Zimmermann-Kogadeeva**

The human microbiota is a complex microbial community living on and in our bodies. Its impact on a host's health is immense, affecting digestion (*1*), the immune system (*2*), behavior (*3*), metabolic diseases (*4*), and responses to drugs (*5*–*7*). Rapid advances in experimental and computational methods have moved the human microbiome field from identifying associations between microbiota composition and host health to unraveling the underlying molecular mechanisms (*8*–*10*). However, exactly how much the microbiota contributes to host health is a very difficult question to answer.

By focusing on mechanistic and quantitative questions about the microbiome's contributions to host metabolism, I leverage my background in applied mathematics and systems biology to develop computational models describing host-microbiota interactions. Good models require good data from controlled experiments—a challenging proposition in complex host-microbiota systems. As a postdoc, I joined Andy Goodman's lab at Yale University and found myself in a perfect position to collect such data.

By combining bacterial genetics with gnotobiotic mouse models, I learned how to modify the microbiome of germ-free, sterile mice. In the Goodman lab, we used these mice to study the contribution of microbiota to host metabolism of a number of pharmaceutical drugs. We found that this was also a good system to quantify host-microbiome interactions in vivo, because the compounds we used can be introduced into the system in a controlled way.

We first focused on brivudine, an antiviral compound that can be converted into a potentially toxic metabolite, bromovinyluracil (BVU), by either a host or its microbiome (*11*). To identify bacteria capable of converting brivudine to BVU, we incubated individual bacterial species with the drug in vitro. One of the most potent brivudine metabolizers was *Bacteroides thetaiotaomicron*, a common gut bacterium with a genetic deletion library readily available. By incubating this library with the drug, we identified one bacterial mutant that had lost the capacity to convert brivudine to BVU. We then colonized germ-free mice with either the wild-type or mutant *B. thetaiotaomicron*, which provided us with a controllable host-microbiome system and two mouse groups that were identical, save for a single bacterial gene.

When we administered brivudine to these two groups, the observed outcome was somewhat puzzling. Although drug levels in the intestine were much higher in mice colonized with the mutant bacterium, serum levels were comparable between the two mouse groups. The metabolite levels showed the opposite pattern: no difference (and very low levels) in the intestine but much higher metabolite levels in the sera of mice colonized with the wild-type bacterium (see the figure). These data could potentially be explained by bacterial conversion of the drug in the intestine and the rapid metabolite absorption into the serum.

To test this explanation, we started with a simple kinetic model with two equations describing host drug metabolism in the liver and bacterial drug metabolism in the intestine. Once solved, this equation system showed that the difference between the amounts of metabolite absorbed into the sera of each of the two mouse groups was determined by the amount of BVU produced by microbes in the gut. This controlled experimental setup allowed us to quantify that the bacterial contribution to the toxic drug metabolite in vivo was about 70% (*12*) (see the figure).

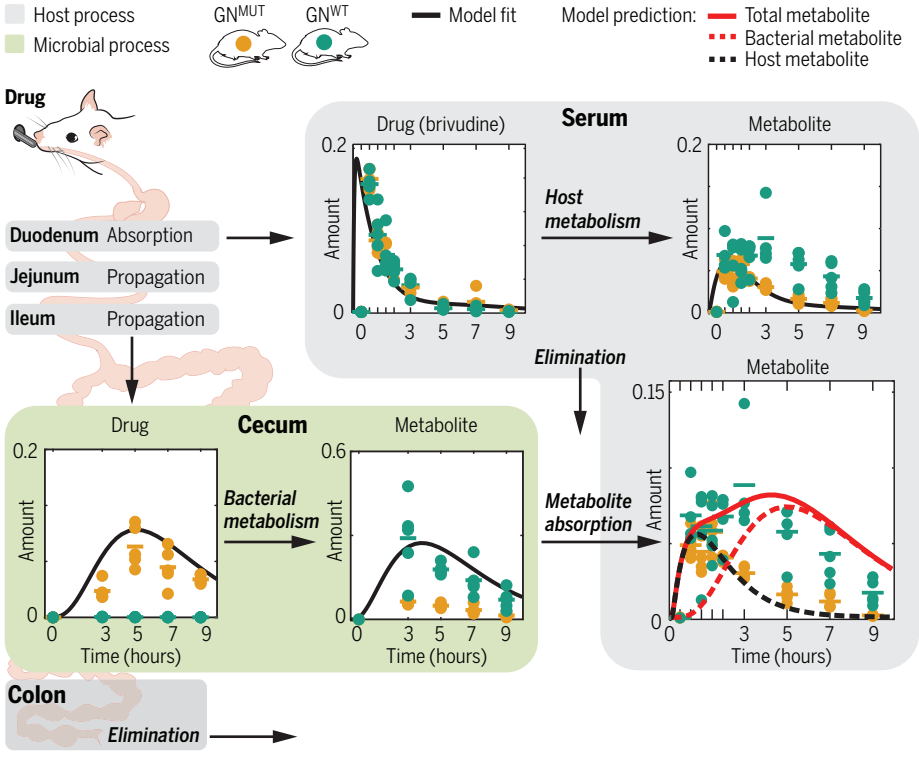
We expanded the model to describe drug metabolism processes in eight different tissues and in enterohepatic circulation (when the drug metabolized in the liver is secreted back into the small intestine via bile). We then demonstrated that our approach can be generalized to estimate the bacterial

Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany.
Email: maria.zimmermann@embl.de

PHOTO: MASSIMO DEL PRETE/EMBL

Experimental and computational approaches that quantify host and microbial contributions to drug metabolism

Oral drugs are administered to gnotobiotic mice that differ in a single microbial drug-metabolizing enzyme (GN^{MUT}, mutant; GN^{WT}, wild type); drug and drug metabolite kinetics are then quantified across tissues. A microbiome-host pharmacokinetic model developed from these measurements accurately predicts serum metabolite exposure and untangles host and microbiome contributions to drug metabolism.



contribution to drug metabolism even if the metabolizing species remain unknown by using data from germ-free mice and mice harboring a complex microbial community. We also showed that microbial contribution to the drug metabolite far exceeds the host for sorivudine, an antiviral drug with different host and microbiome metabolism rates, and for clonazepam, an anxiolytic and anticonvulsant drug converted to multiple metabolites (*12*).

Quantifying the metabolic host-microbiome interactions is not the only purpose of our model. Having a robust model of host-microbiome interaction allows us to study, explain, and predict the system's behavior in different conditions. By analyzing how drug and metabolite profiles change when model parameters are varied, we found that the similarity of drug serum profiles between germ-free and colonized mice can be explained by the fast and microbiota-independent drug absorption from the small intestine. Our model further suggests that even for rapidly absorbed drugs, microbiome contributions to a host's metabolism can be substantial under certain conditions (e.g., a high microbiome to host ratio of drug metabolism or extensive enterohepatic circulation of the drug and its metabolites) (*13*). Such computational models enable us to investigate host-microbiota interactions in silico, guide experimental design, and help reduce the number of experiments needed to confirm model predictions. To systematically investigate microbial capacity to metabolize drugs, we next conducted a high-throughput in vitro screen. We found that microbiota contribution to drug metabolism might even be more widespread than we anticipated—two-thirds (176 out of 271) of the human-targeted drugs we examined were metabolized by at least one of the 76 tested bacteria (*14*).

Although follow-up studies are required to test these microbiota-drug interactions in vivo, our findings emphasize that the microbiota should be considered when developing new drugs, stratifying patients, and choosing the most efficient treatment strategies. In the future, I believe that computational models combined with quantitative experimental data will allow us to measure host-microbiome interactions beyond drug metabolism and to better understand, predict, and control the effect of the microbiome on our health in everyday life. ■

REFERENCES AND NOTES

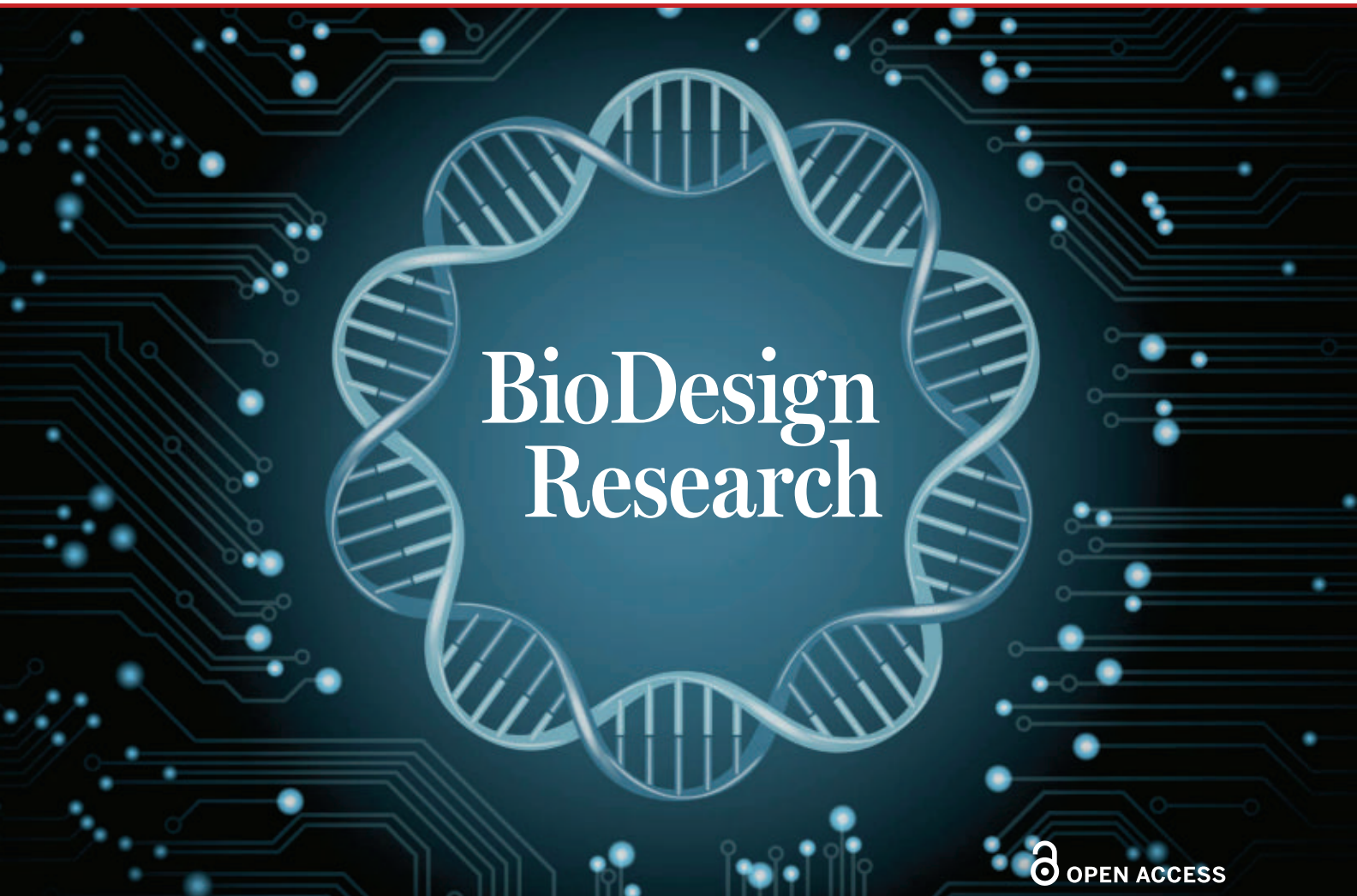
1. H. J. Flint, *Nutr. Rev.* **70**, S10 (2012).
2. A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman, J. I. Gordon, *Nature* **474**, 327 (2011).
3. T. R. Sampson, S. K. Mazmanian, *Cell Host Microbe* **17**, 565 (2015).
4. J. Durack, S. V. Lynch, *J. Exp. Med.* **216**, 20 (2019).
5. P. Spanogiannopoulos, E. N. Bess, R. N. Carmody, P. J. Turnbaugh, *Nat. Rev. Microbiol.* **14**, 273 (2016).
6. N. Koppel, V. Maini Rekdal, E. P. Balskus, *Science* **356**, eaag2770 (2017).
7. I. D. Wilson, J. K. Nicholson, *Transl. Res.* **179**, 204 (2017).
8. T. S. B. Schmidt, J. Raes, P. Bork, *Cell* **172**, 1198 (2018).
9. M. Alexander, P. J. Turnbaugh, *Immunity* **53**, 264 (2020).
10. C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, *Cell Host Microbe* **21**, 433 (2017).
11. H. Machida *et al.*, *Biochem. Pharmacol.* **49**, 763 (1995).
12. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, *Science* **363**, eaat9931 (2019).
13. M. Zimmermann-Kogadeeva, M. Zimmermann, A. L. Goodman, *Gut Microbes* **11**, 587 (2020).
14. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, *Nature* **570**, 462 (2019).

10.1126/science.abi9357

GRAPHIC: ADAPTED FROM M. ZIMMERMANN-KOGADEEVA BY N. CARY/*SCIENCE*

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS



BioDesign Research

OPEN ACCESS

BioDesign Research is a Science Partner Journal published in affiliation with **Nanjing Agricultural University (NAU)** and distributed by the **American Association for the Advancement of Science (AAAS).** *BioDesign Research* publishes high quality breakthrough research, reviews, editorials, and perspectives focusing on in silico biosystems design, genetic or epigenetic modifications, and genome writing or rewriting in any organism.

Submit your research to *BioDesign Research* today!

Learn more at **spj.sciencemag.org/bdr**

The Science Partner Journals (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the non-profit publisher of the *Science* family of journals. The SPJ program features high quality, online-only, editorially independent open-access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS expands its efforts to communicate science broadly and for the benefit of all people by providing a top-tier international research organization with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Learn more at **spj.sciencemag.org**

@SPJournals **@SPJournals**

ARTICLE PROCESSING CHARGES WAIVED UNTIL 2022

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PHOTO: MICHAEL QUINTON/MINDEN PICTURES

Sea otters use thick fur and thermogenic uncoupling in skeletal muscle mitochondria to stay warm in cold waters.

PHYSIOLOGY

Keeping warm when small

Several mammal species live in cold-water environments, enabled by adaptations such as blubber and large size. A notable exception to this rule is the sea otter, a species that is orders of magnitude smaller and skinnier than the others. It is known that the sea otter's unusually thick fur helps, but Wright *et al*. show that they are also internally warmed by thermogenic leak from skeletal muscle, a process that elevates their metabolic rate three times above that expected for their size. This mechanism is present even in infants with immature muscles, providing these animals with internal warmth from birth. —SNV

Science, abf4557, this issue p. 223

INFECTIOUS DISEASE

Anti-infection immunotherapy

Antibiotic-resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) are increasingly prevalent and difficult to treat. Alphonse *et al.* found that a pan-caspase inhibitor called Q-VD-OPH functioned as an effective immunotherapy in mouse models of community-acquired MRSA, *Streptococcus pyogenes*, and *Pseudomonas aeruginosa* skin infections. This inhibitor reduced apoptosis in neutrophils and monocytes and increased necroptosis in macrophages, thereby increasing tumor necrosis factor production and infection clearance in mice. This work suggests a potential strategy to target bacterial infections without the use of antibiotics. —CAC

Sci. Transl. Med. **13**, eabe9887 (2021).

MARTIAN GEOLOGY

Modified clay minerals on Mars

Sedimentary rocks exposed in Gale crater on Mars contain extensive clay minerals. Bristow *et al*. analyzed drill samples collected by the Curiosity rover as it climbed up sedimentary layers in the crater. They found evidence of past reactions with liquid water and sulfate brines, which could have percolated through the clay from an overlying sulfate deposit. Similar sulfate deposits are widespread across the planet and represent some of the last sedimentary rocks to form before the planet lost its surface liquid water, so the results inform our understanding of the geologic processes that occurred as Mars dried out. —KTS

Science, abg5449, this issue p. 198

GEOMORPHOLOGY

The limits of channel offset

The offset of stream channels across a strike-slip fault offer a record of long-term slip rates. The process itself is an interesting case of landscape evolution because the streams will spill straight across the fault at some point and reset the offset. Dascher-Cousineau *et al*. developed a model for this process that they validated using observations from the Carrizo Plain in California. The model leverages transitions from active to abandoned stream channels to determine when and how drainage networks in these regions reorganize and allow quantification of both slip and bedload transport. —BG

Science, abf2320, this issue p. 204

MECHANOCHEMISTRY

Shear selectivity

Chemical reactions typically proceed by distributing energy statistically among all accessible molecular vibrations. Liu *et al*. report that external shear forces can sometimes pry open strained carbon rings without dissipating

energy into adjacent bond rotations. Through careful design and synthesis of polymer-embedded cyclobutyl rings, the authors showed that certain relative substituent geometries are preserved when sonication induces ring opening. Accompanying simulations support the instigation of "flyby" trajectories that channel energy narrowly to cleave the cyclic sigma bonds and then rapidly form acyclic pi bonds. —JSY

Science, abi7609, this issue p. 208

CATALYSIS

Hitting the limits on propene synthesis

The greater abundance of propane from shale gas has spurred efforts to use it as a propylene feedstock. Direct dehydrogenation catalysts consisting of platinum–tin alloy nanoparticles supported on alumina often must run with hydrogen dilution to avoid carbon buildup and excess tin to avoid alloy segregation. Motagamwala *et al*. report that platinum–tin nanoparticles interact more weakly with a silica support and the metals thus do not segregate. The use of undiluted reactants allowed the reaction to run near the thermodynamically limit of about 67% conversion with a selectivity to propylene of more than 99%. This catalyst also does not build up carbon and could run up to 30 hours without deactivation. —PDS

Science, abg7894, this issue p. 217

NEUROSCIENCE

Representing space in past and future

As an organism moves through space, its brain has to remember its most recent location and anticipate its future position, not just its current place in the world. Earlier studies reported so-called retrospective and prospective place coding in rats while they were running along linear tracks. However, it would be advantageous to study an animal that rapidly moves through three-dimensional space with high precision. Dotson and Yartsev recorded from flying bats to investigate whether place cell activity in hippocampus area CA1 represents local (current) or nonlocal positions. They discovered that the hippocampus not only encodes the bat's present location but also signals its positions in the past and future. —PRS

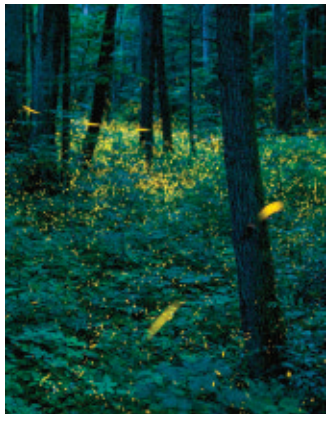
Science, abg1278, this issue p. 242

NETWORK DYNAMICS

How to put on a bioluminescent light show

Synchronous swarms of flashing fireflies are a natural wonder and a canonical example of coordinated group behavior. Using three-dimensional reconstructions from video recordings in Great Smoky Mountains National Park, Sarfati *et al*. reveal that incoherent individual flashes transition to a macroscopic marvel only once a critical density of active fireflies can convey the group pace. Flashing is synchronized over a surprisingly wide range of distances—many meters—leading to the hypothesis that the local interactions are defined not strictly by distance or by nearest neighbor but, for firefly swarms, by line of sight. The results unmask a complex mixture of short- and long-range interactions underlying ensemble behavior. —JSH

Sci. Adv. 10.1126/sciadv.abg9259 (2021).



Synchronous fireflies in Great Smoky Mountain National Park coordinate light flashes locally, depending on high density and visual connections to establish synchrony.

IN OTHER JOURNALS

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

IMAGING

Putting cells in their places

Despite considerable progress in recent years, pathological analysis of tissues and organs often still relies on old-fashioned microscopic techniques. Similarly, genomic and transcriptomic analysis is commonly applied to bulk tissue samples. To address the limitations of these approaches, Zhao *et al.* and van Ineveld *et al*. have developed methods for incorporating both cellular location and molecular information into the analysis of samples from healthy organs and tumors. By adopting a Bayesian statistical approach, the former identified cellular clustering in various carcinomas, and, by using multispectral image analysis, the latter revealed tumor-specific cell populations in pediatric Wilms' tumors. Such resolution will improve our understanding of the biology of multiple tissue types and the functions of cells in their biological contexts. —YN

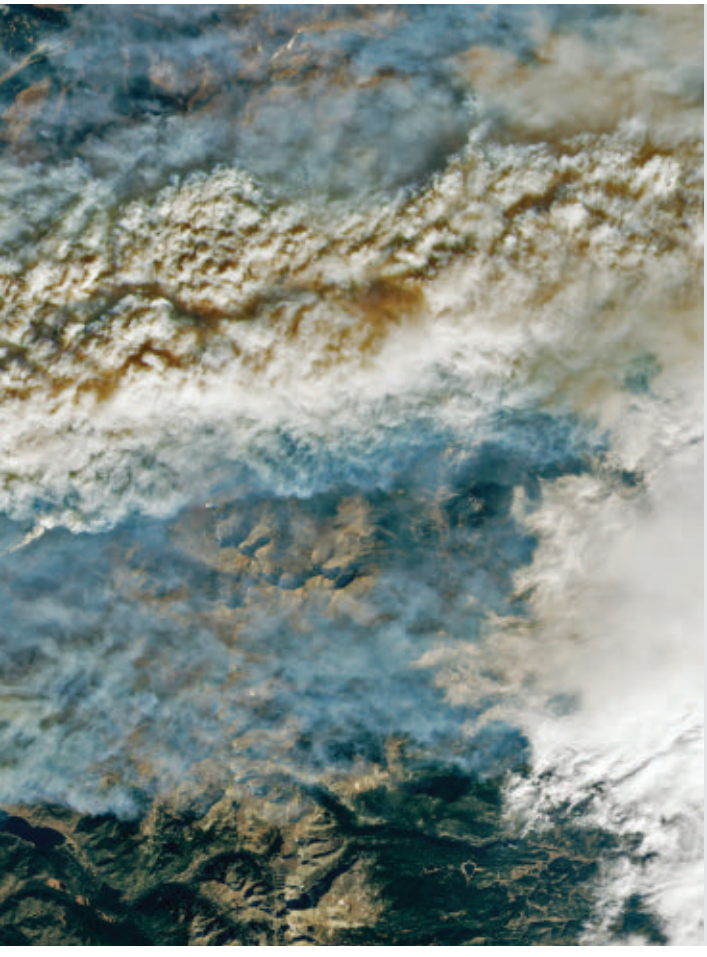
Nat. Biotechnol. 10.1038/s41587-021-00935-2, 10.1038/s41587-021-00926-3 (2021).

QUANTUM OPTICS

Into the valley

The topological features of band structures provide robust propagation paths for classical and quantum states of light. Another ingredient of band structure is the valley degree of freedom, in which the propagation path of light is dependent on its polarization state. Chen *et al.* used both topology and valley-dependent transport to design and fabricate a photonic crystal structure capable of forming the basis of optical quantum circuits. The authors demonstrate the input of left- and right-circularly polarized states of quantum light, their

PHOTOS (TOP TO BOTTOM): NASA EARTH OBSERVATORY IMAGE BY LAUREN DAUPHIN; (DATA) THE U.S. GEOLOGICAL SURVEY; FLORIS VAN BREUGEL/MINDEN PICTURES



WILDFIRES

Fired up

Climate warming has substantially increased wildfire activity in the western United States, particularly during the first two decades of the 21st century. Higuera *et al.* illustrate how destructive this trend has become by using a 2000-year-long record of wildfires in the subalpine forests of the Rocky Mountains. Their data show that these forests now are burning more than at any point in the past 2000 years, capped by the extreme fire season of 2020. The current trajectory of climate means that this situation may be expected only to worsen, with unprecedented rates of burning projected in many western forests by midcentury. —HJS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2103135118 (2021).

Satellite view of the East Troublesome Fire in Grand County, Colorado, on 20 October 2020

separation into different paths, and controlled interference as the single photons are brought back together again. The results provide a route for developing complex optical quantum circuits for on-chip quantum information processing. —ISO

Phys. Rev. Lett. **126**, 230503 (2021).

MAGNETISM

A dilute 2D magnet

Recently discovered two-dimensional (2D) van der Waals magnets exhibit a number of appealing properties, such as a large magnetoresistance. Most of these materials, however, need to be cooled to below room temperature to exhibit these properties and are sensitive to environmental conditions. To address these issues, Chen *et al.* exploited an old idea, doping an oxide with magnetic atoms, but this time in 2D rather than 3D. The researchers used graphitic zinc oxide, a layered compound in which zinc and oxygen atoms are arranged in a hexagonal pattern, and substituted some zinc atoms with cobalt. The samples exhibited ferromagnetism down to monolayer thickness, even at room temperature. Extrinsic mechanisms for magnetism were excluded with careful characterization and control experiments. —JS

Nat. Commun. **12**, 3952 (2021).

NEUROSCIENCE

Breathing in the brain

The effect of respiration on brain activity is much discussed. Neuronal activity is fundamentally rhythmic, so are breathing rhythms entrained by typical neuronal oscillation patterns? Tort *et al.* hypothesized that breathing frequency aids the integration of widespread information similar to the proposed function of other slow-network rhythms in visual or tactile environmental sampling. Systematic analysis revealed strong correlations between instantaneous breathing frequency and the amplitude and frequency of theta and gamma oscillations. Changes in theta activity preceded changes in breathing frequency, which in turn preceded changes in slow-gamma oscillations. —PRS

J. Neurosci. **41**, 5229 (2021).

RESTORATION ECOLOGY

Seeds of success

This year marks the start of the United Nations' Decade of Ecological Restoration. Restoration of damaged habitats to a self-sustaining state (which may not necessarily be identical to their original state) requires close collaboration between ecologists and land managers. Leger *et al.* describe practical aspects of such a partnership in the restoration of degraded grassland habitats in the Great Basin region of the United States. The authors compared the restoration potential of seeds of grass species sourced from remnant native habitat with seeds from commercial sources. The wild-collected seeds showed more promising performance, as measured by phenotypic traits. In turn, these seeds can be selected for enhanced agronomic production, increasing the supply available for more effective restoration. —AMS

Restor. Ecol. **29**, e13260 (2021).

SYMBIOSIS

Costly symbiont inheritance

Long-established microbial symbioses tend toward genome erosion and increasing host dependence. Some relationships are so pivotal to host survival that maintaining a stable partnership is crucial. Koga *et al.* have investigated the stable inheritance of an insect bacterial gut symbiont called *Ishikawaella* between generations of its Plataspidae stinkbug hosts. *Ishikawaella* has a highly reduced genome yet provides essential amino acids for the bug. The bugs lay eggs in rows on plants, and capsules holding symbionts are deposited among the eggs. The capsule also contains a combination of a specialist insect protein the authors identified as posterior midgut dominant protein (PMDP) and a symbiont-derived chaperone molecule, GroEL. Together, these molecules protect the delicate *Ishikawaella* for the 7 to 10 days it takes for the eggs to hatch, at which point the young bugs feed on the capsule to obtain the vital symbionts. However, the production of PMDP shortens the life of the female bugs, which is the price a mother pays to ensure inheritance. —CA

Proc. Natl. Acad. Sci. U.S.A. **118**, e2103957118 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Focusing on transmission risk

Stopping downstream transmission is an important aspect of mitigating the COVID-19 pandemic. Much of the modeling used to understand transmission patterns assumes that everyone has the same risk but this is not the case. Heterogeneity in exposure related to occupation, housing, and economic circumstances can substantially affect the risk of being infected and of infecting others. In a Perspective, Cevik and Baral discuss the factors that drive transmission of severe acute respiratory syndrome coronavirus 2 and argue that people with the highest risk, such as those in economically deprived neighborhoods, those living in large households, and those who cannot work from home, should receive targeted support to allow them to decrease contacts and thus break transmission chains. Without such support, these communities are left at higher risk of infection, death, and economic hardship. —GKA

Science, abg0842, this issue p. 162

LIGHT-MATTER COUPLING

Captivating cavities

Laser technology is a familiar example of how confining light between two mirrors can tune its properties. Quantum mechanics also dictates that even without extraneous light, matter confined in a cavity resonant with its electronic or vibrational transitions can couple with vacuum electromagnetic field fluctuations. Garcia-Vidal *et al*. review the remarkable and still somewhat mysterious implications of this "strong-coupling" regime, with manifestations ranging from enhanced charge transport to site-selective chemical reactivity across a range of molecular and solid-state materials. —JSY

Science, abd0336, this issue p. 178

CORONAVIRUS

Correlates of infectiousness

The role that individuals with asymptomatic or mildly symptomatic severe acute respiratory syndrome coronavirus 2 have in transmission of the virus is not well understood. Jones *et al*. investigated viral load in patients, comparing those showing few, if any, symptoms with hospitalized cases. Approximately 400,000 individuals, mostly from Berlin, were tested from February 2020 to March 2021 and about 6% tested positive. Of the 25,381 positive subjects, about 8% showed very high viral loads. People became infectious within 2 days of infection, and in hospitalized individuals, about 4 days elapsed from the start of virus shedding to the time of peak viral load, which occurred 1 to 3 days before the onset of symptoms. Overall, viral load was highly variable, but was about 10-fold higher in persons infected with the B.1.1.7 variant. Children had slightly lower viral loads than adults, although this difference may not be clinically significant. —CA

Science, abi5273, this issue p. 180

MICROBIOMES

Baboons inform on human gut microbiota

Commensal bacteria are found throughout an organism, but it is not known whether associations between gut bacteria and their host are heritable. Grieneisen *et al*. examined changes in the microbiomes of 585 wild baboons from fecal samples collected over 14 years (see the Perspective by Cortes-Ortiz and Amato). Almost all microbiome traits tested demonstrated some level of statistically significant heritability. Most heritability values were low but varied over time correlating with the age of the host. Baboons live in an environment similar to that postulated for early humans and have a microbiome similar to that of humans. Thus, this heritability of the microbiome may reflect similar genetic determinants in humans, for which similar datasets are not available. —LMZ

Science, aba5483, this issue p. 181; see also abj5287, p. 159

ICE PHYSICS

Ice goes bendy

One well known characteristic of ice is that it fractures instead of bending when strained. This characteristic is caused by the inevitable defects introduced into the ice structure during solidification. Xu *et al*. show that very thin, carefully grown ice microfibers can bend a lot, up to about 11%, and still remain elastic (see the Perspective by Schulson). This value is reasonably close to theoretical limits previously estimated. The fibers are also super clear, allowing for efficient light transmission. —BG

Science, abh3754, this issue p. 187; see also abj4441, p. 158

CANCER

Down with leukemia

Down syndrome is a congenital disorder caused by the trisomy of chromosome 21, and it is associated with a greatly increased risk of leukemia with origins in fetal development. Infants with Down syndrome are often born with a preleukemic condition, which later resolves in most cases. By using gene-edited human cells implanted into mouse models, Wagenblast *et al*. recapitulated the development of preleukemia and leukemia in the context of Down syndrome (see the Perspective by Roberts and Vyas). A specific mutation triggered a preleukemic condition in the context of trisomy 21 as expected, but progression to full-blown leukemia required a different genetic path and was not dependent on trisomy 21. —YN

Science, abf6202, this issue p. 179; see also abj3957, p. 155

SUPERCONDUCTIVITY

Looking for magnetic clues

Thin films of the neodymium nickelate $NdNiO_2$ doped with strontium have recently been found to be superconducting. This materials class bears structural and electronic similarities to the famed cuprate superconductors, but how far the analogy goes remains unclear. Lu *et al*. used resonant inelastic x-ray scattering to look for magnetism, which exists in the cuprates, in $Nd_{1-x}Sr_xNiO_2$ films (see the Perspective by Benckiser). The authors observed magnetic modes in the undoped compound that had a doping evolution consistent with the behavior of a doped Mott insulator. —JS

Science, abd7726, this issue p. 213; see also abi6855, p. 157

SENSORY EVOLUTION

From savory to sweet

Seeing a bird eat nectar from a flower is a common sight in our world. The ability to detect sugars, however, is not ancestral in the bird lineage, where most species were carnivorous. Toda *et al*. looked at receptors within the largest group of birds, the passerines or songbirds, and found that the emergence of sweet detection involved a single shift in a receptor for umami (see the Perspective by Barker). This ancient change facilitated sugar detection not just in nectar feeding birds, but also across the songbird group, and in a way that was different from, though convergent with, that in hummingbirds. —SNV

Science, abf6505, this issue p. 226; see also abj6746, p. 154

ANTIVIRAL DEFENSE

An antiviral Dicer defends stem cells

Stem cells have a pivotal role in maintaining tissue architecture,

integrity, and renewal. Poirier *et al*. demonstrate that mammalian stem cells can protect themselves from some RNA viruses by expressing an alternatively spliced isoform of the enzyme Dicer called aviD, which potentiates antiviral RNA interference (see the Perspective by Shahrudin and Ding). aviD acts by cleaving long, base-paired viral RNAs to generate small interfering RNAs that direct the sequence-specific cleavage of homologous viral RNAs. This process is reminiscent of that in insects and worms, which also use Dicer-dependent RNA interference in antiviral defense, and contrasts with mammalian differentiated cells, which generally rely on the interferon system to combat virus infection. —DJ

Science, abg2264, this issue p. 231; see also abj5673, p. 160

PLANT SCIENCE

Vegetal fractals

Cauliflower, along with dahlias and daisies, develop as phyllotactic spirals. Azpeitia *et al*. combined modeling with experimental investigation to clarify the gene-regulatory network that sets up a multitude of undeveloped flowers to form a cauliflower curd. Irrepressible inflorescence identity genes in the context of dysfunctional meristems and slow internode elongation result in piles of incomplete flowers. If meristem size drifts during organogenesis, then the conical structures of the Romanesco form emerge in fractal formation. —PJH

Science, abg5999, this issue p. 192

CORONAVIRUS

Mind your metals

Iron–sulfur clusters are important cofactors for proteins involved in metabolism and electron transfer but are also sometimes found in enzymes involved in transcription and replication of DNA. In vitro expression of such enzymes can result in faulty cluster assembly and confusion about the composition of the functional enzyme. Using a careful anoxic purification scheme, Maio *et al*. found that the severe acute respiratory syndrome coronavirus 2 RNA–dependent RNA polymerase contains two iron–sulfur clusters at two sites previously observed to bind zinc ions. Mutation of the ligating cysteine residues resulted in loss of polymerase activity. A less severe loss of activity was seen in the zinc-containing enzyme. Treatment with the nitroxide drug TEMPOL resulted in degradation of the clusters, enzyme inhibition, and inhibition of viral replication in cell culture. —MAF

Science, abi5224, this issue p. 236

COVID-19

Sex-specific metabolism and COVID-19

Sex affects immune responses to severe acute respiratory syndrome coronavirus 2 infection, with male sex being a risk factor for mortality, particularly among older individuals. Cai *et al*. identified 17 serum metabolites associated with COVID-19 infection. In male COVID-19 patients only, the amount of the tryptophan metabolite kynurenic acid correlated with age, inflammation, and disease outcome. Kynurenic acid inhibits glutamate release, and glutamate levels were reduced in patients who deteriorated. Thus, kynurenic acid may be linked in a sex-specific manner to immune responses and clinical outcomes in COVID-19. —JFF

Sci. Signal. **14**, eabf8483 (2021).

CYTOKINES

IL-17 makes its m6A marks

Interleukin-17 (IL-17) and related cytokines not only support host defense against certain pathogens but also contribute to the development of autoimmune diseases. A holistic understanding of how IL-17 initiates and propagates inflammation is needed to identify additional therapeutic opportunities in autoimmunity. Bechara *et al.* probed the role of N6-methyladenosine (m6A) modifications of IL-17–induced transcripts in enhancing their stability and translation. Epitranscriptomic m6A marks were detected on the mRNAs for transcription factors involved in IL-17 signaling. The noncanonical m6A reader protein IMP2 was implicated in binding these m6A marks. Mice lacking IMP2 were less susceptible to an autoantibody-induced model of kidney disease driven by IL-17 signaling. These findings establish post-transcriptional modifications involved in proinflammatory cytokine signaling as potential targets for therapeutic intervention. —IRW

Sci. Immunol. **6**, eabd1287 (2021).

REVIEW SUMMARY

LIGHT-MATTER COUPLING

Manipulating matter by strong coupling to vacuum fields

Francisco J. Garcia-Vidal*, Cristiano Ciuti*, Thomas W. Ebbesen*

BACKGROUND: One of the most important phenomena in cavity quantum electrodynamics (cQED) is the so-called strong coupling regime, which appears when the interaction between a photon tightly confined in an optical cavity and a matter excitation creates hybrid light-matter states. When the latter are populated, hybrid particles called polaritons are formed. These particles are very attractive because they combine properties of their constituents, which enables applications ranging from low-threshold lasing in semiconductors to photon quantum information. Since its discovery, most of the investigations on strong coupling have been aimed mainly toward the modification of optical properties. During the past decade, an alternative area of research has emerged that takes advantage of collective strong coupling to take chemistry and materials science into new directions. For this purpose, no external light source is necessary as the hybrid light-matter states are formed even in the dark because the coupling occurs through the zero-point energy of the optical mode (i.e., the vacuum field). The mere presence of the hybrid states has a substantial effect on material properties, as reviewed here.

ADVANCES: Both experimental and theoretical studies have shown changes to photochemical reaction rates under strong coupling between the electronic excitations of molecules and cavity electromagnetic modes. Strong coupling modifies the shape of the potential energy surfaces associated with the excited states of the molecule, allowing for a manipulation of its photophysical properties. Moreover, ground-state chemical reactivity can also be completely modified when molecular vibrations are strongly coupled to infrared cavity modes. Although a detailed picture of the mechanism is still missing, symmetry seems to play a key role. Material properties can also be changed by strong coupling. Charge and energy transport in organic materials and magneto-conductivity in two-dimensional electron gases have been shown to be altered. Thanks to the intrinsic delocalized character of the polaritonic modes, transport properties can be then tuned at a macroscopic scale. It is also feasible to manipulate phases of matter by means of strong coupling. It has been reported that the critical temperature of a superconductor can be substantially enhanced by judiciously exploiting vibrational strong coupling and that the ferromagnetism of nanoparticles can be boosted by orders of magnitude. These examples illustrate the potential of using vacuum fields instead of intense laser fields to induce modification of material properties.

OUTLOOK: There are many classes of organic reactions that are currently being explored under strong coupling. As more results are collected, the underlying physical chemistry will be further clarified and should lead to some general principles to guide chemists and physicists in their use of vibrational strong coupling. The recent demonstrations that water, under vibrational strong coupling, modifies enzyme activity illustrates the potential for manipulating biological activity under strong coupling—an avenue that remains unexplored. Regarding solid-state material properties, the influence of strong coupling in phonon-based phase transitions should also be fully explored, aiming at inducing new condensed phases. Moreover, cavity-controlled magneto-transport might reach the quantum Hall regime. In general, two-dimensional materials are very well suited to be integrated in cavity resonators with deeply subwavelength photon confinement, which provides an intriguing platform to modify electronic properties through vacuum fields. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: fj.garcia@uam.es (F.J.G.-V.); cristiano.ciuti@u-paris.fr (C.C.); ebbesen@unistra.fr (T.W.E.)
Cite this article as F. J. Garcia-Vidal *et al.*, *Science* **373**, eabd0336 (2021). DOI: 10.1126/science.abd0336

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd0336

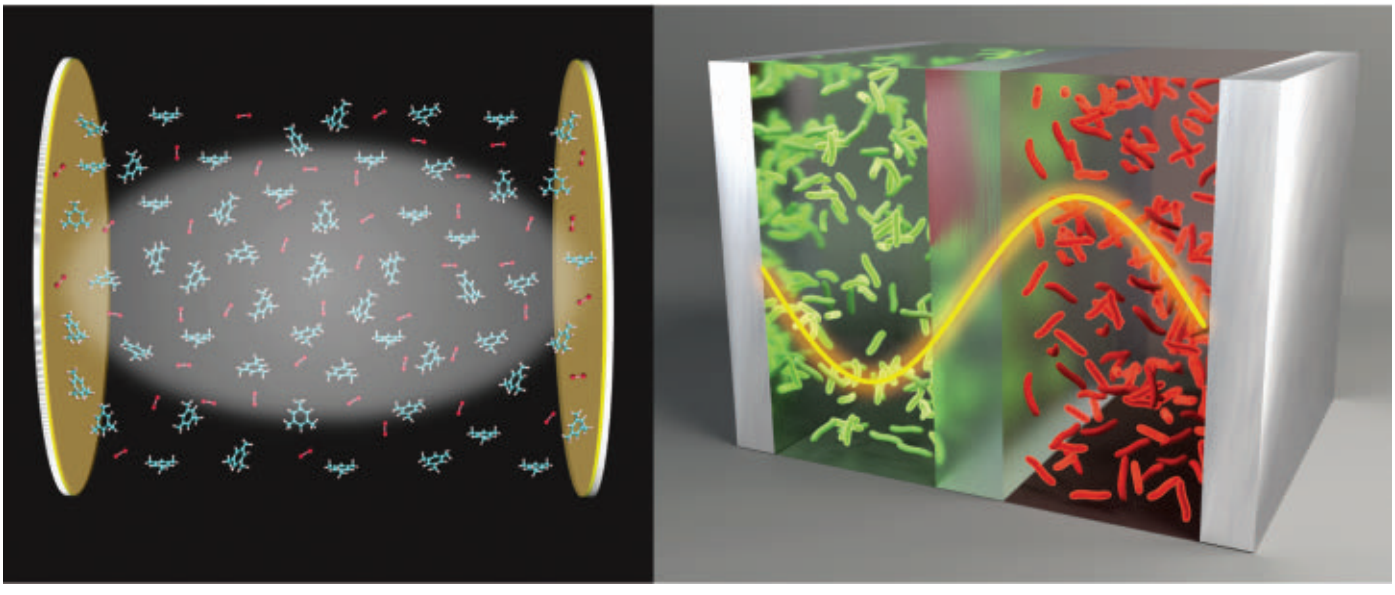


Illustration of modified molecular processes under strong coupling in optical cavities. (Left) Charge transfer complexation between mesitylene and iodide (courtesy of K. Nagarajan). (Right) Energy transfer between donor and acceptor molecules (courtesy of J. Galego).

REVIEW

LIGHT-MATTER COUPLING

Manipulating matter by strong coupling to vacuum fields

Francisco J. Garcia-Vidal[1,2]*, Cristiano Ciuti[3]*, Thomas W. Ebbesen[4]*

Over the past decade, there has been a surge of interest in the ability of hybrid light-matter states to control the properties of matter and chemical reactivity. Such hybrid states can be generated by simply placing a material in the spatially confined electromagnetic field of an optical resonator, such as that provided by two parallel mirrors. This occurs even in the dark because it is electromagnetic fluctuations of the cavity (the vacuum field) that strongly couple with the material. Experimental and theoretical studies have shown that the mere presence of these hybrid states can enhance properties such as transport, magnetism, and superconductivity and modify (bio)chemical reactivity. This emerging field is highly multidisciplinary, and much of its potential has yet to be explored.

It has been well known since the work of Huygens in 17th century that two oscillators, such as pendula, can couple to generate new modes by the exchange of energy (Fig. 1A). Although there are numerous examples of coupled oscillators in nature, what is less well known is that it is possible to couple, in an analogous way, an optical mode and a state transition in a material to generate two new hybrid light-matter states by placing the material inside a resonant optical cavity (Fig. 1B). Perhaps the most notable aspect is that such coupling can also occur in the dark—a natural consequence of the quantum nature of both the electromagnetic (EM) fields and matter—as explained in detail in the next section. The emergence of new hybrid light-matter states, also known as polaritonic states, has the potential to change the material and chemical properties under appropriate conditions, even in the ground state. Furthermore, typical materials encompass a large number of oscillators coupled to a single optical mode, resulting in the formation of collective delocalized states, which also affect their properties. Since the first demonstration of a modified chemical reaction, numerous examples of cavity-controlled changes in properties have been reported from biological functions to solid-state properties. This approach to manipulating material and (bio)chemical properties has been generating tremendous interest during the past decade and is the subject of this Review. The interest is both in the fundamental underpinnings as well as the technological potential.

[1]Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, 28049 Madrid, Spain. [2]Donostia International Physics Center, E-20018 Donostia/San Sebastián, Spain. [3]Université de Paris, Laboratoire Matériaux et Phénomènes Quantiques, CNRS-UMR7162, 75013 Paris, France. [4]University of Strasbourg, CNRS, ISIS, 67000 Strasbourg, France.
*Corresponding author. Email: fj.garcia@uam.es (F.J.G.-V.); cristiano.ciuti@u-paris.fr (C.C.); ebbesen@unistra.fr (T.W.E.)

Fundamentals of light-matter strong coupling

Normally we think of the vacuum as the absence of everything—in particular, the absence of radiation. In reality, empty space is filled with EM fluctuations. These so-called vacuum EM fluctuations originate from the fact that space is full of optical modes, which have zero-point energy as a result of their quantum nature. The presence of these vacuum fields has many consequences for molecular and material properties, such as, for example, spontaneous emission. It was Dirac (*1*) who, in laying the foundations of quantum electrodynamics (QED), provided the explanation that an emitter in an excited state interacts with those vacuum fields, allowing a transfer of energy to an unoccupied EM mode through the spontaneous emission of a photon. In free space, the EM field has a continuum of modes that typically interact weakly with the emitter. One way to enhance such interaction is to place the emitter inside a cavity, such as, for example, the one formed by two parallel metallic mirrors. In such confined geometry, the EM field presents a discrete spectrum of modes. Notably, the field can be much more localized than in free space, thereby increasing the emission quantum yield. This crucial prediction by Purcell in 1946 (*2*) began a whole field of research, today called cavity QED (cQED), which is devoted to exploiting different types of cavities and emitters to enhance and control light-matter interactions.

A very simple model to account for cQED scenarios was introduced by Jaynes and Cummings (JC) in 1963 (*3*). Within this JC model, only a single mode of the cavity with photon energy $\hbar\omega_c$ (where $\hbar$ is Planck's constant h divided by 2π) is taken into account, whereas the considered emitter is a two-level system with transition energy $\hbar\omega_e$. Hybridization between the emitter excited state and a cavity single-photon state leads to the formation of two new eigenstates. When the photon and transition frequencies are equal, the hybrid eigenstates have an energy splitting $2\hbar g$, where g is the rate of energy exchange between light and matter. To describe real systems, one cannot neglect the photon leakage rate, κ, and nonradiative losses for the emitter, γ. Weak and strong light-matter coupling regimes can be quantitatively defined by just comparing $2g$ with κ and γ. When $2g < \kappa, \gamma$, the exchange rate is smaller than the loss rates, and, as a consequence, the excitation is lost before it can be shared between the emitter and the cavity components. In this case, the emitter-cavity system is said to operate in the weak coupling regime. In the opposite case, $2g > \kappa, \gamma$, there is an exchange of energy between the light and matter components, and the system resides in the strong coupling regime. Another paradigmatic quantum model is the one introduced by Hopfield (*4*). This model describes the coupling of the EM quantum field to collective material excitations, like excitons in semiconductors, which are bosonic and so behave as harmonic oscillators.

The hybrid light-matter excited states are known as polaritonic states (P+ and P− in Fig. 1B). As a result, the energy spectrum of the system in this regime displays two distinct peaks (Fig. 1C), separated by the so-called Rabi splitting, $\hbar\Omega_R = \hbar\sqrt{4g^2 - (\kappa - \gamma)^2}$ (Fig. 1B). The huge potential and particular appeal of operating in this regime lie in the composite nature of the polaritonic states because they combine properties of their two constituents: the high coherence of light and the strong interaction of matter. In the extreme case, when the coupling is such that the Rabi frequency Ω_R is a significant fraction of the transition frequency (~>10%), the system enters the so-called ultrastrong coupling regime (*5*–*7*). Within this regime, antiresonant light-matter interaction processes that do not conserve the number of excitations in the system are allowed, which leads to a ground state that contains virtual bare excitations in addition to the emergence of polaritonic states in the excited manifold.

Attaining the strong or ultrastrong coupling regimes is enormously facilitated by the so-called collective coupling, in which a large number N of material oscillators—e.g., molecules—couple to one optical mode. As a consequence, the Rabi splitting increases as $\sqrt{N}$ (*8*). The collective coupling also results in the formation of a manifold of $N-1$ collective states, which are known as dark states (DS) because they cannot be directly excited with light and therefore cannot be seen in an absorption spectrum (Fig. 1B). However, the

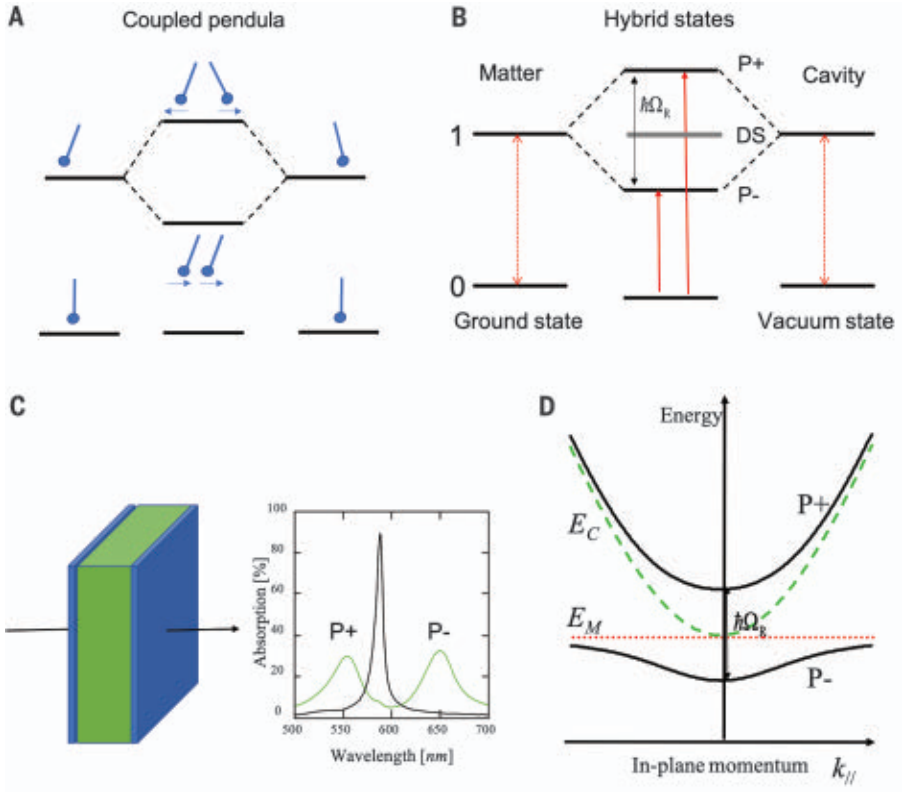


Fig. 1. Coupled oscillators and light-matter strong coupling. (**A**) Classical coupled pendula. (**B**) Strongly coupled material and optical transitions, leading to the formation of the hybrid light-matter or polaritonic states P+ and P– separated in energy by the Rabi splitting $\hbar\Omega_R$, and the DS in the case of collective coupling. (**C**) A FP cavity filled with a material. Under the right conditions for strong coupling, the formation of P+ and P– results in two new peaks in the absorption spectrum (in black, the original material transition; in green, the spectrum after strong coupling). Figure reproduced with permission from (*64*). (**D**) The dispersive optical mode of a FP cavity (E_C; green dashed line) intersects the material transition energy E_M (red dotted line) at normal incidence (where the parallel momentum $k_{//} = 0$).

DS do not have the properties of the states of the uncoupled material, and notably, like the polaritonic states, they can acquire a delocalized character (*9*) that extends over the whole system (Box 1).

The first quantum experiments aiming for the strong coupling of matter excitations were performed using Rydberg atoms in microwave and optical cavities. A detailed account of these experiments can be found in (*10*). In parallel, phonons of inorganic materials were coupled to surface plasmons (SPs) in a more classical approach (*11*, *12*). Later, advances in semiconductor microcavities led to the detection of strong coupling in various solid-state platforms, reaching Rabi splitting values of the order of tens of milli–electron volts (*13*), where experiments needed to be carried out at low temperatures. Within this scheme, it was possible to even achieve polariton lasing, condensation, and superfluidity (*14*, *15*). To enhance light-matter interactions, organic materials provided a great opportunity (*16*–*18*). In fact, organic molecules display large oscillator strengths in the visible spectrum, giving rise to Rabi splitting values more than 10 times as high as those obtained with inorganic semiconductors, thereby enabling lasing and condensation of polaritons at room temperature (*19*–*21*). Molecules also paved the way to explore cavity-modified chemistry (*22*) and other material properties under strong coupling, as discussed in detail in the following sections.

From a practical point of view, strong coupling is achieved by placing a material in the confined EM field, such as typically provided by a Fabry-Perot (FP) cavity (two parallel mirrors), as illustrated in Fig. 1C, which results in two new peaks in the absorption spectrum. Alternatively, SPs or more-complex resonators, such as metallic split rings, can be used (*23*). The choice of the optical resonator can be tailored depending on the material properties to be measured, as detailed in the next sections. Many optical resonators are dispersive—i.e., the resonance energy varies with angle, such as in the case of a FP cavity, where it is minimal at normal incidence (Fig. 1D). Experiments with dispersive optical modes have shown that to maximize the effect of strong coupling on organic material properties, the relevant transition needs to be resonant at the bottom of the dispersion curve—i.e., with the zero in-plane wavevector—probably because this is the equilibrium configuration where the properties are explored. Regarding material excitations, strong coupling can be achieved with different types of transitions that are optically allowed, such as electronic, phonon, and vibrational transitions, which can all affect both material and molecular properties. Most studies with organic molecules have involved coupling to electronic transitions or to vibrational transitions.

Modifying material properties

A large number of properties of solids have been explored under strong coupling to the vacuum field during the past decade, such as electrical conductivity, work-function, energy transfer probability, nonlinear optical response, and phase transitions. Here, we focus on a subset of properties on which much work has been carried out, starting with transport properties.

Enhanced transport of energy and charge

Organic semiconductors have generated great interest for large-scale fabrication of inexpensive and flexible devices. However, charge transport in disordered organic semiconductors displays a very low mobility of charge carriers (*24*), which severely limits their technological applications in electrical devices, such as field-effect transistors (*25*). On the other hand, energy transport in organic materials occurs through the motion of neutral electronic excitations—i.e., excitons—whose propagation is governed by short-range dipole-dipole interactions with a spatial range of only a few nanometers, in a process usually called Förster resonance energy transfer (FRET). Moreover, as most organic systems are disordered and possess relatively large dissipation and dephasing rates, exciton transport typically becomes incoherent and diffusive over long distances, which strongly limits optoelectronic applications of organic materials, such as organic photovoltaic cells (*26*).

As explained above, collective strong coupling leads to the formation of polaritonic states that are delocalized, extending over the whole structure in which the individual excitons and cavity EM mode(s) are interacting. This property is very appealing to enlarge the spatial range of both charge and energy transport in organic materials, as it could overcome the transport limitations linked to short-range interactions and disorder. In a first demonstration of this potential, the electrical conductivities of different n-type organic semiconductors deposited on top of a plasmonic resonator (holey metal Ag surface) were measured (*27*). This surface acts as an open

Box 1. Nomenclature and concepts.

Polaritonic states and DS: The collective coupling of *N* oscillators (e.g., molecules) to one optical mode generates two bright polaritonic states, P+ and P–, and *N* – 1 DS. The DS are also collective states generated by the strong coupling process: They are formed from a linear combination of all possible states with one excited molecule and *N* – 1 molecules in the ground state (*128*). Thus, their properties differ from those of the bare or uncoupled molecules. It has recently been suggested that if one considers the free energy, which depends on entropy, the DS can even have a lower energy than that of the lower polaritonic state (*116*). Moreover, in the presence of disorder and/or spatially inhomogeneous coupling, the *N* – 1 DS actually become gray, acquiring a photonic component that, even if small, can radically change the physical and chemical properties as catalysts or dopants do.
Reservoir of uncoupled excitons: The notion of a reservoir of uncoupled molecular excitons was introduced (*129*) because two types of disorder might lead to a large fraction of uncoupled molecules—namely orientation disorder (dipole moment is not aligned to couple with the cavity field) and spectral inhomogeneous broadening. It has been shown that for an inhomogeneously broadened absorber, the Rabi splitting does not occur from a subset of absorbers exactly resonant with the cavity, as confirmed by experiments (*128*). However, orientation disorder will necessarily lead to a fraction of uncoupled molecules. Recent experiments on chemical systems suggest that this fraction must be very small because it would preclude the observation of a large slowdown in chemical reaction rates and redistribution of products, among other things, as the coupled molecules could not compete with the faster uncoupled molecules.
Markovian versus non-Markovian dissipation regime: In the Markovian regime, the properties of the coupled system can be understood as simply derived from those of the uncoupled entities. In this regime, the dissipation rate of the coupled system is given by the sum of the cavity and exciton dissipation rates, such that the polaritonic states P+ and P– are expected to have the same lifetimes when the cavity mode is resonant to the material transition, contrary to experimental evidence. In the ultrastrong coupling regime, dissipation can be profoundly modified because the polaritonic states experience the effect of the bath (vacuum or thermal) at very different frequencies with respect to the bare modes (*130*). Additionally, the notably large changes in thermodynamics (PES and dynamics) observed in chemical reactions under strong coupling (*131*), and the effects of symmetry, clearly also point to a non-Markovian regime. In this regime, the dynamic properties must be calculated in the coupled basis, which can be challenging.
Cooperative coupling: When the absorption of the material to be coupled is weak as a result of low oscillator strength or low concentration, it is often not possible to attain the strong coupling regime. This limitation can be overcome by cooperative coupling whereby, for example, a solute has a vibrational absorption band at the same frequency as one of the vibrations of the solvent in which it is dissolved. Strong coupling of the solvent then acts on the solute as if it is coupled through intermolecular interactions (*115*, *117*).

EM cavity, and collective ultrastrong coupling between the excitons in the organic molecules and the SPs supported by the metal surface can emerge. At resonance, when the energy of the SP at $k_{//} = 0$ coincides with the exciton energy of the molecule, the electrical current was shown to increase 10-fold at most, reflecting mostly a change in the carrier mobility. The fundamentals of this enhancement in electrical mobility are not straightforward to derive because the excitons that feed strong coupling are neutral quasiparticles. A theoretical model has suggested that, when losses in the cavity are larger than the bandwidth of electronic bands, the system can enter into a "collective dressing regime," in which strong coupling induces a large increase in the hole conductivity, leading to a substantial increase in the electrical mobility (*28*). More recently, p-type semiconductors have been also studied under ultrastrong coupling, revealing both enhanced conductivity and photoconductivity (*29*, *30*). By measuring changes in the photoconductivity of a p-doped organic semiconductor whose excitons were strongly coupled to the EM modes supported by a FP microcavity, it was found that the bare photoinduced electron transfer events can be modified by the presence of the two polaritonic states (upper and lower polaritons) in the energy spectrum: Moving from positive to negative detuning caused the upper and lower polariton photocurrents to swap their field dependence (*29*).

Inspired by the experimental findings reported in (*27*), two theoretical groups independently predicted in 2015 that energy transport in organic materials could also be enhanced when excitons are strongly coupled to a cavity EM field (*31*, *32*). By using a one-dimensional (1D) model system in which disorder was introduced both in the positions of the emitters and the orientations of their transition dipoles (Fig. 2A), they showed how the formation of a collective strongly coupled mode allows excitons to bypass the disordered array of emitters and jump directly from one end of the structure to the other, concluding that polaritonic states can largely extend the spatial range of energy transport in organic materials. This capability has been experimentally verified several times since then. Room temperature ballistic propagation of excitons in an organic semiconductor was demonstrated for distances well beyond 100 micrometers thanks to their hybridization with propagating Bloch surface EM waves that act as open cavities (*33*). These EM modes appear at the surface of a truncated distributed back reflector, which is just a 1D photonic crystal made of two distinct dielectric materials. Similar results for the propagation length of this type of polaritonic modes have been reported more recently (*34*), as shown in Fig. 2B. Additionally, through analysis of the halo-like pattern of the polariton intensity, a coherence length of 20 micrometers was extracted, demonstrating the coherent character of the polariton-assisted transport of energy—a critical difference with the diffusive process occurring in bare organic materials. Along the same lines, long-range transport of organic excitons strongly coupled to EM modes supported by a FP cavity has been also reported (*35*) but with a much shorter propagation length of only a few micrometers, which could be due to the larger losses present in this type of metallic cavity. Such studies point nevertheless to the long lifetime of the polaritonic modes and to the non-Markovian regime of these systems (Box 1). In a properly designed strongly coupled system, it is also possible to induce efficient energy transfer over distances of 100 nm, well beyond the ~10-nm limit in the Förster regime, as shown both theoretically and experimentally (*36*, *37*). This long-range transport process, as illustrated in Fig. 2C, is discussed in more detail in the section devoted to the modification of molecular properties.

Modification of transport properties induced by strong light-matter coupling has also been reported for inorganic materials (*38*). When a 2D electron gas (2DEG) is inserted in a deeply subwavelength split-ring resonator, the 2DEG magneto-resistivity has been altered in the regime of low magnetic fields (*38*). As described theoretically (*39*), in the linear transport regime, the Drude scattering time is modified when the 2DEG cyclotron frequency is quasi-resonant to one mode of the resonator (Fig. 2D). Additionally, antiresonant interaction processes due to ultrastrong light-matter coupling can produce orbital renormalization effects (*39*) that, for example, are expected to modify the effective carrier mass and carrier hopping properties. Related effects have also been predicted for the vertical transport in semiconductor heterostructures with the possibility of controlling the dark current of quantum well infrared (IR) photodetectors (*40*).

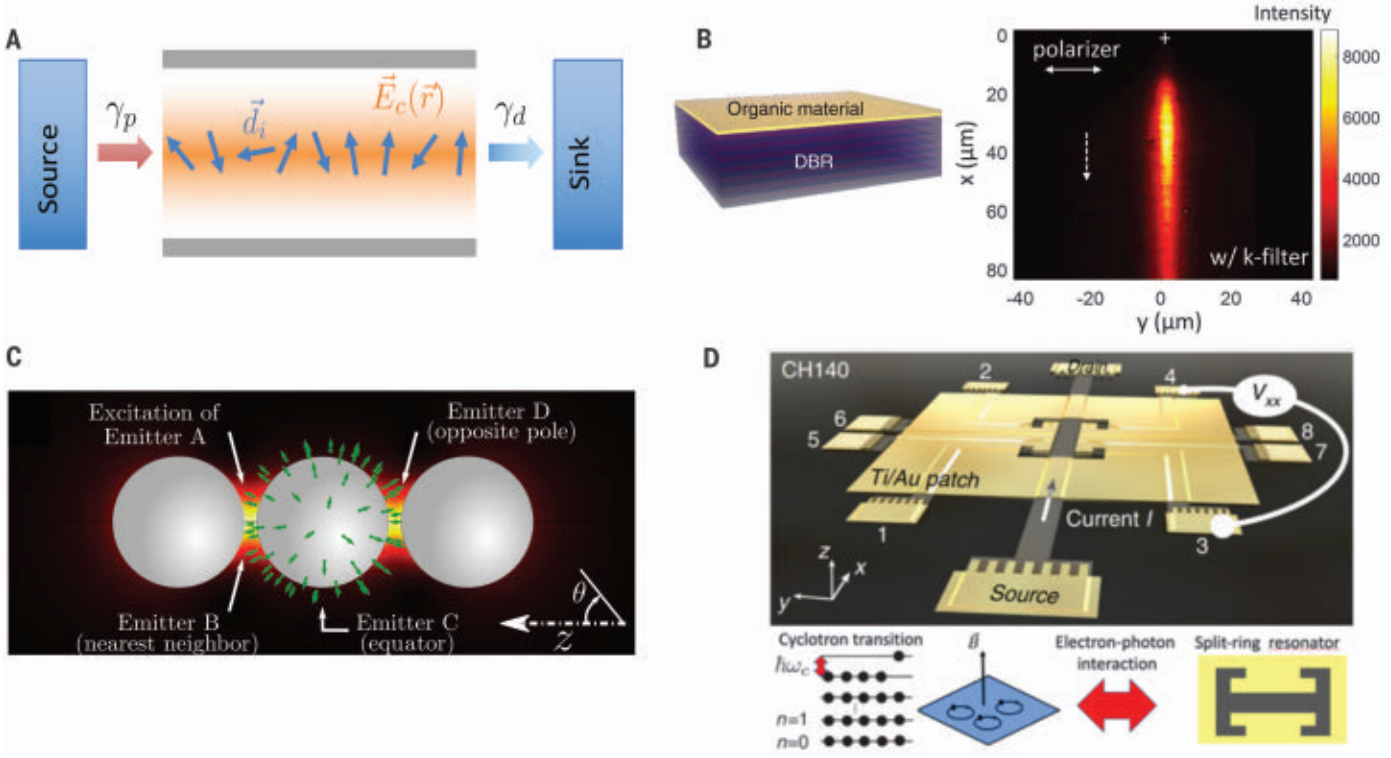


Fig. 2. Cavity-induced modification of transport. (**A**) Sketch of a 1D chain of disordered quantum emitters inside a cavity. Excitons are pumped into the system from the left reservoir, and the exciton current is measured when the excitons reach the sink reservoir on the right. Figure taken from (*31*). (**B**) An organic layer is deposited on a distributed Bragg reflector (DBR) composed of four pairs of ZnS/MgF_2 layers. Real-space photoluminescence image using a k-space filter is shown. The excitation laser spot is located at (0,0), indicated by the plus sign. Figure reproduced with permission from (*34*). (**C**) Illustration of a plasmonic cavity formed by three nanoparticles. A collection of quantum emitters couples strongly with the lowest energy EM mode (colored background), allowing for very efficient transfer of excitons from emitter A to emitter D. Figure reproduced with permission from (*36*). (**D**) The top panel depicts a setup to measure charge magneto-transport of a 2D electronic system coupled to a complementary split-ring resonator. In the presence of a perpendicular magnetic field, the light-matter interaction is responsible for electronic transitions between quantized Landau levels (bottom), whose energy separation is proportional to the cyclotron motion frequency. The bottom-right graphic shows a resonator with a subterahertz "LC" mode with deeply subwavelength spatial confinement, resulting in large vacuum fields coupled to the electrons. Figure reproduced with permission from (*38*).

Manipulating condensed phases of matter

Light-induced modification of material properties in the condensed phase dates back to pioneering experiments by Wyatt and Dayem (*41*, *42*), who observed that the critical current in BCS (Bardeen-Cooper-Schrieffer) (phonon-driven) superconductors could be increased when they are illuminated by coherent microwave radiation. More-recent pump-probe experiments have also shown that materials subjected to very intense terahertz pulses could exhibit transient superconducting properties at much higher temperatures than at equilibrium (*43*). This paradigm has led to the notion of quantum materials and the emergence of new properties through collective interactions (*44*–*46*). With the demonstration that material properties could be modified under strong coupling, several research groups have proposed replacement of the classical high-intensity light field by large vacuum EM fields obtained by tight spatial confinement—i.e., coupling of the system to the EM mode of a cavity. This approach holds the promise of engineering a variety of material properties, not just transport, in the steady state and without the transient nature and excess heating associated with illumination by intense laser fields.

In this context, the effect of a cavity resonator on superconductivity has been explored both theoretically and experimentally. For example, a theoretical study has suggested that the exchange of virtual cavity photons could produce a pairing mechanism and lead to cavity-mediated superconductivity of a 2DEG (*47*). Other studies have discovered that, although strong coupling can lead to an enhancement of the electron-phonon interaction responsible for superconductivity, this change could not always translate to an increase of the superconducting critical temperature (*48*). Experimentally, two different superconductors [Rb_3C_{60} and $YBa_2Cu_3O_{7-x}$ (YBCO)] have been tested in powder form and dispersed in various polymers (*49*). Whereas electron-electron pairing is driven by phonons in the case of Rb_3C_{60}, the origin of superconductivity in YBCO is proposed to be caused by antiferromagnetic spin fluctuations. The direct coupling of the phonon mode of Rb_3C_{60} that drives its superconducting behavior with the IR EM modes of the cavity is very weak. To increase the coupling, a cooperative strong coupling mechanism (Box 1) was used whereby the strong vibrational bands of polystyrene that are resonant with that phonon mode of Rb_3C_{60} act as a mediator in the phonon-EM mode interaction, leading to vibrational strong coupling (VSC). An increase of the critical temperature, T_C, from 30 to 45 K was observed for the case of Rb_3C_{60}. This was interpreted to be a result of the enhancement of the electron-phonon interaction induced by strong coupling, in accord with the theoretical studies discussed before (*47*, *48*). However, T_C in YBCO is decreased from 92 to 86 K when the phonon mode at ~700 cm^{-1} of the apical oxygen atoms of YBCO is coupled to the EM modes of the cavity. Further studies show that this is due to a 700-fold enhancement of the ferromagnetism of the YBCO particles under strong coupling, which competes with the

superconductivity at low temperatures (*50*). Further experiments and accurate theories need to be developed to elucidate the role of strong coupling in modifying such superconducting and magnetic properties.

In addition to superconductivity, the influence of vacuum EM fields on other quantum many-body phenomena has been also analyzed theoretically. It has been suggested that the interplay between electron-electron interaction and electron-photon strong coupling, as realized within a cavity, can lead to the formation of a "superradiant excitonic insulator" (*51*). Superradiant phases are characterized by a simultaneous condensation of excitons in the electronic system and photons in a given EM mode (*52*). On the other hand, it has been shown that resonant coupling between strongly correlated electrons and a single-mode cavity could result in the formation of new types of polaritonic states, which could help manipulate the insulating-conducting character of the electronic system (*53*). Along the same line, intertwined orders of strongly correlated electronic systems, such as charge density waves and different types of superconductivity, may possibly be selectively favored by a proper manipulation of their (strong) coupling to the EM modes of a cavity (*54*). As a final example, the possibility of inducing a phase transition from a paraelectric material to a ferroelectric one by just incorporating the bare material into a FP cavity has also been suggested (*55*). Apart from the exciting avenues for research that all these theoretical proposals have generated in recent times, they have also triggered an intense, sometimes controversial, but fruitful debate on how to model the interaction between a macroscopic quantum many-body system and a macroscopic cavity that can support many EM modes. In particular, we note the problems related to gauge invariance when describing the light-matter coupling in approximated models (*56*, *57*) and the role of spatial inhomogeneity of the cavity EM modes in the occurrence of phase transitions affecting the ground state (*58*, *59*).

Chemistry under strong coupling: QED or polaritonic chemistry

A study of molecules under electronic strong coupling (ESC) was reported as long ago as 1982 (*16*), but it is only in the last decade that the consequences of strong coupling for molecular properties and their ensuing chemical reactivity have been explored. Under ESC, photophysical properties are obviously modified by splitting, for instance, the first excited state into P+, P−, and DS. Emission quantum yields (*60*), intersystem crossing (*61*, *62*), singlet fission (*63*), and lifetimes (*60*, *64*) are affected, and experiments such as pump-probe cannot be carried out as straightforwardly as measurements outside cavities because of the mode structure and the high reflectivity of the cavity optics (*64*). Additionally, resonant Rayleigh scattering is substantially enhanced as a result of the collective delocalized nature of the polaritonic states (*65*) that must be considered in the interpretation of results. Recall that there are $3n - 6$ vibrational modes for nonlinear molecules with n atoms, and, together with their harmonics, they form nearly a continuum of sublevels between P+, DS, and P− (see discussion on this issue in Box 1). As a consequence, at room temperature, only emission from P− is observed (*66*–*68*) because P+ decays too fast through nonradiative channels, mostly driven by internal vibration relaxation and dissipation into the surrounding bath. The P− emission quantum yield is only compatible with a long-lived emitter (*60*). Polariton propagation length also indicates a long lifetime (*34*, *35*). Such findings all point to a non-Markovian regime (Box 1).

Here, we will focus primarily on the consequences of strong coupling for chemical reactivity. The first demonstration of modified chemistry under strong coupling involved coupling an electronic transition of a photochrome (a molecule that photoisomerizes between two forms of different color) in a solid polymer matrix (*22*, *69*). The uncolored spiropyran was dissolved in a polymer solution and spin-coated on an Ag mirror so that when a second mirror was placed on the film, the resulting cavity was tuned so that it had a mode that could be coupled at normal incidence to its isomer, the colored merocyanine molecule, having a peak at 590 nm. As the sample was irradiated at its isobestic point (~326 nm), the evolution toward a photostationary state was monitored. As the reaction proceeded and merocyanine accumulated, the strong coupling increased as $\sqrt{N}$, the reaction slowed down markedly, and its quantum yield increased. This proof of principle that reactivity could be modified under strong coupling came as a surprise and generated much interest that led to numerous additional studies both theoretical and experimental (*70*–*114*). Theoretical studies (*81*, *82*) of model molecules, such as stilbene and azobenzene, agree with the earlier experimental results, which indicates that the distribution of a single excitation over many molecules effectively suppresses such photoisomerization (*81*). These theoretical studies and others (*70*, *71*, *73*–*80*) show that the potential energy surfaces (PES) of the various states at play are modified under strong coupling, which affects conical intersections and internal dynamics, as first pointed out in 2015 (*70*). Not surprisingly, not only chemical dynamics but also bond-lengths and charge density distributions are subject to modification (*73*, *76*). Electron transfer reactions are predicted to be strongly affected—either enhanced or suppressed depending on the exact conditions of the reaction as a result of the decoupling between the electronic and nuclear degrees of freedom (*71*).

Experimentally, the photodegradation rate of different molecules has been studied and found to be suppressed (*84*, *85*). Photodegradation is typically a photo-oxidation process stemming from the reaction of O_2 with the long-lived triplet state of the molecules that must be populated competitively within the lifetime of the polaritonic states. The shorter this lifetime, the more photo-oxidation will be suppressed. Energy transfer between donor and acceptor molecules has also been studied by coupling both the donor and the acceptor to the same optical mode (*36*, *37*, *86*–*89*). It has been shown that the rate and the yields increase substantially in such conditions (*37*, *87*), which has inspired further experiments with vibrational energy transfer (*89*). Because the donor and acceptor molecules become quantum-mechanically entangled in this scheme, the possibility of separating the donor and acceptor molecules by a spacer layer was investigated, and energy transfer remained very efficient—well beyond the 10-nm limit typical of FRET (*37*). In fact, energy transfer enters a new regime that is no longer dependent on distance as long as the strong coupling condition is met (*36*, *37*, *88*). In the experimental study (*37*), it was also suggested that such a scheme could open the door to investigate chemical and molecular processes under entanglement—a prospect that was then analyzed theoretically (*90*).

The work reviewed above confirms the rich possibilities to modify photochemical and perhaps even ground-state thermally driven chemical reactions under ESC because ultrastrong coupling should change the ground state by lending it photonic character (*5*), even if the possibility of modifying ground-state chemistry has been questioned (*83*). Chemistry under ESC suffers nevertheless from one technical limitation—namely, that the cavity has to be resonant in the visible spectrum, which implies, when taking into account the refractive index of the material, that the cavity path length must be in the submicrometer range to avoid technical problems. Solution chemistry thus becomes extremely difficult. Not surprisingly, all of the experimental studies of chemical reactions under ESC have been done in solid solutions.

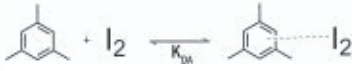
In a 2012 paper (*22*), a way to overcome these limitations was proposed—namely, to couple vibrational transitions in the IR and thereby also influence ground-state chemistry. This wavelength regime is compatible with microfluidic cells that have cavity path lengths in the 10-μm range. Through the use of flexible polymer spacers between the two mirrors, the cavity can be tuned to resonance with a given vibration simply by turning a screwdriver

Fig. 3. QED or polaritonic chemistry in a microfluidic optical cavity. Photograph of a microfluidic IR FP cavity that can be tuned with a screwdriver to put it in resonance with a molecular vibration to achieve strong coupling (courtesy of T. W. Ebbesen).

(Fig. 3). Reactants can be injected into the cavity and the products extracted for further analysis using standard equipment, such as mass spectrometry and nuclear magnetic resonance.

Initial studies focused on establishing VSC (*91*–*94*), and it soon became apparent that VSC could have a substantial effect on chemical reactivity (*95*). Not only did the first reaction studied slow down but, more importantly, the enthalpy and entropy of activation both changed by more than 50 kJ/mole at room temperature, and the entropy changed sign, suggesting a change of reaction mechanism. The large modification of the activation energies came as a surprise because they are ~20 times as large as the collective Rabi splitting on the order of k_BT (2.6 kJ/mole) (where k_B is the Boltzmann constant). Since then, a number of other reactions have been studied that show either deceleration or acceleration of the reaction rate, reflecting very large variations in the activation energies under VSC (*95*–*99*, *115*). Chemical landscapes can be tilted toward a desired product, and only VSC of vibrations closely linked to the bonds breaking induce an effect (*97*). Thus, by tuning cavities across the various vibrational bands of a reactant, VSC can be useful as a tool for elucidating reaction mechanisms. Theoretical studies of chemistry under VSC (*100*–*106*, *111*–*114*) have come to very different conclusions as to why reactions are modified. Notably, the variation in the density of states or the correction to the partition function (*105*, *106*) cannot account for the observed changes or the small magnitude of the Rabi splitting. Something else has to be modified under VSC to explain the large effect on the thermodynamics, and the most likely factor is symmetry. It is well known that symmetry correlation diagrams between reactants and products play a key role in determining the PES of the reactivity landscape (*108*).

To investigate the role of symmetry in VSC, a very simple charge transfer (CT) equilibrium reaction between mesitylene and I_2 (Eq. 1)



(1)

was studied for this purpose (*99*). This type of reaction, which has been investigated since the 1940s, has the advantage that the formation of the CT complex has a distinct and strong peak in the ultraviolet, which can be monitored to extract the equilibrium constant K_{DA}, all the while coupling the vibrational bands of mesitylene in the IR. Furthermore, mesitylene is highly symmetric and rigid, so the symmetry representations of the vibrations are well defined. Under VSC, the CT equilibrium favors complexation or decomplexation depending on the symmetry class of the vibration, as illustrated in Fig. 4A. The shift in K_{DA} is essentially independent of the type of vibration, its energy, and the Rabi splitting. Again, the thermodynamics (now the relative energy of the reactants and product, i.e., the Gibbs free energy, enthalpy, and entropy) are strongly modified by VSC (*99*). This result shows the central role of symmetry in VSC: The coupling acts on the symmetry, thereby modifying the electronic PES. This in turn implies that the vibronic coupling (interaction between the vibrational and electronic manifolds of the molecule) remains very strong under VSC, unlike for ESC, where they are expected to become decoupled (*71*). The complexation reaction also reveals the transition from weak to strong coupling, akin to a phase transition, under variation of the mesitylene concentration, as can be seen in Fig. 4B. There is an abrupt change in the slope, reflecting the new K_{DA} and the change in the absorption cross section of the complex upon VSC. This also indicates that the reservoir of uncoupled molecules must be very small (Box 1).

Further studies confirm that symmetry plays a very important role in chemistry under VSC (*109*), but obviously other factors must also be involved. The modified shape of the PES, as seen in thermodynamic data, will influence the outcome and rate of the reaction. When either the solute, the solvent, or both are coupled under VSC, solvation must also be affected in view of the consequences for crystallization and enzymatic activity (*107*, *110*) and are probably the result of changes in the intermolecular interactions under strong coupling (*73*, *111*, *112*, *114*). Moreover, when molecules are large and floppy with ill-defined symmetry, the signature of symmetry is not obvious and can be obscured by such factors. The effect of VSC will probably be undetectable if it does not influence some limiting step in the reaction trajectory. The role of entropic reordering of energy levels of the DS could very well also play an important role in the chemistry (*116*).

Outlook

Matter owes many of its properties, such as spontaneous photon emission, Lamb spectral shifts, and Casimir and van der Waals forces, to the interaction with the vacuum EM field. The use of EM cavities or plasmonic resonators provides a platform to strengthen this interaction to the point that fundamental properties are modified by the introduction of hybrid light-matter states in the coupled material. This approach to manipulating the properties of matter, including modifying chemical reactivity and processes, appears very promising from the results that have been accumulating over the past decade. The interest is no longer purely fundamental, as technological applications should be quite straight forward to implement if there is an advantage. For example, some industrial chemical reactions are not optimal in terms of yields and reaction conditions despite decades of effort using traditional methods. Chemistry under VSC provides a new approach or tool to control chemical reactivity that could be scaled up with massively parallel microfluidic systems. It should not be very complicated to integrate confined optical fields in devices to modify solid-state properties, such as conductivity or magnetism.

Perhaps the biggest surprise of all recent developments in this field has been the magnitude of the effect of collective VSC on

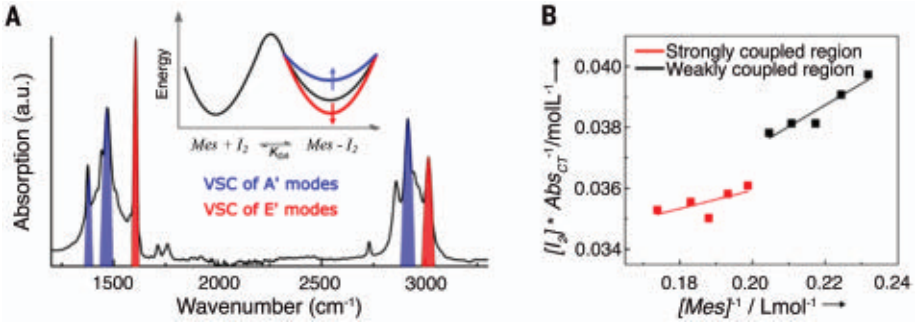


Fig. 4. VSC and symmetry. (**A**) IR absorption spectrum of the mesitylene coupled vibrations with their symmetries *E′* (red) and *A′* (blue), together with the consequences on the CT equilibrium landscape of the mesitylene (Mes)–I_2 complexation process in the inset. a.u., arbitrary units. (**B**) Change in the absorption of CT complex versus inverse mesitylene concentration, revealing the abrupt change at the transition from weak to strong coupling regimes. Figure reproduced with permission from (*99*).

ground-state chemistry. Strong coupling in general, but VSC in particular, is a new approach to chemistry. VSC can also be used as a mechanistic tool (*97*). Although concentration of the coupled species was a limiting factor, the introduction of cooperative strong coupling has opened the door to further advances in the field because reactants can now be coupled at low concentration through, for example, the solvent (*115*, *117*). There are many classes of organic reactions that are currently being explored under strong coupling. As more results are collected, the underlying physical chemistry will be further clarified and should lead to some general principles to guide chemists in their use of VSC. The possibilities of chemistry under entanglement through the cavity field could also open new directions for research. The demonstration that simply coupling water modifies enzyme activity (*96*, *110*) illustrates the potential for studying biological activity under strong coupling that remains otherwise unexplored. Here, the role of coherence in collective coupling (*118*, *119*) could open perspectives for quantum biology (*120*, *121*).

One of the open questions is whether the properties of solvents, such as water, are also modified upon strong coupling. With this idea in mind, selective crystallization of MOFs (metal-organic frameworks) has been demonstrated (*107*), opening an exciting avenue to reduce crystal polymorphism. This approach could potentially be used to favor one supramolecular assembly over another, with direct consequences for the dynamics and hierarchical organization of (bio)molecular materials. The exploration of weak intermolecular interactions under strong coupling with terahertz spectroscopy could reveal such modifications (*122*).

The cavity enhancement of solid-state material properties has already been shown for charge and energy transport, nonlinear processes, magnetism, and superconductivity, as discussed above. In general, phonon-based phase transitions should be good targets if the phonon bands have sufficiently large oscillator strength to be coupled to a cavity mode. The possibility of inducing new phases should also be explored (*55*, *123*). Particularly, 2D materials are well suited to be integrated in cavity resonators with deeply subwavelength photon confinement (*38*, *124*). Among the possible directions, we note quantum Hall systems and superconductivity in 2D van der Waals materials in particular (*125*–*127*). Clearly, strong coupling is a broad frontier, with implications and potential for fundamental science as well as technological applications.

REFERENCES AND NOTES

1. P. A. M. Dirac, The quantum theory of the emission and absorption of radiation. *Proc. R. Soc. Lond. A* **114**, 243–265 (1927). doi: 10.1098/rspa.1927.0039
2. E. M. Purcell, Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
3. E. T. Jaynes, F. W. Cummings, Comparison of quantum and semiclassical radiation theories with application to the beam maser. *Proc. IEEE* **51**, 89–109 (1963). doi: 10.1109/PROC.1963.1664
4. J. J. Hopfield, Theory of the contribution of excitons to the complex dielectric constant of crystals. *Phys. Rev.* **112**, 1555–1567 (1958). doi: 10.1103/PhysRev.112.1555
5. C. Ciuti, G. Bastard, I. Carusotto, Quantum vacuum properties of the intersubband cavity polariton field. *Phys. Rev. B* **72**, 115303 (2005). doi: 10.1103/PhysRevB.72.115303
6. P. Forn-Díaz, L. Lamata, E. Rico, J. Kono, E. Solano, Ultrastrong coupling regimes of light-matter interaction. *Rev. Mod. Phys.* **91**, 025005 (2019). doi: 10.1103/RevModPhys.91.025005
7. A. F. Kockum, A. Miranowicz, S. De Liberato, S. Savasta, F. Nori, Ultrastrong coupling between light and matter. *Nat. Rev. Phys.* **1**, 19–40 (2019). doi: 10.1038/s42254-018-0006-2
8. M. Tavis, F. W. Cummings, Exact solution for an N-molecule—Radiation-field Hamiltonian. *Phys. Rev.* **170**, 379–384 (1968). doi: 10.1103/PhysRev.170.379
9. C. Gonzalez-Ballestero, J. Feist, E. Gonzalo Badía, E. Moreno, F. J. Garcia-Vidal, Uncoupled dark states can inherit polaritonic properties. *Phys. Rev. Lett.* **117**, 156402 (2016). doi: 10.1103/PhysRevLett.117.156402; pmid: 27768353
10. S. Haroche, D. Kleppner, Cavity Quantum Electrodynamics. *Phys. Today* **42**, 24–30 (1989). doi: 10.1063/1.881201
11. V. M. Agranovich, A. G. Malshukov, Surface polariton spectra if the resonance with the transition layer vibrations exist. *Opt. Commun.* **11**, 169–171 (1974). doi: 10.1016/0030-4018(74)90211-9
12. V. A. Yakovlev, V. G. Nazin, G. N. Zhizhin, The surface polariton splitting due to thin surface film LO vibrations. *Opt. Commun.* **15**, 293–295 (1975). doi: 10.1016/0030-4018(75)90306-5
13. C. Weisbuch, M. Nishioka, A. Ishikawa, Y. Arakawa, Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992). doi: 10.1103/PhysRevLett.69.3314; pmid: 10046787
14. D. Snoke, P. Littlewood, Polariton condensates. *Phys. Today* **63**, 42–47 (2010). doi: 10.1063/1.3480075
15. I. Carusotto, C. Ciuti, Quantum fluids of light. *Rev. Mod. Phys.* **85**, 299–366 (2013). doi: 10.1103/RevModPhys.85.299
16. I. Pockrand, A. Brillante, D. Möbius, Exciton–surface plasmon coupling: An experimental investigation. *J. Phys. Chem.* **77**, 6289–6295 (1982). doi: 10.1063/1.443834
17. T. Fujita, Y. Sato, T. Kuitani, T. Ishihara, Tunable polariton absorption of distributed feedback microcavities at room temperature. *Phys. Rev. B* **57**, 12428–12434 (1998). doi: 10.1103/PhysRevB.57.12428
18. D. G. Lidzey *et al*., Strong exciton–photon coupling in an organic semiconductor microcavity. *Nature* **395**, 53–55 (1998). doi: 10.1038/25692
19. S. Kéna-Cohen, S. R. Forrest, Room-temperature polariton lasing in an organic single-crystal microcavity. *Nat. Photonics* **4**, 371–375 (2010). doi: 10.1038/nphoton.2010.86
20. J. D. Plumhof, T. Stöferle, L. Mai, U. Scherf, R. F. Mahrt, Room-temperature Bose-Einstein condensation of cavity exciton-polaritons in a polymer. *Nat. Mater.* **13**, 247–252 (2014). doi: 10.1038/nmat3825; pmid: 24317189
21. K. S. Daskalakis, S. A. Maier, R. Murray, S. Kéna-Cohen, Nonlinear interactions in an organic polariton condensate. *Nat. Mater.* **13**, 271–278 (2014). doi: 10.1038/nmat3874; pmid: 24509602
22. J. A. Hutchison, T. Schwartz, C. Genet, E. Devaux, T. W. Ebbesen, Modifying chemical landscapes by coupling to vacuum fields. *Angew. Chem. Int. Ed.* **51**, 1592–1596 (2012). doi: 10.1002/anie.201107033; pmid: 22234987
23. P. Törmä, W. L. Barnes, Strong coupling between surface plasmon polaritons and emitters: A review. *Rep. Prog. Phys.* **78**, 013901 (2015). doi: 10.1088/0034-4885/78/1/013901; pmid: 25536670
24. N. Tessler, Y. Preezant, N. Rappaport, Y. Roichman, Charge transport in disordered organic materials and its relevance to thin-film devices: A tutorial review. *Adv. Mater.* **21**, 2741–2761 (2009). doi: 10.1002/adma.200803541
25. J. Zaumseil, H. Sirringhaus, Electron and ambipolar transport in organic field-effect transistors. *Chem. Rev.* **107**, 1296–1323 (2007). doi: 10.1021/cr0501543; pmid: 17378616
26. W. Cao, J. Xue, Recent progress in organic photovoltaics: Device architecture and optical design. *Energy Environ. Sci.* **7**, 2123–2144 (2014). doi: 10.1039/c4ee00260a
27. E. Orgiu *et al*., Conductivity in organic semiconductors hybridized with the vacuum field. *Nat. Mater.* **14**, 1123–1129 (2015). doi: 10.1038/nmat4392; pmid: 26366850
28. D. Hagenmüller, J. Schachenmayer, S. Schütz, C. Genes, G. Pupillo, Cavity-enhanced transport of charge. *Phys. Rev. Lett.* **119**, 223601 (2017). doi: 10.1103/PhysRevLett.119.223601; pmid: 29286774
29. N. Krainova, A. J. Grede, D. Tsokkou, N. Banerji, N. C. Giebink, Polaron photoconductivity in the weak and strong light-matter coupling regime. *Phys. Rev. Lett.* **124**, 177401 (2020). doi: 10.1103/PhysRevLett.124.177401; pmid: 32412265
30. K. Nagarajan *et al*., Conductivity and photoconductivity of a p-type organic semiconductor under ultrastrong coupling. *ACS Nano* **14**, 10219–10225 (2020). doi: 10.1021/acsnano.0c03496; pmid: 32806034
31. J. Feist, F. J. Garcia-Vidal, Extraordinary exciton conductance induced by strong coupling. *Phys. Rev. Lett.* **114**, 196402 (2015). doi: 10.1103/PhysRevLett.114.196402; pmid: 26024185
32. J. Schachenmayer, C. Genes, E. Tignone, G. Pupillo, Cavity-enhanced transport of excitons. *Phys. Rev. Lett.* **114**, 196403 (2015). doi: 10.1103/PhysRevLett.114.196403; pmid: 26024186
33. G. Lerario *et al*., High-speed flow of interacting organic polaritons. *Light Sci. Appl.* **6**, e16212 (2017). doi: 10.1038/lsa.2016.212; pmid: 30167229
34. S. Hou *et al*., Ultralong-range energy transport in a disordered organic semiconductor at room temperature via coherent exciton-polariton propagation. *Adv. Mater.* **32**, e2002127 (2020). doi: 10.1002/adma.202002127; pmid: 32484288
35. G. G. Rozenman, K. Akulov, A. Golombek, T. Schwartz, Long-range transport of organic exciton-polaritons revealed

by ultrafast microscopy. *ACS Photonics* **5**, 105–110 (2018). doi: 10.1021/acsphotonics.7b01332

36. C. Gonzalez-Ballestero, J. Feist, E. Moreno, F. J. Garcia-Vidal, Harvesting excitons through plasmonic strong coupling. *Phys. Rev. B* **92**, 121402 (2015). doi: 10.1103/PhysRevB.92.121402
37. X. Zhong *et al*., Energy transfer between spatially separated entangled molecules. *Angew. Chem. Int. Ed.* **56**, 9034–9038 (2017). doi: 10.1002/anie.201703539; pmid: 28598527
38. G. L. Paravicini-Bagliani *et al*., Magneto-transport controlled by Landau polariton states. *Nat. Phys.* **15**, 186–190 (2019). doi: 10.1038/s41567-018-0346-y
39. N. Bartolo, C. Ciuti, Vacuum-dressed cavity magnetotransport of a two-dimensional electron gas. *Phys. Rev. B* **98**, 205301 (2018). doi: 10.1103/PhysRevB.98.205301
40. C. Naudet-Baulieu, N. Bartolo, G. Orso, C. Ciuti, Dark vertical conductance of cavity-embedded semiconductor heterostructures. *New J. Phys.* **21**, 093061 (2019). doi: 10.1088/1367-2630/ab41c2
41. A. F. G. Wyatt, V. M. Dmitriev, W. S. Moore, F. W. Sheard, Microwave-enhanced critical supercurrents in constricted tin films. *Phys. Rev. Lett.* **16**, 1166–1169 (1966). doi: 10.1103/PhysRevLett.16.1166
42. A. H. Dayem, J. J. Wiegand, Behavior of thin-film superconducting bridges in a microwave field. *Phys. Rev.* **155**, 419–428 (1967). doi: 10.1103/PhysRev.155.419
43. R. Mankowsky *et al*., Nonlinear lattice dynamics as a basis for enhanced superconductivity in $YBa_2Cu_3O_{6.5}$. *Nature* **516**, 71–73 (2014). doi: 10.1038/nature13875; pmid: 25471882
44. J. Orenstein, Ultrafast spectroscopy of quantum materials. *Phys. Today* **65**, 44–50 (2012). doi: 10.1063/PT.3.1717
45. Y. Tokura, M. Kawasaki, N. Nagaosa, Emergent functions of quantum materials. *Nat. Phys.* **13**, 1056–1068 (2017). doi: 10.1038/nphys4274
46. D. N. Basov, R. D. Averitt, D. Hsieh, Towards properties on demand in quantum materials. *Nat. Mater.* **16**, 1077–1088 (2017). doi: 10.1038/nmat5017; pmid: 29066824
47. F. Schlawin, A. Cavalleri, D. Jaksch, Cavity-mediated electron-photon superconductivity. *Phys. Rev. Lett.* **122**, 133602 (2019). doi: 10.1103/PhysRevLett.122.133602; pmid: 31012600
48. M. A. Sentef, M. Ruggenthaler, A. Rubio, Cavity quantum-electrodynamical polaritonically enhanced electron-phonon coupling and its influence on superconductivity. *Sci. Adv.* **4**, eaau6969 (2018). doi: 10.1126/sciadv.aau6969; pmid: 30515456
49. A. Thomas, E. Devaux, K. Nagarajan, T. Chervy, M. Seidel, D. Hagenmüller, S. Schütz, J. Schachenmayer, C. Genet, G. Pupillo, T. W. Ebbesen, Exploring Superconductivity under Strong Coupling with the Vacuum Electromagnetic Field. arXiv:1911.01459 [cond-mat.supr-con] (2019).
50. A. Thomas *et al*., Large enhancement of ferromagnetism under collective strong coupling of YBCO nanoparticles. *Nano Lett.* **21**, 4365–4370 (2021). doi: 10.1021/acs.nanolett.1c00973; pmid: 33945283
51. G. Mazza, A. Georges, Superradiant quantum materials. *Phys. Rev. Lett.* **122**, 017401 (2019). doi: 10.1103/PhysRevLett.122.017401; pmid: 31012726
52. R. H. Dicke, Coherence in spontaneous radiation processes. *Phys. Rev.* **93**, 99–110 (1954). doi: 10.1103/PhysRev.93.99
53. M. Kiffner, J. Coulthard, F. Schlawin, A. Ardavan, D. Jaksch, Mott polaritons in cavity-coupled quantum materials. *New J. Phys.* **21**, 073066 (2019). doi: 10.1088/1367-2630/ab31c7
54. J. Li, M. Eckstein, Manipulating intertwined orders in solids with quantum light. *Phys. Rev. Lett.* **125**, 217402 (2020). doi: 10.1103/PhysRevLett.125.217402; pmid: 33275019
55. Y. Ashida *et al*., Quantum electrodynamic control of matter: Cavity-enhanced ferroelectric phase transition. *Phys. Rev. X* **10**, 041027 (2020). doi: 10.1103/PhysRevX.10.041027
56. D. De Bernardis, P. Pilar, T. Jaako, S. De Liberato, P. Rabl, Breakdown of gauge invariance in ultrastrong-coupling cavity QED. *Phys. Rev. A* **98**, 053819 (2018). doi: 10.1103/PhysRevA.98.053819
57. O. Di Stefano *et al*., Resolution of gauge ambiguities in ultrastrong-coupling cavity quantum electrodynamics. *Nat. Phys.* **15**, 803–808 (2019). doi: 10.1038/s41567-019-0534-4
58. G. M. Andolina, F. M. D. Pellegrino, V. Giovannetti, A. H. MacDonald, M. Polini, Theory of photon condensation in a spatially varying electromagnetic field. *Phys. Rev. B* **102**, 125137 (2020). doi: 10.1103/PhysRevB.102.125137
59. D. Guerci, P. Simon, C. Mora, Superradiant phase transition in electronic systems and emergent topological phases. *Phys. Rev. Lett.* **125**, 257604 (2020). doi: 10.1103/PhysRevLett.125.257604; pmid: 33416347
60. S. Wang *et al*., Quantum yield of polariton emission from hybrid light-matter states. *J. Phys. Chem. Lett.* **5**, 1433–1439 (2014). doi: 10.1021/jz5004439; pmid: 26269990
61. K. Stranius, M. Hertzog, K. Börjesson, Selective manipulation of electronically excited states through strong light-matter interactions. *Nat. Commun.* **9**, 2273 (2018). doi: 10.1038/s41467-018-04736-1; pmid: 29891958
62. E. Eizner, L. A. Martínez-Martínez, J. Yuen-Zhou, S. Kéna-Cohen, Inverting singlet and triplet excited states using strong light-matter coupling. *Sci. Adv.* **5**, eaax4482 (2019). doi: 10.1126/sciadv.aax4482; pmid: 31840063
63. D. Polak *et al*., Manipulating molecules with strong coupling: Harvesting triplet excitons in organic exciton microcavities. *Chem. Sci.* **11**, 343–354 (2019). doi: 10.1039/C9SC04950A; pmid: 32190258
64. T. Schwartz *et al*., Polariton dynamics under strong light-molecule coupling. *ChemPhysChem* **14**, 125–131 (2013). doi: 10.1002/cphc.201200734; pmid: 23233286
65. A. Golombek, M. Balasubrahmaniyam, M. Kaeek, K. Hadar, T. Schwartz, Collective Rayleigh scattering from molecular ensembles under strong coupling. *J. Phys. Chem. Lett.* **11**, 3803–3808 (2020). doi: 10.1021/acs.jpclett.0c01012; pmid: 32329347
66. J. Bellessa, C. Bonnand, J. C. Plenet, J. Mugnier, Strong coupling between surface plasmons and excitons in an organic semiconductor. *Phys. Rev. Lett.* **93**, 036404 (2004). doi: 10.1103/PhysRevLett.93.036404; pmid: 15323846
67. J. George *et al*., Ultra-strong coupling of molecular materials: Spectroscopy and dynamics. *Faraday Discuss.* **178**, 281–294 (2015). doi: 10.1039/C4FD00197D; pmid: 25719536
68. S. Baieva, O. Hakamaa, G. Groenhof, T. T. Heikkilä, J. J. Toppari, Dynamics of strongly coupled modes between surface plasmon polaritons and photoactive molecules: The effect of the Stokes shift. *ACS Photonics* **4**, 28–37 (2017). doi: 10.1021/acsphotonics.6b00482
69. T. Schwartz, J. A. Hutchison, C. Genet, T. W. Ebbesen, Reversible switching of ultrastrong light-molecule coupling. *Phys. Rev. Lett.* **106**, 196405 (2011). doi: 10.1103/PhysRevLett.106.196405; pmid: 21668181
70. J. Galego, F. J. Garcia-Vidal, J. Feist, Cavity-induced modifications of molecular structure in the strong coupling regime. *Phys. Rev. X* **5**, 041022 (2015). doi: 10.1103/PhysRevX.5.041022
71. F. Herrera, F. C. Spano, Cavity-controlled chemistry in molecular ensembles. *Phys. Rev. Lett.* **116**, 238301 (2016). doi: 10.1103/PhysRevLett.116.238301; pmid: 27341263
72. P. Zeng *et al*., Photoinduced electron transfer in the strong coupling regime: Waveguide-plasmon polaritons. *Nano Lett.* **16**, 2651–2656 (2016). doi: 10.1021/acs.nanolett.6b00310; pmid: 26963038
73. J. Flick, M. Ruggenthaler, H. Appel, A. Rubio, Atoms and molecules in cavities, from weak to strong coupling in quantum-electrodynamics (QED) chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3026–3034 (2017). doi: 10.1073/pnas.1615509114; pmid: 28275094
74. C. Schäfer, M. Ruggenthaler, H. Appel, A. Rubio, Modification of excitation and charge transfer in cavity quantum-electrodynamical chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4883–4892 (2019). doi: 10.1073/pnas.1814178116; pmid: 30733295
75. J. Feist, J. Galego, F. J. Garcia-Vidal, Polaritonic chemistry with organic molecules. *ACS Photonics* **5**, 205–216 (2018). doi: 10.1021/acsphotonics.7b00680
76. J. Flick, C. Schäfer, M. Ruggenthaler, H. Appel, A. Rubio, Ab initio optimized effective potentials for real molecules in optical cavities: Photon contributions to the molecular ground state. *ACS Photonics* **5**, 992–1005 (2018). doi: 10.1021/acsphotonics.7b01279; pmid: 29594185
77. B. Gu, S. Mukamel, Cooperative conical intersection dynamics of two pyrazine molecules in an optical cavity. *J. Phys. Chem. Lett.* **11**, 5555–5562 (2020). doi: 10.1021/acs.jpclett.0c00381; pmid: 32531166
78. T. Szidarovszky, G. J. Halász, A. G. Császár, L. S. Cederbaum, Á. Vibók, Conical intersections induced by quantum light: Field-dressed spectra from the weak to the ultrastrong coupling regimes. *J. Phys. Chem. Lett.* **9**, 6215–6223 (2018). doi: 10.1021/acs.jpclett.8b02609; pmid: 30296095
79. G. Groenhof, J. J. Toppari, Coherent light harvesting through strong coupling to confined light. *J. Phys. Chem. Lett.* **9**, 4848–4851 (2018). doi: 10.1021/acs.jpclett.8b02032; pmid: 30085671
80. K. Bennett, M. Kowalewski, S. Mukamel, Novel photochemistry of molecular polaritons in optical cavities. *Faraday Discuss.* **194**, 259–282 (2016). doi: 10.1039/C6FD00095A; pmid: 27711849
81. J. Galego, F. J. Garcia-Vidal, J. Feist, Suppressing photochemical reactions with quantized light fields. *Nat. Commun.* **7**, 13841 (2016). doi: 10.1038/ncomms13841; pmid: 27941754
82. J. Fregoni, G. Granucci, M. Persico, S. Corni, Strong coupling with light enhances the photoisomerization quantum yield of azobenzene. *Chem* **6**, 250–265 (2020). doi: 10.1016/j.chempr.2019.11.001
83. L. A. Martínez-Martínez, R. F. Ribeiro, J. Campos-Gonzalez-Angulo, J. Yuen-Zhou, Can ultrastrong coupling change ground-state chemical reactions? *ACS Photonics* **5**, 167–176 (2018). doi: 10.1021/acsphotonics.7b00610
84. B. Munkhbat, M. Wersäll, D. G. Baranov, T. J. Antosiewicz, T. Shegai, Suppression of photo-oxidation of organic chromophores by strong coupling to plasmonic nanoantennas. *Sci. Adv.* **4**, eaas9552 (2018). doi: 10.1126/sciadv.aas9552; pmid: 29984306
85. V. N. Peters *et al*., Effect of strong coupling on photodegradation of the semiconducting polymer P3HT. *Optica* **6**, 318–325 (2019). doi: 10.1364/OPTICA.6.000318
86. D. M. Coles *et al*., Polariton-mediated energy transfer between organic dyes in a strongly coupled optical microcavity. *Nat. Mater.* **13**, 712–719 (2014). doi: 10.1038/nmat3950; pmid: 24793357
87. X. Zhong *et al*., Non-radiative energy transfer mediated by hybrid light-matter states. *Angew. Chem. Int. Ed.* **55**, 6202–6206 (2016). doi: 10.1002/anie.201600428; pmid: 27072296
88. K. Akulov, D. Bochman, A. Golombek, T. Schwartz, Long-distance resonant energy transfer mediated by hybrid plasmonic-photonic modes. *J. Phys. Chem. C* **122**, 15853–15860 (2018). doi: 10.1021/acs.jpcc.8b03030
89. B. Xiang *et al*., Intermolecular vibrational energy transfer enabled by microcavity strong light-matter coupling. *Science* **368**, 665–667 (2020). doi: 10.1126/science.aba3544; pmid: 32381725
90. M. Du, R. F. Ribeiro, J. Yuen-Zhou, Remote control of chemistry in optical cavities. *Chem* **5**, 1167–1181 (2019). doi: 10.1016/j.chempr.2019.02.009
91. A. Shalabney *et al*., Coherent coupling of molecular resonators with a microcavity mode. *Nat. Commun.* **6**, 5981 (2015). doi: 10.1038/ncomms6981; pmid: 25583259
92. J. P. Long, B. S. Simpkins, Coherent coupling between a molecular vibration and a Fabry-Perot optical cavity to give hybridized states in the strong coupling limit. *ACS Photonics* **2**, 130–136 (2015). doi: 10.1021/ph5003347
93. S. R. Casey, J. R. Sparks, Vibrational strong coupling of organometallic complexes. *J. Phys. Chem. C* **120**, 28138–28143 (2016). doi: 10.1021/acs.jpcc.6b10493
94. J. George *et al*., Multiple Rabi splittings under ultrastrong vibrational coupling. *Phys. Rev. Lett.* **117**, 153601 (2016). doi: 10.1103/PhysRevLett.117.153601; pmid: 27768350
95. A. Thomas *et al*., Ground-state chemical reactivity under vibrational coupling to the vacuum electromagnetic field. *Angew. Chem. Int. Ed.* **55**, 11462–11466 (2016). doi: 10.1002/anie.201605504; pmid: 27529831
96. R. M. A. Vergauwe *et al*., Modification of enzyme activity by vibrational strong coupling of water. *Angew. Chem. Int. Ed.* **58**, 15324–15328 (2019). doi: 10.1002/anie.201908876; pmid: 31449707
97. A. Thomas *et al*., Tilting a ground-state reactivity landscape by vibrational strong coupling. *Science* **363**, 615–619 (2019). doi: 10.1126/science.aau7742; pmid: 30733414
98. K. Hirai, R. Takeda, J. A. Hutchison, H. Uji-i, Modulation of Prins cyclization by vibrational strong coupling. *Angew. Chem. Int. Ed.* **59**, 5332–5335 (2020). doi: 10.1002/anie.201915632; pmid: 31970847
99. Y. Pang *et al*., On the role of symmetry in vibrational strong coupling: The case of charge-transfer complexation. *Angew. Chem. Int. Ed.* **59**, 10436–10440 (2020). doi: 10.1002/anie.202002527; pmid: 32220038
100. J. Galego, C. Climent, F. J. Garcia-Vidal, J. Feist, Cavity Casimir-Polder Forces and Their Effects in Ground-State Chemical Reactivity. *Phys. Rev. X* **9**, 021057 (2019). doi: 10.1103/PhysRevX.9.021057
101. C. Climent, J. Galego, F. J. Garcia-Vidal, J. Feist, Plasmonic Nanocavities Enable Self-Induced Electrostatic Catalysis. *Angew. Chem. Int. Ed.* **58**, 8698–8702 (2019). doi: 10.1002/anie.201901926; pmid: 30969014
102. J. A. Campos-Gonzalez-Angulo, R. F. Ribeiro, J. Yuen-Zhou, Resonant catalysis of thermally activated chemical reactions with vibrational polaritons. *Nat. Commun.* **10**, 4685 (2019). doi: 10.1038/s41467-019-12636-1; pmid: 31615990

103. N. T. Phuc, P. Q. Trung, A. Ishizaki, Controlling the nonadiabatic electron-transfer reaction rate through molecular-vibration polaritons in the ultrastrong coupling regime. *Sci. Rep.* **10**, 7318 (2020). doi: 10.1038/s41598-020-62899-8; pmid: 32355233
104. F. J. Hernández, F. Herrera, Multi-level quantum Rabi model for anharmonic vibrational polaritons. *J. Chem. Phys.* **151**, 144116 (2019). doi: 10.1063/1.5121426; pmid: 31615252
105. V. Zhdanov, Vacuum filed in a cavity, light-mediated vibrational coupling, and chemical reactivity. *Chem. Phys.* **535**, 110767 (2020). doi: 10.1016/j.chemphys.2020.110767
106. I. Vurgaftman, B. S. Simpkins, A. D. Dunkelberger, J. C. Owrutsky, Negligible effect of vibrational polaritons on chemical reaction rates via the density of states pathway. *J. Phys. Chem. Lett.* **11**, 3557–3562 (2020). doi: 10.1021/acs.jpclett.0c00841; pmid: 32298585
107. K. Hirai, H. Ishikawa, J. Hutchison, H. Uji-i, Selective crystallization via vibrational strong coupling. ChemRxiv [Preprint] (2020). .doi: 10.26434/chemrxiv.13191617.v1
108. R. B. Woodward, R. Hoffmann, The conservation of orbital symmetry. *Angew. Chem. Int. Ed.* **8**, 781–853 (1969). doi: 10.1002/anie.196907811
109. A. Sau *et al*., Modifying Woodward-Hoffmann stereoselectivity under vibrational strong coupling. *Angew. Chem. Int. Ed.* **60**, 5712–5717 (2021). doi: 10.1002/anie.202013465; pmid: 33305864
110. J. Lather, J. George, Improving enzyme catalytic efficiency by co-operative vibrational strong coupling of water. *J. Phys. Chem. Lett.* **12**, 379–384 (2021). doi: 10.1021/acs.jpclett.0c03003; pmid: 33356291
111. T. S. Haugland, C. Schäfer, E. Ronca, A. Rubio, H. Koch, Intermolecular interactions in optical cavities: An *ab initio* QED study. *J. Chem. Phys.* **154**, 094113 (2021). doi: 10.1063/5.0039256; pmid: 33685159
112. T. E. Li, J. E. Subotnik, A. Nitzan, Cavity molecular dynamics simulations of liquid water under vibrational ultrastrong coupling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18324–18331 (2020). doi: 10.1073/pnas.2009272117; pmid: 32680967
113. X. Li, A. Mandal, P. Huo, Cavity frequency-dependent theory for vibrational polariton chemistry. *Nat. Commun.* **12**, 1315 (2021). doi: 10.1038/s41467-021-21610-9; pmid: 33637720
114. J. F. Triana, F. J. Hernández, F. Herrera, The shape of the electric dipole function determines the sub-picosecond dynamics of anharmonic vibrational polaritons. *J. Chem. Phys.* **152**, 234111 (2020). doi: 10.1063/5.0009869; pmid: 32571050
115. J. Lather, P. Bhatt, A. Thomas, T. W. Ebbesen, J. George, Cavity catalysis by cooperative vibrational strong coupling of reactant and solvent molecules. *Angew. Chem. Int. Ed.* **58**, 10635–10638 (2019). doi: 10.1002/anie.201905407; pmid: 31189028
116. G. D. Scholes, C. A. DelPo, B. Kudisch, Entropy reorders polariton states. *J. Phys. Chem. Lett.* **11**, 6389–6395 (2020). doi: 10.1021/acs.jpclett.0c02000; pmid: 32678609
117. S. Schütz *et al*., Ensemble-induced strong light-matter coupling of a single quantum emitter. *Phys. Rev. Lett.* **124**, 113602 (2020). doi: 10.1103/PhysRevLett.124.113602; pmid: 32242709
118. S. A. Guebrou *et al*., Coherent emission from a disordered organic semiconductor induced by strong coupling with surface plasmons. *Phys. Rev. Lett.* **108**, 066401 (2012). doi: 10.1103/PhysRevLett.108.066401; pmid: 22401091
119. L. Shi *et al*., Spatial coherence properties of organic molecules coupled to plasmonic surface lattice resonances in the weak and strong coupling regimes. *Phys. Rev. Lett.* **112**, 153002 (2014). doi: 10.1103/PhysRevLett.112.153002; pmid: 24785036
120. S. F. Huelga, M. B. Plenio, Vibrations, quanta and biology. *Contemp. Phys.* **54**, 181–207 (2013). doi: 10.1080/00405000.2013.829687
121. J. Cao *et al*., Quantum biology revisited. *Sci. Adv.* **6**, eaaz4888 (2020). doi: 10.1126/sciadv.aaz4888; pmid: 32284982
122. R. Damari *et al*., Strong coupling of collective intermolecular vibrations in organic materials at terahertz frequencies. *Nat. Commun.* **10**, 3248 (2019). doi: 10.1038/s41467-019-11130-y; pmid: 31324768
123. M. Schuler, D. De Bernardis, A. M. Läuchli, P. Rabl, The vacua of dipolar cavity quantum electrodynamics. *SciPost Phys.* **9**, 066 (2020). doi: 10.21468/SciPostPhys.9.5.066
124. G. Scalari *et al*., Ultrastrong coupling of the cyclotron transition of a 2D electron gas to a THz metamaterial. *Science* **335**, 1323–1326 (2012). doi: 10.1126/science.1216022; pmid: 22422976
125. X. Liu *et al*., Strong light-matter coupling in two-dimensional atomic crystals. *Nat. Photonics* **9**, 30–34 (2014). doi: 10.1038/nphoton.2014.304
126. W. Liu *et al*., Strong exciton-plasmon coupling in MoS_2 coupled with plasmonic lattice. *Nano Lett.* **16**, 1262–1269 (2016). doi: 10.1021/acs.nanolett.5b04588; pmid: 26784532
127. S. Wang *et al*., Coherent coupling of WS_2 monolayers with metallic photonic structures at room temperature. *Nano Lett.* **16**, 4368–4374 (2016). doi: 10.1021/acs.nanolett.6b01475; pmid: 27266674
128. R. Houdré, R. P. Stanley, M. Ilegems, Vacuum-field Rabi splitting in the presence of inhomogeneous broadening: Resolution of a homogeneous linewidth in an inhomogeneously broadened system. *Phys. Rev. A* **53**, 2711–2715 (1996). doi: 10.1103/PhysRevA.53.2711; pmid: 9913184
129. V. M. Agranovich, M. Litinskaia, D. G. Lidzey, Cavity polaritons in microcavities containing disordered organic semiconductors. *Phys. Rev. B* **67**, 085311 (2003). doi: 10.1103/PhysRevB.67.085311
130. C. Ciuti, I. Carusotto, Input-output theory of cavities in the ultrastrong coupling regime: The case of time-independent cavity parameters. *Phys. Rev. A* **74**, 033811 (2006). doi: 10.1103/PhysRevA.74.033811
131. A. Canaguier-Durand, C. Genet, A. Lambrecht, T. W. Ebbesen, S. Reynaud, Non-Markovian polariton dynamics in organic strong coupling. *Eur. Phys. J. D* **69**, 24 (2015). doi: 10.1140/epjd/e2014-50539-x

ACKNOWLEDGMENTS

Funding: T.W.E. acknowledges support from the International Center for Frontier Research in Chemistry (icFRC, Strasbourg), the ANR Equipex Union (ANR-10-EQPX-52-01), the Labex NIE (ANR-11-LABX-0058 NIE), CSC (ANR-10- LABX-0026 CSC), and USIAS within the Investissement d'Avenir program ANR-10-IDEX-0002-02, the ERC (project no. 788482 MOLUSC), and QuantERA project RouTe. C.C. acknowledges financial support from FET FLAGSHIP Project PhoQuS (grant agreement ID no. 820392) and from the French agency ANR through the projects UNIQ (ANR-16-CE24-0029), NOMOS (ANR-18-CE24-0026), and TRIANGLE (ANR-20-CE47-0011). F.J.G.-V. acknowledges financial support from the Spanish Agency of Research through grants RTI2018-099737-B-I00, PCI2018-093145, and CEX2018-000805-M (through the Maria de Maeztu program for Units of Excellence in R&D). **Competing interests:** The authors have no competing interests.

10.1126/science.abd0336

ScienceImmunology.org

READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology

RESEARCH ARTICLE SUMMARY

CANCER

Mapping the cellular origin and early evolution of leukemia in Down syndrome

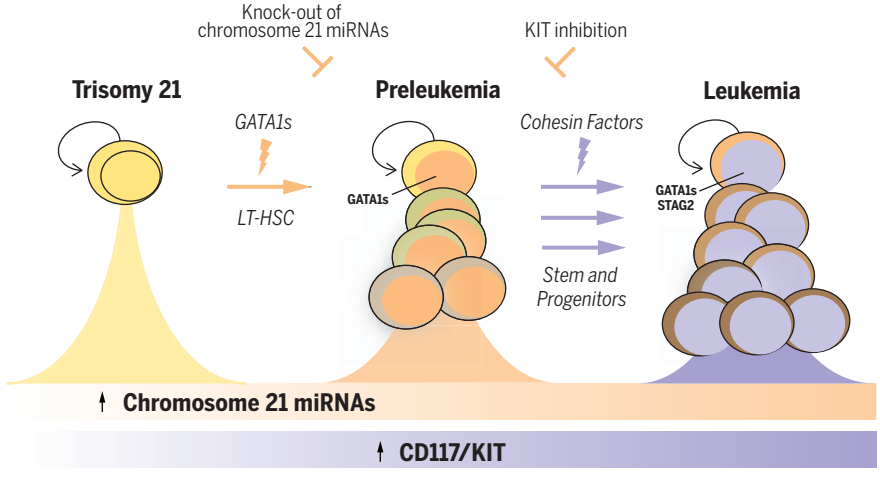
Elvin Wagenblast*, Joana Araújo, Olga I. Gan, Sarah K. Cutting, Alex Murison, Gabriela Krivdova, Maria Azkanaz, Jessica L. McLeod, Sabrina A. Smith, Blaise A. Gratton, Sajid A. Marhon, Martino Gabra, Jessie J. F. Medeiros, Sanaz Manteghi, Jian Chen, Michelle Chan-Seng-Yue, Laura Garcia-Prat, Leonardo Salmena, Daniel D. De Carvalho, Sagi Abelson, Mohamed Abdelhaleem, Karen Chong, Maian Roifman, Patrick Shannon, Jean C. Y. Wang, Johann K. Hitzler, David Chitayat, John E. Dick*, Eric R. Lechman*

INTRODUCTION: Leukemia is the most common cancer in children, with the first genetic alterations often occurring during fetal development. These initiating events generate preleukemic cells, which are the evolutionary ancestors of leukemia that arises after birth. Because of our inability to directly access human fetal preleukemia, the identity of the cell of origin and the steps of leukemia evolution remain largely unknown. Down syndrome leukemogenesis represents a disease setting to study human preleukemia and the evolutionary steps that lead to fully transformed leukemia. Up to 30% of children with Down syndrome (trisomy 21) exhibit a preleukemic transient abnormal myelopoiesis (TAM) and, overall, have a 150-fold increased risk of developing myeloid leukemia within the first 5 years of life. However, the mechanism by which an extra copy of chromosome 21 predisposes to preleukemia and leukemia remains unclear.

RATIONALE: Understanding Down syndrome leukemogenesis requires a humanized model that faithfully recapitulates the full developmental spectrum of premalignant and malignant stages of Down syndrome leukemia. Using CRISPR/Cas9–mediated gene editing in human disomic and trisomic fetal liver–derived hematopoietic stem and progenitor cells and xenotransplantation, we developed a model with which to characterize the genetic events and cellular contexts underlying the preleukemic and leukemic phases of Down syndrome leukemogenesis.

RESULTS: Trisomy 21 hematopoietic stem and progenitor cells (HSPCs) showed reduced proliferation in vitro and generated smaller grafts in xenotransplanted mice, with reduced serial transplant ability, as compared with that of disomic HSPCs. Preleukemia was initiated in trisomy 21, but not disomic, long-term hematopoietic stem cells (LT-HSCs) when mutations in the erythroid-megakaryocyte transcription factor GATA binding protein 1 (*GATA1*) were introduced, which led to exclusive expression of the short isoform (GATA1s). Subsequent leukemic progression could occur in multiple stem and progenitor populations, was independent of trisomy 21, and induced through deletion of cohesin genes, including *STAG2* (STAG2ko). Serial engraftment in mice showed that GATA1s-induced preleukemia underwent spontaneous resolution, which contrasted with the persistent ability of the GATA1s/STAG2ko–induced leukemia to engraft serially in mice. Leukemic progression was developmentally restricted to fetal and early postnatal stages; adult-derived bone marrow HSPCs were unable to undergo GATA1s/STAG2ko-induced leukemic transformation. We identified a molecular mechanism by which three chromosome 21 microRNAs (miRNAs) contributed to the predisposition toward preleukemia initiation. Simultaneous overexpression of *miR-99a*, *miR-125b-2*, and *miR-155* in normal disomic LT-HSCs recapitulated a trisomy 21–like hematopoietic state, as assessed through comparable lineage differentiation, reduced self-renewal capacity, and similar gene expression and open chromatin accessibility profile. Removal of these miRNAs in trisomy 21 LT-HSCs inhibited GATA1s-induced preleukemia development. Using secondary xenotransplantations of defined cell populations, we identified $CD117^{+}$/KIT proto-oncogene (KIT) as a marker of disease-driving cells. Pharmacological KIT inhibition targeted preleukemic stem cells, both in GATA1s-induced preleukemia and in primary Down syndrome preleukemia patient samples.

CONCLUSION: Collectively, our results provide insight into how human preleukemia and leukemia evolve in fetal life and early childhood. We were able to identify distinct cellular origins and effects of trisomy 21 for preleukemia initiation and leukemia progression. Predisposition to preleukemia in Down syndrome is affected by overexpression of distinct chromosome 21 miRNAs, specifically in the preleukemic LT-HSC cell of origin. Our study reveals the relevance of the cellular and developmental status of the cell of origin during leukemogenesis, which begins to explain why genetic drivers can be distinct between pediatric and adult acute myeloid leukemia. KIT inhibitors targeted preleukemic stem cells, providing proof of principle for early prevention strategies in childhood leukemia that may be able to inhibit leukemia progression, and these results encourage further preclinical and clinical assessment. ■



Cell of origin in Down syndrome leukemogenesis. Down syndrome preleukemia originated in long-term hematopoietic stem cells (LT-HSCs) through mutations in GATA1, leading to the expression of the short isoform GATA1s. Progression toward leukemia occurred in various stem and progenitor cells through mutations in cohesin factors such as STAG2. Predisposition to preleukemia was affected by chromosome 21 miRNAs, and pharmacological inhibition of KIT targeted preleukemic stem cells.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: elvin.wagenblast@uhnresearch.ca (E.W.); john.dick@uhnresearch.ca (J.E.D.); eric.lechman@uhnresearch.ca (E.R.L)
Cite this article as E. Wagenblast *et al.*, *Science* **373**, eabf6202 (2021). DOI: 10.1126/science.abf6202

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf6202

RESEARCH ARTICLE

CANCER

Mapping the cellular origin and early evolution of leukemia in Down syndrome

Elvin Wagenblast[1]*, Joana Araújo, [1,2,3,4,5]†, Olga I. Gan[1]†, Sarah K. Cutting[1], Alex Murison[1], Gabriela Krivdova[1,6], Maria Azkanaz[1], Jessica L. McLeod[1], Sabrina A. Smith[1,6], Blaise A. Gratton[1], Sajid A. Marhon[1], Martino Gabra[7], Jessie J. F. Medeiros[1,6,8], Sanaz Manteghi[9], Jian Chen[9], Michelle Chan-Seng-Yue[1,8], Laura Garcia-Prat[1], Leonardo Salmena[1,7], Daniel D. De Carvalho[1,10], Sagi Abelson[6,8], Mohamed Abdelhaleem[11], Karen Chong[12,13], Maian Roifman[12,13], Patrick Shannon[14], Jean C. Y. Wang[1,15,16], Johann K. Hitzler[9,17,18], David Chitayat[12,13], John E. Dick[1,6]*, Eric R. Lechman[1]*

Children with Down syndrome have a 150-fold increased risk of developing myeloid leukemia, but the mechanism of predisposition is unclear. Because Down syndrome leukemogenesis initiates during fetal development, we characterized the cellular and developmental context of preleukemic initiation and leukemic progression using gene editing in human disomic and trisomic fetal hematopoietic cells and xenotransplantation. GATA binding protein 1 (*GATA1*) mutations caused transient preleukemia when introduced into trisomy 21 long-term hematopoietic stem cells, where a subset of chromosome 21 microRNAs affected predisposition to preleukemia. By contrast, progression to leukemia was independent of trisomy 21 and originated in various stem and progenitor cells through additional mutations in cohesin genes. CD117^{+}/KIT proto-oncogene (KIT) cells mediated the propagation of preleukemia and leukemia, and KIT inhibition targeted preleukemic stem cells.

Children with Down syndrome [trisomy 21 (T21)] have a 150-fold increased risk of developing acute myeloid leukemia, which is called myeloid leukemia associated with Down syndrome (ML-DS), in the first 5 years of life (*1*). However, the mechanism by which an extra copy of chromosome 21 (Chr21) predisposes and cooperates with genetic events in Down syndrome leukemogenesis is not known. In pediatric leukemia, the initiating genetic events occur before birth and generate preleukemic cells (*2*), which are the evolutionary ancestors of leukemia that arises after birth. However, characterizing human fetal preleukemia is challenging because of our inability to directly access it, rendering the identity of the cell of origin and the steps of leukemia evolution largely unknown. Up to 30% of newborns with Down syndrome exhibit transient abnormal myelopoiesis (TAM), a preleukemic phase characterized by a clonal proliferation of immature myeloid cells (mostly megakaryoblasts) that carry somatic mutations in the erythroid-megakaryocyte transcription factor GATA binding protein 1 (*GATA1*) (fig. S1A) (*3*–*5*). Mutations in *GATA1* occur in utero, beginning at 21 weeks of gestation (*6*, *7*), and lead to the expression of a truncated isoform [GATA1-short (GATA1s)]. The preleukemia resolves spontaneously in the majority of newborns; however, in 20% of cases ML-DS evolves within 4 years from the GATA1s-mutated preleukemic clone through acquisition of additional mutations, predominantly in genes of the cohesin complex or CCCTC Binding Factor (*CTCF*) (*8*–*10*). Comprehensive sequencing studies have shown that mutations in the cohesin subunit stromal antigen 2 (*STAG2*) are most frequently implicated in ML-DS development (*11*, *12*). On the basis of these observations, it is hypothesized that the evolution of Down syndrome leukemia requires at least three distinct genetic events: T21, GATA1s, and additional mutations such as *STAG2*. However, the identity of the human hematopoietic cell type that acquires GATA1s and originates preleukemia and the cell type in which subsequent mutations accumulate to generate leukemia are unknown. Furthermore, the cellular origin of preleukemic mutations in pediatric leukemia in general is currently unclear, although twin studies of B cell acute lymphoblastic leukemia (ALL) point to a primitive cellular origin (*13*). Down syndrome leukemogenesis offers a disease setting to uncover generalized principles regarding this phenomenon.

Currently, there are no effective strategies to predict the subset of individuals at increased risk for progression from preleukemia to ML-DS. Life-threatening symptoms associated with preleukemia are treated with cytarabine (*14*, *15*); however, this treatment does not prevent subsequent development of leukemia. Despite the generally favorable response of ML-DS to standard chemotherapy, outcomes are dismal for those with refractory or relapsed disease, with an overall survival rate of less than 20% (*16*–*18*). Therefore, therapeutic targeting of preleukemic clones could represent a general concept to prevent development of leukemia. Experimentally, the mechanistic study of T21 and Down syndrome preleukemia has been challenging, primarily because of the fetal origin of the disease and the overall lack of suitable in vivo models (*19*–*21*). To circumvent these limitations, we used disomic and trisomic hematopoietic cells that were isolated from primary human fetal livers, the major hematopoietic organ during prenatal development, to investigate the mechanisms underlying preleukemia and leukemia.

Here, we describe a model that faithfully recapitulates the full spectrum of premalignant and malignant stages of Down syndrome leukemia by use of CRISPR/Cas9 methodology optimized for single human hematopoietic stem cells (HSCs) (*22*) in disomic and trisomic primary human fetal liver-derived HSCs and downstream progenitors (*23*). Using this tool, we uncover insights into the genetic events and cellular contexts underlying the preleukemic and leukemic phases of Down syndrome leukemogenesis and provide proof of concept for targeting the preleukemic stage of the disease.

Results

GATA1s induces a megakaryocytic bias in hematopoietic stem and progenitor cells

To study the initiating events in Down syndrome preleukemia, we first sorted hematopoietic stem and progenitor cells (HSPCs) from normal (disomic) and T21 human fetal livers

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada. [2]Department of Hematology, Centro Hospitalar Universitário de São João, Porto, 4200-319, Portugal. [3]Faculty of Medicine, University of Porto, Porto, 4200-319, Portugal. [4]Instituto de Investigação e Inovação em Saúde, University of Porto, Porto, 4200-135, Portugal. [5]Instituto Nacional de Investigação Biomédica, University of Porto, Porto, 4200-135, Portugal. [6]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada. [7]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada. [8]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada. [9]Program in Developmental and Stem Cell Biology, The Hospital for Sick Children Research Institute, Toronto, ON M5G 1X8, Canada. [10]Department of Medical Biophysics, University of Toronto, Toronto, ON M5S 1A8, Canada. [11]Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada. [12]The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, ON M5S 1A8, Canada. [13]Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON M5S 1A8, Canada. [14]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, University of Toronto, Toronto, ON M5S 1A8, Canada. [15]Department of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [16]Division of Medical Oncology and Hematology, University Health Network, Toronto, Ontario M5G 2M9, Canada. [17]Department of Pediatrics, University of Toronto, Toronto, ON M5G 1X8, Canada. [18]Division of Hematology and Oncology, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada.
*Corresponding author. Email: elvin.wagenblast@uhnresearch.ca (E.W.); john.dick@uhnresearch.ca (J.E.D.); eric.lechman@uhnresearch.ca (E.R.L)
†These authors contributed equally to this work.

(N-FL and T21-FL, respectively) obtained at 16 to 19 weeks of gestation and carried out phenotypic analysis of the HSPC hierarchy (Fig. 1A and fig. S1B). T21 karyotype of sorted HSPCs was confirmed by means of droplet digital polymerase chain reaction (PCR) with a set of probes against Chr21 (fig. S1C). In addition, error-corrected targeted sequencing at 7500× coverage did not reveal any preexisting mutations of GATA1 exon 2 in T21-FL–derived HSPCs (table S1); although rare, preexisting mutations in exon 3 would have been missed (*24*). The impact of T21 on the HSPC hierarchy of fetal liver revealed a 30% increase in the percentage of total phenotypic long-term HSCs (LT-HSCs) and a simultaneous decrease in short-term HSCs (ST-HSCs) compared with that in N-FL. In addition, an expansion of megakaryocyte-erythroid progenitors (MEPs) was seen in T21-FL compared with N-FL, as previously reported (figs. S1, D to E, and S2A) (*25*). Quantification of *GATA1* and its isoforms revealed comparable expression in N-FL and T21-FL HSPC subpopulations, with a gradual

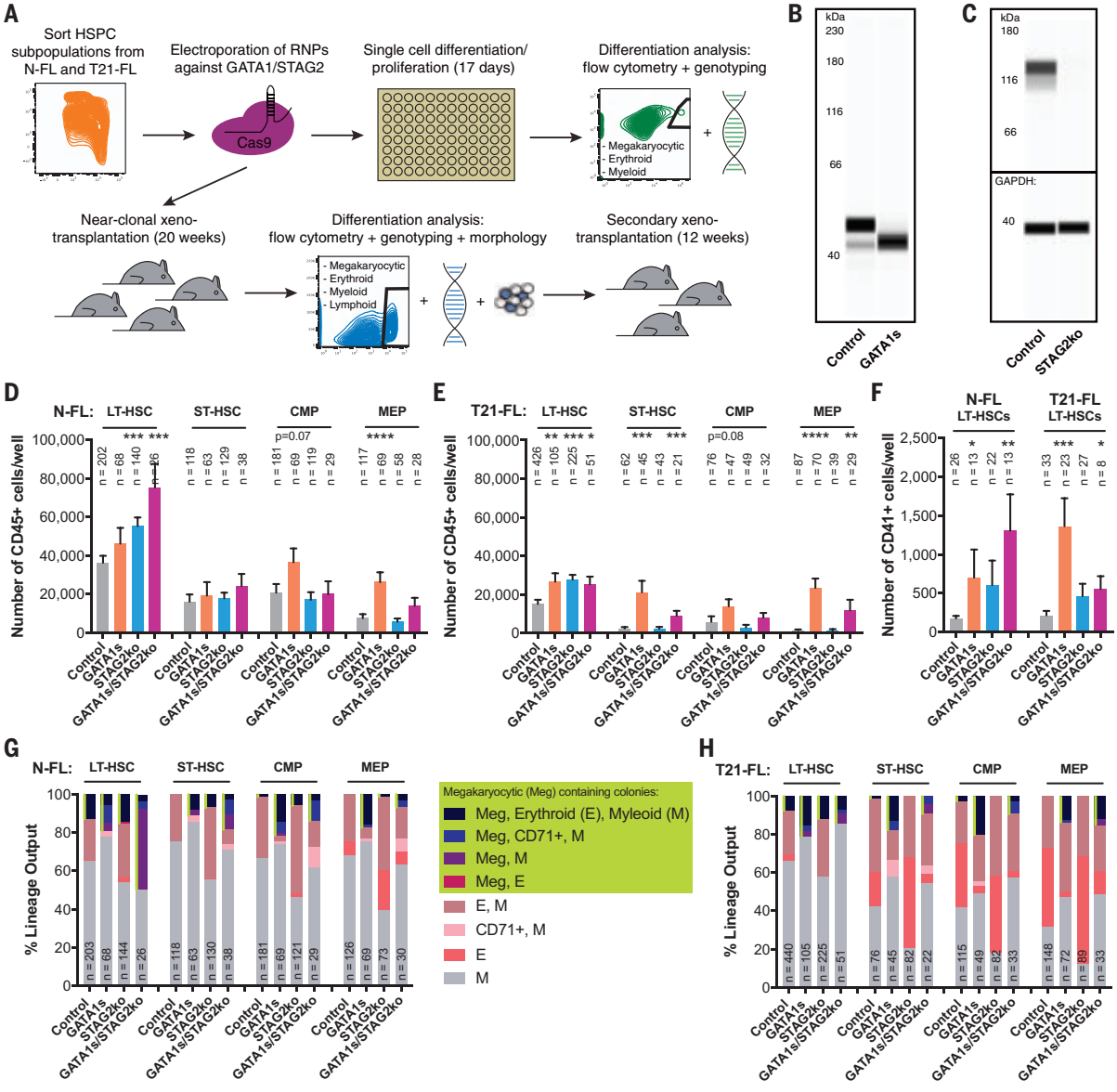


Fig. 1. GATA1s induces a megakaryocytic bias in hematopoietic stem and progenitor cells. (**A**) Experimental overview of in vitro single-cell differentiation/proliferation assay and near-clonal xenotransplantation. (**B**) Western blot assay of GATA1 in combined N-FL CMP and MEP cells, which were CRISPR/Cas9–edited with control and GATA1s gRNAs (n = 2 experiments). (**C**) Western blot assay of STAG2 in combined N-FL CMP and MEP cells, which were CRISPR/Cas9–edited with control and STAG2ko gRNAs (n = 2 experiments). (**D**) Proliferation capacity assessed by the overall number of CD45$^+$ cells from in vitro single-cell assay of individual CRISPR/Cas9–edited cells for N-FL. Numbers of single-cell colonies with appropriate positive genotype are indicated for each condition (n = 2 experiments). (**E**) Proliferation capacity described in (D) for T21-FL (P < 0.0001 for combined T21-FL versus N-FL, n = 2 or 3 experiments). (**F**) Proliferation capacity assessed as the number of CD41$^+$ megakaryocytic cells in all megakaryocyte-containing colonies from the single-cell assays described in (D) and (E). (**G**) Lineage output from in vitro single-cell assay described in (D) [P < 0.05 for Meg in GATA1s versus control, P = 0.12 for Meg in GATA1s/STAG2ko versus control, and P < 0.01 for (E) in GATA1s versus control among all cell types, n = 2 experiments]. (**H**) Lineage output from in vitro single-cell assay as described in (G) for T21-FL [P < 0.001 for Meg in GATA1s versus control, P < 0.01 for Meg in GATA1s/STAG2ko versus control, and P = 0.16 for (E) in GATA1s versus control among all cell types, n = 2 or 3 experiments]. Unpaired Student's t test: *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; error bars represent standard error of the mean.

increase in expression upon differentiation commitment (fig. S2B).

To examine the roles of GATA1s and *STAG2* in leukemogenesis individually or in combination, we performed CRISPR/Cas9 editing of sorted HSPC subpopulations to express the short isoform of *GATA1* (GATA1s) under its endogenous promoter and/or to delete the cohesin subunit STAG2 (STAG2ko). Using our optimized methodology (*22*), editing efficiency exceeded 75% (fig. S2C). Karyotyping analysis of N-FL HSPCs revealed no structural abnormalities after CRISPR/Cas9 editing (fig. S2D), and whole-genome sequencing in N-FL LT-HSCs at 30× coverage revealed either very rare or no off-target indels at sites that were similar to the guide RNA (gRNA) sequence (table S2). In the few cases in which off-target indels were detected, the allelic depth was 6% or lower. Western blot assays of CRISPR/Cas9–edited common myeloid progenitors (CMPs) and MEPs showed exclusive expression of GATA1s and undetectable STAG2 protein (Fig. 1, B and C, and fig. S2, E and F), confirming CRISPR/Cas9 editing of the respective genes at the protein level.

To elucidate the functional consequences of GATA1s and STAG2ko in different HSPC subpopulations, N-FL and T21-FL LT-HSCs, ST-HSCs, CMPs, and MEPs were purified; CRISPR/Cas9–edited for control, GATA1s, STAG2ko, or GATA1s/STAG2ko; and placed into single-cell in vitro differentiation and proliferation assays by using erythro-myeloid-megakaryocytic–promoting medium (*26*). The phenotype and genotype of all ~3000 single-cell–derived colonies used in our study were determined (fig. S2, G and H). Consistent with a previous report (*25*), colony-forming efficiency was higher in CRISPR/Cas9–edited T21-FL LT-HSCs compared with N-FL (fig. S3, A and B). The single-cell CRISPR/Cas9 editing efficiency was above 80% for both control and STAG2ko colonies (fig. S3, C and D). For GATA1s, the CRISPR/Cas9 efficiency was ~40% because only colonies with confirmed complete excision of exon 2 were included in the analysis. No preexisting GATA1 mutations in exon 2 were observed in any of the analyzed ~900 T21-FL control–edited colonies, confirming the results of the error-corrected targeted sequencing of T21-FL–derived HSPCs (fig. S3E).

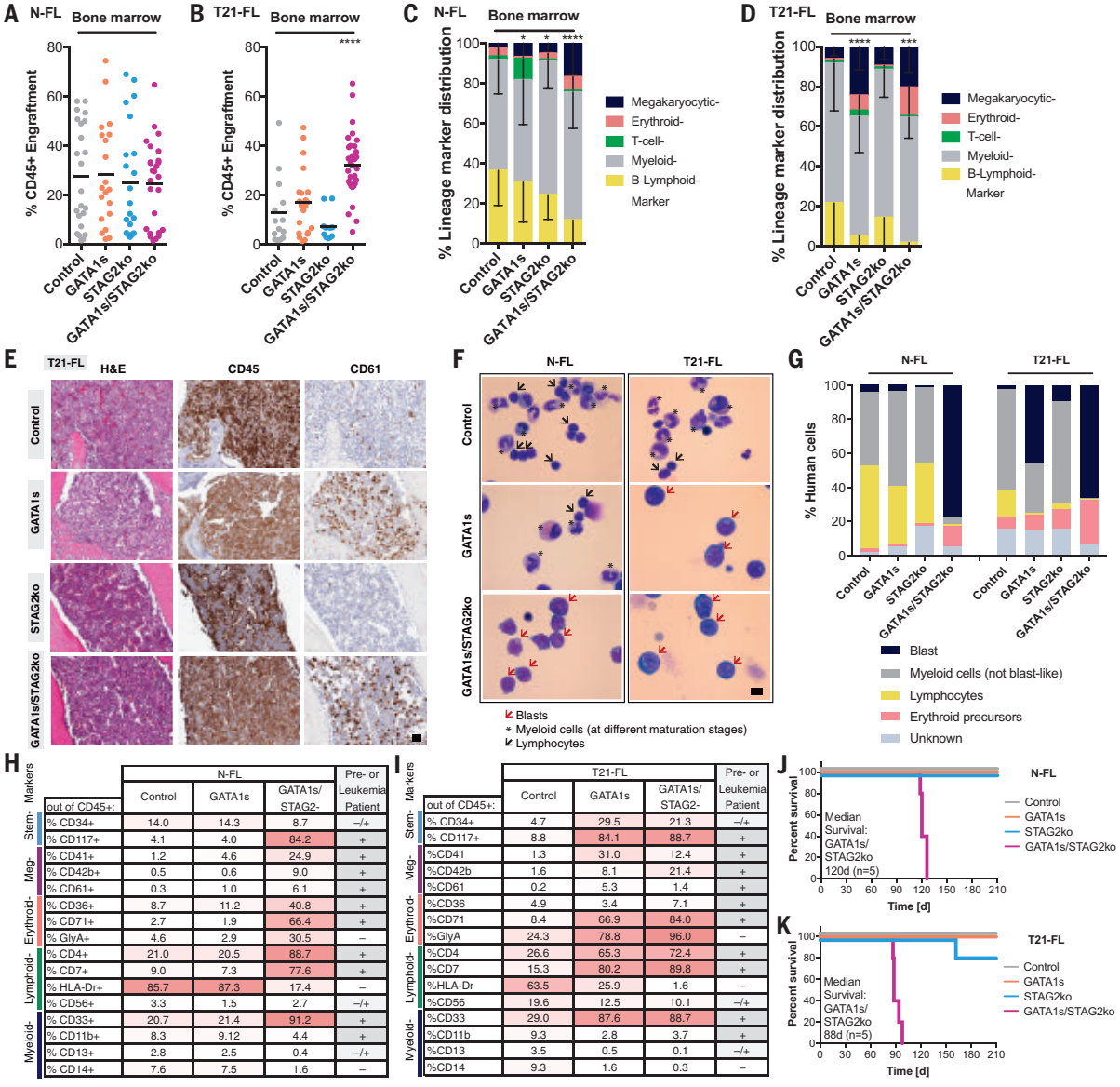
Proliferation measured by total $CD45^+$ cell output was lower in control-edited T21-FL HSPC subpopulations as compared with N-FL (Fig. 1, D and E). However, there was an increase in cell numbers observed in T21-FL GATA1s and GATA1s/STAG2ko subpopulations compared with T21-FL control colonies. A similar pattern was noticed in edited N-FL progenitor subpopulations but to a lesser extent when comparing N-FL GATA1s and GATA1s/STAG2ko colonies with N-FL control colonies. The increase in proliferative capacity in GATA1s and GATA1s/STAG2ko LT-HSCs was accompanied by a significant increase in the production of $CD41^+$ megakaryocytes within all megakaryocyte-containing colonies ($P < 0.05$) (Fig. 1F). To investigate whether the decreased proliferative capacity of control-edited T21-FL was related to a change in the number of cycling or quiescent cells, we performed cell-cycle analysis. T21-FL HSPC subpopulations contained a lower percentage of cells in S phase and a higher frequency of cells arrested in G_0 or G_1 phase compared with those in N-FL (fig. S3F). No difference was observed in the ratio of quiescent G_0 to G_1 cells between N-FL and T21-FL (fig. S3G). Thus, despite an increased proportion of the LT-HSC compartment in T21-FL (fig. S2A) and higher colony-forming capacity (fig. S3, A and B), T21-FL cells exhibited significantly reduced proliferative capacity compared with their N-FL counterparts ($P < 0.0001$) (Fig. 1, D and E). However, the acquisition of a GATA1 mutation increased their proliferative capacity (Fig. 1, D and E), possibly providing a selective advantage to T21-FL HSPCs.

Phenotypic analysis was performed on single-cell–derived colonies, which were cultured in high-cytokine medium that forces terminal differentiation of HSPCs. Colonies derived from single N-FL HSPCs revealed a shift toward megakaryocytic differentiation and a concomitant decrease in erythroid lineage output in GATA1s and GATA1s/STAG2ko colonies compared with controls (Fig. 1G). A subset of these GATA1s and GATA1s/STAG2ko colonies expressed the early erythroid marker CD71 but not the mature erythroid marker GlyA, potentially indicating a hindrance toward erythroid differentiation (fig. S3H). Similar to N-FL, T21-FL GATA1s and GATA1s/STAG2ko HSPC subpopulations displayed a significant megakaryocytic bias ($P < 0.01$) and a decrease in erythroid output compared with controls (Fig. 1H and fig. S3I). By contrast, both N-FL and T21-FL STAG2ko cells exhibited an increase in erythroid output compared with controls. Collectively, these in vitro results indicate that exclusive expression of GATA1s with or without STAG2ko resulted in increased megakaryocytic output in all HSPC subpopulations, with no major differences in terminal differentiation between N-FL and T21-FL.

T21 is required for preleukemia initiation but dispensable for leukemia development

To evaluate the functional effects of GATA1s and STAG2ko in vivo, we used LT-HSCs because these are the only cells that have the ability to permanently repopulate the entire hematopoietic system after transplantation (*23*). We carried out xenotransplantation assays using *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) and *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}Kit^{em1Mvw}/SzJ* (NSGW41) recipients. NSGW41 mice were added to the study for their ability to engraft human cells without the need of irradiation and their increased ability to support the growth of erythroid and megakaryocytic lineages (*27*). The repopulating cell frequency of N-FL LT-HSCs injected into mice evaluated at 20 weeks was approximately 1 in 300 (fig. S3, J and K). We therefore transplanted N-FL and T21-FL control, GATA1s, STAG2ko, and GATA1s/STAG2ko LT-HSCs at cell doses of 300 to 400 into mice to obtain near-clonal grafts. After 20 weeks, human engraftment was analyzed in the bone marrow (BM), and extramedullary hematopoiesis was assessed in the spleen. Only mice bearing confirmed CRISPR/Cas9–edited grafts were used in the subsequent analysis (fig. S4, A to C). To evaluate the clonality of xenografts, BM cells of engrafted mice were plated in methylcellulose colony assays. Sanger sequencing of CRISPR/Cas9–mediated indels in individual colonies showed predominantly clonal engraftment in mice (fig. S4, D and E), validating our in vivo experimental approach.

On average, the human $CD45^+$ engraftment level in BM was ~25% for mice transplanted with N-FL LT-HSCs and lower for T21-FL LT-HSCs, with the exception of mice transplanted with GATA1s/STAG2ko LT-HSCs, which displayed engraftment of ~30% (Fig. 2, A and B). Lineage marker analysis revealed increased myeloid and decreased lymphoid lineage cells in T21-FL control grafts compared with N-FL (Fig. 2, C and D, and fig. S4, F to R). The proportion of human $CD41^+CD45^-$ megakaryocytic lineage cells was at least threefold higher in GATA1s and GATA1s/STAG2ko grafts as compared with control for both N-FL and T21-FL, which is consistent with the results observed in the in vitro single-cell assays. Moreover, immunohistochemistry (IHC) staining for the megakaryocytic marker CD61 in bone sections of humeri revealed an increase in cells expressing megakaryocytic markers in mice engrafted with GATA1s and GATA1s/STAG2ko cells from both N-FL and T21-FL (Fig. 2E and fig. S5, A to D). Engraftment patterns were similar in NSGW41 and NSG recipients of CRISPR/Cas9–edited N-FL and T21-FL (fig. S6, A to I). However, several differences were observable in GATA1s grafts from N- and T21-FL LT-HSCs. Overall, GATA1s LT-HSCs from T21-FL, but not from N-FL, were able to engraft in mice more efficiently than were their control-edited counterparts (fig. S6, J and K). This was further confirmed through enhanced repopulation of GATA1s LT-HSCs when a mixture of control and GATA1s LT-HSCs from T21-FL were transplanted into NSG mice for 6 weeks (fig. S6L). Immunophenotypic analysis of the HSC hierarchy of engrafted mice at 20 weeks after transplantation revealed a distorted LT-HSC/ST-HSC composition in T21-FL GATA1s xenografts but not in N-FL–derived grafts (fig. S6M).



H

Markers	out of CD45+:	N-FL Control	N-FL GATA1s	N-FL GATA1s/STAG2-	Pre- or Leukemia Patient
Stem-	% CD34+	14.0	14.3	8.7	–/+
	% CD117+	4.1	4.0	84.2	+
Meg-	% CD41+	1.2	4.6	24.9	+
	% CD42b+	0.5	0.6	9.0	+
	% CD61+	0.3	1.0	6.1	+
Erythroid-	% CD36+	8.7	11.2	40.8	+
	% CD71+	2.7	1.9	66.4	+
	% GlyA+	4.6	2.9	30.5	–
Lymphoid-	% CD4+	21.0	20.5	88.7	+
	% CD7+	9.0	7.3	77.6	+
	% HLA-Dr+	85.7	87.3	17.4	–
	% CD56+	3.3	1.5	2.7	–/+
Myeloid-	% CD33+	20.7	21.4	91.2	+
	% CD11b+	8.3	9.12	4.4	+
	% CD13+	2.8	2.5	0.4	–/+
	% CD14+	7.6	7.5	1.6	–

I

Markers	out of CD45+:	T21-FL Control	T21-FL GATA1s	T21-FL GATA1s/STAG2-	Pre- or Leukemia Patient
Stem-	% CD34+	4.7	29.5	21.3	–/+
	% CD117+	8.8	84.1	88.7	+
Meg-	%CD41	1.3	31.0	12.4	+
	%CD42b	1.6	8.1	21.4	+
	%CD61	0.2	5.3	1.4	+
Erythroid-	%CD36	4.9	3.4	7.1	+
	%CD71	8.4	66.9	84.0	+
	%GlyA	24.3	78.8	96.0	–
Lymphoid-	%CD4	26.6	65.3	72.4	+
	%CD7	15.3	80.2	89.8	+
	%HLA-Dr	63.5	25.9	1.6	–
	%CD56	19.6	12.5	10.1	–/+
Myeloid-	%CD33	29.0	87.6	88.7	+
	%CD11b	9.3	2.8	3.7	+
	%CD13	3.5	0.5	0.1	–/+
	%CD14	9.3	1.6	0.3	–

Fig. 2. T21 is required for preleukemia initiation but dispensable for leukemia development. (**A**) Engraftment of N-FL LT-HSC grafts in NSG mice. Engraftment was assessed on the basis of human CD45$^+$ expression in BM (only mice with >1% of CD45$^+$ cells in BM and >90% CRISPR/Cas9 efficiency were taken for the analysis; n = 3 cohorts). (**B**) Engraftment as described in (A) for T21-FL (n = 4 cohorts). (**C**) Lineage marker distribution based on cell surface markers in N-FL grafts in NSG mice. (**D**) Lineage marker distribution as described in (C) for T21-FL. $P < 0.05$ for T21-FL control myeloid cells versus N-FL control myeloid cells, and $P < 0.05$ for T21-FL control lymphoid cells versus N-FL control lymphoid cells. (**E**) Hematoxylin and eosin (H&E) and IHC stainings for human CD45 and human megakaryocytic marker CD61 in humeri of T21-FL grafts. Scale bar, 50 μm. (**F**) Morphological analysis of human cells in primary xenografts of N-FL and T21-FL grafts. Human cells were prepared by using cytospin and stained with Giemsa (100× magnification). Scale bar, 10 μm. (**G**) Quantification of cell morphology as seen in (F) (n = 400 cells per condition). (**H**) Percent expression of cell surface markers within the CD45$^+$ blast population in N-FL grafts in NSG mice. Data are from pooled samples of multiple xenografts. (**I**) Percent expression of cell surface markers as described in (H) for T21-FL. Data are from pooled samples of multiple xenografts. (**J**) Survival curve of N-FL LT-HSC grafts in NSGW41 mice (n = 5 mice per condition). (**K**) Survival curve as described in (J) for T21-FL (n = 5 mice per condition). Unpaired Student's t test: *$P < 0.05$; ***$P < 0.001$; ****$P < 0.0001$; error bars indicate standard deviation.

Last, T21-FL GATA1s grafts showed increased infiltration of myeloid lineage cells into the spleen compared with N-FL GATA1s (fig. S4R).

To investigate whether the observed lineage shifts and engraftment patterns were associated with development of preleukemia or malignant transformation to full leukemia, we assessed xenografts for the presence of immature blast cells as assessed from cytomorphology. There was a dramatic increase in blasts to ~30 to 40% in T21-FL GATA1s but not in N-FL GATA1s xenografts (Fig. 2, F and G, and fig. S6N). This was further confirmed through histology, which showed active blast infiltration within the BM of T21-FL GATA1s grafts but not in N-FL GATA1s xenografts (fig. S5D). Higher blast percentages of ~50 to 80% were observed in both N-FL and T21-FL GATA1s/STAG2ko grafts. Subsequently, we carried out a detailed flow cytometric analysis of lineage markers on large nongranulated cells in the blast gate (fig. S7A). N-FL and T21-FL control and N-FL GATA1s grafts had no enrichment of this gated population. The blast population of

T21-FL GATA1s grafts expressed the primitive stem cell markers CD34 and CD117 (KIT), megakaryocytic marker CD41, erythroid markers CD71 and GlyA, and myeloid marker CD33 and also aberrantly expressed lymphoid markers CD4 and CD7 compared with control and N-FL GATA1s grafts. This immunophenotype accurately recapitulates the clinical phenotype seen in patients with preleukemic TAM and meets clinically defined criteria (Fig. 2, H and I) (*14*, *28*–*30*). Blasts in both N-FL and T21-FL GATA1s/STAG2ko grafts had immunophenotypes nearly identical to those of T21-FL GATA1s grafts, which is in keeping with the clinical observation that blasts from patients in the preleukemic and leukemic stages are often indistinguishable (*30*, *31*). The blast immunophenotype of grafts generated in NSGW41 mice followed a comparable pattern (fig. S7, B and C).

We next assessed the survival of NSGW41 mice transplanted with 1300 N-FL or T21-FL control, GATA1s, STAG2ko, or GATA1s/STAG2ko LT-HSCs. No effect on overall survival was found in mice transplanted with control, GATA1s, or STAG2ko LT-HSCs from N-FL and T21-FL during the observation period of 210 days. By contrast, mice transplanted with either N-FL or T21-FL GATA1s/STAG2ko cells had a shorter median survival of 120 and 88 days, respectively (Fig. 2, J and K), highlighting an important difference between the preleukemic and leukemic disease in this model. Thus, in our model, we defined preleukemia on the basis of the GATA1s genotype, characterized by elevated blast counts (>10%) with megakaryocytic features, which is consistent with clinical guidelines (*30*, *32*), whereas leukemia was defined on the basis of the GATA1s/STAG2ko genotype, increased blast counts (>20%) with megakaryocytic features (*30*, *33*), and lethality in humanized mice. Overall, our findings demonstrate that T21 is necessary for preleukemia development driven by GATA1s but dispensable for leukemic progression upon acquisition of STAG2ko.

CD117 marks preleukemia- and leukemia-initiating cells, which possess a more MEP-like chromatin accessibility landscape

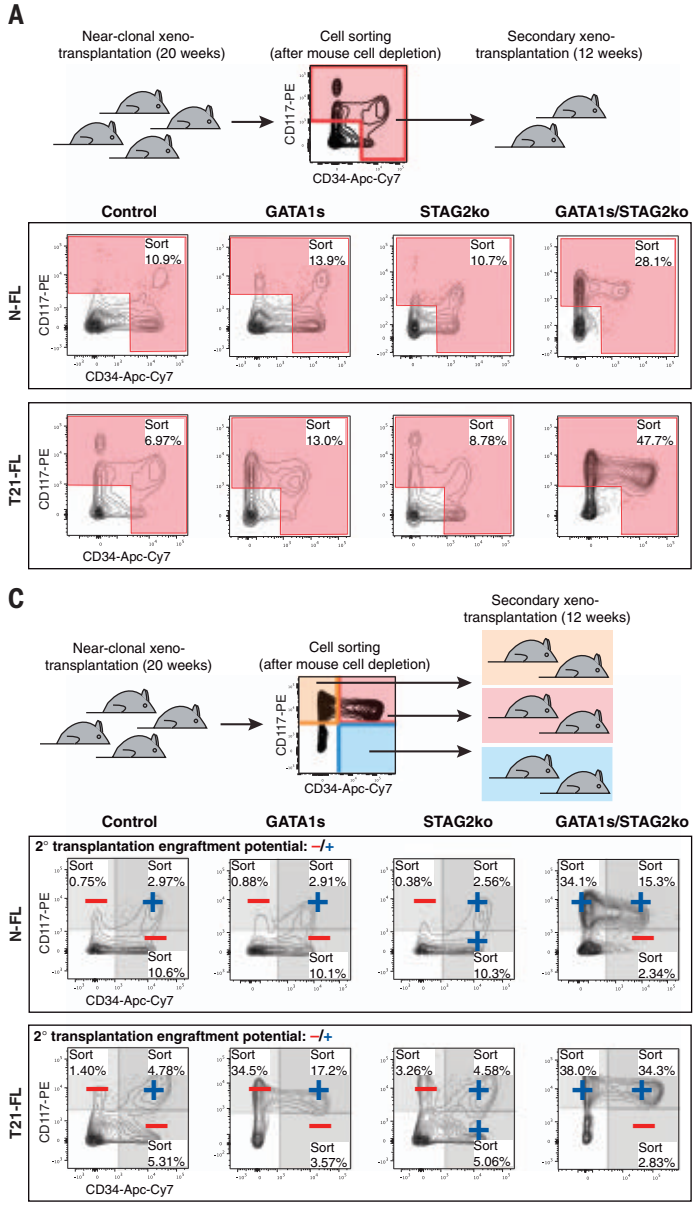
To assess the self-renewal properties of T21 GATA1s–induced preleukemia and GATA1s/STAG2ko–induced leukemia, we carried out secondary xenotransplantation assays. Because CD34 expression is absent in some Down syndrome leukemia cases (*30*), we sorted all primitive $CD34^+$ and $CD117^+$ cells from primary xenografts and transplanted them at defined doses into secondary NSGW41 recipients (Fig. 3A). We observed differences in self-renewal as measured by the ability of N-FL versus T21-FL GATA1s cells to propagate hematopoiesis in secondary recipients (Fig. 3B). Whereas secondary grafts originating from N-FL GATA1s cells were phenotypically similar to control grafts after 12 weeks, preleukemic T21-FL GATA1s secondary grafts contained characteristic blast populations equivalent to those seen in primary recipients (fig. S7, D and E), although with a lower preleukemia-initiating cell frequency of ~1/150,000. Both N-FL and T21-FL STAG2ko grafts had higher initiating-cell frequencies compared with that of controls, albeit lower in T21-FL as compared with N-FL (Fig. 3B), which is consistent with the previously reported increase in HSC self-renewal in a mouse model in which STAG2 was deleted (*34*). Both N-FL and T21-FL GATA1s/STAG2ko cells from primary grafts were able to generate secondary leukemic grafts containing characteristic blast populations (fig. S7, D and E), with initiating-cell frequencies of ~1/45,000 and ~1/90,000, respectively.

To evaluate the relevance of CD34 and CD117 expression independently, cells from primary xenografts were further sorted into $CD34^-CD117^+$, $CD34^+CD117^+$, and $CD34^+CD117^-$ fractions and transplanted at defined doses into secondary NSG recipients (Fig. 3C and table S3). For both N-FL and T21-FL controls, only cells from the $CD34^+CD117^+$ fraction were able to generate serial grafts at 12 weeks, with T21-FL showing a lower stem cell frequency compared with that of N-FL (Fig. 3C and table S3). Similar to control grafts, only $CD34^+CD117^+$ cells from preleukemic T21-FL GATA1s primary grafts were able to generate secondary grafts, with a low preleukemia-initiating cell frequency of ~1/380,000. For STAG2ko primary grafts, both $CD34^+CD117^+$ and $CD34^+CD117^-$ cells were able to engraft in secondary recipients. For leukemic N-FL and T21-FL GATA1s/STAG2ko primary grafts, cells from both $CD34^+CD117^+$ and $CD34^-CD117^+$ fractions propagated engraftment in secondary recipients, indicating that CD117 might be a better marker than CD34 for leukemia-initiating cells in Down syndrome leukemia. Both preleukemic and leukemic engraftments were confirmed by the appearance of characteristic blast populations (fig. S7, F to H). N-FL and T21-FL GATA1s/STAG2ko grafts but not T21-FL GATA1s grafts could be serially reproduced in tertiary mice (Fig. 3D). These findings highlight the transient nature of GATA1s-mediated preleukemia versus leukemia induced by GATA1s/STAG2ko and reflect the spontaneous remission that occurs in most affected individuals with TAM.

To investigate the mechanism by which GATA1s and STAG2 deficiency contribute to leukemogenesis, specifically within the propagating CD34/CD117 cell fractions from primary xenografts, we carried out transcriptional and epigenetic profiling by means of RNA-sequencing (RNA-seq) and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq). CIBERSORTx was used to computationally infer the cell lineage contribution from bulk ATAC-seq (*35*). For this, a signature matrix was generated from normalized read counts over a set of sites specific to individually sorted N-FL HSPC subpopulations (figs. S1D and S8, A and B), including F1, F2, and F3 subgroups of MEPs and CMPs sorted according to CD71 and BAH-1 expression (*26*). Engrafting fractions of T21-FL GATA1s LT-HSCs and N-FL and T21-FL GATA1s/STAG2ko LT-HSCs exhibited an increased MEP-like signature compared with control, with the MEP F3 subgroup being the most prominent (Fig. 3, E and F). Furthermore, these fractions showed enrichment of GATA-binding motifs at promoters (fig. S8C and table S4) associated with an increase in gene expression at these sites (fig. S8D). Gene set enrichment analysis of differentially expressed genes between preleukemic versus control and leukemic versus control populations (table S5) revealed down-regulation of pathways implicated in translation, ribosome biogenesis, and interferon signaling in preleukemic and leukemic populations (fig. S8, E and F, and table S6). Up-regulated genes in T21-FL GATA1s fractions were enriched in Down syndrome leukemia blasts from primary patient samples, whereas down-regulated genes in preleukemic and leukemic fractions were enriched in the stem cell–rich $CD34^+CD38^-$ population of human FL (fig. S8G) (*36*). Our results demonstrate that the preleukemic and leukemic propagating populations possess an open chromatin landscape that is largely driven by GATA1s-mediated transcriptional activation. Both of these propagating populations are identifiable from CD117 expression, which suggests that it could potentially serve as a therapeutic target.

Combined GATA1s and STAG2ko drive leukemic progression in progenitors

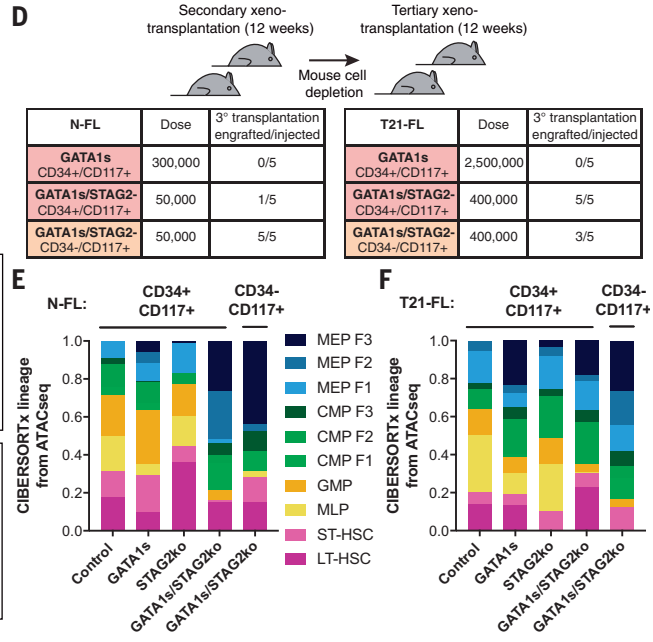
The originating cell type in leukemogenesis is increasingly recognized as playing an essential role in the resulting malignancy (*37*–*39*). To determine whether progeny downstream of LT-HSCs are able to initiate preleukemic or leukemic transformation, we introduced GATA1s and/or STAG2ko into functionally defined subpopulations of ST-HSCs, CMPs, and MEPs and transplanted them into NSGW41 mice at a dose of 1000 cells (Fig. 4A). No consistent human $CD45^+$ engraftment was detected after 12 weeks in mice transplanted with control, GATA1s, or STAG2ko cells from either N-FL or T21-FL progenitors (Fig. 4, B and C), although limited engraftment of early erythroid lineage cells was observed in mice transplanted with N-FL GATA1s cells (fig. S9, A and B). Even higher doses of 5000 control, GATA1s, or STAG2ko progenitors from T21-FL did not produce any $CD45^+$ engraftment (fig. S9, C and D). This is consistent with the limited self-renewal and repopulation potential



B

N-FL	Dose	2° transplantation engrafted/injected	Normal stem cell frequency	Preleukemia initiating cell frequency	Leukemia initiating cell frequency
Control	11,111	1/5	1/96,525	-	-
	33,333	0/5			
	100,000	4/5			
GATA1s	11,111	2/5	1/51,618	-	-
	33,333	2/5			
	100,000	4/5			
STAG2ko	11,111	4/5	1/6,575	-	-
	33,333	5/5			
	100,000	5/5			
GATA1s/STAG2ko	11,111	3/5	-	-	1/45,865
	33,333	2/5			
	100,000	3/4			

T21-FL	Dose	2° transplantation engrafted/injected	Normal stem cell frequency	Preleukemia initiating cell frequency	Leukemia initiating cell frequency
Control	11,111	0/5	1/75,224	-	-
	33,333	2/2			
	100,000	1/2			
GATA1s	11,111	2/5	-	1/149,986	-
	33,333	0/5			
	100,000	2/5			
STAG2ko	11,111	1/5	1/49,755	-	-
	33,333	3/5			
	100,000	4/5			
GATA1s/STAG2ko	11,111	1/5	-	-	1/90,923
	33,333	3/5			
	100,000	2/5			



D

N-FL	Dose	3° transplantation engrafted/injected
GATA1s CD34+/CD117+	300,000	0/5
GATA1s/STAG2- CD34+/CD117+	50,000	1/5
GATA1s/STAG2- CD34-/CD117+	50,000	5/5

T21-FL	Dose	3° transplantation engrafted/injected
GATA1s CD34+/CD117+	2,500,000	0/5
GATA1s/STAG2- CD34+/CD117+	400,000	5/5
GATA1s/STAG2- CD34-/CD117+	400,000	3/5

Fig. 3. CD117 marks preleukemia and leukemia initiating cells. (**A**) Experimental overview of secondary xenotransplantation experiments. Flow cytometry plots of sorted human fractions from primary grafts are depicted. (**B**) Stem cell frequencies based on secondary xenotransplantations as described in (A). Limiting dilution analysis was used to assess normal, preleukemia-initiating, and leukemia-initiating cell frequencies (>0.1% CD45$^+$ cells in BM was defined as engraftment, n = 2 to 5 mice for each condition and dose, total 113 mice). (**C**) Experimental overview of secondary xenotransplantations using sorted fractions of CD34$^+$CD117$^-$, CD34$^+$CD117$^+$, and CD34$^-$CD117$^+$ cells from primary grafts. Flow cytometry plots of sorted human fractions are shown, and highlighted cells were transplanted at defined doses into NSG mice. A blue plus sign indicates engraftment with CD45$^+$ cells, and a red minus sign indicates no engraftment in secondarily transplanted mice (n = 2 to 5 mice for each condition and dose, in total 333 mice) (stem cell frequencies are provided in table S3). (**D**) Tertiary xenotransplantations of N-FL and T21-FL grafts in NSG mice for 12 weeks (>0.1% CD45$^+$ cells in BM was defined as engraftment, n = 5 mice per condition). (**E**) CIBERSORTx analysis to computationally quantify cell type lineage in the sorted fractions from primary N-FL grafts (n = 3 replicates per condition). (**F**) CIBERSORTx analysis as described in (E) for T21-FL (n = 3 replicates per condition).

of progenitors but also highlights the inability of T21-FL GATA1s progenitor cells compared with LT-HSCs to initiate preleukemia. By contrast, human CD45$^+$ engraftment was observed in mice transplanted with N-FL or T21-FL GATA1s/STAG2ko cells, regardless of the differentiation stage of the progenitors (Fig. 4, B and C). Only progenitors and stem cells could initiate leukemic engraftment because CD34$^-$ mature cells from T21-FL failed to initiate any CD45$^+$ grafts upon GATA1s/STAG2ko, even when transplanted at a high dose of 125,000 cells for 12 weeks (fig. S9E). All grafts generated by N-FL and T21-FL GATA1s/STAG2ko progenitors contained high proportions of CD117$^+$ blasts (Fig. 4, D and E, and fig. S9F) accompanied by other phenotypic markers typical of Down syndrome leukemia (fig. S9, G and H). Last, cells harvested from grafts generated by N-FL and T21-FL GATA1s/STAG2ko progenitors were able to propagate the leukemia

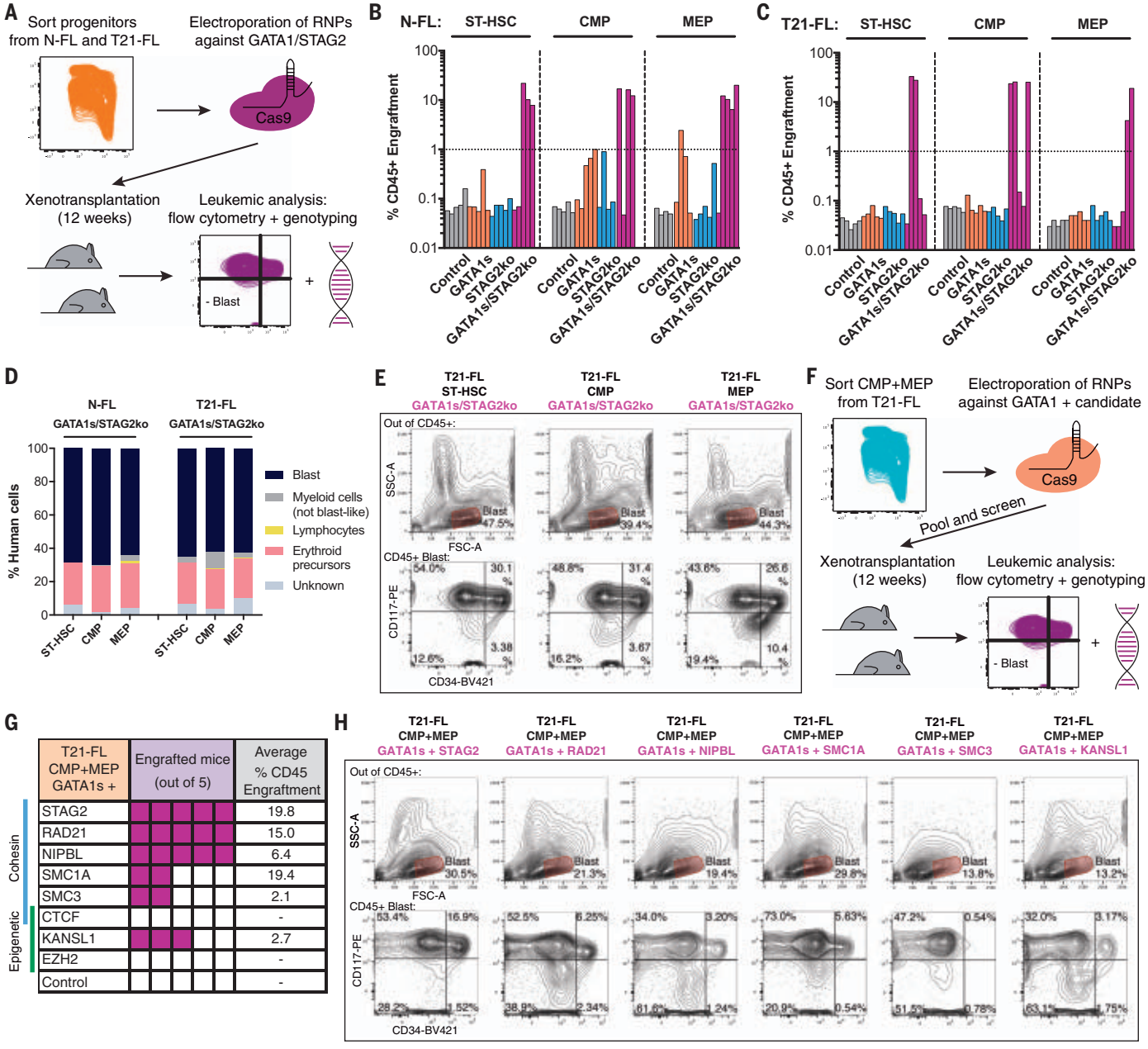


Fig. 4. Combined GATA1s and STAG2ko drive leukemic progression in progenitors. (**A**) Experimental overview of sorting N-FL and T21-FL derived progenitor cells for CRISPR/Cas9 editing and transplanting into NSGW41 mice. (**B**) Engraftment of N-FL ST-HSC, CMP, and MEP grafts in NSGW41 mice (all mice are shown regardless of $CD45^+$ engraftment, n = 4 or 5 mice per condition). (**C**) Engraftment as described in (B) for T21-FL (n = 5 mice per condition). (**D**) Quantification of cell morphology of human cells prepared by using cytospin in N-FL and T21-FL GATA1s/STAG2ko grafts in NSGW41 mice from (B) and (C) (n = 400 cells per condition). (**E**) Flow cytometry plots depicting the blast population out of $CD45^+$ cells in primary xenografts of T21-FL GATA1s/STAG2ko progenitors in NSGW41 mice, as described in (C). The CD34/CD117 profiles out of the $CD45^+$ blast populations are depicted below. (**F**) Experimental overview of sorting T21-FL CMPs and MEPs to conduct a loss-of-function screen to identify genes that endow leukemic progression in combination with GATA1s. (**G**) Result of screen described in (F) showing the number of mice with leukemic phenotypes based on average $CD45^+$ engraftment in BM and blast appearance (>1% $CD45^+$ in BM, n = 5 mice per condition). (**H**) Flow cytometry plots of blast populations out of $CD45^+$ cells in grafts of T21-FL CMPs and MEPs edited with GATA1s and candidate gene gRNAs as described in (F).

in secondary recipients (fig. S9, I and J). Taken together, our results, on the basis of the assessment of these defined HSPC subpopulations, suggest that the GATA1s preleukemia-initiating event likely occurs in LT-HSCs and not downstream progenitors. However, subsequent *STAG2* mutations are not limited to LT-HSCs but can be acquired further downstream in the expanded pool of GATA1s-primed progenitor cells, highlighting that the preleukemic and leukemic events could occur in distinct cells of origin.

To verify whether leukemic transformation can be induced in a stepwise transplantation setting and to confirm whether progenitor-like

cells could be responsible for leukemic progression, we sorted CD34$^+$CD38$^+$ progenitor-enriched and CD34$^+$CD38$^-$ stem cell–enriched HSPCs from T21-FL GATA1s primary xenografts, induced STAG2ko, and transplanted them into NSGW41 secondary recipients for 12 weeks (fig. S10A). Stepwise introduction of GATA1s followed by STAG2ko in both subpopulations of T21-FL HSPCs elicited leukemic transformation, which was evident by the higher percentage of blasts as compared with that of the GATA1s-edited preleukemic control (fig. S10, B to E), confirming that progenitor-like cells with acquired *STAG2* mutations could drive leukemic progression. Furthermore, to understand whether leukemic transformation with GATA1s and STAG2ko is developmentally restricted, we introduced GATA1s/STAG2ko in normal disomic FL-, postnatal umbilical cord blood (CB)–, or adult BM–derived CD34$^+$ enriched HSPCs and transplanted them into NSGW41 mice for 12 weeks (fig. S11A). CD34$^+$ cells from only FL and CB, but not BM, were able to induce leukemic transformation as assessed from blast accumulation and their characteristic immunophenotype (fig. S11, B to G). Therefore, the potential for leukemic transformation is developmentally restricted to a time window during fetal and early postnatal development.

To explore whether mutations in genes other than *STAG2* can drive leukemic transformation, we carried out a focused loss-of-function screen to evaluate the effects of deleting seven additional genes in T21-FL GATA1s CMPs and MEPs, including four additional cohesin genes and three genes encoding epigenetic regulators that are frequently mutated in Down syndrome leukemia (*11*, *12*). For each gene, four gRNAs were individually introduced into T21-FL progenitor cells together with GATA1s, and cells were pooled after CRISPR/Cas9 editing and transplanted at a dose of 20,000 cells into NSGW41 mice (Fig. 4F and table S7). After 12 weeks, all five cohesin gene mutations, each in combination with GATA1s, drove leukemic engraftment in mice (average level of CD45$^+$ engraftment 2 to 20%), with *STAG2*, *RAD21*, and *NIPBL* being the most consistent (Fig. 4G). Of the three targeted epigenetic regulators, mutations in only *KANSL1* drove leukemic transformation with GATA1s, implying that additional events are needed in the case of *CTCF* and *EZH2* mutations. As expected, control-edited T21-FL progenitor cells with GATA1s did not produce any CD45$^+$ grafts. All leukemic grafts contained CD117$^+$ blasts with varying degrees of CD34$^+$ expression (Fig. 4H). The blast immunophenotype in the leukemic grafts was similar regardless of the underlying mutation (fig. S12A), suggesting that the mutations converge on a common pathway for leukemic transformation. By contrast, a previous loss-of-function screen in a mouse model of TAM did not identify cohesin mutations as drivers of Down syndrome leukemogenesis (*11*), implying marked differences between mouse and human systems in their susceptibility to particular mutations. Altogether, our results show that specific cell types within a particular developmental time window are susceptible to GATA1s-induced preleukemia and GATA1s- and cohesin mutation–induced leukemia, underscoring the importance of the cellular and developmental context during leukemogenesis.

Chr21 microRNAs predispose to preleukemia

To investigate the mechanism underlying the cooperation between T21 and GATA1s in driving preleukemia development, we analyzed the binding occupancy of GATA1. To do this, we performed Cut&Run assays (*40*) to profile genome-wide GATA1 binding sites to quantify binding changes upon GATA1s editing in N-FL and T21-FL CD34$^+$-enriched HSPCs. GATA1s retained many of the binding sites of full-length GATA1, as evidenced by the large number of shared peaks (fig. S12B), which is consistent with previously reported findings in a mouse cell line and a mouse model of *Gata1s* (*41*, *42*). GATA-binding motifs were highly enriched in these peaks, as were motifs for ETS family members (fig. S12C), suggesting binding cooperativity with GATA1. Pathway enrichment analysis of GATA1s-specific peaks in T21-FL compared with either control-edited full-length GATA1 peaks in T21-FL or GATA1s peaks in N-FL revealed a 13-fold enrichment of promoter sites of genes involved in microRNA (miRNA) loading (fig. S12, D and E, and table S8), which was confirmed through gene expression of *AGO1*, *AGO2*, *TARBP2*, and *ADAR* (fig. S12F). These results suggest that GATA1s binding to these miRNA biogenesis genes increases their respective expression in the T21 context, possibly further influencing miRNA-mediated silencing and posttranscriptional regulation.

To explore this idea further, we investigated miRNA expression in T21-FL. We profiled miRNAs from N-FL and T21-FL CD34$^+$-enriched HSPCs with next-generation sequencing. Differential expression of miRNAs on Chr21 was not observed (fig. S12, G and H). However, when Chr21 miRNAs were profiled by means of quantitative PCR in highly purified LT-HSCs, as compared with bulk CD34$^+$ cells, differences were found (Fig. 5A). *MiR-99a*, *miR-125b-2*, *miR-155*, and *let-7c* were up-regulated in T21-FL LT-HSCs compared with N-FL, with the first three having the greatest differential expression.

To investigate whether our observed T21-specific phenotypes could be recapitulated upon enforced expression of these differentially expressed Chr21 miRNAs in N-FL LT-HSCs, we used lentiviral transduction to overexpress *miR-99a*, *miR-125b-2*, and *miR-155* (Chr21 miRNAs) together with the fluorescent marker mOrange. Transduced cells were transplanted into NSG and NSGW41 mice, and lineage output was assessed at 12 weeks (Fig. 5B and fig. S12I). Cells transduced with Chr21 miRNAs generated twofold higher engraftment in BM and spleen as compared with control-transduced cells (Fig. 5C). Chr21 miRNA grafts displayed a significant bias toward increased myeloid and decreased lymphoid differentiation in the transplanted BM ($P < 0.0001$) (Fig. 5D and fig. S13, A to E), similar to the lineage output of control CRISPR/Cas9–edited T21-FL cells in transplanted NSG mice (Fig. 2D). Similar grafts were seen in NSGW41 recipients of Chr21 miRNAs N-FL LT-HSCs (fig. S13, F and G). No elevated or abnormal blast populations were detected in any of these grafts with morphologic or flow cytometric analysis (Fig. 5, E and F, and fig. S13H). Immunophenotypic analysis of the HSC hierarchy of engrafted mice revealed reduced LT-HSCs in Chr21 miRNA grafts, which resulted in a lower ability of transduced CD45$^+$ cells to engraft in secondary NSG mice (fig. S13, I and J). ATAC-seq and RNA-seq analysis of Chr21 miRNA LT-HSCs cultured in vitro and compared with analogously cultured T21-FL control-edited LT-HSCs showed similar chromatin accessibility profiles and enrichment of similar down-regulated genes (fig. S13, K and L, and table S9). These results demonstrate that simultaneous overexpression of *miR-99a*, *miR-125b-2*, and *miR-155* in N-FL LT-HSCs recapitulates features of a T21-like hematopoietic state.

Next, to examine the role of Chr21 miRNAs in preleukemic initiation and leukemic transformation, we first deleted Chr21 miRNAs in T21-FL LT-HSCs and then followed with CRISPR/Cas9 editing for GATA1s, with or without STAG2ko, and transplanted into NSG mice (Fig. 5G and fig. S13 M to O). Deletion of Chr21 miRNAs combined with GATA1s resulted in a significant reduction in the blast population, including CD117$^+$CD45$^+$ blasts, at 20 weeks after transplantation ($P < 0.001$) (Fig. 5, H to K). However, leukemic engraftment or blast accumulation in mice with T21-FL GATA1s/STAG2ko grafts with or without deletion of Chr21 miRNAs were similar. Taken together, these results suggest that Chr21 miRNAs seem to play a major role in preleukemic initiation but are dispensable for leukemic progression.

CD117/KIT inhibition targets preleukemic-initiating cells and inhibits leukemic progression

Currently, there are no effective strategies to prevent progression from preleukemia to leukemia in individuals with Down syndrome. Our results indicate that CD117/KIT expression marked the cells that mediated the propagation of the GATA1s-induced preleukemia and GATA1s/STAG2ko–induced leukemia. Thus, it is possible that both the preleukemia and leukemia are dependent on KIT signaling for maintenance and progression.

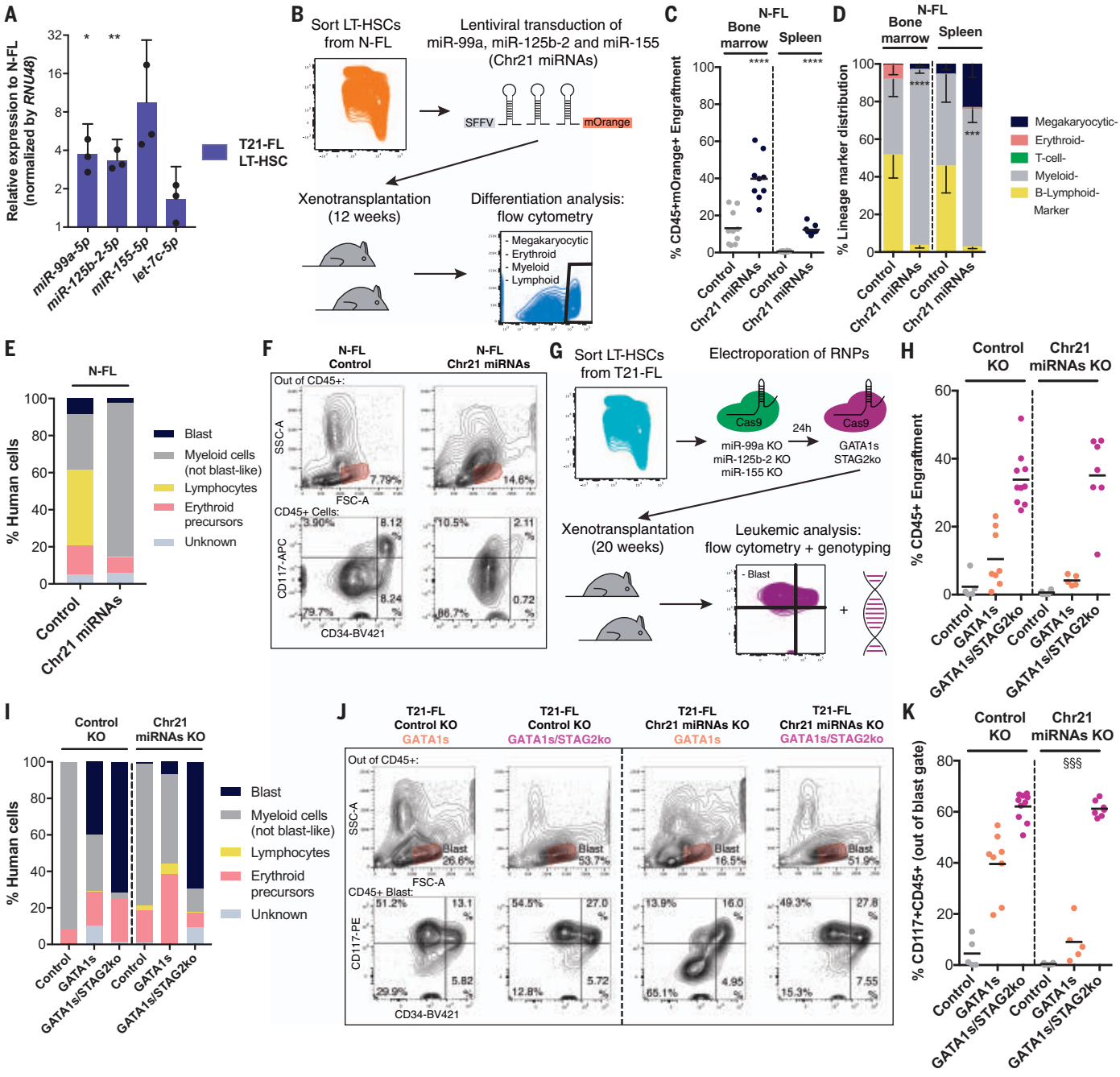


Fig. 5. Chr21 miRNAs predispose to preleukemia. (**A**) Relative expression of Chr21 miRNAs in T21-FL LT-HSCs compared with N-FL as measured with reverse-transcription quantitative PCR (n = 3 replicates per condition). (**B**) Overview of lentiviral mediated overexpression of Chr21 miRNAs in N-FL LT-HSCs used for primary xenotransplantation into NSG and NSGW41 mice. (**C**) Engraftment of transduced control and Chr21 miRNAs in N-FL LT-HSCs transplanted into NSG mice (only mice with >1% CD45$^+$ cells in BM were analyzed, n = 9 or 10 mice per condition). (**D**) Lineage marker distribution based on cell surface markers of engrafted NSG mice in (C). (**E**) Quantification of cell morphology of human cells prepared by using cytospin in grafts described in (C) (n = 400 cells per condition). (**F**) Flow cytometry plots out of CD45$^+$ cells in primary xenografts described in (C). (**G**) Experimental overview of sorting T21-FL LT-HSCs for CRISPR/Cas9 editing with miR-99a, miR-125b-2, and miR-155 gRNAs and subsequently with GATA1s and STAG2 gRNAs for primary xenotransplantation into NSG mice. (**H**) Engraftment of control knockout (KO) and Chr21 miRNAs KO in T21-FL LT-HSCs transplanted into NSG mice. Each subgroup was additionally edited with control, GATA1s, and GATA1s/STAG2 gRNAs (only mice with >1% CD45$^+$ cells in BM and >90% CRISPR/Cas9 efficiency are depicted, n = 5 to 10 mice per condition). (**I**) Quantification of cell morphology of human cells prepared by using cytospin in transplanted NSG mice described in (H) (n = 400 cells per condition). (**J**) Flow cytometry plots showing blast populations out of CD45$^+$ cells in primary xenografts described in (H). (**K**) Quantification of CD117$^+$CD45$^+$ blasts of transplanted NSG mice described in (H) (§§§ indicate significance in relation to GATA1s control KO). Unpaired Student's t test: *P < 0.05; **P < 0.01; ***/§§§P < 0.001; ****P < 0.0001; error bars indicate standard deviation.

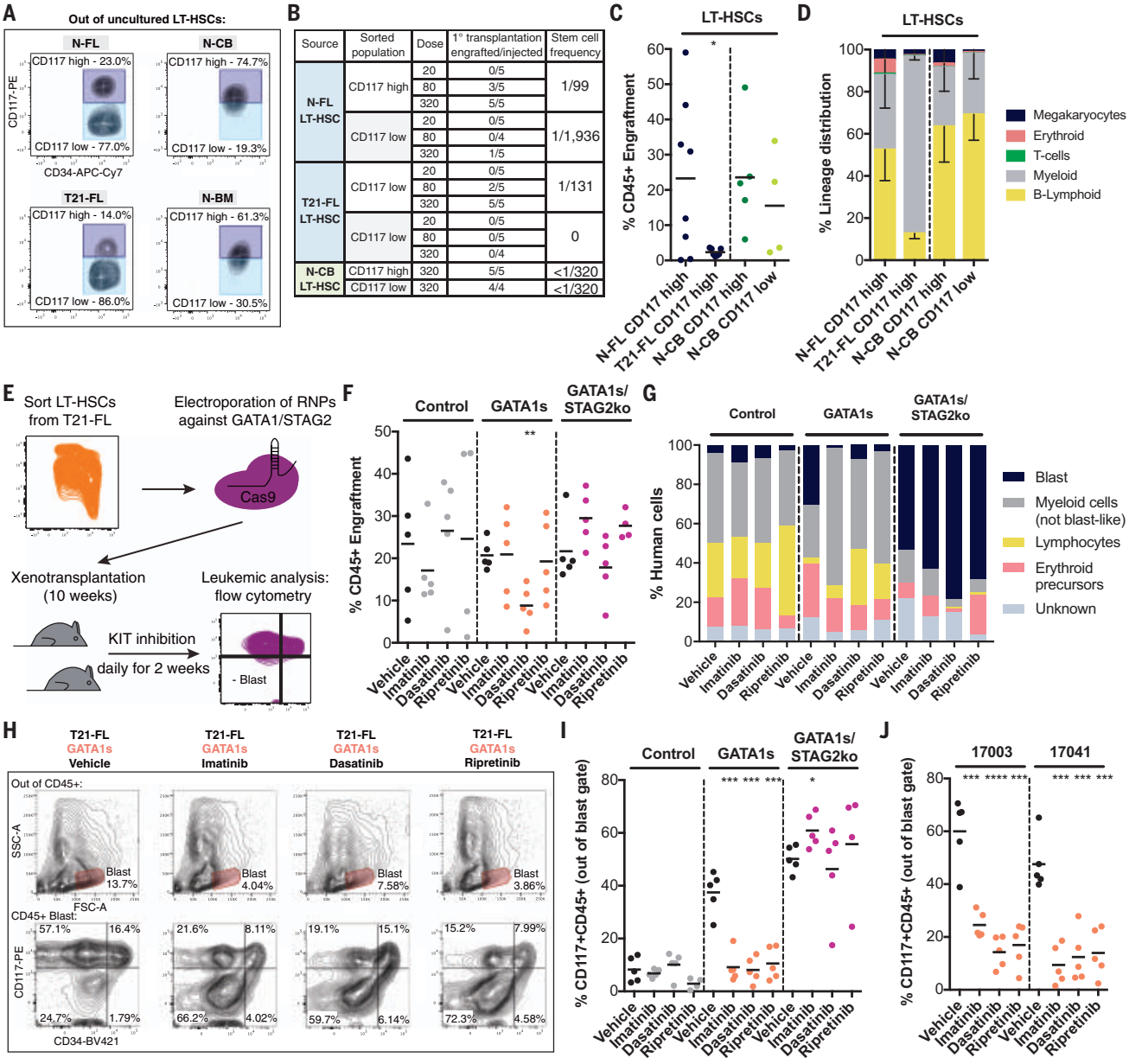
CD117/KIT is a receptor tyrosine kinase that regulates HSC proliferation, maintenance, and survival after binding to its ligand, stem cell factor (*43*). To analyze KIT expression in normal hematopoiesis, FL-, CB-, and BM-derived LT-HSCs were immunophenotypically profiled for CD117 expression. N-FL and T21-FL LT-HSCs contained distinct CD117-low and CD117-high populations (Fig. 6A). By contrast, LT-HSCs from N-CB and N-BM showed a single population of cells with a continuum of low to high CD117 expression. After transplantation at limiting cell dose into NSG mice, only the CD117-high populations and not the CD117-low populations from N-FL and T21-FL LT-HSCs were able to generate grafts at 20 weeks (Fig. 6, B and C). By contrast, both CD117-low and CD117-high N-CB LT-HSCs were able to generate long-term engraftment, albeit with different lineage outputs (Fig. 6D). These results suggest that KIT signaling plays an essential role in LT-HSC function during fetal development. On the basis of the engraftment results of disomic CB (Fig. 6, B and C), KIT signaling may be less dependent, at least transiently, for LT-HSC function after birth.

To investigate whether pharmacological inhibition of KIT can target and eliminate preleukemic and leukemic blasts, mice engrafted with 1300 T21-FL control, GATA1s, or GATA1s/STAG2ko LT-HSCs were treated with first-, second-, and third-generation KIT inhibitors (50 mg/kg imatinib, 20 mg/kg dasatinib, or 7.5 mg/kg ripretinib) starting 10 weeks after transplantation with twice-daily dosing for 2 weeks (Fig. 6E) (*44*–*46*). KIT inhibition did not have a significant effect on the overall level of $CD45^+$ engraftment for any group, except for dasatinib-treated mice bearing T21-FL GATA1s preleukemic grafts ($P < 0.01$) (Fig. 6F). KIT inhibition had no effect on the blast population of leukemic grafts generated by T21-FL GATA1s/STAG2ko cells (Fig. 6G). GATA1s preleukemic grafts from mice treated with any of the KIT inhibitors contained significantly lower proportions of blasts as compared with that of vehicle-treated mice ($P < 0.001$), with reduction of $CD117^+CD45^+$ blast populations to levels seen in controls (Fig. 6, H and I, and fig. S14A). KIT inhibitor–treated mice revealed an increase in granular $CD33^+$, $CD11b^+$, and $CD13^+$ myeloid cells, suggesting differentiation of blast cells toward more mature myeloid cells (fig. S14, B and C). Because some residual $CD117^+$ blasts remained detectable in mice with preleukemic GATA1s grafts (Fig. 6H), cells harvested from primary mice were serially transplanted at defined doses into secondary NSG recipients to determine whether preleukemia-initiating cells were affected. Cells from vehicle-treated mice showed a 32-fold higher ability to generate secondary grafts at 8 weeks compared with cells from KIT inhibitor–treated mice (fig. S14D). This lies in contrast to the secondary grafts generated from KIT inhibitor–treated GATA1s/STAG2ko–induced leukemia, which did not show any difference in their ability to generate secondary grafts (fig. S14E). To further validate the sensitivity of KIT inhibition in GATA1s-induced preleukemia, two primary TAM samples were phenotypically characterized (fig. S14, F and G, and table S10), subsequently transplanted into mice, and treated with KIT inhibitors starting at 6 weeks after transplantation for 2 weeks (fig. S15A). KIT inhibition resulted in reduced $CD45^+$ engraftment with significantly reduced $CD117^+CD45^+$ blast populations ($P < 0.001$) and increased granular $CD33^+$ myeloid cells (Fig. 6J and fig. S15, B to H). However, we cannot rule out that the effects of pharmacological KIT inhibition could also be mediated, at least partly, through other receptor tyrosine kinases. Altogether, our results demonstrate as a proof of principle that KIT inhibition targets preleukemic expansion (fig. S15I), supporting further clinical evaluation of the concept that preleukemic intervention could inhibit progression to leukemia.

Discussion

Our study provides insight into the cellular and molecular mechanism of Down syndrome leukemogenesis, from atypical hematopoiesis associated with T21 to preleukemia initiation and ultimately to leukemic progression. We confirmed that the T21-FL hematopoietic system exhibits an altered phenotypic HSPC hierarchy as previously described (*25*, *47*, *48*). Although earlier reports have proposed that T21 enhances self-renewal in vitro (*48*), our functional studies revealed the opposite; individual T21 HSPC subpopulations exhibited reduced proliferation in vitro and generated smaller grafts in xenotransplanted mice with myeloid and megakaryocytic bias and reduced serial transplant ability. These are likely cell-autonomous effects and may be the basis for the higher incidence of hematopoietic abnormalities such as isolated cytopenias, myelodysplasia, and BM failure seen in adults with Down syndrome (*49*). Despite the reduced proliferative capacity of T21 LT-HSCs, our data demonstrate that preleukemia can be initiated in this cellular compartment, contrary to previous hypotheses that megakaryocytic-erythroid progenitor cells are the cell of origin for preleukemia, a prediction derived from their expansion in the HSPC hierarchy of T21-FL (*50*). The reduced proliferative capacity of T21 LT-HSCs is offset by the acquisition of *GATA1* mutations, providing a possible explanation for the observed selection of *GATA1* mutations in the context of T21. However, the increased function provided by GATA1 mutation comes at a cost, including the development of preleukemia and a hindrance in erythrocyte maturation—a result consistent with the anemia seen in GATA1-deficient mice and humans (*51*–*54*). However, because we based our study on well-characterized functional HSPC subpopulations, we cannot exclude the possibility that undefined populations may also be able to initiate preleukemia (*55*). In contrast to the LT-HSC origin for preleukemia, leukemic progression can occur in multiple types of downstream progenitors in addition to LT-HSCs. The overall pool of progenitors is vastly expanded owing to GATA1s-priming, providing a large reservoir for acquisition of secondary mutations in genes such as *STAG2* and thereby increasing the probability of leukemic progression. Enhanced self-renewal mediated by STAG2 deficiency could explain why it is subsequently selected for during leukemic evolution. Selection could also arise from STAG2 deficiency resulting in a temporary increase in mature erythroid output, as seen in our in vitro single-cell differentiation assays. Erythroid cells make up the vast majority of the numerical daily output of the blood system, raising the possibility that there may be strong evolutionary pressure for the erythropoiesis-defective GATA1s-mutated clones to reacquire erythroid potential through additional *STAG2* mutation. Further, this leukemic transformation can only occur during fetal and early postnatal development but not in adult BM stem cells. Thus, our study reveals how critical it is to understand the identity and the developmental stage of the cell type that acquires genetic drivers during leukemogenesis. Moreover, genetic drivers of leukemia are typically distinct between pediatric and adult acute myeloid leukemia (*56*), so our findings uncover that the basis for the differential leukemic potential could possibly be the developmental status of the cell of origin.

Our findings establish that initiation of GATA1s-induced preleukemia is dependent on T21, which exerts its effects at least in part through up-regulation of Chr21 miRNAs—specifically, *miR-99a*, *miR-125b-2*, and *miR-155*—exclusively within the LT-HSC compartment. This result refines previous suggestions that deregulated expression of many Chr21 genes contributes to Down syndrome leukemogenesis and extends them to noncoding RNAs (*57*). These three miRNAs are highly expressed in leukemia-initiating populations of adult acute myeloid leukemia (*58*), and miR-125b-2 has been shown to be a potential oncomiR (*59*). Children with Down syndrome are also at high risk for developing B cell ALL (*60*), and it will be important to determine whether this mechanism of predisposition also affects ALL development. Although preleukemic initiation is dependent on T21, we made the unexpected finding that progression to leukemia is independent of T21 and can be induced by deficiency of STAG2 in combination with GATA1s. We found that preleukemic and leukemic populations were similar with respect to expression of lineage markers on blasts, enrichment



Source	Sorted population	Dose	1° transplantation engrafted/injected	Stem cell frequency
N-FL LT-HSC	CD117 high	20	0/5	1/99
		80	3/5	
		320	5/5	
	CD117 low	20	0/5	1/1,936
		80	0/4	
		320	1/5	
T21-FL LT-HSC	CD117 low	20	0/5	1/131
		80	2/5	
		320	5/5	
	CD117 low	20	0/5	0
		80	0/5	
		320	0/4	
N-CB LT-HSC	CD117 high	320	5/5	<1/320
	CD117 low	320	4/4	<1/320

Fig. 6. KIT inhibition targets preleukemic-initiating cells and inhibits leukemic progression. (**A**) Immunophenotypic profile of CD117 and CD34 expression of isolated LT-HSCs from N-FL and T21-FL, normal disomic CB (N-CB), and normal disomic BM (N-BM). (**B**) CD117-low and CD117-high LT-HSCs were transplanted at defined doses into NSG mice for 20 weeks. Resulting stem cell frequencies are depicted (>0.1% $CD45^+$ cells in BM was defined as engraftment, $n = 4$ or 5 mice per condition, total 67 mice). (**C**) Engraftment of N-FL CD117-high, T21-FL CD117-high, N-CB CD117-high, and N-CB CD117-low LT-HSCs transplanted into NSG mice ($n = 4$ to 8 mice per condition). (**D**) Lineage marker distribution based on cell surface markers of engrafted NSG mice from (C). (**E**) Experimental overview of control-, GATA1s-, and GATA1s/STAG2ko–edited T21-FL LT-HSCs transplanted into NSG mice, which were subsequently treated twice daily with small-molecule inhibitors against KIT for 2 weeks. (**F**) Engraftment of T21-FL LT-HSCs transplanted into NSG mice treated with vehicle, imatinib, dasatinib, or ripretinib ($n = 4$ or 5 mice per condition). (**G**) Quantification of cell morphology of human cells prepared by using cytospin in transplanted NSG mice described in (F) ($n = 400$ cells per condition). (**H**) Flow cytometry plots showing blast populations out of $CD45^+$ cells in primary xenografts described in (F). (**I**) Quantification of $CD117^+CD45^+$ blasts in transplanted NSG mice described in (F). (**J**) Quantification of $CD117^+CD45^+$ blasts in NSGW41 mice transplanted with primary sample TAM 17003 and NSG mice transplanted with primary sample TAM 17041 and treated with vehicle, imatinib, dasatinib, or ripretinib ($n = 5$ mice per condition). Unpaired Student's t test: $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; error bars indicate standard deviation.

of GATA1-binding sites at their promoters, and down-regulated pathways compared with T21 controls. Nevertheless, the addition of STAG2ko to GATA1s-bearing cells led to enhanced self-renewal, as evident in the increased frequency of leukemia-initiating cells enriched in $CD117^+$ cells. Similar leukemic transformation was observed upon induced deficiency of other cohesin genes in combination with GATA1s, and thus, it is possible that the effects of these mutations converge on an increase in self-renewal and stemness programs in general. LSC17 stemness signature being strongly associated with survival across a wide spectrum of acute myeloid leukemia patients irrespective

of genetic drivers (*61*) lends support to the concept of convergence. Last, although our results demonstrate that Down syndrome leukemogenesis can be modeled by a sequence of cell-intrinsic mechanisms, we cannot exclude an additional role for the T21 microenvironment in leukemic evolution in individuals with Down syndrome.

Through analysis of individual preleukemic and leukemic cell fractions isolated on the basis of primitive stem cell markers and propagated in serial transplants, we identified CD117/KIT as a marker of disease-driving cells. Further, pharmacological KIT inhibition targeted preleukemic stem cells, and this proof of concept could open an approach for therapy at the preleukemic stage. This would limit the pool of GATA1s-primed progenitors that can acquire additional mutations and ultimately inhibit progression to ML-DS. Such a scenario could foreseeably work in combination with refined clinical scores capable of identifying patients with TAM who have risk of progression toward leukemia subtypes that have poor outcome (*62*). However, this concept requires further in-depth preclinical and clinical assessment. Our findings not only provide insight into Down syndrome leukemogenesis but also have implications for pediatric leukemia in general. Sequencing data from newborns indicated that the first genetic alterations for many subtypes of childhood leukemia occur during fetal development (*63*–*65*). Our results potentially suggest that the cellular origin of preleukemic mutations in other pediatric leukemia may also be LT-HSCs, which is supported by that it can take years after birth until leukemia is diagnosed (*66*). Early targeting of preleukemia during the newborn phase could become a clinical paradigm for pediatric leukemia. Last, numerical and structural alterations of Chr21 are extremely common in hematological malignancies in both children and adults (*67*). Gains of Chr21 are seen in up to 35% of cases in several types of acute leukemia (*68*). Thus, it will be important to identify mechanisms including but not limited to dysregulation of Chr21 miRNAs that are shared among those cases.

Materials and methods summary

All mouse experiments were approved by the University Health Network (UHN) Animal Care Committee, and we confirm that all experiments conform to the relevant regulatory and ethical standards. All xenotransplantations were performed in 8- to 12-week-old female *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) mice (JAX) that were sublethally irradiated with 225cGy, 24 hours before transplantation, or in 8- to 12-week-old female *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}Kit^{em1Mvw}/SzJ* (NSGW41) mice that were not irradiated. Human fetal liver samples were obtained from elective pregnancy terminations at Mount Sinai Hospital with informed consent in accordance to guidelines approved by the Mount Sinai Hospital Research Ethics Board (18-0093-E) and the UHN Research Ethics Board (02-0763). Fetal liver samples of normal disomic karyotype and T21 were collected at 16 to 19 weeks gestation from either sex. Fetal liver samples were processed within 1 to 3 hours, and CD34$^+$ cells were isolated by use of the human CD34 MicroBead kit (Miltenyi Biotec) according to the manufacturer's protocol (table S11). For all in vitro and in vivo experiments, the full stem and progenitor hierarchy was used to sort LT-HSCs, ST-HSCs, CMPs, and MEPs (*26*). CRISPR/Cas9 RNP electroporations were carried out by using chemically synthesized gRNAs (IDT), recombinant Cas9 nuclease (IDT), and the 4D-Nucleofector (Lonza) as previously described (*22*). Single-cell in vitro assays were performed in erythro-myeloid-megakaryocytic promoting medium. Xenotransplantations of CRISPR/Cas9–edited LT-HSCs were performed at cell doses of 300 to 400 per mouse, unless otherwise specified, to obtain near-clonal xenotransplantations.

REFERENCES AND NOTES

1. H. Hasle, I. H. Clemmensen, M. Mikkelsen, Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet* **355**, 165–169 (2000). doi: 10.1016/S0140-6736(99)05264-2; pmid: 10675114
2. M. F. Greaves, A. T. Maia, J. L. Wiemels, A. M. Ford, Leukemia in twins: Lessons in natural history. *Blood* **102**, 2321–2333 (2003). doi: 10.1182/blood-2002-12-3817; pmid: 12791663
3. J. K. Hitzler, J. Cheung, Y. Li, S. W. Scherer, A. Zipursky, GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome. *Blood* **101**, 4301–4304 (2003). doi: 10.1182/blood-2003-01-0013; pmid: 12586620
4. I. Roberts *et al*., GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: Identification of a population at risk of leukemia. *Blood* **122**, 3908–3917 (2013). doi: 10.1182/blood-2013-07-515148; pmid: 24021668
5. J. Wechsler *et al*., Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* **32**, 148–152 (2002). doi: 10.1038/ng955; pmid: 12172547
6. S. Hoeller *et al*., Morphologic and GATA1 sequencing analysis of hematopoiesis in fetuses with trisomy 21. *Hum. Pathol.* **45**, 1003–1009 (2014). doi: 10.1016/j.humpath.2013.12.014; pmid: 24746204
7. J. W. Taub *et al*., Prenatal origin of GATA1 mutations may be an initiating step in the development of megakaryocytic leukemia in Down syndrome. *Blood* **104**, 1588–1589 (2004). doi: 10.1182/blood-2004-04-1563; pmid: 15317736
8. M. Uffmann *et al*., Therapy reduction in patients with Down syndrome and myeloid leukemia: The international ML-DS 2006 trial. *Blood* **129**, 3314–3321 (2017). doi: 10.1182/blood-2017-01-765057; pmid: 28400376
9. J. W. Taub *et al*., Improved outcomes for myeloid leukemia of Down syndrome: A report from the Children's Oncology Group AAML0431 trial. *Blood* **129**, 3304–3313 (2017). doi: 10.1182/blood-2017-01-764324; pmid: 28389462
10. A. S. Gamis *et al*., Natural history of transient myeloproliferative disorder clinically diagnosed in Down syndrome neonates: A report from the Children's Oncology Group Study A2971. *Blood* **118**, 6752–6759, quiz 6996 (2011). doi: 10.1182/blood-2011-04-350017; pmid: 21849481
11. M. Labuhn *et al*., Mechanisms of progression of myeloid preleukemia to transformed myeloid leukemia in children with Down syndrome. *Cancer Cell* **36**, 340 (2019). doi: 10.1016/j.ccell.2019.08.014; pmid: 31526763
12. K. Yoshida *et al*., The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* **45**, 1293–1299 (2013). doi: 10.1038/ng.2759; pmid: 24056718
13. D. Hong *et al*., Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* **319**, 336–339 (2008). doi: 10.1126/science.1150648; pmid: 18202291
14. J.-H. Klusmann *et al*., Treatment and prognostic impact of transient leukemia in neonates with Down syndrome. *Blood* **111**, 2991–2998 (2008). doi: 10.1182/blood-2007-10-118810; pmid: 18182574
15. A. D. Sorrell *et al*., Favorable survival maintained in children who have myeloid leukemia associated with Down syndrome using reduced-dose chemotherapy on Children's Oncology Group trial A2971: A report from the Children's Oncology Group. *Cancer* **118**, 4806–4814 (2012). doi: 10.1002/cncr.27484; pmid: 22392565
16. T. Taga *et al*., Clinical characteristics and outcome of refractory/relapsed myeloid leukemia in children with Down syndrome. *Blood* **120**, 1810–1815 (2012). doi: 10.1182/blood-2012-03-414755; pmid: 22776818
17. J. K. Hitzler *et al*., Outcome of transplantation for acute myelogenous leukemia in children with Down syndrome. *Biol. Blood Marrow Transplant.* **19**, 893–897 (2013). doi: 10.1016/j.bbmt.2013.02.017; pmid: 23467128
18. J. T. Caldwell, Y. Ge, J. W. Taub, Prognosis and management of acute myeloid leukemia in patients with Down syndrome. *Expert Rev. Hematol.* **7**, 831–840 (2014). doi: 10.1586/17474086.2014.959923; pmid: 25231553
19. K. A. Alford *et al*., Perturbed hematopoiesis in the Tc1 mouse model of Down syndrome. *Blood* **115**, 2928–2937 (2010). doi: 10.1182/blood-2009-06-227629; pmid: 20154221
20. G. Kirsammer *et al*., Highly penetrant myeloproliferative disease in the Ts65Dn mouse model of Down syndrome. *Blood* **111**, 767–775 (2008). doi: 10.1182/blood-2007-04-085670; pmid: 17901249
21. S. Malinge *et al*., Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *J. Clin. Invest.* **122**, 948–962 (2012). doi: 10.1172/JCI60455; pmid: 22354171
22. E. Wagenblast *et al*., Functional profiling of single CRISPR/Cas9-edited human long-term hematopoietic stem cells. *Nat. Commun.* **10**, 4730–11 (2019). doi: 10.1038/s41467-019-12726-0; pmid: 31628330
23. F. Notta *et al*., Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* **469**, 362–367 (2011). doi: 10.1038/nature09733; pmid: 21248843
24. K. A. Alford *et al*., Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. *Blood* **118**, 2222–2238 (2011). doi: 10.1182/blood-2011-03-342774; pmid: 21715302
25. A. Roy *et al*., Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17579–17584 (2012). doi: 10.1073/pnas.1211405109; pmid: 23045701
26. F. Notta *et al*., Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016). doi: 10.1126/science.aab2116; pmid: 26541609
27. S. Rahmig *et al*., Improved human erythropoiesis and platelet formation in humanized NSGW41 mice. *Stem Cell Reports* **7**, 591–601 (2016). doi: 10.1016/j.stemcr.2016.08.005; pmid: 27618723
28. C. Langebrake, U. Creutzig, D. Reinhardt, Immunophenotype of Down syndrome acute myeloid leukemia and transient myeloproliferative disease differs significantly from other diseases with morphologically identical or similar blasts. *Klin. Padiatr.* **217**, 126–134 (2005). doi: 10.1055/s-2005-836510; pmid: 15858703
29. L. Wang *et al*., Acute megakaryoblastic leukemia associated with trisomy 21 demonstrates a distinct immunophenotype. *Cytometry B Clin. Cytom.* **88**, 244–252 (2015). doi: 10.1002/cytob.21198; pmid: 25361478
30. WHO, *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues* (World Health Organization, 2008).
31. N. Bhatnagar, L. Nizery, O. Tunstall, P. Vyas, I. Roberts, Transient abnormal myelopoiesis and AML in Down syndrome: An update. *Curr. Hematol. Malig. Rep.* **11**, 333–341 (2016). doi: 10.1007/s11899-016-0338-x; pmid: 27510823
32. O. Tunstall *et al*., Guidelines for the investigation and management of transient leukaemia of Down syndrome. *Br. J. Haematol.* **182**, 200–211 (2018). doi: 10.1111/bjh.15390; pmid: 29916557
33. H. Döhner *et al*., Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017). doi: 10.1182/blood-2016-08-733196; pmid: 27895058
34. A. D. Viny *et al*., Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control

of HSC self-renewal and differentiation. *Cell Stem Cell* **25**, 682–696.e8 (2019). doi: 10.1016/j.stem.2019.08.003; pmid: 31495782
35. A. M. Newman *et al*., Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019). doi: 10.1038/s41587-019-0114-2; pmid: 31061481
36. A. Schwarzer *et al*., The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat. Commun.* **8**, 218–17 (2017). doi: 10.1038/s41467-017-00212-4; pmid: 28794406
37. A. V. Krivtsov *et al*., Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006). doi: 10.1038/nature04980; pmid: 16862118
38. Y. Wang *et al*., The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. *Science* **327**, 1650–1653 (2010). doi: 10.1126/science.1186624; pmid: 20339075
39. M. Ye *et al*., Hematopoietic differentiation is required for initiation of acute myeloid leukemia. *Cell Stem Cell* **17**, 611–623 (2015). doi: 10.1016/j.stem.2015.08.011; pmid: 26412561
40. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017). doi: 10.7554/eLife.21856; pmid: 28079019
41. T. Ling *et al*., Chromatin occupancy and epigenetic analysis reveal new insights into the function of the GATA1 N terminus in erythropoiesis. *Blood* **134**, 1619–1631 (2019). doi: 10.1182/blood.2019001234; pmid: 31409672
42. T. M. Chlon, M. McNulty, B. Goldenson, A. Rosinski, J. D. Crispino, Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression. *Haematologica* **100**, 575–584 (2015). doi: 10.3324/haematol.2014.112714; pmid: 25682601
43. J. Domen, I. L. Weissman, Hematopoietic stem cells need two signals to prevent apoptosis; BCL-2 can provide one of these, Kitl/c-Kit signaling the other. *J. Exp. Med.* **192**, 1707–1718 (2000). doi: 10.1084/jem.192.12.1707; pmid: 11120768
44. H. Kantarjian *et al*., Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *N. Engl. J. Med.* **362**, 2260–2270 (2010). doi: 10.1056/NEJMoa1002315; pmid: 20525995
45. B. D. Smith *et al*., Ripretinib (DCC-2618) is a switch control kinase inhibitor of a broad spectrum of oncogenic and drug-resistant KIT and PDGFRA variants. *Cancer Cell* **35**, 738–751.e9 (2019). doi: 10.1016/j.ccell.2019.04.006; pmid: 31085175
46. M. C. Heinrich *et al*., Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor. *Blood* **96**, 925–932 (2000). doi: 10.1182/blood.V96.3.925; pmid: 10910906
47. S. T. Chou *et al*., Trisomy 21 enhances human fetal erythro-megakaryocytic development. *Blood* **112**, 4503–4506 (2008). doi: 10.1182/blood-2008-05-157859; pmid: 18812473
48. O. Tunstall-Pedoe *et al*., Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. *Blood* **112**, 4507–4511 (2008). doi: 10.1182/blood-2008-04-152967; pmid: 18689547
49. S. McLean, C. McHale, H. Enright, Hematological abnormalities in adult patients with Down's syndrome. *Ir. J. Med. Sci.* **178**, 35–38 (2009). doi: 10.1007/s11845-008-0223-2; pmid: 19020924
50. B. Liu, S. Filippi, A. Roy, I. Roberts, Stem and progenitor cell dysfunction in human trisomies. *EMBO Rep.* **16**, 44–62 (2015). doi: 10.15252/embr.201439583; pmid: 25520324
51. L. Gutiérrez *et al*., Ablation of Gata1 in adult mice results in aplastic crisis, revealing its essential role in steady-state and stress erythropoiesis. *Blood* **111**, 4375–4385 (2008). doi: 10.1182/blood-2007-09-115121; pmid: 18258797
52. L. M. Hollanda *et al*., An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat. Genet.* **38**, 807–812 (2006). doi: 10.1038/ng1825; pmid: 16783379
53. L. S. Ludwig *et al*., Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat. Med.* **20**, 748–753 (2014). doi: 10.1038/nm.3557; pmid: 24952648
54. V. G. Sankaran *et al*., Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J. Clin. Invest.* **122**, 2439–2443 (2012). doi: 10.1172/JCI63597; pmid: 22706301
55. Z. Li *et al*., Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat. Genet.* **37**, 613–619 (2005). doi: 10.1038/ng1566; pmid: 15895080
56. H. Bolouri *et al*., The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **24**, 103–112 (2018). doi: 10.1038/nm.4439; pmid: 29227476
57. I. Roberts, S. Izraeli, Haematopoietic development and leukaemia in Down syndrome. *Br. J. Haematol.* **167**, 587–599 (2014). doi: 10.1111/bjh.13096; pmid: 25155832
58. E. R. Lechman *et al*., miR-126 Regulates distinct self-renewal outcomes in normal and malignant hematopoietic stem cells. *Cancer Cell* **29**, 214–228 (2016). doi: 10.1016/j.ccell.2015.12.011; pmid: 26832662
59. J.-H. Klusmann *et al*., miR-125b-2 is a potential oncomiR on human chromosome 21 in megakaryoblastic leukemia. *Genes Dev.* **24**, 478–490 (2010). doi: 10.1101/gad.1856210; pmid: 20194440
60. T. D. Buitenkamp *et al*., Acute lymphoblastic leukemia in children with Down syndrome: A retrospective analysis from the Ponte di Legno study group. *Blood* **123**, 70–77 (2014). doi: 10.1182/blood-2013-06-509463; pmid: 24222333
61. S. W. K. Ng *et al*., A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016). doi: 10.1038/nature20598; pmid: 27926740
62. D. Cruz Hernandez *et al*., Sensitive, rapid diagnostic test for transient abnormal myelopoiesis and myeloid leukemia of Down syndrome. *Blood* **136**, 1460–1465 (2020). doi: 10.1182/blood.2020005610; pmid: 32556129
63. C. M. McHale *et al*., Prenatal origin of childhood acute myeloid leukemias harboring chromosomal rearrangements t(15;17) and inv(16). *Blood* **101**, 4640–4641 (2003). doi: 10.1182/blood-2003-01-0313; pmid: 12756163
64. C. M. McHale *et al*., Prenatal origin of TEL-AML1-positive acute lymphoblastic leukemia in children born in California. *Genes Chromosomes Cancer* **37**, 36–43 (2003). doi: 10.1002/gcc.10199; pmid: 12661004
65. J. L. Wiemels *et al*., In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* **99**, 3801–3805 (2002). doi: 10.1182/blood.V99.10.3801; pmid: 11986239
66. Y. Wang *et al*., Impact of age on the survival of pediatric leukemia: An analysis of 15083 children in the SEER database. *Oncotarget* **7**, 83767–83774 (2016). doi: 10.18632/oncotarget.11765; pmid: 27590519
67. P. H. G. Duijf, N. Schultz, R. Benezra, Cancer cells preferentially lose small chromosomes. *Int. J. Cancer* **132**, 2316–2326 (2013). doi: 10.1002/ijc.27924; pmid: 23124507
68. A. P. Laurent, R. S. Kotecha, S. Malinge, Gain of chromosome 21 in hematological malignancies: Lessons from studying leukemia in children with Down syndrome. *Leukemia* **34**, 1984–1999 (2020). doi: 10.1038/s41375-020-0854-5; pmid: 32433508

ACKNOWLEDGMENTS

We thank D. Curovic, R. Kelly, J. Law, and M. Niit at the Research Centre for Women's and Infants' Health Biobank (Mount Sinai Hospital) for sample coordination; B. Chow at the Pathology and Laboratory Medicine (Mount Sinai Hospital) for assistance with pathology; M. DSouza and R. Lopez at the Animal Resources Centre (UHN) for support with mouse work; M. Bergeret, S. Boddeda, S. Ng, A. Srinath, O. Subedar, A. AuYeung, and S. Zhao at the SickKids-UHN Flow and Mass Cytometry Facility for assistance with flow cytometry; B. Apresto at The Centre for Applied Genomics (SickKids) for sanger sequencing; K. Ho at the Centre for Applied Genomics (SickKids) for next-generation sequencing; A. Smith at the Cancer Cytogenetics Laboratory (UHN) for karyotyping; K. Asoyan, C. Cimafranca, J. Mouatt, M. Peralta, and Y. Yang at the Pathology Research Program (UHN) and N. Law at the Sttarr Innovation Centre (UHN) for assistance with histology; M. Bartolini at the Advanced Optical Microscopy Facility (UHN) for slide scanning; J. Moffat, K. Brown, and C. Ross at the Donnelly Centre for supplying the gRNA sequences for the in vivo screen; S. Henikoff at the Fred Hutchinson Cancer Research Center for supplying the Protein A-Micrococcal nuclease fusion protein for the Cut&Run assay; L. Shultz at the Jackson Laboratory for providing NSGW41 mice; the laboratories of S. Chan and F. Notta for sharing equipment; N. Mbong and A. Mitchell for technical assistance; and H. Hasle for sharing unpublished work. We thank C. Jones, A. Tikhonova, and members of the Dick laboratory for comments on the manuscript. **Funding:** This work was supported by funds from Human Frontier Science Program (LT-000601); Alex's Lemonade Stand Foundation (19-16679); The Leukemia & Lymphoma Society and The Leukemia & Lymphoma Society of Canada (3404-21); Portuguese Foundation for Science and Technology (SFRH/BD/136200/2018); Princess Margaret Cancer Centre Foundation; Ontario Institute for Cancer Research through funding provided by the Government of Ontario; Canadian Institutes for Health Research (Foundation: 154293, Operating Grant 130412, Operating Grant 89932, and Canada-Japan CEEHRC Teams in Epigenetics of Stem Cells 127882); International Development Research Centre, Canadian Cancer Society (703212); Terry Fox Research Institute Program Project Grant; University of Toronto's Medicine by Design initiative, which receives funding from the Canada First Research Excellence Fund; and a Canada Research Chair. **Author contributions:** E.W., J.E.D., and E.R.L. conceived the project, supervised research, and wrote the paper. J.A., O.I.G., G.K., and J.C.Y.W. edited the paper. E.W., J.A., O.I.G., and E.R.L. analyzed experiments. E.W., J.A., S.K.C., M.A., S.A.S., and B.A.G. performed in vitro and in vivo experiments. O.I.G and E.R.L. assisted with mouse work. J.A. performed morphological analysis. O.I.G. assisted with single-cell assays. A.M. analyzed ATAC-seq and RNA-seq data. G.K. performed Western blot assays and Cut&Run assays and prepared miRNA libraries. J.L.M. assisted with intrafemoral injections. S.A.M. and D.D.D.C. performed Cut&Run analysis. M.G. and L.S. analyzed miRNA sequencing data. J.J.F.M. generated smMIP libraries. S.M., J.C., and J.K.H. supplied primary TAM samples. M.C.-S.-Y. performed CRISPR/Cas9 off-target analysis. L.G.-P. assisted with ATAC-seq library preparations. S.A. performed smMIP analysis. M.A. performed histopathological analysis. K.C., M.R., P.S., and D.C. coordinated patient consent and sample collection. J.C.Y.W. and J.K.H. provided study consultation. J.E.D. secured funding for this study. **Competing interests:** D.D.D.C.: Pfizer and Nektar Therapeutics, research funding; DNAMx, cofounder and shareholder. J.E.D.: Celgene, research funding; Trillium Therapeutics, advisory board. All other authors declare no competing interests. **Data and materials availability:** Raw sequence data are available at European Genome-phenome Archive (EGAS00001004780) and processed data are available at Gene Expression Omnibus (GSE160096). All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/eabf6202/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S9
References (*69*–*87*)
MDAR Reproducibility Checklist

7 November 2020; resubmitted 9 March 2021
Accepted 21 May 2021
10.1126/science.abf6202

Science Webinars

AAAS

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

Get alerts about upcoming free webinars.

Sign up at: webinar.sciencemag.org/stayinformed

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Estimating infectiousness throughout SARS-CoV-2 infection course

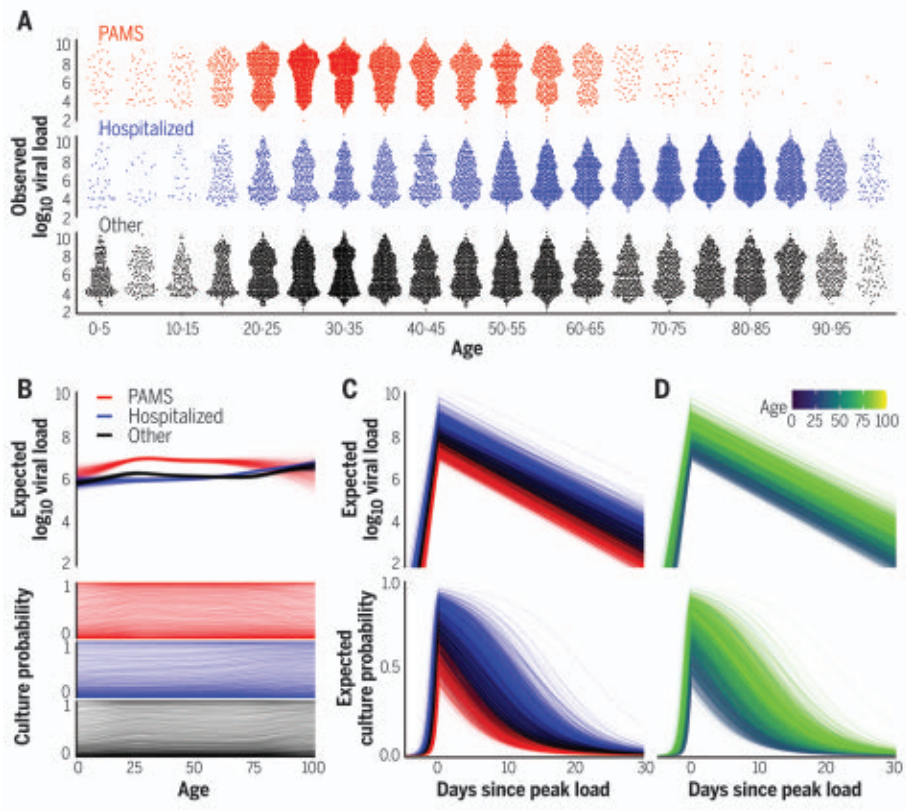
Terry C. Jones[1,2,3]†, Guido Biele[4,5]†, Barbara Mühlemann[1,2], Talitha Veith[1,2], Julia Schneider[1,2], Jörn Beheim-Schwarzbach[1], Tobias Bleicker[1], Julia Tesch,[1] Marie Luisa Schmidt[1], Leif Erik Sander[6], Florian Kurth[6,7], Peter Menzel[8], Rolf Schwarzer[8], Marta Zuchowski[8], Jörg Hofmann[8], Andi Krumbholz[9,10], Angela Stein[8], Anke Edelmann[8], Victor Max Corman[1,2], Christian Drosten[1,2]*

INTRODUCTION: Although post facto studies have revealed the importance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission from presymptomatic, asymptomatic, and mildly symptomatic (PAMS) cases, the virological basis of their infectiousness remains largely unquantified. The reasons for the rapid spread of variant lineages of concern, such as B.1.1.7, have yet to be fully determined.

RATIONALE: Viral load (viral RNA concentration) in patient samples and the rate of isolation success of virus from clinical specimens in cell culture are the clinical parameters most directly relevant to infectiousness and hence to transmission. To increase our understanding of the infectiousness of SARS-CoV-2, especially in PAMS cases and those infected with the B.1.1.7 variant, we analyzed viral load data from 25,381 German cases, including 9519 hospitalized patients, 6110 PAMS cases from walk-in test centers, 1533 B.1.1.7 variant infections, and the viral load time series of 4434 (mainly hospitalized) patients. Viral load results were then combined with estimated cell culture isolation probabilities, producing a clinical proxy estimate of infectiousness.

RESULTS: PAMS subjects had, at the first positive test, viral loads and estimated infectiousness only slightly less than hospitalized patients. Similarly, children were found to have mean viral loads only slightly lower (0.5 $\log_{10}$ units or less) than those of adults and ~78% of the adult peak cell culture isolation probability. Eight percent of first-positive viral loads were 10^9 copies per swab or higher, across a wide age range (mean 37.6 years, standard deviation 13.4 years), representing a likely highly infectious minority, one-third of whom were PAMS. Relative to non-B.1.1.7 cases, patients with the B.1.1.7 variant had viral loads that were higher by a factor of 10 and estimated cell culture infectivity that was higher by a factor of 2.6. Similar ranges of viral loads from B.1.1.7 and B.1.177 samples were shown to be capable of causing infection in Caco-2 cell culture. A time-course analysis estimates that a peak viral load of $10^{8.1}$ copies per swab is reached 4.3 days after onset of shedding and shows that, across the course of infection, hospitalized patients have slightly higher viral loads than nonhospitalized cases, who in turn have viral loads slightly higher than PAMS cases. Higher viral loads are observed in first-positive tests of PAMS subjects, likely as a result of systematic earlier testing. Mean culture isolation probability declines to 0.5 at 5 days after peak viral load and to 0.3 at 10 days after peak viral load. We estimate a rate of viral load decline of 0.17 $\log_{10}$ units per day, which, combined with reported estimates of incubation time and time to loss of successful cell culture isolation, suggests that viral load peaks 1 to 3 days before onset of symptoms (in symptomatic cases).

CONCLUSION: PAMS subjects who test positive at walk-in test centers can be expected to be approximately as infectious as hospitalized patients. The level of expected infectious viral shedding of PAMS people is of high importance because they are circulating in the community at the time of detection of infection. Although viral load and cell culture infectivity cannot be translated directly to transmission probability, it is likely that the rapid spread of the B.1.1.7 variant is partly attributable to higher viral load in these cases. Easily measured virological parameters can be used, for example, to estimate transmission risk from different groups (by age, gender, clinical status, etc.), to quantify variance, to show differences in virus variants, to highlight and quantify overdispersion, and to inform quarantine, containment, and elimination strategies. ■



Viral load and cell culture infectivity in 25,381 SARS-CoV-2 infections. (**A**) Viral loads in presymptomatic, asymptomatic, and mildly symptomatic cases (PAMS; red), hospitalized patients (blue), and other subjects (black). (**B**) Expected first-positive viral load and cell culture isolation probability, colored as in (A). (**C**) Temporal estimation with lines representing patients, colored as in (A). (**D**) As in (C), but colored by age.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: christian.drosten@charite.de
†These authors contributed equally to this work.
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as T. C. Jones *et al.*, *Science* **373**, eabi5273 (2021). DOI: 10.1126/science.abi5273

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi5273

RESEARCH ARTICLE

CORONAVIRUS

Estimating infectiousness throughout SARS-CoV-2 infection course

Terry C. Jones[1,2,3]†, Guido Biele[4,5]†, Barbara Mühlemann[1,2], Talitha Veith[1,2], Julia Schneider[1,2], Jörn Beheim-Schwarzbach[1], Tobias Bleicker[1], Julia Tesch[1], Marie Luisa Schmidt[1], Leif Erik Sander[6], Florian Kurth[6,7], Peter Menzel[8], Rolf Schwarzer[8], Marta Zuchowski[8], Jörg Hofmann[8], Andi Krumbholz[9,10], Angela Stein[8], Anke Edelmann[8], Victor Max Corman[1,2], Christian Drosten[1,2]*

Two elementary parameters for quantifying viral infection and shedding are viral load and whether samples yield a replicating virus isolate in cell culture. We examined 25,381 cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Germany, including 6110 from test centers attended by presymptomatic, asymptomatic, and mildly symptomatic (PAMS) subjects, 9519 who were hospitalized, and 1533 B.1.1.7 lineage infections. The viral load of the youngest subjects was lower than that of the older subjects by 0.5 (or fewer) $\log_{10}$ units, and they displayed an estimated ~78% of the peak cell culture replication probability; in part this was due to smaller swab sizes and unlikely to be clinically relevant. Viral loads above 10^9 copies per swab were found in 8% of subjects, one-third of whom were PAMS, with a mean age of 37.6 years. We estimate 4.3 days from onset of shedding to peak viral load ($10^{8.1}$ RNA copies per swab) and peak cell culture isolation probability (0.75). B.1.1.7 subjects had mean $\log_{10}$ viral load 1.05 higher than that of non-B.1.1.7 subjects, and the estimated cell culture replication probability of B.1.1.7 subjects was higher by a factor of 2.6.

Respiratory disease transmission is highly context-dependent and difficult to quantify or predict at the individual level. This is especially the case when transmission from presymptomatic, asymptomatic, and mildly symptomatic (PAMS) subjects is frequent, as with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*1*–*8*). Transmission is therefore typically inferred from population-level information and summarized as a single overall average, known as the basic reproductive number, R_0. Although R_0 is an essential and critical parameter for understanding and managing population-level disease dynamics, it is a resultant, downstream characterization of transmission. With regard to SARS-CoV-2, many finer-grained upstream questions regarding infectiousness remain unresolved or unaddressed. Three categories of uncertainty are (i) differences in infectiousness among individuals or groups such as PAMS subjects, according to age, gender, vaccination status, etc.; (ii) timing and degree of peak infectiousness, timing of loss of infectiousness, rates of infectiousness increase and decrease, and how these relate to onset of symptoms (when present); and (iii) differences in infectiousness due to inherent properties of virus variants.

These interrelated issues can all be addressed through the combined study of two clinical virological parameters: the viral load (viral RNA concentration) in patient samples, and virus isolation success in cell culture trials. Viral load and cell culture infectivity cannot be translated directly to in vivo infectiousness, and the impact of social context and behavior on transmission is very high; nonetheless, these quantifiable parameters can generally be expected to be those most closely associated with transmission likelihood. A strong relationship between SARS-CoV-2 viral load and transmission has been reported (*9*), comparing favorably with the situation with influenza virus, where the association is less clear (*10*, *11*).

The emergence of more transmissible SARS-CoV-2 variants, such as the B.1.1.7 lineage (UK Variant of Concern 202012/01), emphasizes the importance of correlates of shedding and transmission. The scarcity of viral load data in people with recent variants, and in PAMS subjects of all ages (*12*), is a blind spot of key importance because many outbreaks have clearly been triggered and fueled by these subjects (*2*, *13*–*17*). Viral load data from PAMS cases are rarely available, greatly reducing the number of studies with information from both symptomatic and PAMS subjects and that span the course of infections (*12*, *18*). Making matters worse, it is not possible to place positive reverse transcription polymerase chain reaction (RT-PCR) results from asymptomatic subjects in time relative to a nonexistent day of symptom onset, so these cases cannot be included in studies focused on incubation period. Additionally, viral load time courses relative to the day of symptom onset rely on patient recall, a suboptimal measure that is subject to human error and that overlooks infections from presymptomatic or asymptomatic contacts (*12*). An alternative and more fundamental parameter, the day of peak viral load, can be estimated from dated viral load time-series data, drawn from the entire period of viral load rise and fall and the full range of symptomatic statuses.

To better understand SARS-CoV-2 infectiousness, we analyzed viral load, cell culture isolation, and genome sequencing data from a diagnostic laboratory in Berlin (Charité–Universitätsmedizin Berlin Institute of Virology and Labor Berlin). We first address a set of questions regarding infectiousness at the moment of disease detection, especially in PAMS subjects whose infections were detected at walk-in community test centers. Because these people are circulating in the general community before their infections are detected, and are healthy enough to present themselves at such centers, their prevalence and shedding are of key importance to the understanding and prevention of transmission. In addition to PAMS subjects, we consider the infectiousness suggested by first-positive tests from hospitalized patients, including differences according to age, virus variant, and gender. A further set of temporal questions are then addressed by studying how infectiousness changes during the infection course. Using viral load measurements from patients with at least three RT-PCR tests, we estimate the onset of infectious viral shedding, peak viral load, and the rates of viral load increase and decline. Knowledge of these parameters enables fundamental comparisons between groups of subjects and between virus strains, and highlights the misleading impression created by viral loads from first-positive RT-PCR tests if the time of testing in the infection course is not considered.

Study composition

We examined 936,423 SARS-CoV-2 routine diagnostic RT-PCR results from 415,935 subjects aged 0 to 100 years from 24 February 2020 to 2 April 2021. Samples were collected at test centers and medical practices mostly in and around Berlin, Germany, and analyzed with LightCycler 480 and cobas 6800/8800 systems from Roche. Of all tested subjects, 25,381 (6.1%)

[1]Institute of Virology, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 10117 Berlin, Germany. [2]German Centre for Infection Research (DZIF), partner site Charité, 10117 Berlin, Germany. [3]Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, U.K. [4]Norwegian Institute of Public Health, 0473 Oslo, Norway. [5]University of Oslo, 0315 Oslo, Norway. [6]Department of Infectious Diseases and Respiratory Medicine, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 10117 Berlin, Germany. [7]Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, and Department of Medicine I, University Medical Centre Hamburg-Eppendorf, 20359 Hamburg, Germany. [8]Labor Berlin–Charité Vivantes GmbH, Sylter Straße 2, 13353 Berlin, Germany. [9]Institute for Infection Medicine, Christian-Albrechts-Universität zu Kiel and University Medical Center Schleswig-Holstein, Campus Kiel, 24105 Kiel, Germany. [10]Labor Dr. Krause und Kollegen MVZ GmbH, 24106 Kiel, Germany.
*Corresponding author. Email: christian.drosten@charite.de
†These authors contributed equally to this work.

Table 1. Age stratification of first-positive RT-PCR tests and viral load for 25,381 positive cases.
N, number of subjects with a positive test result; Pos. %, percentage of positive subjects; Load (SD), mean $\log_{10}$(viral load) and standard deviation; ≥3 tests, number of subjects with at least three RT-PCR test results, as used in the viral load time course analysis. Age ranges (in years) are open-closed intervals.

	All cases				PAMS cases			Hospitalized cases		
Age	*N*	Pos. %	Load (SD)	≥3 tests	*N*	Pos. %	Load (SD)	*N*	Pos. %	Load (SD)
0–5	330	1.8	5.9 (1.84)	16	36	5.1	6.6 (1.87)	32	0.9	5.6 (2.22)
5–10	185	1.8	6.0 (1.73)	12	39	6.2	6.1 (1.83)	18	1.4	5.8 (1.97)
10–15	227	2.2	6.0 (1.76)	8	51	6.9	6.4 (1.92)	22	1.4	6.0 (2.02)
15–20	643	3.0	6.3 (1.87)	39	192	5.1	6.7 (1.77)	121	2.5	6.1 (1.95)
20–25	1637	3.2	6.5 (1.89)	110	696	4.0	6.9 (1.86)	246	2.7	5.9 (1.92)
25–35	4452	3.0	6.6 (1.90)	320	1988	3.9	7.0 (1.83)	614	2.2	6.0 (1.88)
35–45	3393	2.7	6.4 (1.84)	323	1277	3.5	6.9 (1.79)	576	2.0	6.0 (1.90)
45–55	3341	3.1	6.4 (1.81)	401	1012	3.4	6.9 (1.83)	733	2.3	5.9 (1.77)
55–65	3322	2.7	6.3 (1.78)	623	674	3.0	6.8 (1.82)	1039	2.1	5.9 (1.80)
>65	7851	3.0	6.4 (1.79)	2492	145	5.8	6.8 (1.87)	3434	2.3	6.2 (1.86)

had at least one positive RT-PCR test (Table 1). Positive subjects had a mean age of 51.7 years with high standard deviation (SD) of 22.7 years, and a mean of 4.5 RT-PCR tests (SD 5.7), of which 1.7 (SD 1.4) were positive. Of the positive subjects, 4344 had tests on at least 3 days (with at least two tests positive) and were included in a time-series analysis.

We divided the 25,381 positive subjects into three groups (Fig. 1). The Hospitalized group (9519 subjects, 37.5%) included all those who tested positive in an in-patient hospitalized context at any point in their infection. The PAMS group (6110 subjects, 24.1%) included people whose first positive sample was obtained in any of 24 Berlin COVID-19 walk-in community test centers, provided they were not in the Hospitalized category. The Other group (9752 subjects, 38.4%) included everyone not in the first two categories (table S1). As Fig. 1 shows, there were relatively low numbers of young subjects in all three groups, and very few elderly PAMS subjects. The validity of the PAMS classification is supported by the fact that of the overall 6159 infections detected at walk-in test centers, only 49 subjects (0.8%) were later hospitalized. Subjects testing positive at these centers are almost certainly receiving their first positive test because they are instructed to immediately self-isolate, and our data confirm that such subjects are rarely retested: Only 4.6% of people with at least three test results had their first test at a walk-in test center. Of the 9519 subjects who were ever hospitalized, 6835 were already in hospital at the time of their first positive test. PAMS subjects had a mean age of 38.0 years (SD 13.7), typically younger than Other subjects (mean 49.1 years, SD 23.5), with Hospitalized the oldest group (mean 63.2 years, SD 20.7). Typing RT-PCR indicated that 1533 subjects were infected with a strain belonging to the B.1.1.7 lineage, as confirmed by full genomes from next-generation sequencing (see materials and methods).

First-positive viral load

Across all subjects, the mean viral load [given as $\log_{10}$(RNA copies per swab)] in the first positive-testing sample was 6.39 (SD 1.83). The PAMS subjects had viral loads higher than those of the Hospitalized subjects for ages up to 70 years, as exemplified by a 6.9 mean for PAMS compared to a 6.0 mean in Hospitalized adult subjects of 20 to 65 years. Crude comparisons of viral loads in age groups showed no substantial difference in first-positive viral load between groups of people older than 20 years (Table 1). Children and adolescents had mean first-positive viral load differences ranging between −0.49 (−0.69, −0.29) and −0.16 (−0.31, −0.01) relative to adults aged 20 to 65 (Table 2). Here and below, parameter differences between age groups show the younger value minus the older, so a negative difference indicates a lower value in the younger group. Ranges given in parentheses are 90% credible intervals.

We used a Bayesian thin-plate spline regression to estimate the relationship among age, clinical status, and viral load from the first positive RT-PCR of each subject, adjusting for gender, type of test center, and PCR system used. The Bayesian model well represents the observed data (Fig. 1B, Table 2, and fig. S1). The raw data and the Bayesian estimation (Fig. 2A) suggest consideration of subjects in three age categories: young (ages 0 to 20 years, grouped into 5-year brackets), adult (20 to 65 years), and elderly (over 65 years). We estimated an average first-positive viral load of 6.40 (6.37, 6.42) for adults and a similar mean of 6.35 (6.32, 6.39) for the elderly (Fig. 2A). Younger age groups had lower mean viral loads than adults, with the difference falling steadily from −0.50 (−0.62, −0.37) for the very youngest (0 to 5 years) to −0.18 (−0.23, −0.12) for older adolescents (15 to 20 years) (Table 2). Young age groups of PAMS subjects had lower estimated viral loads than older PAMS subjects, with differences ranging from −0.18 (−0.29, −0.07) to −0.63 (−0.96, −0.32). Among Hospitalized subjects these differences were smaller, ranging from −0.18 (−0.45, 0.07) to −0.11 (−0.22, 0.01) (Table 2 and Fig. 2B). Viral loads of subjects younger than 65 years were ~0.75 higher for PAMS subjects than for Hospitalized subjects (Fig. 2A), likely because of a systematic difference in RT-PCR test timing, discussed below.

Associating viral load with cell culture infectivity

We estimated the association between viral load and successful cell culture isolation probability (hereafter "culture probability") by combining the viral load estimated from the Bayesian regression with cell culture isolation data from our own laboratory (*19*) and from Perera *et al.* (*20*) (Fig. 2C). Across all ages, the average estimated culture probability at the time of first positive RT-PCR was 0.35 (0.01, 0.94). The mean culture probability for PAMS cases, 0.44 (0.01, 0.98), was higher than for Hospitalized cases, 0.32 (0.00, 0.92) (Fig. 2D). Comparing PAMS cases, we found differences, in particular for children aged 0 to 5 compared to adults aged 20 to 65, with average culture probabilities of 0.329 (0.003, 0.950) and 0.441 (0.008, 0.981) respectively, and a difference of −0.112 (−0.279, −0.003). Age group differences in Hospitalized cases ranged from −0.028 (−0.104, 0.009) to −0.018 (−0.055, 0) (Table 2).

First-positive viral loads are weakly bimodally distributed (Figs. 1A and 2A), which is not reflected in age-specific means. The resultant distribution includes a majority of subjects with relatively low culture probability and a minority with very high culture probability (Fig. 2E and fig. S2). The highly infectious subset includes 2228 of 25,381 positive subjects (8.78%) with a first-positive viral load of at least 9.0, corresponding to an estimated culture probability of ~0.92 to 1.0. Of these 2228 subjects, 804 (36.09%) were PAMS at the time of testing, with a mean (median) age of 37.6 (34.0) and SD of 13.4 years. PAMS subjects are overrepresented in this highly infectious group among people aged 20 to 80 years, and Hospitalized subjects are overrepresented in people aged 80 to 100 years (fig. S3).

Estimating B.1.1.7 infectiousness at first-positive test

The 1533 subjects infected with a B.1.1.7 virus in our dataset had an observed mean first-positive viral load of 7.38 (SD 1.54), which is 1.05 higher (0.97, 1.13) than non-B.1.1.7 subjects in the full dataset. To increase specificity, we compared 1453 B.1.1.7 cases with

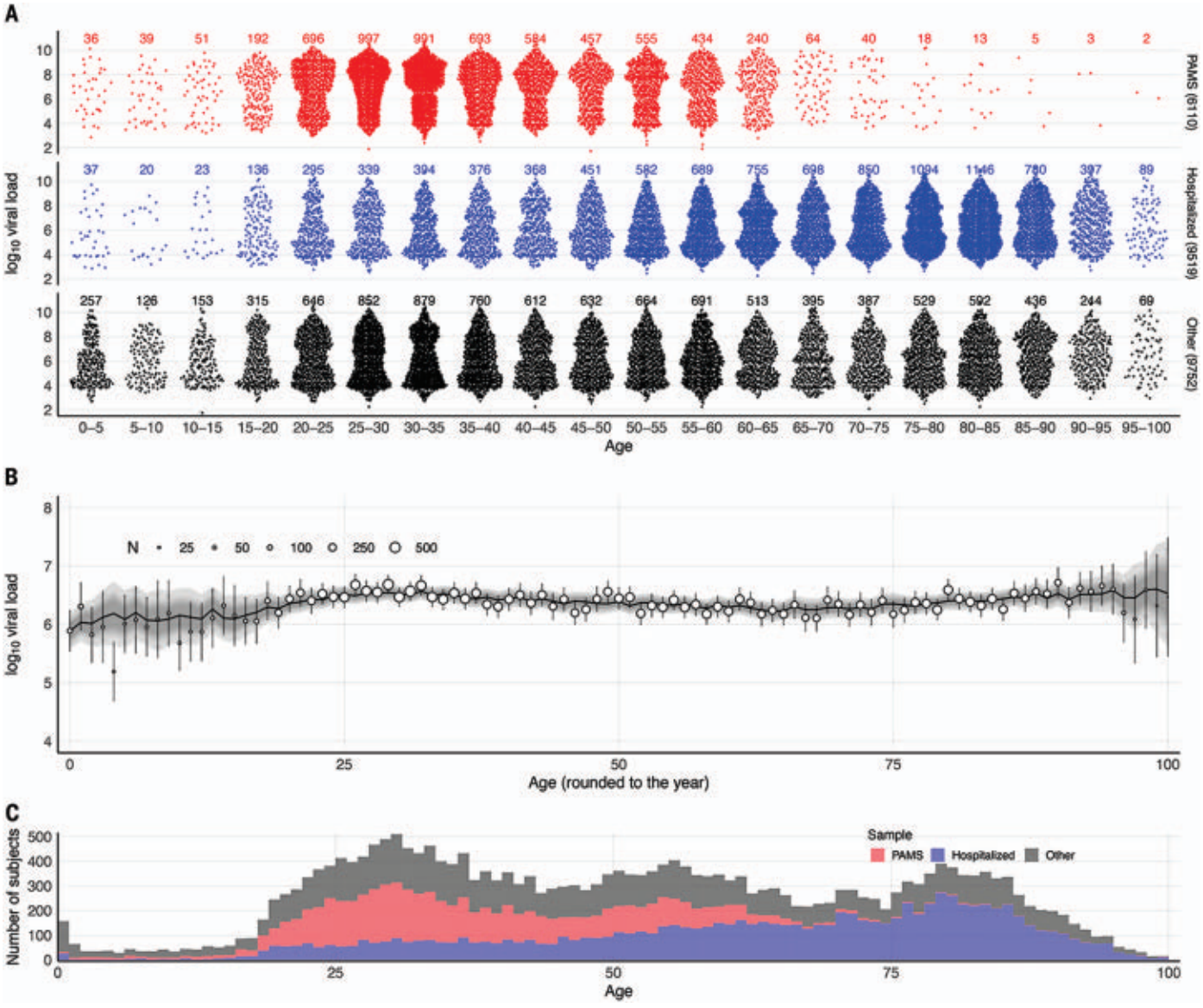


Fig. 1. Distribution of age and first-positive viral load in PAMS, Hospitalized, and Other subjects. (**A**) Distribution of observed first-positive viral loads for 25,381 subjects according to clinical status (6110 PAMS, 9519 Hospitalized, 9752 Other) and age group. (**B**) Age–viral load association. Observed viral loads are shown as circles (circle size indicates subject count) with vertical lines denoting confidence intervals; model-predicted viral loads are shown as a black, roughly horizontal line, with gray shading denoting credible intervals. (**C**) Stacked age histograms according to subject clinical status. Because inclusion in the study required a positive RT-PCR test result, and because testing is in many cases symptom-dependent, the study may have a proportion of PAMS cases that differs from the proportion in the general population.

977 non-B.1.1.7 cases using viral loads only from centers with B.1.1.7 and non-B.1.1.7 cases, and only from the same day or 1 day before or after the B.1.1.7 sample was taken. This analysis adjusted for clinical status, gender, RT-PCR system, and subject age, and also modeled random test center effects. The results show that B.1.1.7 cases are associated with a 1.0 (0.9, 1.1) higher viral load (Fig. 3 and table S2). This results in a mean estimated B.1.1.7 subject culture probability of 0.50 (0.03, 0.97), considerably higher than the overall figure of 0.31 (0.00, 0.94) for the non-B.1.1.7 subjects in the comparison, corresponding to a median factor of 2.6 (50% credible interval: 1.4, 5.1) higher culture probability for samples from B.1.1.7 cases. To investigate whether there might be a difference in cell culture infectivity due to a factor other than viral load, we isolated virus from 105 samples (22 B.1.1.7, 83 B.1.177) in Caco-2 cells from a collection of 223 samples with matched viral loads. Although no statistical difference was seen in the distribution of viral loads that resulted in successful isolation (fig. S4), uncertainty attributable to the routine diagnostic laboratory context—including uncontrolled preanalytical parameters such as transportation time and temperature, together with the small isolation-positive sample sizes—are insufficient to support a conclusion that the distributions do not differ (see materials and methods).

Estimating infectiousness over time

To investigate viral load over the course of the infection, we estimated the slopes of a model of linear increase and then decline of $\log_{10}$ viral load using a Bayesian hierarchical model. The analysis used the time series of the 4344 subjects who had RT-PCR results on at least 3 days (with at least two tests being positive). The number of subjects with multiple test

Table 2. Pairwise age comparisons of first-positive RT-PCR viral load and estimated culture probability calculated from spline regression or raw data. Only the spline-based regression adjusts for effects of the test center and RT-PCR system. Differences are mean differences, with 90% credible intervals or confidence intervals from null-hypothesis significance testing given in parentheses. *P* values are from Mann-Whitney *U* tests (*96*).

		Spline-based regression (adjusted)		Raw data (unadjusted)	
Sample	**Comparison**	**Culture probability difference**	**log_{10}(load difference)**	**log_{10}(load difference)**	***P***
All	0–5 vs. 20–65	−0.067 (−0.167, −0.002)	−0.50 (−0.62, −0.37)	−0.49 (−0.69, −0.29)	<0.001
All	5–10 vs. 20–65	−0.054 (−0.132, −0.002)	−0.40 (−0.50, −0.30)	−0.38 (−0.64, −0.13)	0.004
All	10–15 vs. 20–65	−0.045 (−0.111, −0.002)	−0.30 (−0.39, −0.22)	−0.42 (−0.65, −0.18)	<0.001
All	15–20 vs. 20–65	−0.033 (−0.076, −0.001)	−0.18 (−0.23, −0.12)	−0.16 (−0.31, −0.01)	0.033
PAMS	0–5 vs. 20–65	−0.067 (−0.167, −0.002)	−0.50 (−0.62, −0.37)	−0.49 (−0.69, −0.29)	<0.001
PAMS	5–10 vs. 20–65	−0.112 (−0.279, −0.003)	−0.63 (−0.96, −0.32)	−0.37 (−1.00, 0.26)	0.213
PAMS	10–15 vs. 20–65	−0.092 (−0.228, −0.003)	−0.51 (−0.77, −0.26)	−0.86 (−1.46, −0.26)	0.004
PAMS	15–20 vs. 20–65	−0.064 (−0.162, −0.002)	−0.35 (−0.54, −0.17)	−0.56 (−1.10, −0.02)	0.034
Hospitalized	0–5 vs. 20–65	−0.033 (−0.087, −0.001)	−0.18 (−0.29, −0.07)	−0.26 (−0.52, −0.01)	0.046
Hospitalized	5–10 vs. 20–65	−0.028 (−0.104, 0.009)	−0.18 (−0.45, 0.07)	−0.36 (−1.10, 0.37)	0.115
Hospitalized	10–15 vs. 20–65	−0.025 (−0.084, 0.003)	−0.16 (−0.36, 0.03)	−0.48 (−1.38, 0.43)	0.172
Hospitalized	15–20 vs. 20–65	−0.022 (−0.071, 0.001)	−0.14 (−0.29, 0.02)	−0.11 (−0.97, 0.74)	0.625
Other	0–5 vs. 20–65	−0.018 (−0.055, 0.000)	−0.11 (−0.22, 0.01)	0.00 (−0.33, 0.33)	0.845
Other	5–10 vs. 20–65	−0.058 (−0.148, −0.001)	−0.36 (−0.51, −0.20)	−0.33 (−0.55, −0.10)	0.004
Other	10–15 vs. 20–65	−0.044 (−0.110, −0.001)	−0.27 (−0.39, −0.15)	−0.10 (−0.40, 0.20)	0.586
Other	15–20 vs. 20–65	−0.026 (−0.072, −0.001)	−0.16 (−0.27, −0.06)	−0.31 (−0.58, −0.04)	0.045

results skews heavily toward older subjects, with very few below the age of 20 meeting the criterion (Fig. 4A). We estimated time from onset of shedding to peak viral load of 4.31 (4.04, 4.60) days, mean peak viral load of 8.1 (8.0, 8.3), and mean decreasing viral load slope of −0.168 (−0.171, −0.165) per day (fig. S5). Figure S6 shows that while Hospitalized patients are estimated to be uniformly highly infectious at peak viral load, the infectiousness of PAMS subjects at peak load is more variable.

The temporal placement of the full 18,136 RT-PCR results from these 4344 subjects (80% of whom were hospitalized with COVID-19 at some point in their infections) is shown in fig. S7. Per-subject trajectories can differ considerably from that described by the mean parameters (Fig. 4B and fig. S8). Across all subjects, PAMS cases were on average detected 5.1 (4.5, 5.7) days after peak load, 2.4 (1.7, 3.0) days before non-PAMS cases, which were on average detected 7.4 (7.2, 7.6) days after peak load. We estimate that 962 (914, 1010) of the 4344 subjects [22.14% (21.04, 23.25)] had a first positive test before the time of their peak viral load, with a mean of 1.4 (1.3, 1.5) days before reaching peak viral load. Among the infections detected after peak viral load, the timing of the first positive RT-PCR test is estimated at 9.8 (9.6, 10.0) days after peak viral load, with SD of 6.9 (6.8, 7.0) days, reflecting a broad time range of infection detection. Estimated peak viral loads were higher in Hospitalized subjects than in Other subjects, and higher in Other subjects than in PAMS subjects, with differences of 0.68 (0.83, 0.52) and 0.96 (0.33, 1.53) respectively (fig. S9 and table S3). No differences according to gender were seen. Viral load time courses were similar across age groups, although younger subjects had lower peak viral load than adults aged 45 to 55 (Fig. 5, A and C, fig. S10, and table S4). Model parameters suggest a slightly longer time to peak, a higher peak, and a more rapid decline in viral load when the analysis is restricted to subjects with successively higher numbers of RT-PCR results (fig. S11 and table S5), with an increasing percentage of hospitalized subjects. Differences in model parameters according to the number of tests in subjects may reflect increased parameter accuracy due to additional data, although other factors associated with being tested more frequently may be responsible. The Bayesian estimation of the model agrees well with a separate second implementation based on simulated annealing (fig. S12, table S5, and supplementary text).

We estimate that the rise from near-zero to peak culture probability takes 1.8 (1.3, 2.6) days, with a mean peak culture probability of 0.74 (0.61, 0.85). Mean culture probability then declines to 0.52 (0.40, 0.64) at 5 days and to 0.29 (0.19, 0.40) at 10 days after peak viral load. Subject-level time courses can deviate substantially from these mean estimates (Fig. 4C). Peak culture probabilities for age groups range from a low of 0.54 (0.39, 0.71) for 0- to 5-year-olds to 0.80 (0.67, 0.90) for subjects more than 65 years old. The least infectious youngest children have 78% (61, 94) of the peak culture probability of adults aged 45 to 55 (Fig. 5, B and D, and table S4). An insufficient amount of data precludes a reliable B.1.1.7 viral load time-series analysis at this point.

Discussion

Limitations

Our analysis attempted to account for the effects of gender, PCR system, and test center type. Although we could not incorporate inter-run variability or the variability in the sample preanalytic (such as type of swab or initial sample volume) in our conversion of RT-PCR cycle threshold values to log_{10}(viral load) values, these variabilities apply to all age groups and do not affect the interpretation of data for the purpose of our study. If the proportion of subjects with a certain clinical status differs between age groups in the study sample, this could lead to over- or underestimation of differences in viral load between age groups. However, as our study compares viral load between age groups stratified by clinical status, it appears unlikely that differential testing biases our results.

Interpreting first-positive viral loads

Viral loads and their differences are not easy to interpret without knowledge of when in the disease course the samples were taken, and of the correspondence between viral load and shedding. The higher first-positive viral loads in PAMS subjects than in Hospitalized subjects are likely due to time of detection. This is suggested in the first place by the estimated difference of 2.4 (1.7, 3.0) days in test timing, which would produce a viral load difference of ~0.4 using the −0.168 daily viral load decline gradient from the (mainly hospitalized) time-series subjects. Additionally, from the time series of PAMS, Other, and Hospitalized subjects, we can estimate that throughout the infection

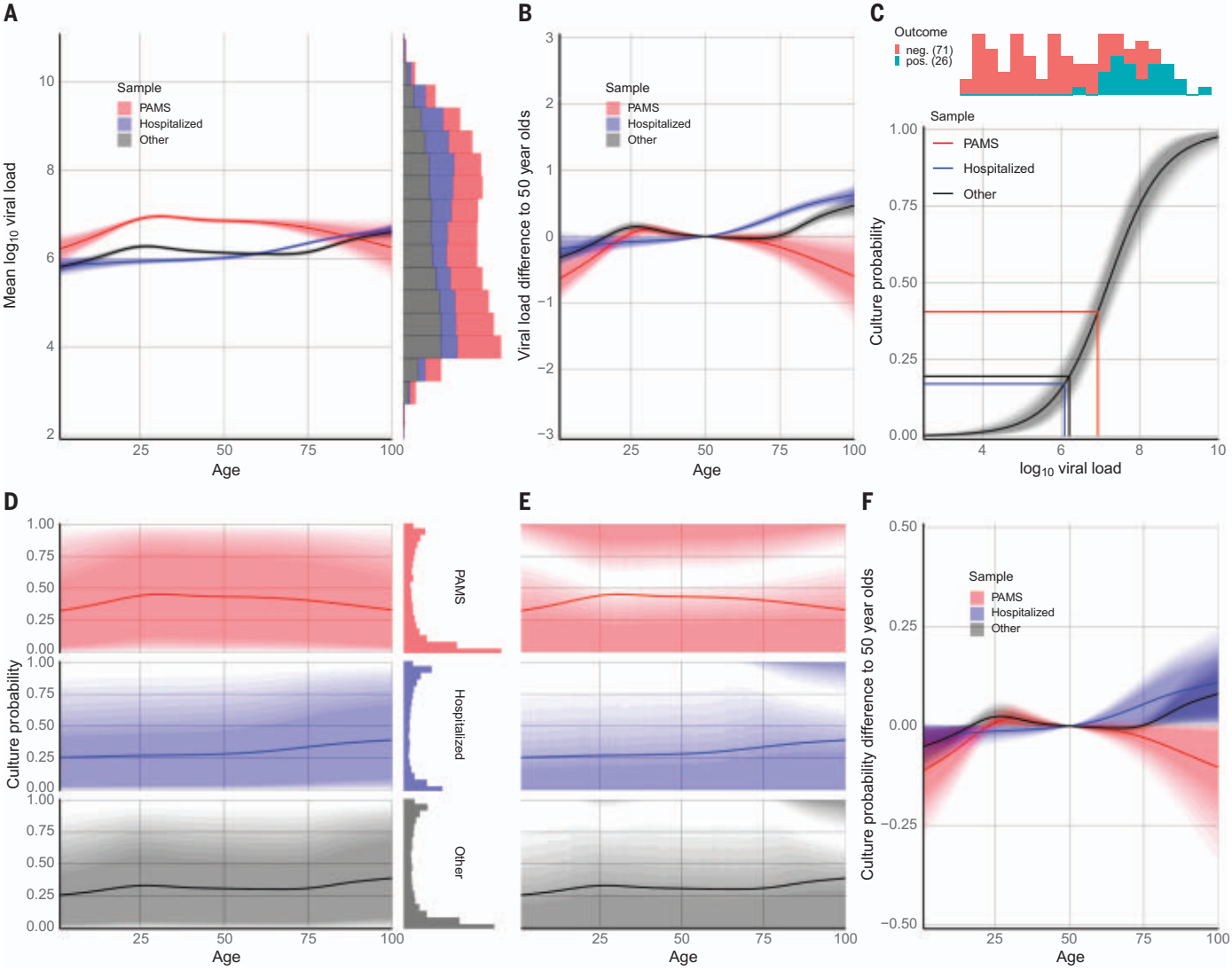


Fig. 2. Estimated viral load and culture probability at time of first positive RT-PCR test. Shaded regions denote 90% credible intervals in all panels. To indicate change within each 90% region, shading decreases in intensity from a narrow 50% credibility interval level to the full 90%. (**A**) Estimated mean viral load in first-positive RT-PCR tests according to age and status. The stacked histogram (right) shows the observed viral load distribution. Because the shaded region shows the 90% credible interval for the mean, it does not include the higher values shown in the histogram on the right. (**B**) Differences in estimated first-positive viral load according to age and status. Each colored line is specific to a particular subset of subjects (PAMS, Hospitalized, Other). Each line shows how viral load differs by age for subjects of the corresponding status from that of 50-year-old (rounded age) subjects of the same status. The comparison against 50-year-olds avoids comparing any subset of the subjects against a value (such as the overall mean) that is computed in part on the basis of that subset, thereby partially comparing data to the same data. The mean first-positive viral loads for 50-year-old PAMS and Hospitalized subjects are 7.2 and 6.2, respectively, allowing relative *y*-axis differences to be translated to approximate viral loads. (**C**) Estimation of the association between viral load and cell culture isolation success rate based on data from our own laboratory (*19*) and Perera *et al*. (*20*). Viral load differences in the $\log_{10}$ range ~6 to ~9 have a large impact on culture probability, whereas the impact is negligible for differences outside that range. The vertical lines indicate the observed mean first-positive viral loads for different subject groups; the horizontal lines show the corresponding expected probabilities of a positive culture. (**D**) Estimated culture probability at time of first-positive RT-PCR according to age and status, obtained by combining the results in (A) and (C). Culture probability is calculated from posterior predictions [i.e., the posterior means shown in (A) plus error variance]. The histogram at right shows that mean culture probabilities calculated from observed viral loads are not well matched by credible intervals, which do not include the most probable estimated culture probabilities. (**E**) Culture probability with highest–posterior density regions, which do include the most probable estimated culture probabilities and match the histograms in (D) well. The *y* axis is the same as in (D). (**F**) Differences of estimated expected culture probability at time of first-positive RT-PCR for age groups, with plot elements as described for (B).

course, the Hospitalized group has higher viral loads than the Other group, whose viral loads are in turn higher than those of the PAMS group (fig. S9 and table S3). This relationship holds across age groups (fig. S13) and also in a fine-grained split of test centers by clinical severity (fig. S14). Similarly, the lower first-positive viral loads in elderly PAMS subjects may be due to these subjects being less likely to be tested as early because they are more likely to be house-bound, less likely to be employed, less mobile, more cautious (therefore disinclined to get tested with only mild symptoms), etc. The impact on infectiousness of differences in viral load must be informed by

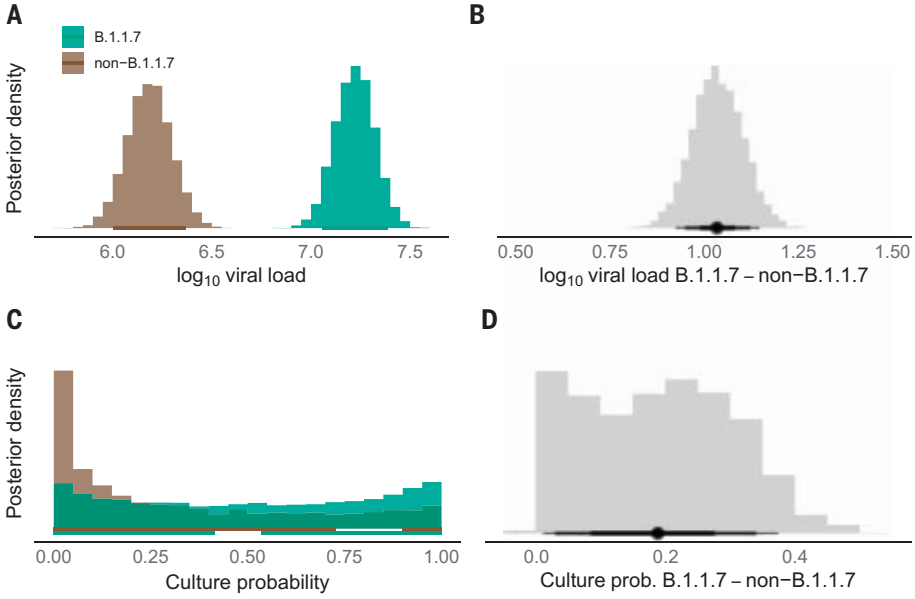


Fig. 3. Posterior distributions of estimated viral loads and culture probabilities for B.1.1.7 and non-B.1.1.7 subjects, and their differences. Viral loads and estimated culture probabilities of 1387 B.1.1.7 subjects and 977 non-B.1.1.7 subjects are represented. To select a comparable subset of non-B.1.1.7 viral loads for the comparison, we included only non-B.1.1.7 subjects from test centers that had detected a B.1.1.7 variant as well as at least one non-B.1.1.7 subject, and only if the non-B.1.1.7 infection was detected on the same day as a B.1.1.7 infection was detected, plus or minus 1 day. Similar differences exist when viral loads from larger, less restrictive, subsets of non-B.1.1.7 subjects are used in the comparison (table S2; see materials and methods). (**A**) Posterior distribution of viral load. (**B**) Posterior distribution of difference of average viral load between B.1.1.7 and non-B.1.1.7 cases. (**C**) Posterior distribution of the estimated culture probability. See also fig. S2. (**D**) Difference of mean culture probability between B.1.1.7 and non-B.1.1.7 cases. Horizontal lines indicate 90% credible intervals in (A), (B), and (D) and the highest posterior density intervals in (C).

where the viral loads fall on the viral load–culture probability curve. In our data, the viral loads involved in the difference between means in children and adults and the difference between means in B.1.1.7 and non-B.1.1.7 subjects result in quite different corresponding culture probabilities (see below).

A highly infectious minority and overdispersion

The bimodal distribution of culture probabilities (Fig. 2, D and E) shows a small group of 8.78% of highly infectious subjects. This qualitatively agrees with a model (*21*) and a study (*22*) concluding that 10% and 15% of index cases, respectively, may be responsible for 80% of transmission. Other studies reported that 8 to 9% of individuals harbored 90% of total viral load (*23*), and that in cases from India (*24*) and Hong Kong (*6*) ~70% of index cases had no secondary cases. PAMS subjects can be construed to pose a risk for several reasons: 36.1% of the highly infectious subjects in our study were PAMS at the time of the detection of their infection, their mean age was 37.6 years with a high standard deviation of 13.4 years (figs. S2 and S3), and we estimate that infectiousness peaks 1 to 3 days before onset of symptoms (if any).

Comparison with influenza virus

Without direct knowledge from a large number of SARS-CoV-2 transmission events, we could try to draw conclusions regarding infectiousness from studies of other respiratory viruses, such

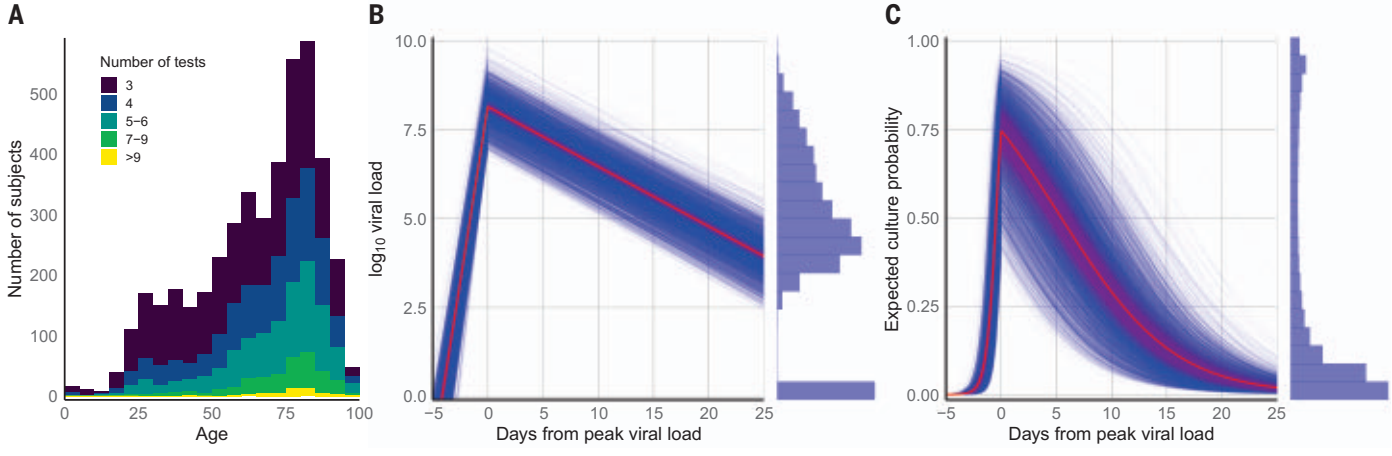


Fig. 4. Viral load and estimated infectious virus shedding time series. Of 25,381 positive subjects, 4344 had three or more RT-PCR test results available, and these were used in a viral load time-series analysis. Subjects with only one result cannot be placed in time because of inherent ambiguity (given that the model has both an increasing and a decreasing phase), and those with only two test results are excluded from the time-series analysis because of insufficient data for temporal placement (their number of data points is less than the number of model parameters being estimated). (**A**) Number of subjects with three or more RT-PCR test results available, at least two of which were positive, according to age. (**B**) Estimated time course of viral load for 18,136 RT-PCR results from the 4344 subjects with at least three RT-PCR results. Blue lines are expected complete time courses for individual cases. The sample mean is shown in red, with its 90% credible interval as a shaded area. The histogram at right shows the distribution of all observed viral loads. The histogram values at zero correspond to the initial and trailing negative tests in subject timelines. Figure S8 shows raw viral load time series, per subject and split by number of RT-PCR tests. (**C**) Estimated time course of positive cell culture probability, calculated by applying the results shown in Fig. 2C to the estimated viral load time courses in (B). Blue lines are expected time courses for individual subjects. The sample average is shown in red, with its 90% credible interval as a shaded area. The histogram at right shows the distribution of culture probabilities in the sample and was obtained by applying the curve in Fig. 2C to the data in the histogram in (B).

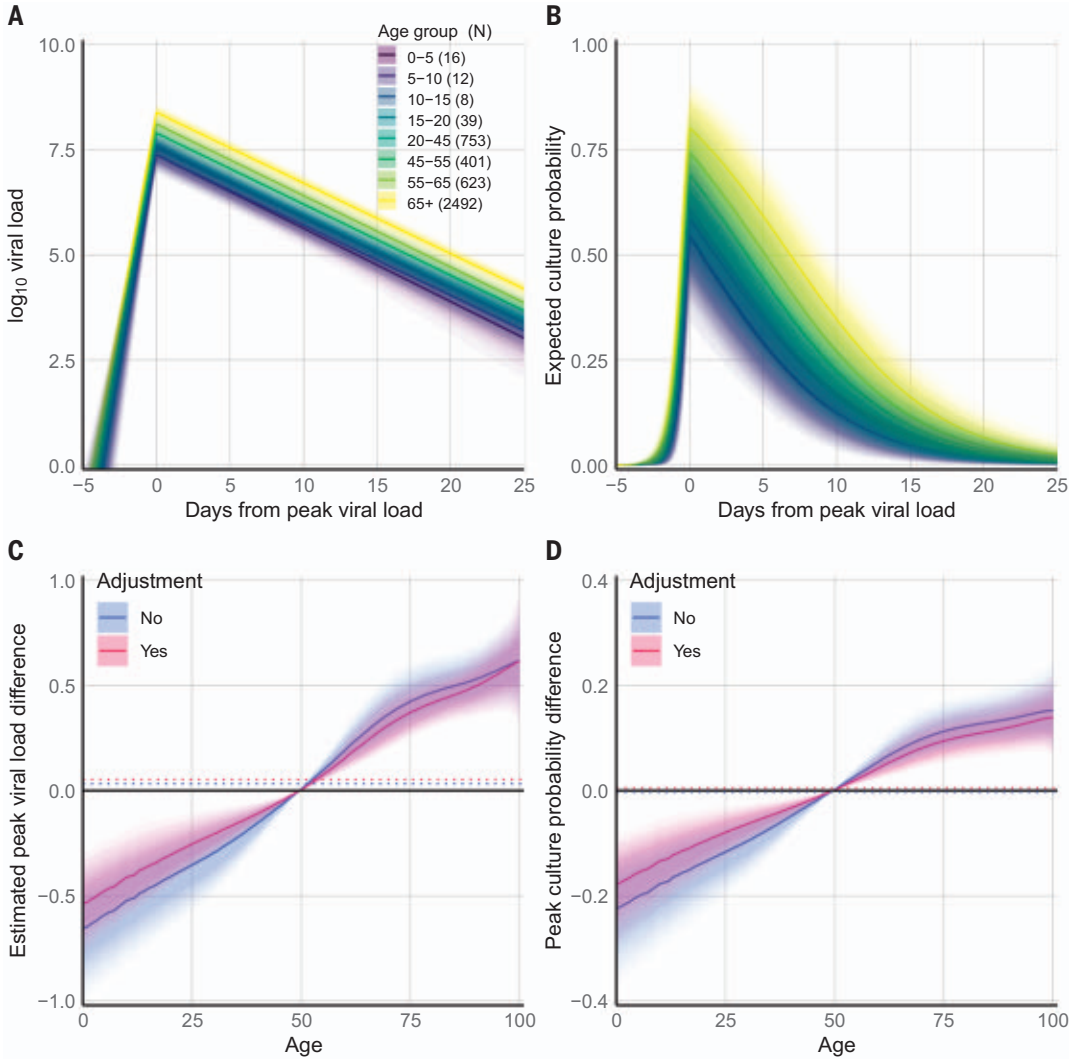


Fig. 5. Estimated expected viral load and culture probability for age groups by time. (**A**) Change in estimated viral load over time according to age group for 4344 subjects with at least three RT-PCR tests, at least two of which were positive. Shading indicates the 90% credible interval of the mean. (**B**) Change in estimated culture probability over time according to age. Age groups, coloring, and shading are as in (A). (**C**) Estimated age group differences in mean peak viral load, corresponding to the values at day zero in (A). (**D**) Estimated age group differences in mean peak culture probability, corresponding to the values at day zero in (B). In (C) and (D), adjusted differences account for variations by age in clinical status and gender. Dotted lines indicate grand means for the 4344 subjects.

as influenza. However, it has become clear that there are important differences and uncertainties that would cast doubt on such a comparison. Influenza may have later onset of viral shedding; shedding finishes earlier; there may be a lower secondary attack rate; viral loads are much lower; there is variation between virus subtypes; the role of asymptomatic subjects in transmission is uncertain or thought to be reduced; and the frequency of asymptomatic infections is uncertain, especially in children (*10*, *11*, *25*–*29*). Age-specific behavioral differences do, however, make a large contribution to the established higher shedding of children relative to adults in influenza. This should be an important consideration for SARS-CoV-2, as shown by studies indicating higher transmission between children of similar ages (*6*, *24*) and high transmission heterogeneity (*22*). Despite many decades of close study of influenza virus, the relationship between viral load and transmission is unclear (*10*, *11*). The situation with respiratory syncytial virus is even less clear (*30*). Understanding SARS-CoV-2 transmission will likely be at least as challenging, given the high frequency of transmission from PAMS subjects (*1*–*8*). This suggests an important role for clinical parameters, given the apparently strong association between viral load and transmission, independent of symptoms (*9*).

Estimated infectiousness in the young

The differences we observe in first-positive RT-PCR viral load between groups based on age are minor, as in other studies (*31*–*35*), and the viral loads in question—in the range of 5.9 to 6.6 (Table 1)—are in a region of the viral load–culture probability association where changes in viral load have relatively little impact on estimated culture probability (Fig. 2C). Comparisons between adult viral loads and those of children, and the relative infectious risks they pose, are impeded by the likely influence of nonviral factors. Nasopharyngeal swab samples, which often carry higher viral loads, are rarely taken from young children because they can be painful, and the sample volume carried by smaller pediatric swab devices is lower than in larger swabs used for adults (*36*). Infections in mildly symptomatic children may be initially missed and only detected later (*37*), resulting in lower first-positive viral loads. Our results of similar viral load trajectories for children and adults (Fig. 5), and the numeric range of the viral load values in question (Fig. 2C), suggest that viral load differences between children and adults are too small to be solely responsible for large differences in infectiousness. The impact on transmission of general age-related physiological differences, such as different innate immune responses (*38*), may be small relative to the impact of large differences in frequency of close contacts and transmission opportunities.

Timing of estimated peak infectiousness relative to onset of symptoms

We estimated the time from onset of shedding to peak viral load at 4.3 days. Previous studies

and reviews of COVID-19 report mean incubation times of 4.8 to 6.7 days (*4*, *39*–*44*), which suggests that, on average, a period of high infectivity can start several days before the onset of symptoms. Viral load rise may vary between individuals, and limitations of the available data suggest that our analysis may underestimate interindividual variation in viral load increase. The failure to isolate virus in cell culture beyond 10 days from symptom onset (*19*, *20*, *35*, *45*, *46*), together with our estimated slope of viral load decline, also suggest that peak viral load occurs 1 to 3 days before symptom onset (supplementary text). Data from 171 hospitalized patients from a Charité-Universitätsmedizin cohort suggest a figure of 4.3 days (fig. S15 and supplementary text).

Estimated infectiousness of the B.1.1.7 variant

We found that people infected with a B.1.1.7 virus had a first-positive viral load that was ~1 higher than in people infected with a wild-type virus. The scale of the viral load difference, and its presence in the comparison between B.1.1.7-infected and non–B.1.1.7-infected subjects drawn from the same test centers at the same times, argue that the difference is not due to a systematic difference in time of sampling. The higher B.1.1.7 viral load can be compared to the findings of two large and closely controlled UK studies, a mortality study (*47*) and a vaccine trial (*48*), which imply higher B.1.1.7 viral loads by a factor of 5 to 10 (based on RT-PCR cycle threshold differences of 2.3 and ~3, respectively). Several other studies also appear to point to a higher B.1.1.7 viral load (*49*–*52*) (supplementary text).

The mean B.1.1.7 viral load value in our study falls in a region of the viral load–culture probability curve with a steep gradient (Fig. 2C), resulting in an estimated culture probability considerably higher than for non-B.1.1.7 subjects. Although a strong correlation has been observed between SARS-CoV-2 viral load and transmission (*9*), here we are estimating infectivity probability from cell culture trials. Any impact of a change in viral load on transmission will be highly dependent on context, so the large difference in estimated culture probability in our data is only a proxy indication of potentially higher transmissibility of the B.1.1.7 strain. We estimate that B.1.1.7-infected subjects' mean culture probability is higher than that of non–B.1.1.7-infected subjects by a factor of 2.6. This can be compared to a UK study that found a factor of 1.3 relative increase in secondary attack rates for B.1.1.7 index cases in ~60,000 household contacts (*53*), a UK study estimating a factor of 1.7 to 1.8 increase in transmission (*54*), and an estimate of a 43% to 90% higher reproductive number (*55*).

Summary

Our results indicate that PAMS subjects in apparently healthy groups can be expected to be as infectious as hospitalized patients at the time of detection. The relative levels of expected infectious virus shedding of PAMS subjects (including children) is of high importance because these people are circulating in the community and it is clear that they can trigger and fuel outbreaks (*56*). The results from our time-series analysis, and their generally good agreement with results from studies based on other metrics (often epidemiological), show that accurate estimations can be directly obtained from two easily measured virological parameters, viral load and sample cell culture infectivity. Such results can be put to many uses: to estimate transmission risk from different groups (by age, gender, clinical status, etc.), to quantify variance, to show differences in virus variants, to highlight and quantify overdispersion, and to inform quarantine, containment, and elimination strategies. Our understanding of the timing and magnitude of change in viral load and infectiousness, including the impact of influencing factors, will continue to improve as data from large studies accumulate and are analyzed. A major ongoing challenge is to connect what we learn about estimated infectiousness from these clinical parameters to highly context-dependent in vivo transmission. On the basis of our estimates of infectiousness of PAMS subjects and the higher viral load found in subjects infected with the B.1.1.7 variant, we can safely assume that nonpharmaceutical interventions such as social distancing and mask wearing have been key in preventing many additional outbreaks. Such measures should be used in all social settings and across all age groups wherever the virus is present.

Materials and methods

Age ranges

Age categories for the analysis of the first-positive test results mentioned in the text indicate mathematically open-closed ranges of years (e.g., 0-5 signifies (0-5] years). We group subjects up to 20 years old into age categories spanning 5 years, subjects from 20 to 65 years into an adult group, and elderly subjects into a 65+ category. This categorization is motivated by the observed data and the Bayesian estimation of viral load differences between children of different ages and adults. The age groupings used in the viral load time-series analysis are broader in the younger categories to increase the cardinality of those groups, because few young people have at least three RT-PCR tests (Fig. 4A).

Viral loads

Viral load is semiquantitative, estimating RNA copies per entire swab sample, whereas only a fraction of the volume can reach the test tube. The quantification is based on a standard preparation tested in multiple diluted replicates to generate a standard curve and derive a formula in which RT-PCR cycle threshold values are converted to viral loads. This approach does not reflect inter-run variability or the variability in the sample preanalytic, such as type of swab or initial sample volume (varying between 2.0 and 4.3 ml). However, these variabilities apply to all age groups and do not affect the interpretation of data for the purpose of the present study.

Viral load figures are given as the logarithm base 10. Viral load is estimated from the cycle threshold (Ct) value using the empirical formulae $14.159 - (Ct \times 0.297)$ for the Roche Light Cycler 480 system and $15.043 - (Ct \times 0.296)$ for the Roche cobas 6800/8800 systems. The formulae are derived from testing standard curves and cannot be transferred to calculate viral load in other laboratory settings. Calibration of the systems and chemistries in actual use is required.

B.1.1.7 viral load analysis

No analysis regarding symptomatic status was made for B.1.1.7 subjects because of uncertainties regarding exact operational protocols at outbreak hospitals. B.1.1.7 assignment to samples was initially made according to typing RT-PCR tests that detect the N501Y and 69/70 deletion in the amino acid sequence of the virus spike protein. Examination of the complete viral genome of 49 samples confirmed that the subjects were in fact infected with the B.1.1.7 variant, with all variant-defining substitutions and deletions (*57*) found in all cases. No consistent additional mutations or deletions/insertions were found in the sequences.

Sequencing read mapping was performed with Bowtie, with alignment using MAFFT and visual inspection using Geneious Prime (all version numbers given below). For the statistical comparison of B.1.1.7 and non-B.1.1.7 subjects, we identified test centers (hospital departments or wards, or organizations outside hospitals) that reported B.1.1.7 cases, and chose as comparison groups non-B.1.1.7 cases that were detected in these test centers on the same day or 1 day earlier or later. By modeling random effects for test centers, we estimate the expected viral load difference as the average of the within-test center differences. The consistent effect of B.1.1.7 throughout a range of comparison scenarios is shown in table S2.

Sample type

An estimated 3% of our samples were from the lower respiratory tract. These were not removed from the dataset because of their low frequency and the fact that the first samples for patients are almost universally swab samples. Samples from the lower respiratory tract are generally taken from patients only after intubation, by which point viral loads have typically fallen.

PAMS status

Metadata needed to discriminate patients into subcohorts on the basis of underlying diseases,

outcome, or indications for diagnostic test application, including symptomatic status, were not always available. In the absence of subject-level data, we inferred PAMS status using the type of submitting test center as an indicator, classifying subjects as PAMS at the time of testing if their first-positive sample was taken from a walk-in COVID-19 test center and the subject had no later RT-PCR test done in a hospitalized context (e.g., in a ward or an intensive care unit). The correspondence between viral load and PAMS status derived herein may therefore be less accurate than in studies with subject-level symptom data. However, we make no formal claims regarding symptomatic status, and instead emphasize the fact that these PAMS subjects were healthy enough to be presenting at walk-in COVID-19 test centers, and were therefore capable to some extent, at that time, of circulating in the general community.

Bayesian analysis of age–viral load associations

We estimated associations of viral load and age with a thin-plate spline regression using the brms package (*58*, *59*) in R (*60*). Spline coefficients were allowed to vary between groups determined by the clinical status (PAMS, Hospitalized, or Other), and random intercepts captured effects of test centers. To reduce the impact of outliers, we used Student *t*–distributed error terms. The analysis additionally accounted for baseline differences between subject groups, B.1.1.7 status, gender, and for the effect of the RT-PCR system. We also estimated the association between viral load and culture probability in order to calculate the expected culture probability at different age levels. This analysis used weakly informative priors and was estimated using four chains with 1000 warm-up samples and 2000 post–warm-up samples. Convergence of MCMC chains was examined by checking that potential scale reduction factors (R-hat) values were below 1.1. All calculations of age averages and group differences are based on posterior predictions generated from estimated model parameters. Expected probabilities of positive cultures (and their differences) were calculated by applying the posterior distribution of model parameters from the culture probability model to posterior predictions from the age association model.

Combining culture probability data

To estimate the association between viral load and culture probability, we used data previously described by Wölfel (*19*) and Perera (*20*). Four other datasets could not be included because Ct values were not converted to viral loads (*35*, *46*, *61*, *62*). The data from the study by van Kampen *et al*. (*63*) were not included because they differed (by viral load of ~1.0) from the data used for the current analysis (*97*); this is likely due to a combination of factors including many patients who were in critical or immunocompromised condition, a high proportion of samples obtained from the lower respiratory tract (including late in the infectious course), and likely differences in cell culture trials. It is unsurprising that these data result in a shifted viral load/culture probability curve, and we excluded them because our focus was largely on first positive RT-PCR results from the upper respiratory tract, including from many subjects who were PAMS. [See (*97*) for a figure comparing the plot of the van Kampen dataset to the two we used.] To calculate the expected culture probability, by age (as in Fig. 2D) or by day from peak viral load (as in Fig. 4C), we combined the estimated viral loads (Figs. 2A and 4B) with the results of the regression of culture probability shown in Fig. 2C. We used posterior predictions from the age regression model, which reflect the variation of viral load within age groups, to estimate culture probabilities by age. For instance, to obtain the culture probability for a specific age and group, we look up the estimated (expected) viral load for that group, add an error term according to the estimated error variance, and, using the association shown in Fig. 2C, determine the expected culture probability. We used expected time courses (i.e., the model's best guess for a time course) to estimate culture probability time courses.

B.1.1.7 isolation data

The Institute of Virology at Charité–Universitätsmedizin Berlin routinely receives SARS-CoV-2–positive samples for confirmatory testing and sequencing. For this study we used anonymized remainder samples from a large laboratory in northern Germany, which were all stored in phosphate-buffered saline (PBS) and therefore suitable for cell culture isolation trials. Sample transport to the originating lab and later to Berlin was unrefrigerated, via road. As part of the routine testing, these samples were classified by typing RT-PCR and complete genome sequencing (*64*); 113 B.1.1.7 lineage samples and 110 B.1.177 lineage samples were selected, with approximately matched (pre-inoculation) SARS-CoV-2 RNA concentrations. Caco-2 (human colon carcinoma) cell cultures (*65*) were inoculated twice from each sample, once with undiluted material and once with a 1:10 dilution. The diluted inoculant was used to reduce the probability of culturing failure due to the possible presence of host immune factors (antibodies, cytokines, etc.) that might have a negative impact on isolation success, and to reduce the possibility of other unrelated agents (bacteria, fungi, etc.) resulting in cytopathic effect in the culture system. For cell culture isolation trials, 1.6×10^5 cells were seeded per well in a 24-well plate. Cells were inoculated with swab suspensions for 1 hour at 37°C, subsequently rinsed with PBS, and fed with 1 ml of fresh Dulbecco's modified Eagle's minimum essential medium (DMEM; ThermoFisher Scientific) supplemented with 2% fetal bovine serum (FBS; Gibco), penicillin and streptomycin (P/S; 100 U/ml and 100 μg/ml, respectively; ThermoFisher Scientific), and amphotericin B (2.5 μg/ml; Biomol), then incubated for 5 days before harvesting supernatant for RT-PCR testing. Positive cell culture isolation was defined by a minimum 10× higher SARS-CoV-2 RNA load in the supernatant compared to the inoculant and signs of a typical SARS-CoV-2 cytopathic effect. Culture isolation was successful for 22 B.1.1.7 and 61 B.1.177 samples. Because of uncertainty regarding sample handling before arrival at the originating diagnostic laboratory and the unrefrigerated transport, it was not possible to determine whether isolation failures were due to samples containing no infectious particles (due to sample degradation) or for other reasons. Such reasons could include systematic handling differences according to variant type or a difference in virion stability and durability regarding environmental factors such as temperature. Therefore, samples with negative isolation outcome were excluded from analysis. The strong likelihood of many cases of complete sample degradation is evident from the isolation failure of many samples with high pre-inoculation viral load, with the viral load in these cases merely indicating the presence of noninfectious SARS-CoV-2 RNA (fig. S4). Given this context, we were reduced to questioning whether there might be a difference in the range of viral loads that were able to result in isolation between B.1.1.7 and non-B.1.1.7 variants. Such a difference could result from a difference in the ratio of viral RNA to infectious particles produced by the variants, or from a difference other than viral load in the variants. We examined the distribution of pre-inoculation viral loads from isolation-positive samples from both variants for a difference. No statistically significant difference was found, but in the converse, the isolation-positive sample sizes are too low to support the assertion that the distributions do not differ.

Estimating viral load time course

Each RT-PCR test in our dataset has a date, but no information regarding the suspected date of subject infection or onset of symptoms (if any). Although determining the day of peak viral load for a single person based on a series of dated RT-PCR results would not in general be feasible because of individual variation, data from a large enough set of people would enable the inference of a clear and consistent model of viral load change over time with very few assumptions.

We included a single leading and/or trailing negative RT-PCR result, if dated within 7 days of the closest positive RT-PCR. To produce a model of typical viral load decline on a reasonable single-infection time scale, we excluded

subjects whose full time series contains positive RT-PCRs spread over a period exceeding 30 days. Such time series may be attributable to contamination, to later swabbing that picks up residual RNA fragments in tonsillar tissue (*66*), or to re-infection (*67*–*69*), or they may represent atypical infection courses (such as in immunocompromised or severely ill elderly patients) (*70*). We excluded data from subjects with an infection delimited by both an initial and a trailing negative test when there was only a single positive RT-PCR result between them.

We estimated the slopes for a model of linear increase and then decline of $\log_{10}$(viral load). To compensate for the absence of information regarding time of infection, we also estimated the number of days from infection to the first positive test for each participant, so as to position the observed time series relative to the day of peak viral load. The analysis was implemented in two ways. Initially, simulated annealing was used to find an optimized fit of the parameters, minimizing a least-squares error function. Second, a Bayesian hierarchical model estimated subject-specific time courses, imputed the viral load assigned to each initial or trailing negative test, and captured effects of age, gender, clinical status, and RT-PCR system with model parameters. We tested both methods on data subsets ranging from subjects with at least three to at least nine RT-PCR results. The two methods produced results that were in generally good agreement (table S5). The finer-grained Bayesian approach appears more sensitive than the simulated annealing; its results, for subjects with at least three RT-PCR results, are those described in the main text.

Simulated annealing approach: A simulated annealing optimization algorithm (*71*) was used to adjust the time series for each subject slightly earlier or later in time, by amounts drawn from a normal distribution with mean 0.0 and standard deviation 0.1 days. The error function was the sum of squares of distances of each viral load from a viral load decline line whose slope was also adjusted as part of the annealing process. In the error calculation, negative test results were assigned a viral load of 2.0, in accordance with our SARS-CoV-2 assay limit of detection and sample dilution (*19*). The initial slope of the decline line was set to –2.0 and was varied using $N(0, 0.01)$. A second, optional, increase line initialized with a slope of 2.0, adjusted using an $N(0, 0.01)$ random variable, was included in the error computation if the day of a RT-PCR test was moved earlier than day zero (the modeled day of peak viral load). The height of the intercept (i.e., the estimated peak viral load) between the increase line (if any) and the decline line was also allowed to vary randomly [starting value 10.0, varied using $N(0, 0.1)$]. The full time series for each subject was initialized with the first positive result positioned at day 2 + $N(0.0, 0.5)$ after peak viral load. The random-move step of the simulated annealing modified either of the two slopes or the intercept, each with probability 0.01, otherwise (with probability 0.97) one subject's time series was randomly chosen to be adjusted earlier or later in time. After the simulated annealing stage, each time series was adjusted to an improved fit (when possible) based on the optimized increase and decline lines. Linear regression lines were then fitted through the results occurring before and after the peak viral load ($x = 0$) and compared to the lines with slopes optimized by the simulated annealing alone. This final step helped to fine-tune the simulated annealing, in particular sometimes placing a time series much earlier or much later in time after it had stochastically moved initially in a direction that later (when the increase and decline line slopes had converged) proved to be suboptimal. The slopes of the lines fitted via linear regression after this final step were in all cases very similar (generally ±0.1) to those produced by the initial simulated annealing step. The final adjustments can be regarded as a last step in the optimization, using a steepest-descent movement operator instead of an uninformed random one. A representative optimization run for subjects with at least three RT-PCR results is shown in fig. S12.

Bayesian approach: The Bayesian analysis of viral load time course implements the same basic model, and additionally estimates associations of model parameters with covariates age, gender, B.1.1.7 status, and clinical status, estimates subject-level parameters (slope of $\log_{10}$ viral load increase, peak viral load, slope of $\log_{10}$ viral load decrease) as random effects, and accounts for effects of PCR system and test center types with random effects. To estimate the number of days from infection to the first test (henceforth "shift"), we constrained the possible shift values from –10 to 20 days and used a uniform prior on the support. In contrast to the other subject-level parameters, we estimated subject-level shifts independently (i.e., without a hierarchical structure). Figure S7 shows the placement in time of individual viral loads after shifting for subjects with RT-PCR results from at least 3 days. Model parameters changed gradually when subsets of subjects with an increasing minimum number of RT-PCR results, from three to nine, were examined (fig. S11 and table S5). The viral load assigned to negative test results (which may include viral loads below the level of detection) is estimated with a uniform prior on the support from –Inf to 3 (see also the caption of fig. S7). Using prior predictive simulations, we specified (weakly) informative priors for this analysis. This analysis was implemented in Stan (*72*), as described in (*97*).

Checking convergence of the model parameters showed that although 99.3% of all parameters converged with an R-hat value below 1.1, some subject-level parameters of 118 subjects (among 4344 subjects with at least three RT-PCR results) showed R-hat values between 1.1 and 1.74. Inspection of these parameters showed that these convergence difficulties were due to observed time courses that could arguably be placed equally well at the beginning or a later stage of the infection. Figure S16 shows a set of 81 randomly selected posterior predictions, to give an impression of time-series placement; fig. S17 shows the 49 participants with the parameters with the highest R-hat values. Although the high R-hat values could be removed by using a mixture approach to model shift for these participants, in light of their low frequency we retained the simpler model to avoid additional complexity. Alternatively, constraining the shift parameter to negative numbers would also improve R-hat values for these subjects, at the cost of the additional assumption that infections are generally not detected weeks after infection.

Sensitivity analysis: In addition to examining the viral load time series of subjects with RT-PCR results on at least 3 days, we tested both approaches on data from subjects with results from a minimum of 4 to 9 days. Given the degree of temporal viral load variation seen in other studies (*18*–*20*, *35*, *41*, *46*, *63*, *73*, *74*) and in our own data, our expectation was that a relatively high minimum number of results might be required before reliable parameter estimates with small variance would be obtained, but this proved not to be the case. The simulated annealing approach was tested with a wide range of initial slopes and intercept heights as well as seven different methods for the initial placement of time series. In general, maximum viral load and decline slopes were robust to data subset and initial time-series location, although there was variation in the length of the time to peak viral load, depending on how early in time the time series were initially positioned, the initial slopes of the increase and decrease lines and height of the maximum viral load. This is as expected, as the settings of these parameters can be used to bias the probability that a time series is initially positioned early or late in time and how difficult it is for it to subsequently move to the other side of the peak viral load at day zero. Table S5 shows parameter values for both approaches on the various data subsets.

Onset of shedding: We define the onset of shedding as the time point at which the increasing viral load crosses zero of the $\log_{10}$ y axis—that is, when just one viral particle was estimated to be present. Because the estimated time of infection depends on the estimated peak viral load and the slope with which viral load increases, the data should optimally include multiple pre-peak viral load test results for each individual. If, as in the

current dataset, only a subset of subjects have test results from pre-peak viral load, a hierarchical modeling approach still allows calculating subject-level estimates. Intuitively, this approach uses data from all subjects to calculate an average slope parameter for increasing viral load. In addition, it models subject-level parameters as varying around the group-level parameter. To further refine the estimation of slope parameters, the model also uses the age (see fig. S10), gender, and clinical status as covariates. Because negative test results could be false negatives, viral loads for these tests are imputed (with an upper bound of 3). Subject-level peak viral load and declining slope are modeled with the same approach. More generally, using a hierarchical model and shrinkage priors for the effects of covariates results in more accurate predictions in terms of expected squared error (*75*) compared to analyzing each subject in isolation, but the overall improvement introduces a slight bias toward the group mean, resulting in an underestimation of the true variability of subject-level parameters. This is especially the case if, as in the current dataset, subject-level data are sparse.

Onset of symptoms: The 317 onset-of-symptoms dates for hospitalized patients were collected as part of the Pa-COVID-19 study, a prospective observational cohort study at Charité–Universitätsmedizin Berlin (*76*, *77*), approved by the local ethics committee (EA2/066/20), conducted according to the Declaration of Helsinki and Good Clinical Practice principles (ICH 1996), and registered in the German and WHO international clinical trials registry (DRKS00021688).

Software

The following Python (version 3.8.2) software packages were used in the data analysis and in the production of figures: Scipy (version 1.4.1) (*78*), pandas (version 1.0.3) (*79*), statsmodels (version 0.11.1) (*80*), matplotlib (version 3.2.1) (*81*), numpy (1.18.3) (*82*), seaborn_sinaplot (*83*), simanneal (version 0.5.0) (*71*), and seaborn (version 0.10.1) (*84*). Sequence analysis used Bowtie2 (2.4.1) (*85*), bcftools and samtools (1.9) (*86*, *87*), Geneious Prime (2021.0.3) (*88*), ivar (1.2.2) (*89*), and MAFFT (4.475) (*90*). Analyses in R (4.0.2) (*60*) were conducted using the following main packages: brms (2.13.9) (*58*, *59*), rstanarm (2.21.1) (*91*), rstan (2.21.2) (*92*), data.table (1.13.3) (*93*), and ggplot2 (3.3.2) (*94*). Bayesian analysis in R was based on Stan (2.25) (*72*). Parallel execution was performed with GNU Parallel [20201122 ('Biden') (*95*)].

Data curation and anonymization

Research clearance for the use of routine data from anonymized subjects is provided under paragraph 25 of the Berlin *Landeskrankenhausgesetz*. All data are anonymized before processing to ensure that it is not possible to infer patient identity from any processing result. All patient information is securely combined into a token that is then replaced with a value from a strong one-way hash function prior to the distribution of data for analysis. Viral loads are calculated from RT-PCR cycle threshold values that have only one decimal place of precision.

REFERENCES AND NOTES

1. S. Lee *et al*., Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Intern. Med.* **180**, 1447–1452 (2020). doi: 10.1001/jamainternmed.2020.3862; pmid: 32780793
2. C. M. Szablewski *et al*., SARS-CoV-2 Transmission and Infection Among Attendees of an Overnight Camp - Georgia, June 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1023–1025 (2020). doi: 10.15585/mmwr.mm6931e1; pmid: 32759921
3. Q.-X. Long *et al*., Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020). doi: 10.1038/s41591-020-0965-6; pmid: 32555424
4. Q. Bi *et al*., Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020). doi: 10.1016/S1473-3099(20)30287-5; pmid: 32353347
5. T. Waterfield *et al*., Seroprevalence of SARS-CoV-2 antibodies in children: A prospective multicentre cohort study. *Arch. Dis. Child.* 10.1136/archdischild-2020-320558 (2020). doi: 10.1136/archdischild-2020-320558; pmid: 33172887
6. D. C. Adam *et al*., Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020). pmid: 32943787
7. M. Hippich *et al*., A public health antibody screening indicates a six-fold higher SARS-CoV-2 exposure rate than reported cases in children. *Med* **2**, 149–163.e4 (2020). doi: 10.1038/s41586-020-2488-1; pmid: 32604404
8. E. Lavezzo *et al*., Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429 (2020). doi: 10.1038/s41586-020-2488-1; pmid: 32604404
9. M. Marks *et al*., Transmission of COVID-19 in 282 clusters in Catalonia, Spain: A cohort study. *Lancet Infect. Dis.* **21**, 629–636 (2021). doi: 10.1016/S1473-3099(20)30985-3; pmid: 33545090
10. L. L. H. Lau *et al*., Viral shedding and clinical illness in naturally acquired influenza virus infections. *J. Infect. Dis.* **201**, 1509–1516 (2010). doi: 10.1086/652241; pmid: 20377412
11. T. K. Tsang *et al*., Individual Correlates of Infectivity of Influenza A Virus Infections in Households. *PLOS ONE* **11**, e0154418 (2016). doi: 10.1371/journal.pone.0154418; pmid: 27153194
12. E. A. Meyerowitz, A. Richterman, I. I. Bogoch, N. Low, M. Cevik, Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. *Lancet Infect. Dis.* S1473-3099(20)30837-9 (2020). doi: 10.1016/S1473-3099(20)30837-9; pmid: 33301725
13. A. Fontanet *et al*., SARS-CoV-2 infection in schools in a northern French city: A retrospective serological cohort study in an area of high transmission, France, January to April 2020. *Euro Surveill.* **26**, 2001695 (2021). doi: 10.2807/1560-7917.ES.2021.26.15.2001695
14. C. Stein-Zamir *et al*., A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. *Euro Surveill.* **25**, (2020). doi: 10.2807/1560-7917.ES.2020.25.29.2001352; pmid: 32720636
15. I. W. Pray *et al*., COVID-19 Outbreak at an Overnight Summer School Retreat - Wisconsin, July-August 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1600–1604 (2020). doi: 10.15585/mmwr.mm6943a4; pmid: 33119558
16. J. P. Torres *et al*., SARS-CoV-2 antibody prevalence in blood in a large school community subject to a Covid-19 outbreak: A cross-sectional study. *Clin. Infect. Dis.* ciaa955 (2020). doi: 10.1093/cid/ciaa955; pmid: 32649743
17. M. H. Ebell, C. Chupp, M. Bentivegna, A high proportion of SARS-CoV-2-infected university students are asymptomatic. *J. Fam. Pract.* **69**, 428–429 (2020). pmid: 33176345
18. S. M. Kissler *et al*., Viral dynamics of SARS-CoV-2 infection and the predictive value of repeat testing. medRxiv 20217042 [preprint]. 23 October 2020. pmid: 20217042
19. R. Wölfel *et al*., Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020). doi: 10.1038/s41586-020-2196-x; pmid: 32235945
20. R. A. P. M. Perera *et al*., SARS-CoV-2 Virus Culture and Subgenomic RNA for Respiratory Specimens from Patients with Mild Coronavirus Disease. *Emerg. Infect. Dis.* **26**, 2701–2704 (2020). doi: 10.3201/eid2611.203219; pmid: 32749957
21. A. Endo, S. Abbott, A. J. Kucharski, S. Funk, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020). doi: 10.12688/wellcomeopenres.15842.3; pmid: 32685698
22. K. Sun *et al*., Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, eabe2424 (2021). pmid: 33234698
23. N. J. Lennon *et al*., Comparison of viral levels in individuals with or without symptoms at time of COVID-19 testing among 32,480 residents and staff of nursing homes and assisted living facilities in Massachusetts. *Open Forum Infect. Dis.* **7** (suppl. 1), 848–849 (2020). doi: 10.1093/ofid/ofaa515.1908
24. R. Laxminarayan *et al*., Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020). doi: 10.1126/science.abd7672; pmid: 33154136
25. T. Suess *et al*., Comparison of shedding characteristics of seasonal influenza virus (sub)types and influenza A(H1N1)pdm09; Germany, 2007-2011. *PLOS ONE* **7**, e51653 (2012). doi: 10.1371/journal.pone.0051653; pmid: 23240050
26. M. Loeb *et al*., Longitudinal study of influenza molecular viral shedding in Hutterite communities. *J. Infect. Dis.* **206**, 1078–1084 (2012). doi: 10.1093/infdis/jis450; pmid: 22837493
27. B. J. Cowling *et al*., Comparative epidemiology of pandemic and seasonal influenza A in households. *N. Engl. J. Med.* **362**, 2175–2184 (2010). doi: 10.1056/NEJMoa0911530; pmid: 20558368
28. D. K. M. Ip *et al*., Viral Shedding and Transmission Potential of Asymptomatic and Paucisymptomatic Influenza Virus Infections in the Community. *Clin. Infect. Dis.* **64**, 736–742 (2017). pmid: 28011603
29. N. H. L. Leung, C. Xu, D. K. M. Ip, B. J. Cowling, The Fraction of Influenza Virus Infections That Are Asymptomatic: A Systematic Review and Meta-analysis. *Epidemiology* **26**, 862–872 (2015). doi: 10.1097/EDE.0000000000000340; pmid: 26133025
30. L. P. Moreira *et al*., Respiratory syncytial virus evaluation among asymptomatic and symptomatic subjects in a university hospital in Sao Paulo, Brazil, in the period of 2009-2013. *Influenza Other Respir. Viruses* **12**, 326–330 (2018). doi: 10.1111/irv.12518; pmid: 29078028
31. D. Jacot, G. Greub, K. Jaton, O. Opota, Viral load of SARS-CoV-2 across patients and compared to other respiratory viruses. *Microbes Infect.* **22**, 617–621 (2020). doi: 10.1016/j.micinf.2020.08.004; pmid: 32911086
32. T. Heald-Sargent *et al*., Age-Related Differences in Nasopharyngeal Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Levels in Patients With Mild to Moderate Coronavirus Disease 2019 (COVID-19). *JAMA Pediatr.* **174**, 902–903 (2020). doi: 10.1001/jamapediatrics.2020.3651; pmid: 32745201
33. L. M. Yonker *et al*., Pediatric SARS-CoV-2: Clinical Presentation, Infectivity, and Immune Responses. *J. Pediatr.* **227**, 45–52.e5 (2020). doi: 10.1016/j.jpeds.2020.08.037; pmid: 32827525
34. S. Baggio *et al*., SARS-CoV-2 viral load in the upper respiratory tract of children and adults with early acute COVID-19. *Clin. Infect. Dis.* ciaa1157 (2020). doi: 10.1093/cid/ciaa1157; pmid: 32761228
35. A. Singanayagam *et al*., Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill.* **25**, (2020). doi: 10.2807/1560-7917.ES.2020.25.32.2001483; pmid: 32794447
36. V. M. Corman *et al*., Comparison of seven commercial SARS-CoV-2 rapid point-of-care antigen tests: a single-centre

laboratory evaluation study. *Lancet Microbe* 10.1016/S2666-5247(21)00056-2 (2021). doi: 10.1001/jamapediatrics.2020.3988; pmid: 32857112
37. M. S. Han *et al*., Clinical Characteristics and Viral RNA Detection in Children With Coronavirus Disease, 2019 in the Republic of Korea. *JAMA Pediatr.* **175**, 73–80 (2021). doi: 10.1001/jamapediatrics.2020.3988; pmid: 32857112
38. C. A. Pierce *et al*., Natural mucosal barriers and COVID-19 in children. *JCI Insight* **6**, e148694 (2021). doi: 10.1172/jci.insight.148694; pmid: 33822777
39. Q. Li *et al*., Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020). doi: 10.1056/NEJMoa2001316; pmid: 31995857
40. L. Ferretti *et al*., The timing of COVID-19 transmission. medRxiv [preprint]. 7 September 2020. pmid: 20188516
41. X. He *et al*., Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020). doi: 10.1038/s41591-020-0869-5; pmid: 32296168
42. C. McAloon *et al*., Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open* **10**, e039652 (2020). doi: 10.1136/bmjopen-2020-039652; pmid: 32801208
43. P. Banka, C. Comiskey, The incubation period of COVID-19: A scoping review and meta-analysis to aid modelling and planning. MedRxiv 20216143 [preprint]. 3 November 2020.
44. B. Rai, A. Shukla, L. K. Dwivedi, Incubation period for COVID-19: A systematic review and meta-analysis. *J. Public Health* 10.1007/s10389-021-01478-1 (2021). doi: 10.1007/s10389-021-01478-1; pmid: 33643779
45. J. Bullard *et al*., Predicting infectious SARS-CoV-2 from diagnostic samples. *Clin. Infect. Dis.* **71**, 2663–2666 (2020). doi: 10.1093/cid/ciaa638
46. M. M. Arons *et al*., Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020). doi: 10.1056/NEJMoa2008457; pmid: 32329971
47. R. Challen *et al*., Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *BMJ* **372**, n579 (2021). pmid: 33687922
48. K. R. W. Emary, T. Golubchik, Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 VOC 202012/01 (B.1.1.7). SSRN [preprint]. 4 February 2021; https://ssrn.com/abstract=3779160.
49. M. D. Parker *et al*., Altered subgenomic RNA expression in SARS-CoV-2 B.1.1.7 infections. bioRxiv 433156 [preprint]. 4 March 2021. pmid: 433156
50. M. Kidd *et al*., S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples tested by TaqPath Polymerase Chain Reaction. *J. Infect. Dis.* jiab082 (2021). doi: 10.1093/infdis/jiab082; pmid: 33580259
51. T. Golubchik *et al*., COVID-19 Genomics UK (COG-UK) Consortium, Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. medRxiv 20249080 [preprint]. 15 January 2021. pmid: 20249080
52. S. Kissler *et al*., "Densely sampled viral trajectories suggest longer duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2" (Harvard T. H. Chan School of Public Health, 2021); https://dash.harvard.edu/handle/1/37366884.
53. Public Health England, "Investigation of novel SARS-CoV-2 Variant of Concern 202012/01: Technical briefing 5" (2021).
54. K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.* **26**, (2021). doi: 10.2807/1560-7917.ES.2020.26.1.2002106; pmid: 33413740
55. N. G. Davies *et al*., Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021). doi: 10.1126/science.abg3055; pmid: 33658326
56. R. Li *et al*., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020). doi: 10.1126/science.abb3221; pmid: 32179701
57. Public Health England, "Investigation of novel SARS-CoV-2 Variant of Concern 202012/01: Technical briefing 1" (2020).
58. P.-C. Bürkner, brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, (2017). doi: 10.18637/jss.v080.i01
59. P.-C. Bürkner, Advanced Bayesian Multilevel Modeling with the R Package brms. *R J.* **10**, 395 (2018). doi: 10.32614/RJ-2018-017
60. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020); www.R-project.org/.
61. K. Basile *et al*., Cell-based culture of SARS-CoV-2 informs infectivity and safe de-isolation assessments during COVID-19. *Clin. Infect. Dis.* ciaa1579 (2020). doi: 10.1093/cid/ciaa1579; pmid: 33098412
62. B. La Scola *et al*., Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *Eur. J. Clin. Microbiol. Infect. Dis.* **39**, 1059–1061 (2020). doi: 10.1007/s10096-020-03913-9; pmid: 32342252
63. J. J. A. van Kampen *et al*., Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nat. Commun.* **12**, 267 (2021). doi: 10.1038/s41467-020-20568-4; pmid: 33431879
64. M. Widera *et al*., Surveillance of SARS-CoV-2 in Frankfurt am Main from October to December 2020 Reveals High Viral Diversity Including Spike Mutation N501Y in B.1.1.70 and B.1.1.7. *Microorganisms* **9**, 748 (2021). doi: 10.3390/microorganisms9040748; pmid: 33918332
65. T. Toptan *et al*., Evaluation of a SARS-CoV-2 rapid antigen test: Potential to help reduce community spread? *J. Clin. Virol.* **135**, 104713 (2021). doi: 10.1016/j.jcv.2020.104713; pmid: 33352470
66. S. Herberhold, A.-M. Eis-Hübinger, M. Panning, Frequent detection of respiratory viruses by real-time PCR in adenoid samples from asymptomatic children. *J. Clin. Microbiol.* **47**, 2682–2683 (2009). doi: 10.1128/JCM.00899-09; pmid: 19494063
67. F. M. Liotti *et al*., Assessment of SARS-CoV-2 RNA Test Results Among Patients Who Recovered From COVID-19 With Prior Negative Results. *JAMA Intern. Med.* **181**, 702–704 (2021). doi: 10.1001/jamainternmed.2020.7570; pmid: 33180119
68. R. L. Tillett *et al*., Genomic evidence for reinfection with SARS-CoV-2: A case study. *Lancet Infect. Dis.* **21**, 52–58 (2021). doi: 10.1016/S1473-3099(20)30764-7; pmid: 33058797
69. K. K.-W. To *et al*., Coronavirus Disease 2019 (COVID-19) Re-infection by a Phylogenetically Distinct Severe Acute Respiratory Syndrome Coronavirus 2 Strain Confirmed by Whole Genome Sequencing. *Clin. Infect. Dis.* **9**, 1664 (2020).
70. P. Simmonds, S. Williams, H. Harvala, Understanding the outcomes of COVID-19 – does the current model of an acute respiratory infection really fit? *J. Gen. Virol.* **102**, 10.1099/jgv.0.001545 (2020). doi: 10.1099/jgv.0.001545; pmid: 33331810
71. M. Perry, simanneal: A Python Module for Simulated Annealing Optimization; https://github.com/perrygeo/simanneal.
72. Stan Development Team, Stan Modeling Language Users Guide and Reference Manual (version 2.25); https://mc-stan.org.
73. Y. Liu *et al*., Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect. Dis.* **20**, 656–657 (2020). doi: 10.1016/S1473-3099(20)30232-2; pmid: 32199493
74. K. K.-W. To *et al*., Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: An observational cohort study. *Lancet Infect. Dis.* **20**, 565–574 (2020). doi: 10.1016/S1473-3099(20)30196-1; pmid: 32213337
75. S. Greenland, Principles of multilevel modelling. *Int. J. Epidemiol.* **29**, 158–167 (2000). doi: 10.1093/ije/29.1.158; pmid: 10750618
76. F. Kurth *et al*., Studying the pathophysiology of coronavirus disease 2019: A protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). *Infection* **48**, 619–626 (2020). doi: 10.1007/s15010-020-01464-x; pmid: 32535877
77. C. Thibeault *et al*., Clinical and virological characteristics of hospitalised COVID-19 patients in a German tertiary care centre during the first wave of the SARS-CoV-2 pandemic: A prospective observational study. *Infection* 10.1007/s15010-021-01594-w (2021). doi: 10.1007/s15010-021-01594-w; pmid: 33890243
78. P. Virtanen *et al*., SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020). doi: 10.1038/s41592-019-0686-2; pmid: 32015543
79. W. McKinney, Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* (2010). doi: 10.25080/majora-92bf1922-00a
80. S. Seabold, J. Perktold, Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference* (2010). doi: 10.25080/majora-92bf1922-011
81. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007). doi: 10.1109/MCSE.2007.55
82. T. Oliphant, *Guide to NumPy* (CreateSpace, ed. 2, 2015).
83. M. Parker, seaborn_sinaplot; https://github.com/mparker2/seaborn_sinaplot.
84. M. Waskom *et al*., seaborn: v0.5.0 (2014); DOI: 10.5281/zenodo.12710.
85. 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
86. P. Danecek *et al*., Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021). doi: 10.1093/gigascience/giab008; pmid: 33590861
87. J. K. Bonfield *et al*., HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **10**, giab007 (2021). doi: 10.1093/gigascience/giab007; pmid: 33594436
88. G. Dick, *Genomic Approaches in Earth and Environmental Sciences* (Wiley, 2018).
89. N. D. Grubaugh *et al*., An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019). doi: 10.1186/s13059-018-1618-7; pmid: 30621750
90. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). doi: 10.1093/molbev/mst010; pmid: 23329690
91. B. Goodrich, J. Gabry, I. Ali, S. Brilleman, rstanarm: Bayesian applied regression modeling via Stan (2020); https://mc-stan.org/rstanarm.
92. B. Carpenter *et al*., Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **76**, (2017). doi: 10.18637/jss.v076.i01
93. M. Dowle, A. Srinivasan, data.table: Extension of 'data.frame' (2020).
94. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
95. O. Tange, GNU Parallel 20201122 ('Biden') (2020); www.gnu.org/software/parallel/.
96. H. B. Mann, D. R. Whitney, On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* **18**, 50–60 (1947). doi: 10.1214/aoms/1177730491
97. Additional statistical information and the R code and data to reproduce the results, figures, and tables are available at https://doi.org/10.5281/zenodo.4774226.

ACKNOWLEDGMENTS

Computation was performed on the HPC for Research/Clinic cluster of the Berlin Institute of Health, supported by D. Beule, M. Holtgrewe, and O. Stolpe. We thank U. Gieraths and L. Meiners for careful commentary on the manuscript, T. D. Best for compiling cell culture isolation data, the Charité–Universitätsmedizin Pa-COVID-19 collaborative study group for providing additional onset of symptoms data, and S. Kissler for providing additional details regarding their NBA study. The conditions allowing the work to be done with no need for consent are given at https://gesetze.berlin.de/bsbe/document/jlr-KHGBE2011V4P25. **Funding:** Work at Charité–Universitätsmedizin Institute of Virology is funded by European Commission via project ReCoVer, German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) through projects DZIF (301-4-7-01.703) to C.D.; VARIPath (01KI2021) to V.M.C.; PROVID (FKZ 01KI20160C) to C.D., V.M.C., and L.E.S.; and NaFoUniMedCovid19 (NUM)–COVIM (FKZ 01KX2021) to C.D., V.M.C., and L.E.S. The Pa-COVID 19 Study is supported by grants from the Berlin Institute of Health. This study was supported in part by the German Ministry of Health (Konsiliarlabor für Coronaviren and SeCoV) to C.D. and V.M.C. T.C.J. is in part funded through NIAID-NIH CEIRS contract HHSN272201400008C.
Author contributions: T.C.J., G.B., B.M.: bioinformatic processing, statistical analysis, interpretation of results, writing of original draft and final text; T.V.: statistical analysis, interpretation of results, writing of original draft and final text, next-generation sequencing; J.S., J.B.-S., T.B., J.T., M.L.S.: sample preparation, virus isolation and culturing, RT-PCR, next-generation sequencing; L.E.S., F.K.: collection of symptom onset data; P.M., R.S., M.Z., J.H., A.K., A.S., A.E.: diagnostic work and collection of raw data; V.M.C.: diagnostic data collection, viral load calibration, supervision of laboratory work, interpretation of results; C.D.: project concept, interpretation of results, writing of original draft and final text.
Competing interests: The authors declare that they have no

competing interests. **Data and materials availability:** Additional statistical information and the R code and data to reproduce the results, figures, and tables are available (*97*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

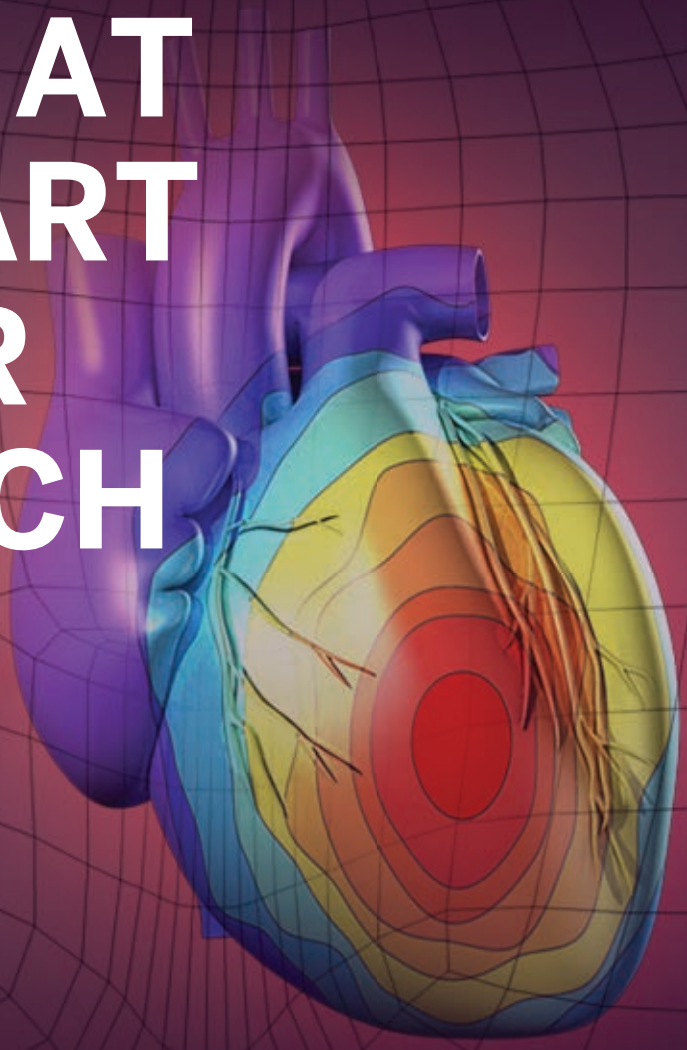
SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/eabi5273/suppl/DC1
Supplementary Text
Figs. S1 to S17
Tables S1 to S5
References (*98*–*102*)
MDAR Reproducibility Checklist

15 March 2021; accepted 21 May 2021
Published online 25 May 2021
10.1126/science.abi5273

ScienceTranslationalMedicine.org

PUT HUMAN HEALTH AT THE HEART OF YOUR RESEARCH



Submit your research:
cts.ScienceMag.org

Science Translational Medicine
AAAS

Twitter: @ScienceTM
Facebook: @ScienceTranslationalMedicine

RESEARCH ARTICLES

MICROBIOMES

Gut microbiome heritability is nearly universal but environmentally contingent

Laura Grieneisen[1]*, Mauna Dasari[2], Trevor J. Gould[1], Johannes R. Björk[2], Jean-Christophe Grenier[3,4], Vania Yotova[3], David Jansen[2], Neil Gottel[5], Jacob B. Gordon[6], Niki H. Learn[7], Laurence R. Gesquiere[6], Tim L. Wango[8,9], Raphael S. Mututua[8], J. Kinyua Warutere[8], Long'ida Siodi[8], Jack A. Gilbert[5], Luis B. Barreiro[3,10], Susan C. Alberts[6,11,12], Jenny Tung[6,11,12,13]†*, Elizabeth A. Archie[2]†*, Ran Blekhman[1,14]†*

Relatives have more similar gut microbiomes than nonrelatives, but the degree to which this similarity results from shared genotypes versus shared environments has been controversial. Here, we leveraged 16,234 gut microbiome profiles, collected over 14 years from 585 wild baboons, to reveal that host genetic effects on the gut microbiome are nearly universal. Controlling for diet, age, and socioecological variation, 97% of microbiome phenotypes were significantly heritable, including several reported as heritable in humans. Heritability was typically low (mean = 0.068) but was systematically greater in the dry season, with low diet diversity, and in older hosts. We show that longitudinal profiles and large sample sizes are crucial to quantifying microbiome heritability, and indicate scope for selection on microbiome characteristics as a host phenotype.

An important goal of microbiome research is to determine the heritability of gut microbiome traits (*1*–*8*). Linking microbiome variation to host genetic variation can reveal which aspects of the microbiome are capable of responding to selection on the host, suggest which microbiome traits are under host control, and connect microbial abundance to host pathways and disease states (*1*, *7*). However, current research suggests that heritable gut microbiome taxa are uncommon. In humans, only 3 to 13% of gut microbes have nonzero heritability, and one study estimated that overall microbiome heritability may be as low as 0.019 (*1*, *2*, *4*, *6*, *7*). Furthermore, the few heritable microbiome phenotypes in humans, such as the abundance of the family Christensenellaceae, exhibit widely varying heritability estimates across studies [narrow-sense heritability (h^2) = 0.31 to 0.64 (*1*, *2*, *4*, *6*, *7*, *9*)].

There are challenges in accurately estimating h^2, the proportion of phenotypic variance explained by additive genetic variance, for the human microbiome. First, relatives, especially twins and other first-degree relatives, which are the basis for most microbiome heritability studies, often share diets, behaviors, and built environments, which can cause heritability to be overestimated (*10*). Controlling for gene-environment correlations requires fine-grained, individual-based environmental and behavioral data, which have not been available in previous studies (*1*, *2*, *4*–*6*). Second, all current estimates of microbiome heritability in humans rely on cross-sectional microbiome sampling even though microbial abundances are dynamic and difficult to accurately phenotype from one-time measures (*1*, *2*, *4*, *6*, *7*, *11*). Further, h^2 can change over a host's lifetime because of shifting environmental conditions and host attributes [e.g., h^2 for body mass index decreases with age, as dietary and behavioral effects increase relative to the effects of genotype: (*12*, *13*)]. To date, no studies of gut microbiome heritability have fully accounted for this temporal variability or its dependence on the environment.

Estimating microbiome heritability in a natural primate population

To overcome these challenges, we estimated h^2 for gut microbiome traits in 585 wild baboons (*Papio cynocephalus*, the yellow baboon, with some admixture from anubis baboons, *Papio anubis*; Fig. 1A). To do so, we used 16,234 16*S* rRNA gene sequencing–based microbiome profiles from fecal samples collected longitudinally over 14 years. These samples were collected from the Amboseli baboon population (Fig. 1B), which has been the subject of individual-based research since 1971 (*14*). Each study subject had on average 28 samples collected across 4.5 years (range = 1 to 177 samples per baboon; median days between samples = 28; Fig. 1A).

Baboons lead shorter lives than humans, so these time series often span a substantial fraction of the baboon life span [female life expectancy at birth is 10 years; females and males achieve sexual maturity at 4.5 and 5.7 years respectively; (*15*)]. Each microbiome sample is accompanied by detailed information on the pedigree relationships of its donor (fig. S1), as well as fine-grained data on environmental conditions, social behavior, demography, and group-level diet composition at the time of sampling (Fig. 1C and tables S1 and S2). These complementary data allowed us to achieve precise estimates of heritability and quantify the impact of shifting environmental and social conditions on heritability. They also break apart gene-environment correlations: Baboon social groups contain a wide range of maternal relatives, paternal relatives, and nonrelatives (median within-group relatedness in a given year = 0.055, SD = 0.11), yet all group members travel in a coordinated fashion across the landscape and feed on the same seasonally available foods (*14*). Additionally, groups exhibit substantial home range overlap [Fig. 1B; (*16*)]. Here, we studied 10 social groups that varied in size from 17 to 118 members (mean = 58).

Each 16*S* gut microbiome profile was generated from a fecal sample collected from an individually recognized baboon and processed as described previously (*17*) (figs. S2 to S4 and table S3). Similar to other primates [Fig. 1D; (*18*–*21*)], the most common gut microbial phyla were Firmicutes, Bacteroidetes, and Actinobacteria (Fig. 1E). Both the abundances of these phyla and the composition of baboon diets showed cyclic fluctuations (Fig. 1, C and E), which reflect Amboseli's wet-dry seasonal dynamics (*14*).

Using these microbiome profiles, we estimated the h^2 of 1034 gut microbiome phenotypes. These included seven community phenotypes, or measures of microbiome community composition [amplicon sequence variant (ASV) richness, ASV Shannon's H index, and the first five principal coordinates (PCs) of a Bray-Curtis dissimilarity matrix], and 283 single-taxon phenotypes representing the relative abundance of individual microbiome taxa, from ASVs through phyla, found in >50% of samples [figs. S5 and S6; (*1*–*6*, *17*, *22*)]. We also estimated h^2 for 744 presence/absence phenotypes, which reflect

[1]Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455, USA. [2]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. [3]Department of Genetics, CHU Sainte Justine Research Center, Montréal, Quebec H3T 1C5, Canada. [4]Research Center, Montreal Heart Institute, Montréal, Quebec H1T 1C8, Canada. [5]Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093. [6]Department of Biology, Duke University, Durham, NC 27708, USA. [7]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [8]Amboseli Baboon Research Project, Amboseli National Park, Kenya. [9]The Department of Veterinary Anatomy and Animal Physiology, University of Nairobi, Kenya. [10]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637, USA. [11]Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA. [12]Duke Population Research Institute, Duke University, Durham, NC 27708, USA. [13]Canadian Institute for Advanced Research, Toronto, Ontario M5G 1M1, Canada. [14]Department of Ecology, Evolution, and Behavior, University of Minnesota, Minneapolis, MN 55455, USA.
*Corresponding author. Email: lgrienei@umn.edu (L.G.); blekhman@umn.edu (R.B.); earchie@nd.edu (E.A.A.); jenny.tung@duke.edu (J.T.)
†These authors contributed equally to this work.

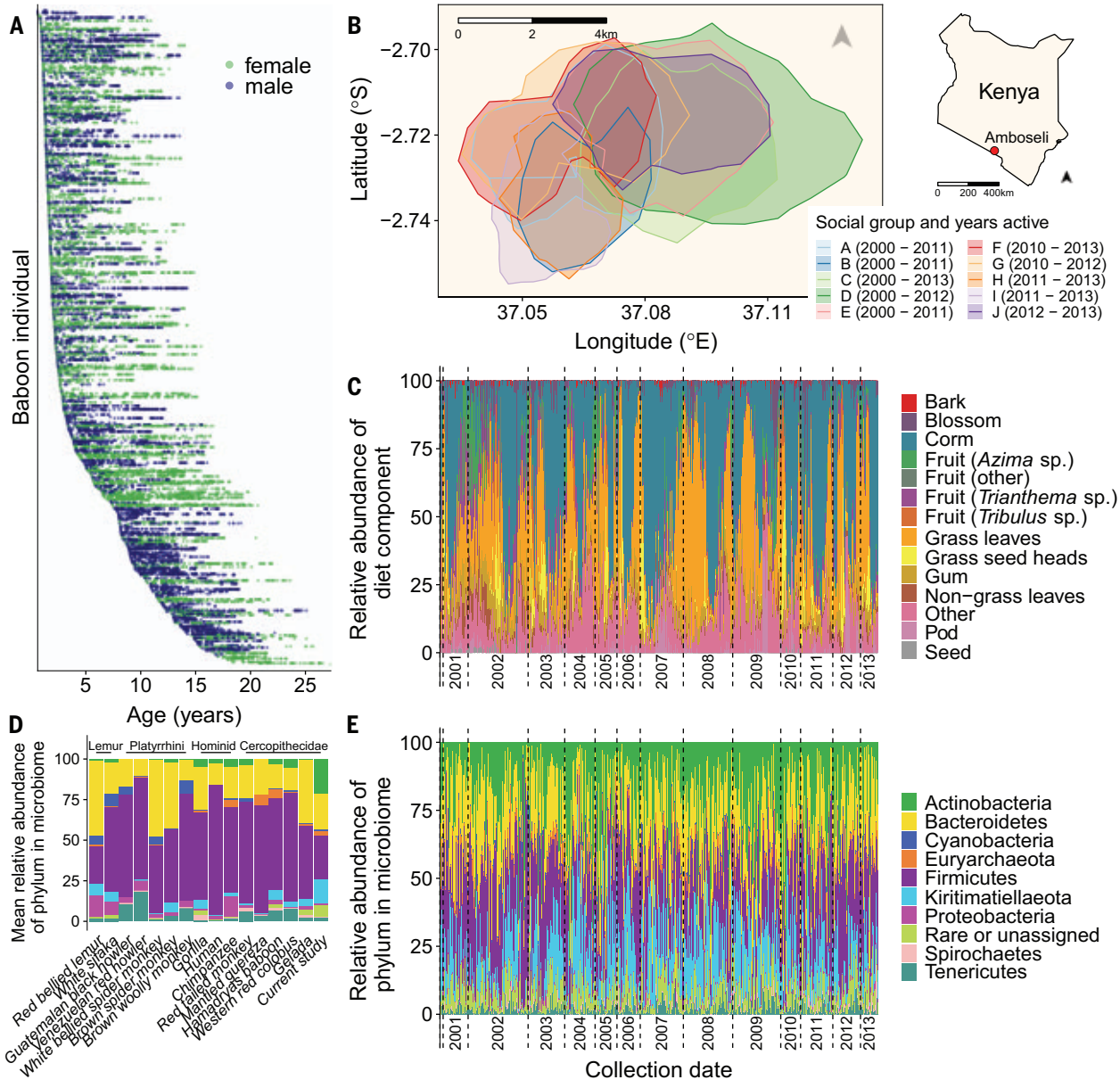


Fig. 1. Time-series data used to estimate microbiome heritability. (**A**) Dataset consisting of 16,234 microbiome samples collected from 585 individually recognized baboons. Each point represents a sample; the *y*-axis is ordered by baboon age at first sample collection. (**B**) Map of the 90% kernel density estimate (KDE) home ranges and active dates for the 10 baboon social groups sampled over the study period based on 71,645 GPS points collected during group monitoring. (**C**) For each microbiome sample, we had data on the diet consumed by members of the corresponding social group in the 30 days before sample collection. Each vertical bar represents one sample, ordered by collection date. Colors represent diet components (see table S1). (**D**) The relative abundances of microbial phyla in the current study are similar to prior primate studies (*18*–*21*). (**E**) Relative abundance of microbial phyla in all 16,234 samples ordered by collection date. "Rare" taxa: <0.5% mean relative abundance per sample. In (C) and (E), the *x*-axis starts at the year 2000. The same legend applies for (D) and (E).

whether a taxon is present or absent in a sample, respectively (limited to taxa found in 10 to 90% of samples).

We estimated h^2 separately for each phenotype using the animal model implemented in ASReml-R v3 [tables S4 and S5; (*23*)]. This mixed-effects model estimates each individual's additive genetic value as a random effect based on the expected covariance in additive genetic effects between relatives in a pedigree (*24*–*26*). It also partitions phenotypic variance across additive genetic variance and other random effects, after conditioning on fixed effects. Following the typical approach in human genetics and plant and animal breeding, we estimated total phenotypic variance (the denominator of h^2) after correcting for fixed effects (*12*, *24*, *27*). This allowed us to exclude the effects of environmentally variable traits such as diet and rainfall, technical effects, and demographic variables such as sex and age (figs. S7 and S8).

Genetic effects on the gut microbiome are nearly universal

We found that 97% of single-taxon and community phenotypes were significantly heritable, including all seven community phenotypes and 93% (273/283) of single-taxon phenotypes [likelihood ratio test; false discovery rate (FDR) threshold = 0.1; Fig. 2, A and B; figs. S9 and S10; and table S6]. Heritability was not limited to prevalent taxa because 95% of the 744 presence/absence phenotypes were also

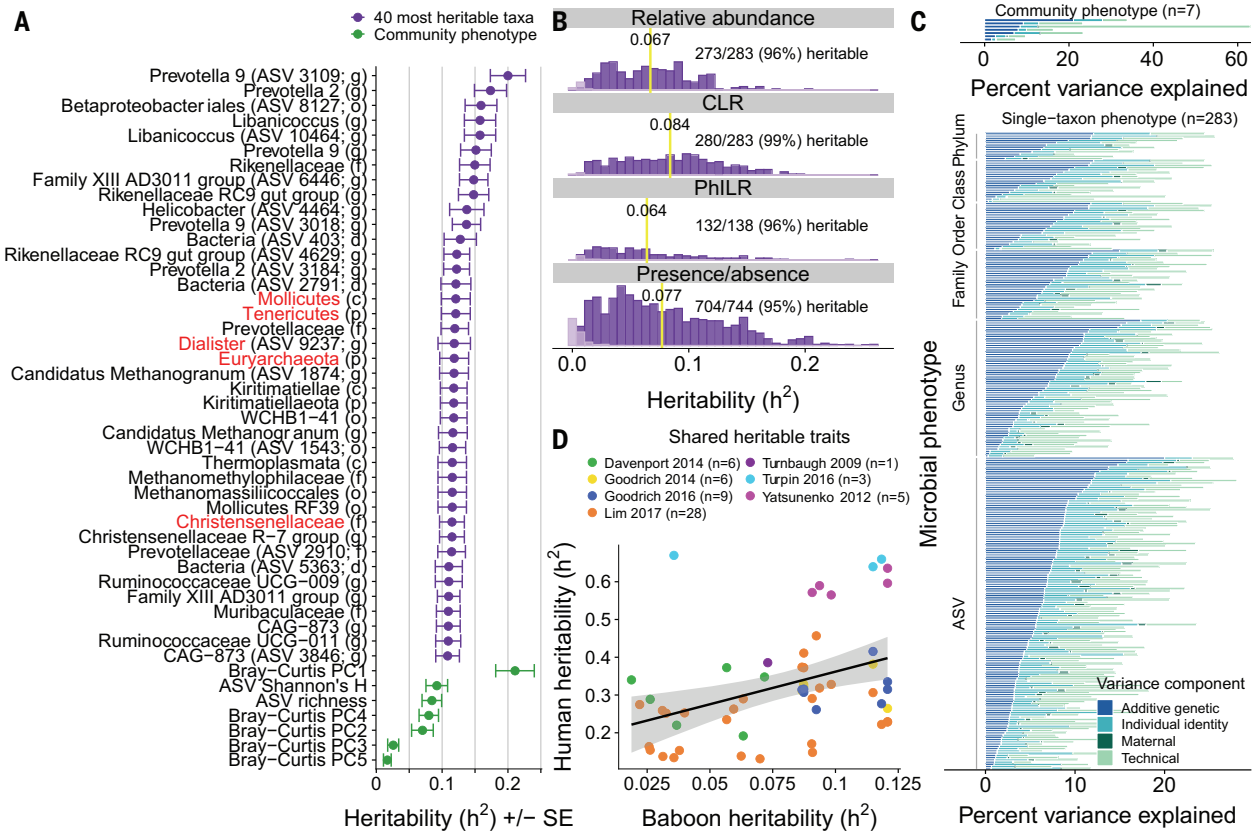


Fig. 2. Most microbiome phenotypes are heritable. (**A**) Heritability estimates for the 40 most heritable single-taxon phenotypes and all seven community phenotypes. Red text indicates taxa that are also heritable in humans (*1*, *2*, *4*–*6*). (**B**) Heritability estimates were robust across data transformations. Dark purple bars show significantly heritable phenotypes; thin yellow bars indicate mean heritability. (**C**) Additive genetic variance explained significantly more variance in microbiome phenotypes than host identity or maternal effects. The *y*-axis is ordered by taxonomic level and h^2, as given in table S7. (**D**) For the 32 microbial taxa heritable in our study (*x*-axis) and at least one human study [*y*-axis; (*1*, *2*, *4*–*6*, *32*, *33*)], h^2 was correlated between baboons and humans (Pearson's R = 0.52, P = 0.002).

significantly heritable, some of which were found in only 10% of samples (Fig. 2B; fig. S9, A to C; and table S7). However, more prevalent taxa tended to have higher h^2 (Pearson's R = 0.28, $P = 2.3 \times 10^{-15}$; fig. S9D). The proportion of significantly heritable single-taxon phenotypes was robust across phylogenetically and compositionally aware data transformations [phylogenetic isometric log-ratio (PhILR) transformation = 96% heritable; centered log-ratio (CLR) transformation = 99% heritable; FDR threshold = 0.1; Fig. 2B and tables S8 and S9]. Heritability estimates were correlated between single-taxon phenotypes and CLR-transformed single-taxon phenotypes (Pearson's R = 0.82, $P = 2.3 \times 10^{-69}$), and between single-taxon phenotypes and presence/absence phenotypes (Pearson's R = 0.68, $P = 3.2 \times 10^{-29}$; fig. S9, E and F).

The most heritable phenotype among the single-taxon and community phenotypes was the first PC of a principal coordinates analysis of Bray-Curtis dissimilarities, which captures a global summary of variation in the baboon gut microbiome [h^2 = 0.21; $P = 5.7 \times 10^{-15}$; Fig. 2A; Bray-Curtis PC1 explained 19% of the variance in microbiome composition overall (*17*)]. More closely related ASVs tended to have similar h^2 (Moran's I = 0.0996, P = 0.001; and Pagel's lambda = 0.73, P = 0.001), especially ASVs belonging to the families Prevotellaceae, Lachnospiraceae, and Ruminococcaceae (local Moran's I, $P < 0.05$; fig. S11), suggesting a phylogenetic signal in microbe heritability.

Although h^2 for single-taxon and community phenotypes tended to be low to modest (mean h^2 among the 280 significant phenotypes = 0.068; range = 0.008 to 0.21; Fig. 2C), heritability values for presence/absence traits were significantly higher (paired t test $P = 2.2 \times 10^{-27}$; mean h^2 = 0.077, maximum h^2 = 0.26; Fig. 2B and fig. S9, A, B, C, and F), as were heritability estimates from compositionally aware abundance transformations (paired t test $P = 2.7 \times 10^{-27}$; mean h^2 = 0.084, maximum h^2 = 0.20; Fig. 2B and fig. S9E). Overall, these values are similar to the heritability of social behavioral traits in nonhuman primates (fig. S12 and table S10) and traits with strong social components in humans (*28*, *29*) but exceed most available estimates for fitness in animal populations (*30*).

Across traits, host genotype explained more variance than host identity (paired t test $P = 2.4 \times 10^{-27}$) or maternal effects (paired t test $P = 5.5 \times 10^{-86}$). These results suggest that host genotype is more important in creating familial similarity in baboon microbiome composition than matrilines, even though matrilines form the core kinship units in baboon societies (Fig. 2C and fig. S9, B and C). Further, we found no evidence that microbial transmission between relatives or assortative mating inflates h^2. Parent pairs did not have more similar microbiome composition than non-parent female-male pairs, as would be expected under assortative mating by microbiome composition (Mantel test r = 0.004, P = 0.22; fig. S7B). In addition, accounting for grooming-based social interaction networks (in the subset of models where such networks could be robustly estimated; n = 500) decreased h^2 by only 0.0051 on average and did not significantly improve any models (fig. S7C and table S11). The weak effects of social networks on microbiome similarity were likely due to the longitudinal nature of this dataset. In our population, social effects on microbiome composition are strongest between samples collected in the same month, and samples from social partners separated by long time periods are not especially similar (*18*, *31*).

Humans and baboons share heritable taxa

We next investigated whether similar gut microbiome taxa are influenced by host genotype across baboons and humans, which would suggest that trait heritability in the microbiome is conserved. Heritability estimates were correlated for the 32 microbiome taxa found to be heritable in both our study and in at least one of seven human datasets from five studies [n = 3511 aggregate sample size in total; (*1*, *2*, *4*–*6*, *17*, *32*, *33*)] despite substantial methodological variation in data collection and methods for h^2 estimation (Pearson's R = 0.52, P = 0.002; results are consistent using a linear mixed model that controls for study: b = 0.91, P = 0.014; Fig. 2D and table S12). Shared, heritable taxa include the family Christensenellaceae, one of the most consistently heritable phenotypes in humans [baboon: single-taxon h^2 = 0.12; presence/absence h^2 = 0.20; humans: 0.31 to 0.64; Fig. 2A and fig. S9A; (*1*, *2*, *4*, *6*, *7*, *9*, *34*)]. In contrast to a previous study in humans, heritable taxa did not co-occur more frequently than expected within hosts (*1*). However, more heritable taxa did exhibit higher connectivity in taxon co-occurrence networks (Pearson's R = 0.58, P = 0.006; fig. S13).

Year, season, and host age modify heritability estimates

To understand why microbiome h^2 estimates often vary across studies (*1*, *2*, *4*–*6*), we then investigated social and environmental factors that systematically influence trait heritability. Here, we focused on a refined set of 100 collapsed phenotypes, including the seven community phenotypes and 93 single-taxon phenotypes in which we collapsed phylogenetically nested taxa to the lowest taxonomic level [as described previously (*1*, *2*, *6*); fig. S14 and table S13]. We found that host traits and environmental conditions had substantial effects on h^2. Across years, h^2 calculated for a single year can differ by up to 0.24 compared with h^2 calculated using all years (n = 15 most heritable collapsed phenotypes, evaluated in years with at least 150 individuals and 1000 samples; table S11 and Fig. 3A). For example, although h^2 for the *Christensenellaceae* R-7 group (the collapsed phenotype for Christensenellaceae) was 0.12 across all years, its

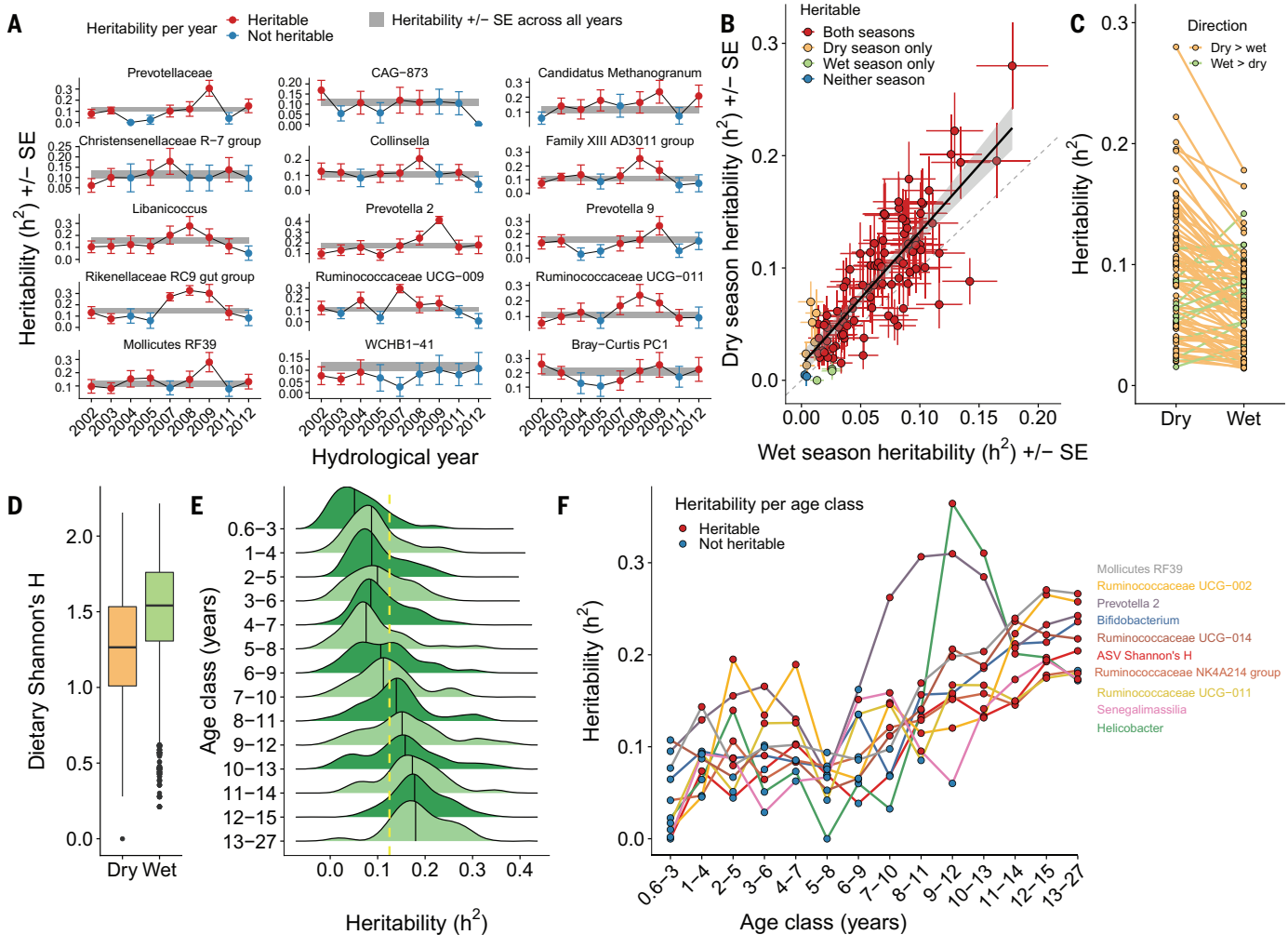


Fig. 3. Heritability estimates are affected by year, season, and host age. (**A**) Heritability estimates varied across years. Panels show h^2 for the 15 most heritable collapsed phenotypes in years with sufficient sample size (>150 baboons and >1000 total samples). (**B**) Heritability estimates for all 100 collapsed phenotypes were highly correlated between seasons (black line; R = 0.83, P = 4.7 × 10^{-27}). Dashed line indicates x = y. (**C**) Heritability estimates for collapsed phenotypes were higher in the dry season than in the wet season (n = 89 taxa heritable in both seasons; paired t test P = 4.4 × 10^{-12}). (**D**) Dietary diversity was higher in the wet season (paired t test P = 4.2 × 10^{-5}). (**E**) Heritability increased with age for 29/100 collapsed phenotypes. Each density plot represents the observed h^2 for these 29 collapsed phenotypes across 3-year sliding age classes. The dashed yellow line indicates mean h^2 across all age classes. (**F**) Heritability estimates per age window for the 10 collapsed phenotypes with the steepest increase in h^2 with host age.

annual h^2 estimates ranged from 0.06 (in 2002) to 0.18 (in 2007).

Within years, we also observed systematic effects of wet/dry seasonal dynamics on microbiome heritability. On the basis of the 89 collapsed phenotypes that were heritable in both dry and wet season samples (estimated separately; red points in Fig. 3B), we found that h^2 was, on average, 48% higher in the dry season than in the wet season (paired t test $P = 4.4 \times 10^{-12}$; Fig. 3C) even though h^2 estimates were strongly correlated between seasons (Pearson's $R = 0.81$, $P = 3.5 \times 10^{-22}$; Fig. 3B). These seasonal differences in h^2 may be explained by seasonal changes in phenotypic variance (V_p): Weather in Amboseli is highly variable during the 7-month wet season, with periods of intense rain followed by several weeks with little or no rain, compared with the near invariant dry season. In support of this, V_p for microbiome phenotypes was higher in wet versus dry seasons (paired t test $P = 4.2 \times 10^{-5}$; fig. S15A). Baboons also consume a greater diversity and evenness of food types in the wet season compared with the dry season (linear mixed model; $b = 0.15$, $P = 5.9 \times 10^{-114}$; Fig. 3D). Although diet composition and rainfall per se are included in our models, individuals who eat diverse diets may also experience season-dependent environmental variation that our model does not capture. To test this hypothesis, we stratified the data by dietary diversity and found that heritability estimates were higher in the low–diet-diversity dataset (paired t test $P = 1.0 \times 10^{-11}$; fig. S15, B and C; 72% of samples in the high–diet-diversity dataset were collected in the wet season).

Host characteristics such as age can also modify trait heritability (*13*, *35*). Indeed, we found that for many of the microbiome phenotypes, h^2 increased with host age. When we stratified the 100 collapsed phenotypes into overlapping 3-year age classes of similar sample size (table S14), we found that h^2 changed significantly with age for 32% of phenotypes, and 91% of these phenotypes (29 of 32) resulted in higher h^2 in older animals (linear models $P < 0.05$; Fig. 3E), with a total increase in h^2 of up to 0.24 (Fig. 3F). This observation is driven by both increasing genetic contributions to gut microbiome variation with host age (i.e., increased V_A; linear mixed model, $b = 1.7 \times 10^{-5}$, $P = 0.0085$) and decreasing contributions from residual environmental variance (i.e., decreased V_R; linear mixed model, $b = -2.9 \times 10^{-5}$, $P = 1.5 \times 10^{-4}$). Older baboons ate less diverse diets than younger baboons regardless of season (linear mixed model; effect of age on diet diversity in the wet season: $b = -1.6 \times 10^{-2}$, $P = 1.6 \times 10^{-24}$; effect of age on diet diversity in the dry season: $b = -1.2 \times 10^{-2}$, $P = 2.0 \times 10^{-11}$; fig. S16, A and B). In addition, females exhibited reduced social partner diversity with age (linear mixed model; $b = -0.35$, $P = 1.4 \times 10^{-19}$; fig. S16, C and D). Moreover, microbiome diversity (Shannon's H) also decreased slightly with age (linear mixed model; $b = -0.0063$, $P = 0.024$; fig. S16E) and its h^2 exhibited the sixth strongest increase with age (linear model; $b = 0.013$, $P = 2.5 \times 10^{-5}$; Fig. 3F). A possible explanation for this pattern is behavioral canalization that is not fully captured by the diet composition effects in our models, whereby older baboons increase in behavioral conservatism with age.

Longitudinal sampling affects heritability estimation

Together, our results qualitatively differ from similar research on humans: Instead of a very small number of heritable microbiome phenotypes, we found nearly universal heritability (*1*, *2*, *4*, *6*, *7*, *9*). Further, we explain systematic variation in h^2 on the basis of temporal, environmental, and individual characteristics. These findings suggest that deep, longitudinal

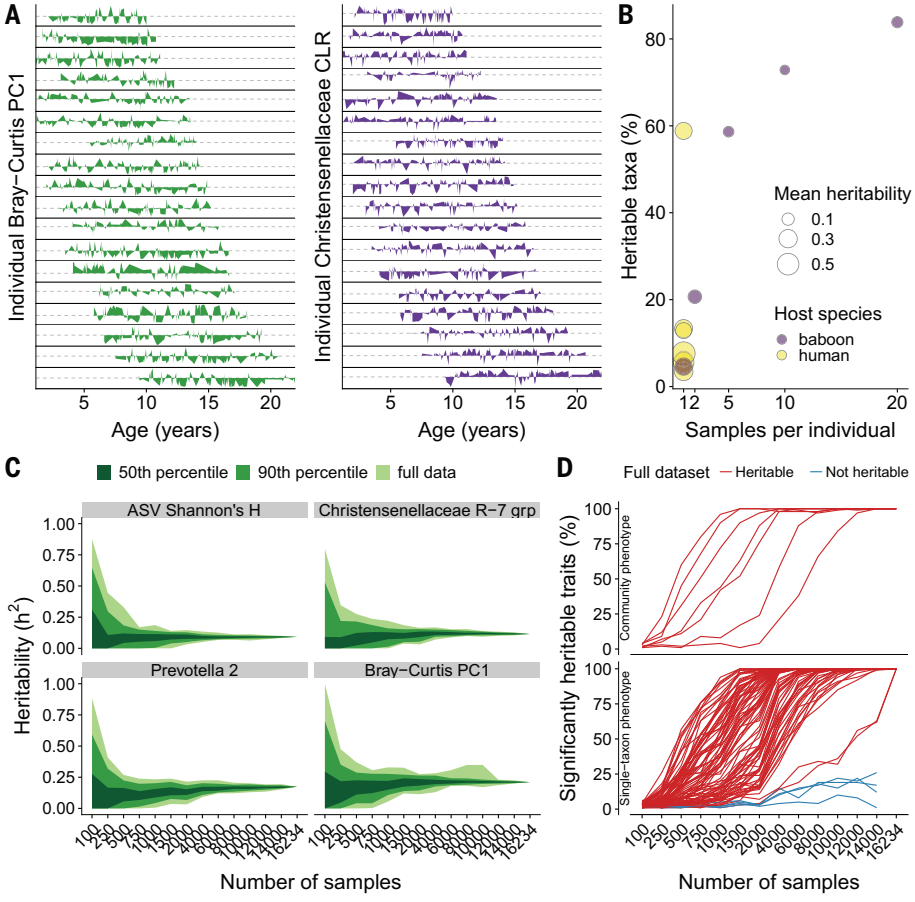


Fig. 4. Microbiome phenotypes are dynamic and sampling design affects heritability estimates. (**A**) Highly heritable microbiome phenotypes fluctuate in abundance (*y*-axis) in individual hosts over time (*x*-axis), as shown by Bray-Curtis PC1 and Christensenellaceae. Each row represents a baboon with >100 samples. (**B**) Longitudinal sampling improves the detection of heritable phenotypes. Purple circles indicate the percent of significantly heritable taxa in our dataset when subset from 1 to 20 samples per individual. Yellow circles are the percentage of significantly heritable microbiome phenotypes in seven human datasets from five studies (*1*, *2*, *4*–*6*, *32*, *33*); note that the plotted points from (*33*) and (*32*) show nearly perfect overlap. (**C**) Heritability varies widely at lower sampling depths, even for highly heritable phenotypes (*x*-axis). The range of h^2 from 100 random subsets at each sampling depth is shown on the *y*-axis. (**D**) The percentage of significantly heritable traits rises with increasing sample size. Plot shows the percentage of models (out of 100 subsamples) that were improved by adding pedigree information. Each line represents one of the 100 collapsed phenotypes.

sampling is required to accurately characterize microbiome heritability and account for potentially extensive temporal variation [Fig. 4A; note that trait heritability was not correlated with its variability across the life course (coefficient of variation in abundance): $R = -0.14$, $P = 0.15$ for $n = 357$ individuals with >10 samples; fig. S17].

In support of the importance of deep longitudinal sampling, we found that sample size and longitudinal sampling affected both our ability to detect heritable microbiome phenotypes and the heritability estimates themselves. Specifically, if we simulated cross-sectional data by randomly subsetting our collapsed phenotype dataset to one sample per individual ($n = 585$ samples, repeated 100 times), we found that <5% of phenotypes were significantly heritable, on average (mean = 4.6%; 95% confidence interval = 3.1 to 6.1%; Fig. 4B and table S12). This proportion is comparable to that described in most human studies but increases with more longitudinal samples per individual (Fig. 4B). Further, when we randomly subsetted our collapsed phenotype dataset to 1000 samples (including repeated samples for some individuals), h^2 estimates fell outside their standard error in the full dataset in an average of 74% of cases (across 100 random subsamples; Fig. 4C, fig. S18, and table S15). Increasing the subset size to 10,000 samples dropped this percentage to 11% (Fig. 4C and fig. S18) and increased the number of significantly heritable phenotypes. With 1000 samples, heritable microbiome phenotypes detected in the full dataset were significantly heritable in only 38% of 100 subsamples, on average (Fig. 4D), but at 10,000 samples, this concordance rose to 98%.

Conclusions

Nearly all gut microbiome taxa are heritable in baboons, including both prevalent and rare taxa. Although the magnitude of these heritability estimates is typically small, some traits exhibit $h^2 > 0.15$ ($n = 59/744$ presence/absence phenotypes; 6/283 single-taxon phenotypes; 1/7 community phenotypes). The universal role played by host genetic variation in our dataset contrasts with previous work in humans finding few heritable taxa (*1*, *2*, *4*, *6*, *7*). These datasets may have had limited power because all human studies to date have been cross-sectional and may have lacked data on key environmental variables that mask or modify heritability levels (*1*, *2*, *4*, *6*, *7*). Further, h^2 for traits detected in both baboons and humans are correlated (Fig. 2D), suggesting that traits with low h^2 in baboons may also be heritable but have gone undetected in humans.

Our findings do, however, agree with the observation that environmental effects on gut microbiome variation are larger than additive genetic effects (*7*). Future work will help to refine our understanding of these environmental influences, including whether they mediate and/or interact with the effects of host genotype. Additionally, as 16*S* rRNA-sequencing data have limited resolution, large-scale metagenomic data will be important for understanding whether individual microbial strains or gene content are also heritable and, perhaps more interestingly, whether microbial genotype affects host heritability. Our work argues for a qualitative change in perspective, from a microbial landscape largely unaffected by host genotype to one in which host genetics play a consistent and sometimes appreciable role. These qualities imply that microbiome traits are therefore visible to natural selection on the host genome.

REFERENCES AND NOTES

1. J. K. Goodrich *et al*., *Cell* **159**, 789–799 (2014).
2. J. K. Goodrich *et al*., *Cell Host Microbe* **19**, 731–743 (2016).
3. R. Blekhman *et al*., *Genome Biol.* **16**, 191 (2015).
4. E. R. Davenport *et al*., *PLOS ONE* **10**, e0140301 (2015).
5. M. Y. Lim *et al*., *Gut* **66**, 1031–1038 (2017).
6. W. Turpin *et al*., *Nat. Genet.* **48**, 1413–1417 (2016).
7. D. Rothschild *et al*., *Nature* **555**, 210–215 (2018).
8. A. Kurilshikov *et al*., Large-scale association analyses identify host factors influencing human gut microbiome composition. bioRxiv 173724 [Preprint]. 16 December 2020. https://doi.org/10.1101/2020.06.26.173724.
9. J. K. Goodrich, E. R. Davenport, A. G. Clark, R. E. Ley, *Annu. Rev. Genet.* **51**, 413–433 (2017).
10. S. Lax *et al*., *Science* **345**, 1048–1052 (2014).
11. B. H. Schlomann, R. Parthasarathy, *Curr. Opin. Microbiol.* **50**, 56–63 (2019).
12. P. M. Visscher, W. G. Hill, N. R. Wray, *Nat. Rev. Genet.* **9**, 255–266 (2008).
13. T. Ge, C.-Y. Chen, B. M. Neale, M. R. Sabuncu, J. W. Smoller, *PLOS Genet.* **13**, e1006711 (2017).
14. S. C. Alberts, J. Altmann, "The Amboseli Baboon Research Project: 40 years of continuity and change," in *Long-Term Field Studies of Primates*, P. M. Kappeler, D. P. Watts, Eds. (Springer, 2012), pp. 261–287.
15. A. M. Bronikowski *et al*., *Sci. Data* **3**, 160006 (2016).
16. A. C. Markham, V. Guttal, S. C. Alberts, J. Altmann, *Behav. Ecol. Sociobiol.* **67**, 875–884 (2013).
17. Materials and methods are available as supplementary materials.
18. J. Tung *et al*., *eLife* **4**, e05224 (2015).
19. L. E. Grieneisen, J. Livermore, S. Alberts, J. Tung, E. A. Archie, *Integr. Comp. Biol.* **57**, 770–785 (2017).
20. K. Berer *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10719–10724 (2017).
21. A. E. Mann *et al*., *ISME J.* **14**, 609–622 (2020).
22. W. A. Walters *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7368–7373 (2018).
23. A. R. Gilmour, "ASREML for testing fixed effects and estimating multiple trait variance components," in Proceedings of the Association for the Advancement of Animal Breeding and Genetics (AABG, 1997), vol. 12, pp. 386–390.
24. L. E. B. Kruuk, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 873–890 (2004).
25. A. J. Wilson *et al*., *J. Anim. Ecol.* **79**, 13–26 (2010).
26. A. J. Wilson, *J. Evol. Biol.* **21**, 647–650 (2008).
27. L. E. B. Kruuk, J. D. Hadfield, *J. Evol. Biol.* **20**, 1890–1903 (2007).
28. N. Barban *et al*., *Nat. Genet.* **48**, 1462–1472 (2016).
29. D. Cesarini, P. M. Visscher, *NPJ Sci. Learn.* **2**, 4 (2017).
30. A. P. Hendry, D. J. Schoen, M. E. Wolak, J. M. Reid, *Annu. Rev. Ecol. Evol. Syst.* **49**, 457–476 (2018).
31. T. Ren, L. E. Grieneisen, S. C. Alberts, E. A. Archie, M. Wu, *Environ. Microbiol.* **18**, 1312–1325 (2016).
32. T. Yatsunenko *et al*., *Nature* **486**, 222–227 (2012).
33. P. J. Turnbaugh *et al*., *Nature* **457**, 480–484 (2009).
34. J. L. Waters, R. E. Ley, *BMC Biol.* **17**, 83 (2019).
35. R. Plomin, I. J. Deary, *Mol. Psychiatry* **20**, 98–108 (2015).
36. A. Gonzalez *et al*., *Nat. Methods* **15**, 796–798 (2018).
37. L. Grieneisen *et al*., Data for: Gut microbiome heritability is nearly universal but environmentally contingent, Zenodo (2021); https://doi.org/10.5281/zenodo.4662081.

ACKNOWLEDGMENTS

We thank J. Altmann for her stewardship of the Amboseli Baboon Research Project (ABRP) and for collecting the fecal samples used in this manuscript (see complete ABRP acknowledgments at https://amboselibaboons.nd.edu/acknowledgements/); K. Pinc for ABRP database design; T. Voyles, A. Dumaine, Y. Zhang, M. Rao, T. Vilgalys, A. Lea, N. Snyder-Mackler, P. Durst, J. Zussman, G. Chavez, S. Mukherjee, and R. Debray for fecal sample processing; and three anonymous reviewers for their constructive comments. We also thank the Kenya Wildlife Service, the National Council for Science, Technology, and Innovation, and the National Environment Management Authority for permission to conduct research and collect biological samples in Kenya. We thank the University of Nairobi, Institute of Primate Research, National Museums of Kenya, the Amboseli-Longido pastoralist communities, the Enduimet Wildlife Management Area, Ker & Downey Safaris, Air Kenya, and Safarilink for support in Kenya. The research in this study was approved by the institutional animal care and use committees at Duke University, Princeton University, and the University of Notre Dame, and adhered to the laws and guidelines of the Kenyan government. **Funding:** This work was directly supported by NSF DEB 1840223 (E.A.A., J.A.G.), NIH R21 AG055777 (E.A.A., R.B.), NIH R01 AG053330 (E.A.A.), and NIGMS R35 GM128716 (R.B.). We also acknowledge support from the University of Minnesota Grand Challenges in Biology Postdoctoral Fellowship (to L.G.), the Duke University Population Research Institute P2C-HD065563 (pilot award to J.T.), and Notre Dame's Eck Institute for Global Health (E.A.A.) and Environmental Change Initiative (E.A.A.). Since 2000, ABRP has been supported by NSF and NIH, including IOS 1456832 (S.C.A.), IOS 1053461 (E.A.A.), DEB 1405308 (J.T.), IOS 0919200 (S.C.A.), DEB 0846286 (S.C.A.), DEB 0846532 (S.C.A.), IBN 0322781 (S.C.A.), IBN 0322613 (S.C.A.), BCS 0323553 (S.C.A.), BCS 0323596 (S.C.A.), P01AG031719 (S.C.A.), R21AG049936 (J.T., S.C.A.), R03AG045459 (J.T., S.C.A.), R01AG034513 (S.C.A.), R01HD088558 (J.T.), and P30AG024361 (S.C.A.). We also thank Princeton University, the Chicago Zoological Society, the Max Planck Institute for Demographic Research, the L.S.B. Leakey Foundation, and the National Geographic Society. **Author contributions:** L.G., R.B., E.A.A., L.B.B., J.A.G., and J.T. designed the research; S.C.A., E.A.A., J.T., R.B., L.B.B., M.D., T.J.G., V.Y., D.J., N.G., J.B.G., N.H.L., L.R.G., T.L.W., R.S.M., J.K.W., L.S., and J.A.G. produced the data; L.G., J.R.B., M.D., T.J.G., and D.J. analyzed the data; L.G., R.B., J.T., and E.A.A. wrote the manuscript with important contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Our data and code are publicly available, but the original biological and DNA samples cannot be shared due to restrictions on third-party sharing for CITES-regulated samples exported from Kenya. The fecal samples and DNA extracts used in this study are subject to material transfer agreements between Duke University and the University of Notre Dame in the United States and the Kenya Wildlife Service in Kenya. These biological materials are maintained at J.T.'s laboratory at Duke University and E.A.A.'s laboratory at the University of Notre Dame and can only be shared with third parties with prior written authorization from the Kenya Wildlife Service. 16*S* rRNA gene sequences are deposited on EBI-ENA (project ERP119849) and Qiita [study 12949; (*36*)]. Analyzed data and code are available on Zenodo (*37*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/181/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S15
References (*38*–*129*)
MDAR Reproducibility Checklist

18 February 2020; resubmitted 25 January 2021
Accepted 17 May 2021
10.1126/science.aba5483

ICE PHYSICS

Elastic ice microfibers

Peizhen Xu[1]†, Bowen Cui[1]†, Yeqiang Bu[2], Hongtao Wang[2], Xin Guo[1]*, Pan Wang[1], Y. Ron Shen[3], Limin Tong[1,4]*

Ice is known to be a rigid and brittle crystal that fractures when deformed. We demonstrate that ice grown as single-crystal ice microfibers (IMFs) with diameters ranging from 10 micrometers to less than 800 nanometers is highly elastic. Under cryotemperature, we could reversibly bend the IMFs up to a maximum strain of 10.9%, which approaches the theoretical elastic limit. We also observed a pressure-induced phase transition of ice from Ih to II on the compressive side of sharply bent IMFs. The high optical quality allows for low-loss optical waveguiding and whispering-gallery-mode resonance in our IMFs. The discovery of these flexible ice fibers opens opportunities for exploring ice physics and ice-related technology on micro- and nanometer scales.

Ice is one of the most abundant and important crystalline solids on Earth's surface and plays an essential role across a diverse range of topics in chemical physics, life science, geophysics, astronomy, and other disciplines (*1*–*5*). As a result, ice has been extensively studied in the past centuries (*1*, *2*, *6*–*10*). Bulk ice is rigid and fragile, leading to natural phenomena such as avalanches, glacier sliding, and sea ice fragmentation (*1*, *11*, *12*). The maximum elastic strain is experimentally found to be much lower than the theoretical limit of >10% (*13*). This difference is mostly driven by structural imperfection of real ice crystals. Materials in low-dimensional forms, such as nanoscale crystals (*14*–*17*), nanowires (*18*), and microfibers (*19*), can exhibit far superior mechanical properties than their bulk counterparts, because of lower defect density and more uniform stress distribution (*14*). Nano- and microstructures of ice exist naturally in the forms of snowflakes and ice whiskers; these should be expected to also have better mechanical properties than bulk ice. Although low-dimensional ice structures such as whiskers and needles have been grown in laboratories (*8*, *20*, *21*), the focus was on growth and morphology rather than investigating mechanical properties.

We found that ice microfibers (IMFs) have exceptional mechanical properties. By adopting an electric field–enhanced growth method with a growth temperature much lower than what has previously been used (*20*), we succeeded in growing IMFs of high quality and small diameter (down to hundreds of nanometers). We show that the as-grown IMFs are hexagonal single crystals with the hexagonal axis along the core and have very smooth surfaces and excellent uniform cross section over their length. Our IMFs can be bent with a strain up to 10.9%, which is much higher than previously reported maximum strains (*22*, *23*) and is a value approaching the theoretical elastic limit (14 to 16.2%) (*13*, *24*). We conducted Raman spectroscopy measurement on bent IMFs and detected a reversible phase transition between ice Ih and II around a critical strain of 3% at −70°C. To show that the IMFs are of good optical quality, we demonstrate that they can be used to guide visible light with low optical loss and support whispering gallery modes (WGMs) around their circumference.

Growth and morphology of IMFs

The IMFs were fabricated using an electric field–enhanced growth method (*20*, *25*) (sketched in Fig. 1A and fig. S1) (*26*). We grew multiple IMFs

[1]State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China. [2]Center for X-Mechanics, Zhejiang University, Hangzhou 310027, China. [3]Department of Physics, University of California, Berkeley, CA 94720, USA. [4]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China.
*Corresponding author. Email: guoxin@zju.edu.cn (X.G.); phytong@zju.edu.cn (L.T.)
†These authors contributed equally to this work.

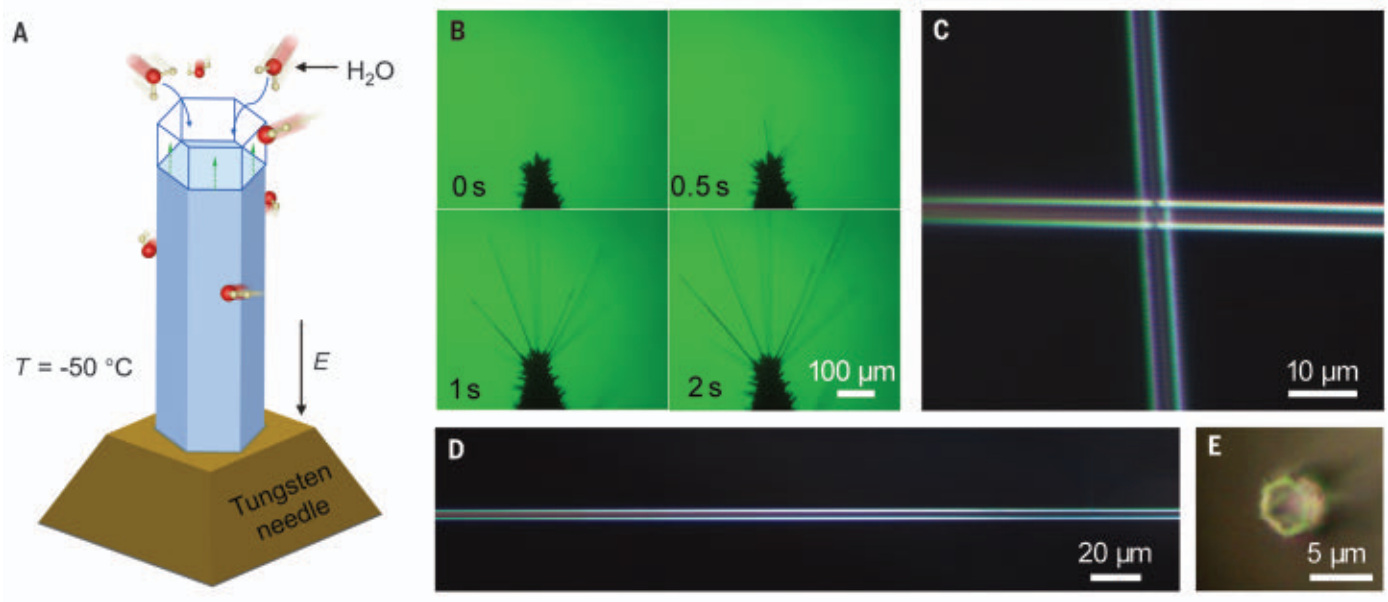


Fig. 1. Growth and optical microscopic morphology of IMFs. (**A**) Schematic illustration of the electrical field–enhanced growth of IMFs. A 2-kV direct-current voltage is applied to a tungsten needle in a cold chamber with a temperature (*T*) of −50°C. The partially shown tip of the tungsten needle serves as the base for growth of a single IMF upward along the electric field (*E*) direction. (**B**) Optical microscopic snapshots of growth in length of multiple IMFs from the tip of a tungsten needle with a rate of ~200 μm/s. (**C**) Optical microscopic image of two crossed 3-μm-diameter IMFs, one on top of the other. (**D**) Optical microscopic image of a long IMF showing a uniform diameter of ~3.3 μm along its length. (**E**) Optical microscopic image of the end face of a 4.3-μm-diameter IMF, showing a hexagonal cross section.

from the tip of a tungsten needle. We modified the method developed earlier by choosing a much lower temperature for growth [−50°C instead of −5°C as in (*20*); fig. S2] (*26*), which reduced the lateral growth rate and enabled the IMF to grow into a fiber with a smaller and more uniform diameter. We applied a 2-kV voltage to the tungsten needle to enhance diffusion of water gas molecules to the tip of the needle and to accelerate the fiber length growth. An IMF longer than 400 μm could be grown in 2 s (Fig. 1B and movie S1). After growth, we transferred the IMFs to a cold stage for characterization in different chambers (figs. S1B and S3) (*26*). One example we show is of two IMFs with a diameter of ~3 μm forming a cross after two successive transfer operations (Fig. 1C). We also imaged a segment of a 3.3-μm-diameter IMF that illustrates the excellent lateral uniformity (Fig. 1D). Our IMFs have a hexagonal cross section (Fig. 1E), similar to what was reported earlier (*20*, *27*).

We also used cryo–transmission electron microscopy (cryo-TEM) and cryo–focused ion beam microscopy (cryo-FIBM) to characterize morphology and crystalline structure of IMFs (fig. S4) (*26*). We carried out the measurements at around −170°C. The diameter of the as-grown IMFs was typically a few micrometers (Fig. 2, A and B) but could be as small as hundreds of nanometers (Fig. 2B). A typical example of our IMFs has a tapered end with a hexagonal cross section (Fig. 2C). The root-mean-square surface roughness estimated from a high-magnification TEM image of a 2.6-μm-diameter IMF (Fig. 2D) was <1 nm (fig. S5) (*26*), and the electron diffraction pattern (Fig. 2E, inset) reveals a <0001> crystal orientation, indicating that the hexagonal *c* axis is aligned along the fiber length.

Elastic bending of IMFs

The bending of a fiber exerts tensile and compressive stresses on it. We used micromanipulation to bend IMFs in a cold chamber (fig. S6 and movies S2 and S3) (*26*). Unlike bulk ice, thin IMFs were highly flexible and could be elastically bent and readily restored to the unbent form. We show a set of snapshots during the bending of a 4.7-μm-diameter IMF to a minimum radius of curvature (R_c) of 63 μm (Fig. 3A, panels A1 to A4) at −70°C. When we released the bending pressure, the IMF returned to its original shape (Fig. 3A, panels A5 and A6, and movie S3). This sequence indicates a reversible elastic deformation during the bending process. To measure the strain induced during the bending of an IMF, we fit the central axis of the bent IMF with a circle and obtained the strain maxima at the inner and outer sides of the IMF (Fig. 3B) (*26*). Owing to the low transverse dimension and high crystal quality of the IMF, we assumed the central axis to be zero-strain and the elastic strain maxima at both sides (i.e., tensile and compressive) to be equal in absolute value. We measured IMFs of different diameters that were bent under two different temperatures, −70° and −150°C, with maximum elastic strains of ~4.6% (in a 4.6-μm-diameter IMF) and ~10.9% (in a 4.4-μm-diameter IMF), respectively; a typical set of results are shown

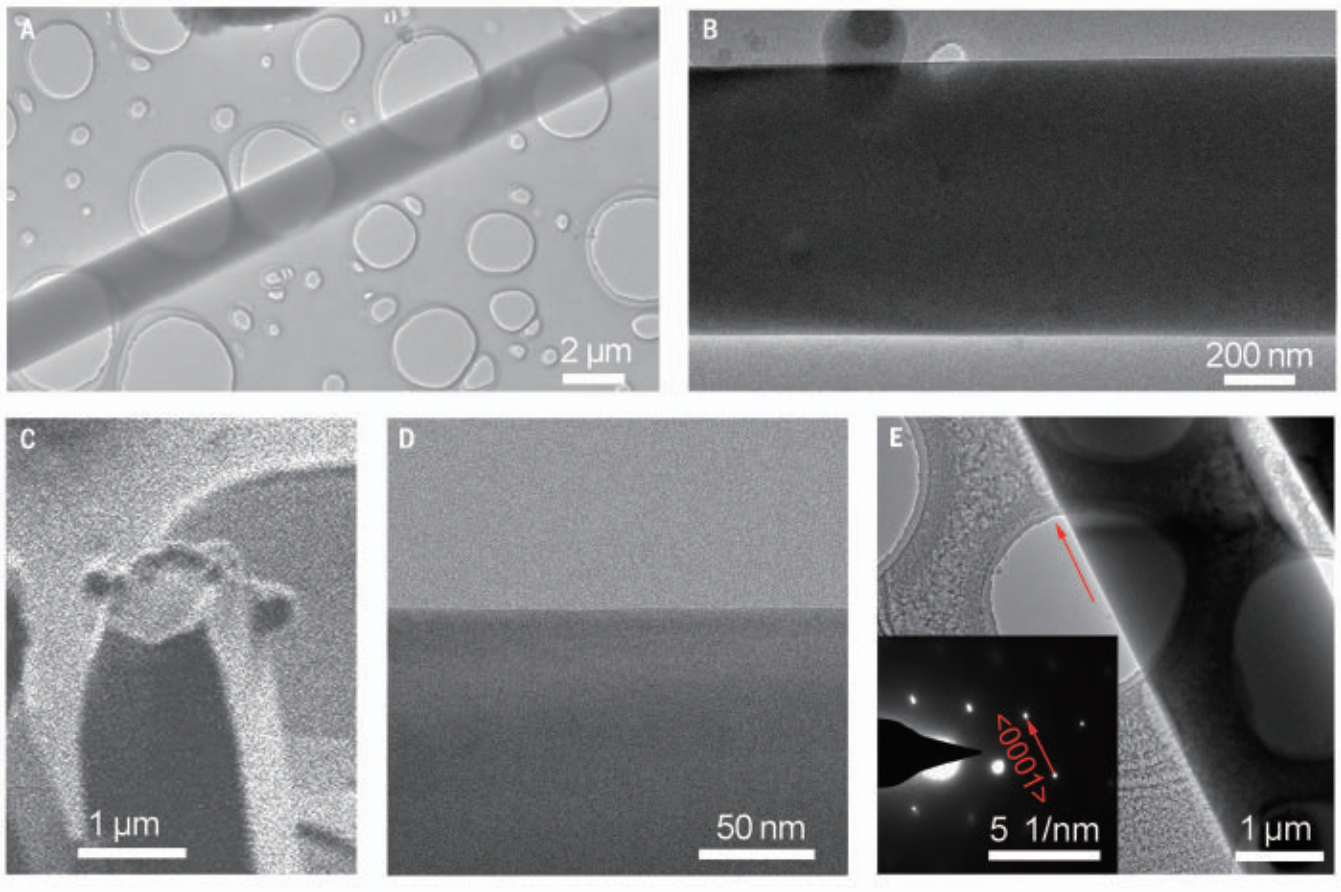


Fig. 2. Cryo-electron or ion microscopy study of IMFs. (**A** and **B**) Cryo-TEM images of IMFs with a diameter of (A) 2.1 μm and (B) 780 nm. (**C**) Cryo-FIBM image of a tapered end face of an as-grown IMF showing a hexagonal cross section. (**D**) Close-up cryo-TEM image of the surface of a 2.6-μm-diameter IMF. (**E**) Cryo-TEM image of a 2.3-μm-diameter IMF. (Inset) Electron diffraction pattern revealing single hexagonal crystalline structure of the IMF with <0001> orientation along the fiber. The red arrows indicate the growth direction of the IMF along the *c* axis. Disks and small speckles in (A) and (E) are holes on the carbon film that covers the copper grid on the TEM sample holder.

in Fig. 3C. These values are much higher than those reported in other forms of ice (e.g., <0.3%) (*12*, *22*, *23*), and the strain maxima measured at −150°C (>10%) are approaching the theoretical elastic limit (*13*, *24*). Even having experienced ultra-large strain, all the bent IMFs returned to their original shape, indicating the absence of the type of creep that ice Ih in other forms generally suffers from under high strain (*23*, *28*, *29*). We attribute this lack of creep to the low transverse dimension and high crystal quality of IMFs and perhaps also to the relatively fast loading-unloading process we used (e.g., strain rate of ~10^{-4} s^{-1} for loading, and 10^{3} s^{-1} for unloading; see movie S3), although we did not observe any strain rate dependence. The maximum attainable strain increased with decreasing temperature, e.g., 10.9% at −150°C versus 4.6% at −70°C (Fig. 3C), similar to the trend reported on bulk ice under much lower strain (*12*) and other micro- or nanometer-scale brittle solids (*14*).

Ice phase transition in sharply bent IMFs

Bulk ice Ih undergoes a phase transition to ice II across a range of temperatures below −20°C under sufficiently high pressure (*30*, *31*). At −70°C, the critical pressure is ~0.2 GPa (*2*, *23*). For our bent IMF (Fig. 4A) with a maximum compressive deformation of 3%, the corresponding maximum stress calculated (using a finite element method) from the Young's modulus of ice Ih along the *c* axis (12.7 GPa) (*32*) was ~0.38 GPa at the inner wall of the IMF (Fig. 4B). It is therefore possible that the inner wall of the bent IMF might have already experienced the phase transition. To check this possibility, we used Raman spectroscopy (fig. S7) to probe the bent IMF (Fig. 4A) in the region where the strain is largest (*26*). We compared the Raman spectra before and after bending (Fig. 4C) and found extra peaks at 158 and 3225 cm^{-1} in the spectrum of the bent IMF. For comparison, we included the Raman spectra for ice Ih and II (*30*, *31*) with mode frequencies corrected by their temperature and pressure dependence (table S1) (*26*). The two extra peaks in the spectrum for the bent IMF coincide well with the two most prominent spectral peaks of ice II (*30*, *31*), providing evidence of the partial presence of ice II in the bent IMF. The peaks are weak because only a small fraction of the illuminated volume of the IMF had a stress above the critical value for the Ih-to-II phase transition. The spectra were reversible on bending and unbending the IMF and were also reproducible for different IMFs with sharp bending. We observed the Raman peak of ice II within tens of seconds after bending, indicating that the

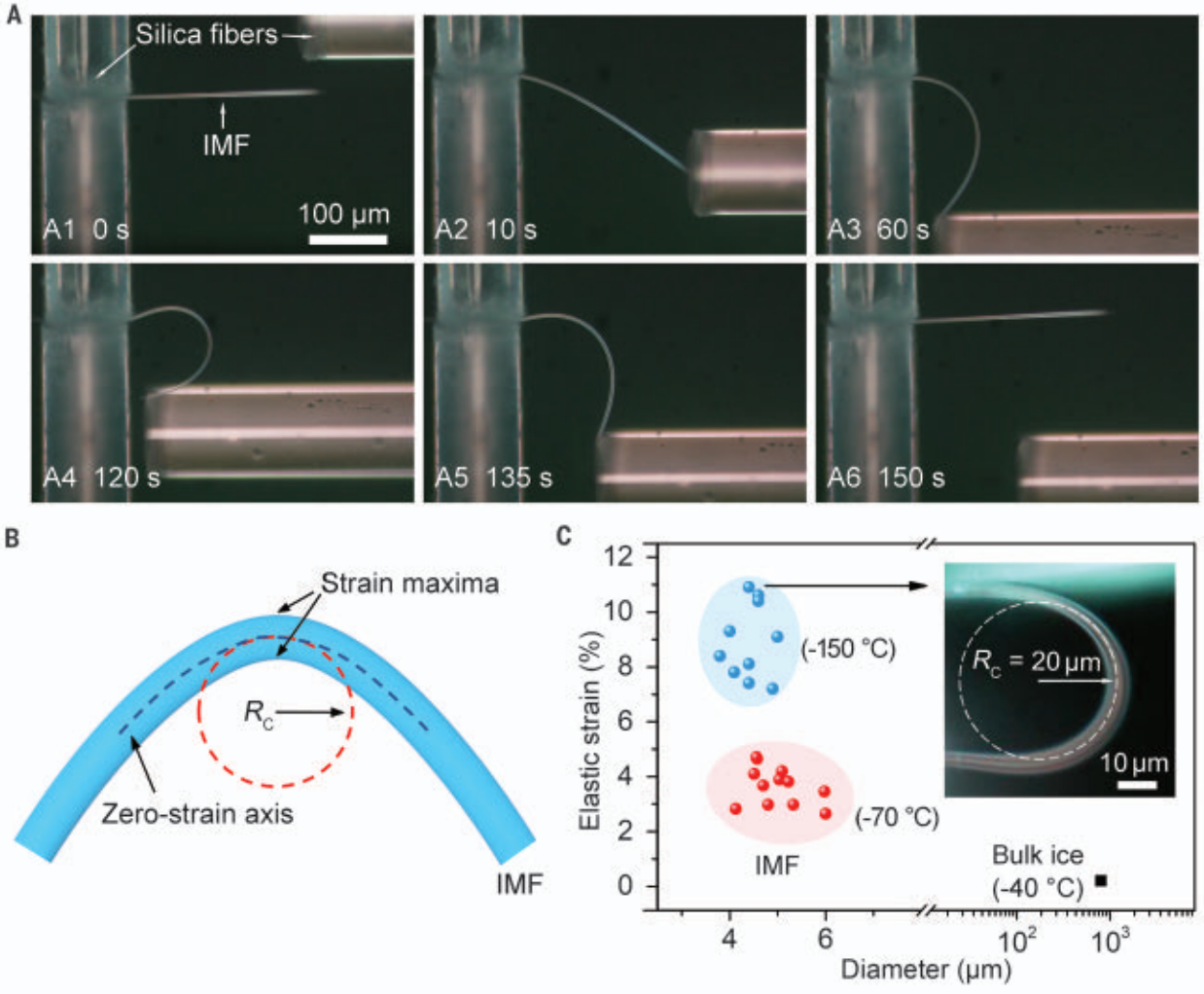


Fig. 3. Elastic bending of individual IMFs. (**A**) Optical microscopic snapshots illustrating how an IMF with a 4.7-μm diameter was elastically bent and unbent. (**B**) Schematic of the bending model of an IMF. (**C**) Maximum elastic strains obtained for various IMFs of different diameters bent under two different temperatures, −70°C (red circles) and −150°C (blue circles), in comparison with that previously reported for bulk ice Ih (black square) (*22*, *23*). The inset shows an optical microscopic image of a bent IMF (4.4 μm in diameter and 20 μm in radius of curvature) at −150°C experiencing a maximum strain of 10.9%.

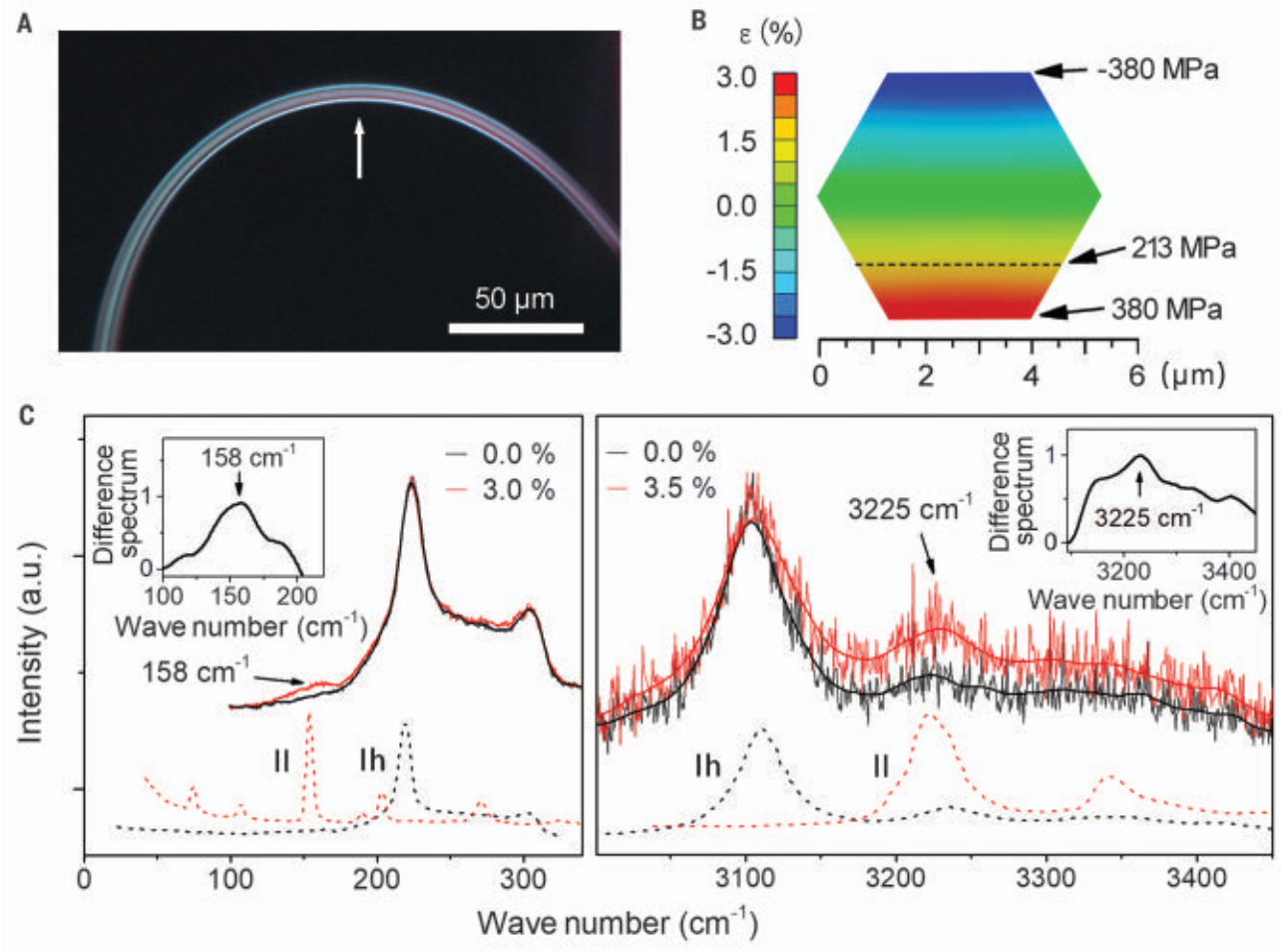


Fig. 4. Phase transition in elastic bent IMFs. (**A**) Optical microscopic image and (**B**) calculated cross-sectional strain distribution of a 5.2-μm-diameter IMF bent to a radius of curvature of ~87 μm. The white arrow in (A) points at the region of maximum compressive strain. The dashed line in (B) indicates the critical pressure of phase transition for bulk ice from phase Ih to II (*30*, *31*). (**C**) Raman spectra of the IMF in (A), before (black line) and after bending (red line), obtained with probe light focused on the maximum strain region. New spectral features at ~158 and 3225 cm^{-1}, characteristic of ice II, appear in the spectrum for the bent IMF, indicating the presence of ice II. The insets show the difference spectra of the Raman spectra with and without strain around 158 and 3225 cm^{-1}. For reference, the Raman spectra of ice Ih (black dashed line) and II (red dashed line), adapted from (*31*), are shown. a.u., arbitrary units.

Ih-to-II transition of micrometer-size ice occurs rapidly, as suggested by Schulson and Fortt (*33*).

Optical characterization of IMFs

The excellent diameter uniformity and surface smoothness allow us to use IMFs to guide light at low temperature. By evanescently coupling visible light into one end of an IMF (Fig. 5A), we readily observed waveguiding of light of different wavelengths along the fiber length (Fig. 5B). From analysis of the position-dependent scattered light intensity along the IMF (Fig. 5C), we obtained a waveguiding loss of ~0.2 dB/cm at 525-nm wavelength (*26*), which is of the same order as those of the state-of-the-art on-chip waveguides (*34*). The measured loss is much higher than the absorption loss in the IMF because the absorption coefficient for ice (*35*) is known to be much smaller than that of typical waveguide materials such as Si_3N_4 and silica (fig. S8) (*26*, *36*, *37*). The waveguide loss we observed must have mainly come from scattering loss, and given that surface roughness of our IMFs was at the subnanometer level, we attribute the scattering loss to structural defects in IMFs. In the bent region of an IMF, the field of the guided light is redistributed and considerably enhanced at the outer part of the bend (fig. S9). Thus, IMFs have the potential to serve as flexible waveguides with much lower loss than other optical waveguides within the visible spectral range and to be used for optical applications such as high-sensitivity microfiber-based optical sensing at low temperature (*38*). To explore the possibility of exciting WGMs on the circumference of IMFs, we focused a white light beam perpendicularly on an IMF and collected the scattered signal with an objective in the dark-field configuration (fig. S10) (*26*). We found a typical spectrum of a transverse electric (TE) WGM with a quality factor of ~60 and a free spectral range of ~20.9 nm at around 605-nm wavelength, which agrees well with the calculated value (~22.0 nm) of the TE_{26} WGM in the 4.4-μm-diameter IMF (*26*) (Fig. 5D). The WGM has enhanced field intensity on the IMF surface (Fig. 5E), making it also sensitive to surface or environmental changes.

Summary and prospects

We have succeeded in fabricating high-quality single-crystalline IMFs with a diameter down to a few hundreds of nanometers and demonstrated that they are highly flexible and can be reversibly deformed by bending to a radius of a few tens of micrometers. We obtained a lattice compression (tension) of 10.9% on the inner (outer) side of a bent IMF, which is more than one order of magnitude larger than those reported in other forms of ice and approaches the theoretical elastic limit (*13*, *24*). In sharply bent IMFs, we observed phase transition between ice Ih and II when the compressive strain

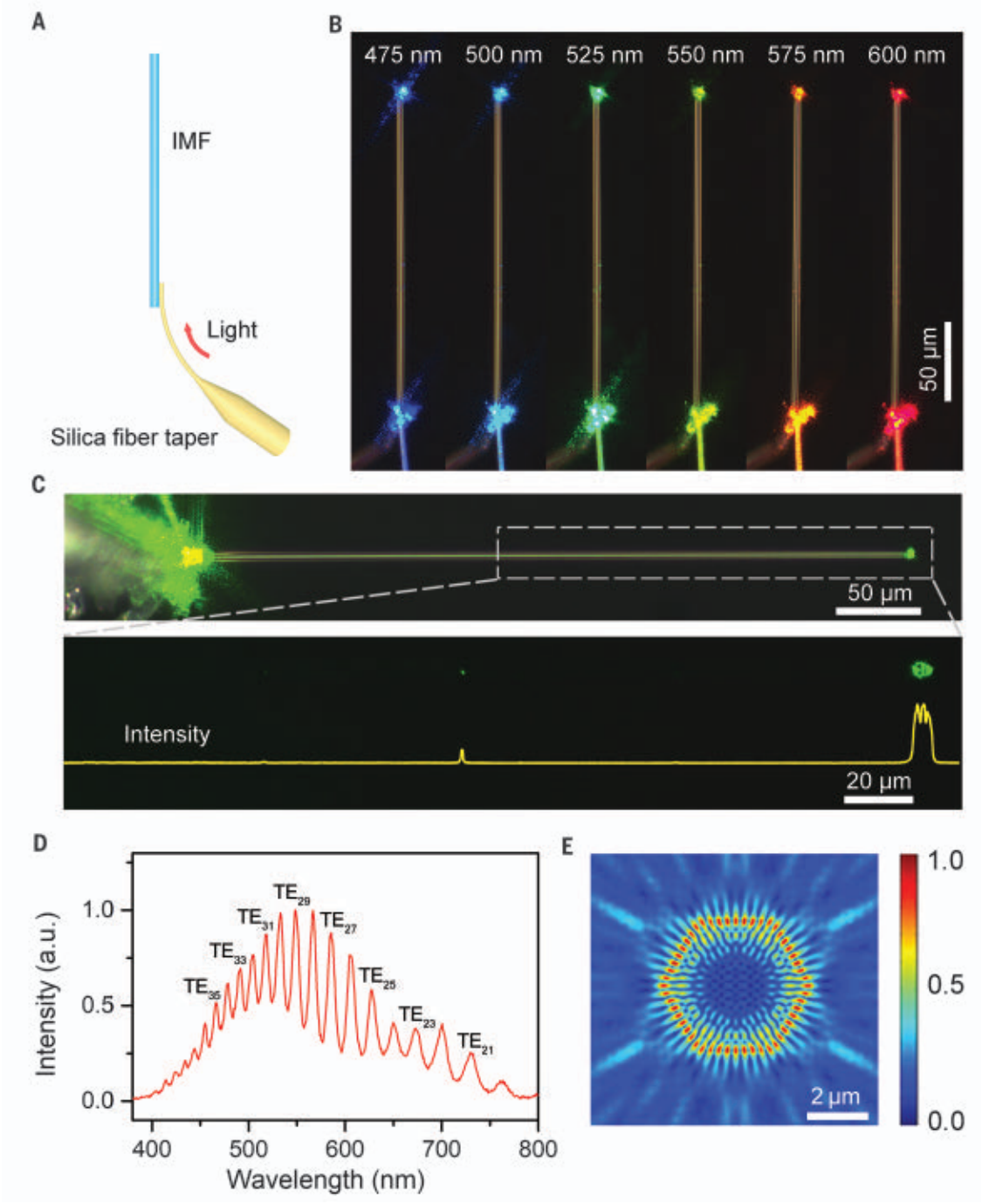


Fig. 5. Optical characterization of IMFs. (**A**) Schematic diagram of launching light into an IMF by the evanescent wave coupling method. (**B**) Microscopic images of an IMF (4.4 μm in diameter and 200 μm in length) guiding light of different wavelengths. (**C**) (Top) Bright-field microscopic image of a 5.4-μm-diameter IMF guiding 525-nm light. The section included in the dashed box was used in collecting scattered light. (Bottom) Dark-field microscopic image of the boxed section of the IMF in the top frame, showing very weak intensity of scattered light from the IMF except for a spot near the center. (**D**) Spectrum of TE WGM of a 4.4-μm-diameter IMF detected from light scattering from the IMF. (**E**) Simulated electric field distribution of the TE_{26} WGM on the IMF in (D).

exceeded 3%, suggesting a simple means to study phase transitions of ice using the bending of IMFs. Higher strains are achievable with sharper bending of IMFs at lower temperature, allowing study of ice-phase transitions of Ih not only to II but also to phases III, V, VI, and IX (*10*). Furthermore, the outer side of a bent IMF should experience a tensile stress (negative pressure) of equal magnitude to the compressive stress on the inner side and can also be used to study ice phases and transition dynamics under large negative pressure (*39*, *40*), which are known to be experimentally challenging (*41*). We have also demonstrated that IMFs can serve as optical waveguides with very low loss and support WGMs in the visible spectrum, both of which are means to enhance light interaction with matter. We could imagine the use of IMFs as low-temperature sensors to study, for example, molecular adsorption on ice, environmental changes, structural variation,

and surface deformation of ice [e.g., surface premelting (*42*)]. In short, the elastic IMFs demonstrated here may offer an alternative platform for exploring ice physics and open previously unexplored opportunities for ice-related technology in various disciplines.

REFERENCES AND NOTES

1. V. F. Petrenko, R. W. Whitworth, *Physics of Ice* (Oxford Univ. Press, 2002).
2. T. Bartels-Rausch *et al*., *Rev. Mod. Phys.* **84**, 885–944 (2012).
3. C. Hoose, O. Möhler, *Atmos. Chem. Phys.* **12**, 9817–9854 (2012).
4. R. Staroszczyk, *Ice Mechanics for Geophysical and Civil Engineering Applications* (Springer Nature, 2019).
5. L. O'Rourke *et al*., *Nature* **586**, 697–701 (2020).
6. J. Kepler, *The Six-Cornered Snowflake* (Oxford Clarendon Press, 1966).
7. P. W. Bridgman, *Proc. Am. Acad. Arts Sci.* **47**, 441–558 (1912).
8. T. W. Zawidzki, H. M. Papee, *Nature* **196**, 568–569 (1962).
9. G. Malenkov, *J. Phys. Condens. Matter* **21**, 283101 (2009).
10. C. G. Salzmann, *J. Chem. Phys.* **150**, 060901 (2019).
11. N. Urabe, T. Iwasaki, A. Yoshitake, *Cold Reg. Sci. Technol.* **3**, 29–37 (1980).
12. J. J. Petrovic, *J. Mater. Sci.* **38**, 1–6 (2003).
13. P. A. Santos-Flórez, C. J. Ruestes, M. de Koning, *J. Chem. Phys.* **149**, 164711 (2018).
14. T. Zhu, J. Li, *Prog. Mater. Sci.* **55**, 710–757 (2010).
15. A. Banerjee *et al*., *Science* **360**, 300–302 (2018).
16. A. Nie *et al*., *Nat. Commun.* **10**, 5533 (2019).
17. C. Dang *et al*., *Science* **371**, 76–78 (2021).
18. G. Brambilla, D. N. Payne, *Nano Lett.* **9**, 831–835 (2009).
19. L. Tong *et al*., *Nature* **426**, 816–819 (2003).
20. K. G. Libbrecht, T. Crosby, M. Swanson, *J. Cryst. Growth* **240**, 241–254 (2002).
21. R. Ma *et al*., *Nature* **577**, 60–63 (2020).
22. E. M. Schulson, *Eng. Fract. Mech.* **68**, 1839–1887 (2001).
23. E. M. Schulson, P. Duval, *Creep and Fracture of Ice* (Cambridge Univ. Press, 2009).
24. Z. M. Jendi, P. Servio, A. D. Rey, *Cryst. Growth Des.* **15**, 5301–5309 (2015).
25. K. G. Libbrecht, V. M. Tanusheva, *Phys. Rev. Lett.* **81**, 176–179 (1998).
26. Materials and methods are available as supplementary materials.
27. K. G. Libbrecht, *Annu. Rev. Mater. Res.* **47**, 271–295 (2017).
28. M. Mellor, D. M. Cole, *Cold Reg. Sci. Technol.* **6**, 207–230 (1983).
29. J. Weertman, *Annu. Rev. Earth Planet. Sci.* **11**, 215–240 (1983).
30. B. Minceva-Sukarova, W. F. Sherman, G. R. Wilkinson, *J. Phys. C Solid State Phys.* **17**, 5833–5850 (1984).
31. A. K. Garg, *Phys. Status Solidi A* **110**, 467–480 (1988).
32. J. J. Neumeier, *J. Phys. Chem. Ref. Data* **47**, 033101 (2018).
33. E. M. Schulson, A. L. Fortt, *J. Geophys. Res. Solid Earth* **117**, B12204 (2012).
34. T. Horikawa, D. Shimura, T. Mogami, *MRS Commun.* **6**, 9–15 (2016).
35. S. G. Warren, R. E. Brandt, *J. Geophys. Res. Atmos.* **113**, D14220 (2008).
36. H. R. Philipp, *J. Electrochem. Soc.* **120**, 295–300 (1973).
37. R. Kitamura, L. Pilon, M. Jonasz, *Appl. Opt.* **46**, 8118–8133 (2007).
38. J. Lou, Y. Wang, L. Tong, *Sensors* **14**, 5823–5844 (2014).
39. Y. Huang *et al*., *Sci. Adv.* **2**, e1501010 (2016).
40. T. Matsui, T. Yagasaki, M. Matsumoto, H. Tanaka, *J. Chem. Phys.* **150**, 041102 (2019).
41. A. Falenty, T. C. Hansen, W. F. Kuhs, *Nature* **516**, 231–233 (2014).
42. Y. M. Li, G. A. Somorjai, *J. Phys. Chem. C* **111**, 9631–9637 (2007).

ACKNOWLEDGMENTS

We thank X. K. Zhang, L. Y. Wu, and S. H. Chang at the Center of Cryo-Electron Microscopy (CCEM), Zhejiang University, for their technical assistance on cryo-TEM; X. J. Wang at the cryo-EM facility of Westlake University for the technical assistance on cryo-FIBM; and L. Yang and Y. Y. Jin for assistance with scattering spectra analysis. This work was supported by the National Key Research and Development Project of China (2018YFB2200404), the National Natural Science Foundation of China (11527901), the Natural Science Foundation of Zhejiang Province (LR21F050002), and Fundamental Research Funds for the Central Universities.
Author contributions: L.T., X.G., and Y.R.S conceived of and supervised the project. P.X. and B.C. performed the experiments. Y.B., P.X., and B.C. performed the cryo-TEM measurements. All authors analyzed the data. L.T., X.G., P.X., B.C., and Y.R.S. wrote the manuscript. All authors discussed the results and commented on the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/187/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (*43*–*46*)
Movies S1 to S3

6 March 2021; accepted 2 June 2021
10.1126/science.abh3754

PLANT SCIENCE

Cauliflower fractal forms arise from perturbations of floral gene networks

Eugenio Azpeitia[1]†, Gabrielle Tichtinsky[2], Marie Le Masson[2], Antonio Serrano-Mislata[3], Jérémy Lucas[2], Veronica Gregis[4], Carlos Gimenez[3], Nathanaël Prunet[5,6], Etienne Farcot[7], Martin M. Kater[4], Desmond Bradley[8], Francisco Madueño[3], Christophe Godin[1]*, Francois Parcy[2]*

Throughout development, plant meristems regularly produce organs in defined spiral, opposite, or whorl patterns. Cauliflowers present an unusual organ arrangement with a multitude of spirals nested over a wide range of scales. How such a fractal, self-similar organization emerges from developmental mechanisms has remained elusive. Combining experimental analyses in an *Arabidopsis thaliana* cauliflower-like mutant with modeling, we found that curd self-similarity arises because the meristems fail to form flowers but keep the "memory" of their transient passage in a floral state. Additional mutations affecting meristem growth can induce the production of conical structures reminiscent of the conspicuous fractal Romanesco shape. This study reveals how fractal-like forms may emerge from the combination of key, defined perturbations of floral developmental programs and growth dynamics.

Above-ground plant architectures arise from the activity of shoot apical meristems (SAMs), pools of stem cells that give rise to organs such as leaves, shoots, or flowers. The arrangement of organs on stems is called phyllotaxis. Plants with a spiral phyllotaxis usually form two families of organ spirals that are visible on compact structures such as flower heads, pine cones, or cacti (Fig. 1, A to C). These two families of spirals turn in opposite directions and come in two consecutive numbers of the Fibonacci series (Fig. 1A) (*1*). In cauliflowers, spiral families are visible not only at one but at several scales (Fig. 1, D to F). This self-similar organization culminates in the Romanesco cultivar in which the spirals appear in relief because of their conical shape at all scales, a geometrical feature conferring the whole curd a marked fractal-like aspect (Fig. 1G).

Cauliflowers (*Brassica oleracea* var. *botrytis)* were domesticated from cabbages (*2*). The cauliflower inflorescence (the flower-bearing shoot) takes a curd shape because each emerging flower primordia never matures to the floral stage but instead generates more curd-shaped inflorescences (*2*, *3*). In *B. oleracea,* the genetic modifications causing curd development are still debated and likely affect multiple genes (*2*–*5*). However, cauliflower-like structures also exist in the model Brassicaceae *Arabidopsis thaliana* and are caused by a double mutation in *APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*) (Fig. 1, H and I), two paralogous genes encoding MADS-box transcription factors (TFs) promoting floral development (*6*, *7*). The *Arabidopsis* molecular regulators governing the development of shoots and flowers have been largely identified (*8*–*10*) (table S1). Network models based on these regulators have been proposed to explain flower and inflorescence development (*11*–*14*). However, whether variants of these networks are able to account for the development of *Arabidopsis ap1 cal* curds is unknown.

[1]Laboratoire de Reproduction et Développement des Plantes, Univ. Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRAE, Inria, F-69364 Lyon, France. [2]Laboratoire Physiologie Cellulaire et Végétale, Univ. Grenoble Alpes, CNRS, CEA, INRAE, IRIG-DBSCI-LPCV, F-38054 Grenoble, France. [3]Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas (CSIC) - Universidad Politécnica de Valencia (UPV), 46022 Valencia, Spain. [4]Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milan, Italy. [5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [6]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA. [7]School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK. [8]Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK.
***Corresponding author. Email: christophe.godin@inria.fr (C.G.); francois.parcy@cea.fr (F.P.)** †Present address: Centro de Ciencias Matemáticas, Universidad Nacional Autónoma de México, Morelia, México.

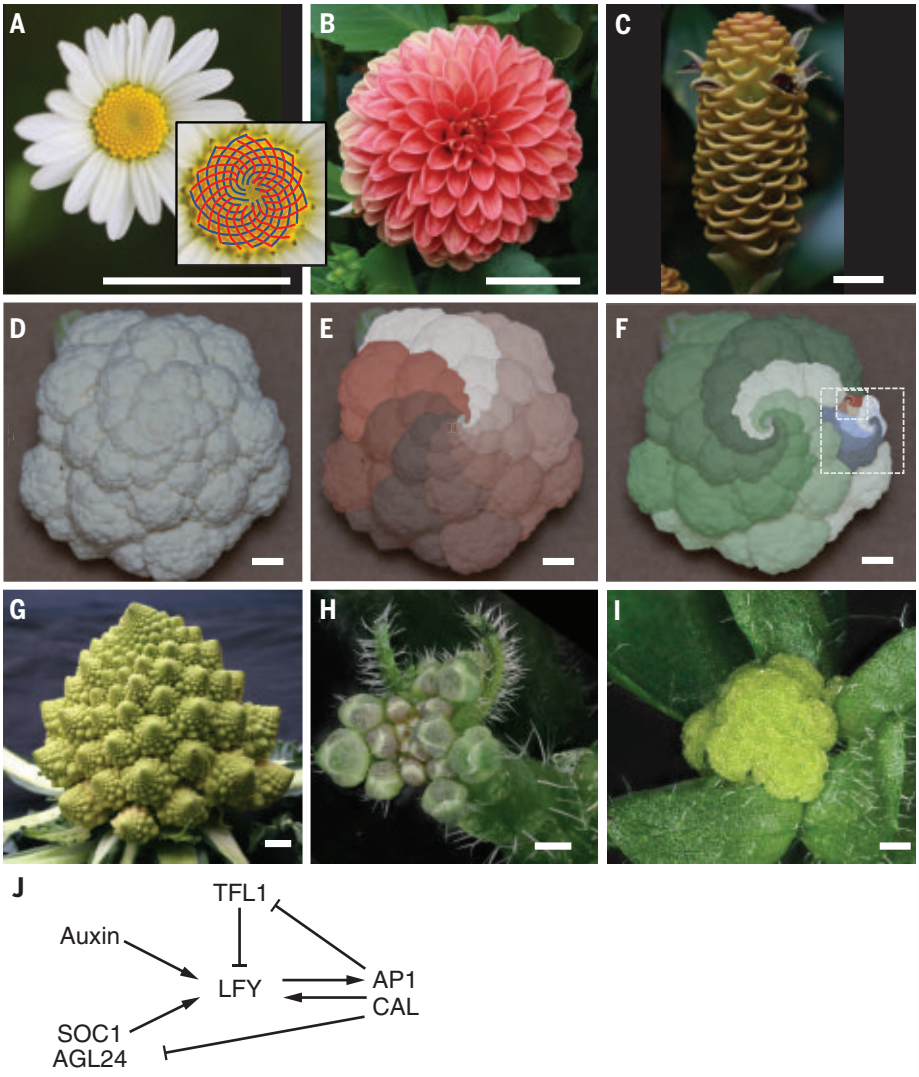


Fig. 1. Illustrations of phyllotactic spirals on plant inflorescences. (**A**) Daisy capitulum. The two families of spirals are indicated in the close-up (13 blue spirals and 21 red). (**B**) Dahlia composite flower. (**C**) Zingiber inflorescence. (**D** to **F**) *B. oleracea* var. *botrytis* cauliflower with eight counterclockwise [(E); brown family] and five clockwise [(F); green family] main spirals. Dashed rectangles show families of spirals nested over several scales. (**G** to **I**) Romanesco curd (G), *Arabidopsis* WT inflorescence (H), and *ap1 cal* curd (I). Scale bars, 2 cm [(A) to (G)]; 500 μm [(H) and (I)]. (**J**) Interactions between major floral regulators; arrows depict activation and barred lines indicate repression.

To address this question, we first built a network of the main regulators involved in both flower and curd development. Then, we embedded this network within a three-dimensional (3D) computational model of plant development to understand how mutations could transform wild-type (WT) inflorescences into curds.

Genetic basis of cauliflower curds

In *Arabidopsis*, flowers are initiated by the TF LEAFY (LFY) (Fig. 1J) (table S1). *LFY* is up-regulated by the SUPPRESSOR-OF-OVEREXPRESSION-OF-CO 1 (SOC1) and AGAMOUS-LIKE 24 (AGL24) MADS-box proteins (induced throughout the inflorescence meristem by environmental and endogenous cues) and by auxin phytohormone maxima that mark floral meristem initiation sites. *LFY* is expressed specifically in floral primordia because its induction in the SAM is repressed by the TFL1 inflorescence identity protein. In the floral primordium, LFY induces *AP1* and *CAL* (*AP1/CAL*), which positively feedback on *LFY* and repress both *SOC1/AGL24* and *TFL1*, thereby stabilizing the floral fate of the new meristem. In the *ap1 cal* cauliflower mutant, the AP1/LFY positive feedback is absent and *TFL1* is not repressed by AP1/CAL in the nascent floral meristem. Consequently, young flower primordia cannot maintain *LFY* expression and start expressing *TFL1*. As a result, they lose their floral identity and become inflorescence meristems (*6*). Whereas *TFL1* repression in nascent flower primordia is well understood, the factors directly responsible for its up-regulation in *ap1 cal* and inflorescence meristems are unknown.

To complete our network, we thus searched for direct positive regulators of *TFL1* other than LFY [which induces *TFL1* (*15*) but is not active in inflorescence meristems]. *TFL1* is indirectly regulated by day length (*16*): During long days (LDs), *TFL1* is up-regulated by CONSTANS (CO) and FLOWERING LOCUS T (FT), two key upstream effectors of the LD pathway (*11*, *17*–*19*) (fig. S1). To search for direct regulators, we examined SOC1 and AGL24, which act downstream of CO and FT in the LD pathway (*9*). Loss- and gain-of-function experiments demonstrated that both SOC1 and AGL24 induce *TFL1* (Fig. 2, A to I) and chromatin immunoprecipitation (ChIP) showed that these two TFs bind to the *TFL1* regions that regulate its expression in the SAM (*20*) (Fig. 2, J to L). These regions were sufficient to activate a *TFL1* reporter construct by SOC1 and AGL24 in a transient assay (Fig. 2, M and N), confirming that both MADS-box TFs are direct regulators of *TFL1*. Because XAANTAL2 (XAL2), a homolog of SOC1 and AGL24, also bound to and induced *TFL1* (*21*), we aggregated the activities of SOC1, AGL24, and XAL2 into a SAX proxy acting as a *TFL1* positive regulator (Fig. 3A).

We thus created the SALT network (for SAX, AP1/CAL, LFY, and TFL1; Fig. 3A) made of these four regulator sets, auxin (*22*), and F, a flower-inducing signal (a proxy for the FT florigen) that increases when the plant ages or is exposed to flower-inducing environmental conditions (*23*, *24*). We also added a short-lived transient early Repressor of *TFL1* (eREP) as a proxy for *TFL1* early repression in the young flower bud performed by the redundant activities of SOC1, AGL24, SHORT VEGETATIVE PHASE, and SEPALLATA4 (*25*).

The steady states of the SALT network correspond to the gene expression patterns observed in WT vegetative (low SALT values), inflorescence (high TFL1/SAX, low AP1/CAL/LFY), and flower (low TFL1/SAX, high AP1/CAL/LFY) meristems (Fig. 3, B and C, and fig. S2). Above a certain F threshold value, the network generates a flower or an inflorescence state depending on F and auxin values. Simulations of *tfl1*, *lfy*, *ap1 cal* mutants produced the expected outputs consistent with experimentally reported gene expressions (*6*, *16*, *26*, *27*) (Fig. 3, B and C). The simulated *sax* mutant did not reach a floral state, consistent with the late-flowering behavior of the *soc1 agl24* double mutant (*28*).

The modeled gene expression dynamics (Fig. 3D) illuminate the fundamental differences between WT and cauliflower meristems: In a WT flower primordium, F induces *SAX*. SAX and auxin induce *LFY*, which, together with F,

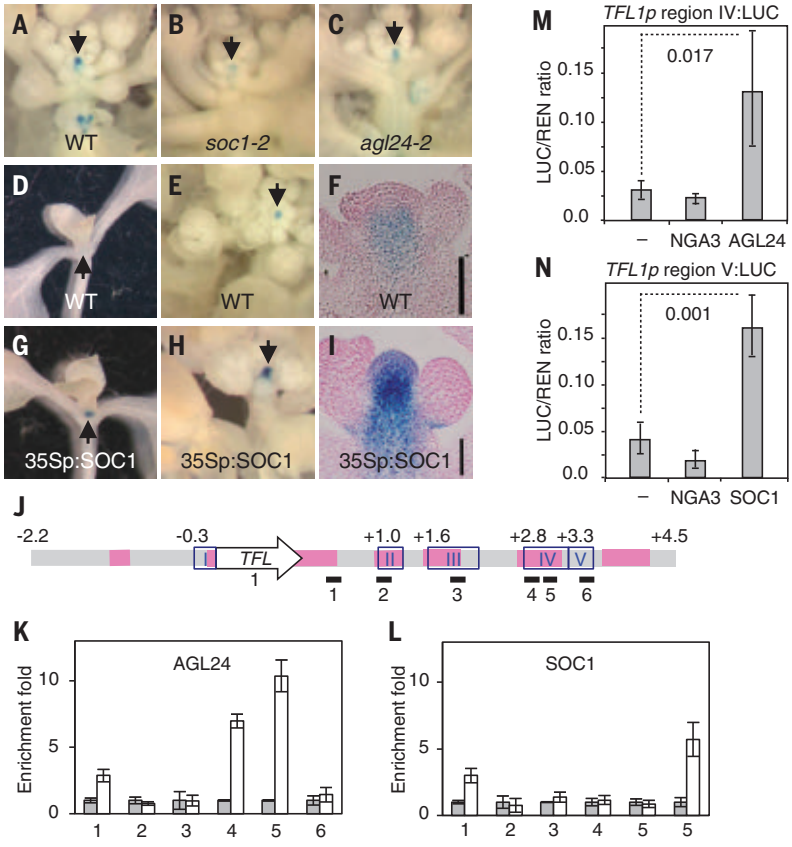


Fig. 2. AGL24 and SOC1 are direct positive regulators of *TFL1*. (**A** to **C**) TFL1p:GUS activity in WT (A), *soc1-2* (B), and *agl24-2* (C) inflorescence apices. (**D** to **I**) TFL1p:GUS activity (blue signal) in WT [(D) to (F)] and *35Sp:SOC1* [(G) to (I)] apices at the vegetative [(D) and (G)] and flowering [(E), (F), (H), and (I)] stages. (F) to (I) are longitudinal sections through flowering shoots. Arrows mark the SAM. Scale bars in (F) and (I), 40 μm. (**J** to **L**) Structure of *TFL1* locus, with regions conserved in Brassicaceae (pink lines), regulatory regions (*20*) (blue boxes I to V), and fragments used in ChIP (black lines 1 to 6). ChIP experiments on plants expressing a tagged version of AGL24 [(K), white bars] or the WT SOC1 protein [(L), white bars] or on control plants [(K) and (L), gray bars; see the supplementary materials and methods] showed that AGL24 binds region IV [(K), fragments 4 and 5] and SOC1 region V [(L), fragment 6]. A representative biological replicate is shown with the mean ± SE for three technical replicates. (**M** and **N**) Transient assays showing transactivation of the LUCIFERASE (LUC) reporter driven by region IV (activation by 35Sp:AGL24) and region V (activation by 35Sp:SOC1). NGA3 is an unrelated TF used as a negative control. Bars denote the mean ± SD of three independent biological replicates. *P* values are for the equality of means (Student's *t* test).

induces *AP1/CAL*. AP1 positively feeds back on *LFY* and represses *SAX* (Fig. 3D). *TFL1* expression, which could be induced by SAX and LFY in early floral stages, is constantly repressed, first by eREP and later by SAX plus AP1/CAL. High AP1/CAL and LFY with low TFL1 and SAX expression stabilize the floral fate. By contrast, in the *ap1 cal* flower primordia, the absence of AP1/CAL activity has two consequences: (i) *LFY* expression is up-regulated only transiently because AP1/CAL positive feedback is missing (Fig. 3D) and (ii) *SAX* genes are not repressed by AP1 and thus induce *TFL1* in nascent flower meristems. TFL1 represses *LFY* even further and the meristem returns to a shoot meristem state (Fig. 3D). The early LFY induction would likely be reinforced (while remaining transient) by incorporating the recently discovered direct induction of LFY by the F partner protein FD (*29*). The SALT model predicts that *SAX* expression should extend over the entire cauliflower. We analyzed a SOC1-GFP reporter line and indeed observed expansion of its expression domain in *ap1 cal* compared with WT (Fig. 3, E and F).

The SALT network thus recapitulates realistic gene expressions driving meristem fates. However, a plant architecture depends not just on meristem fates but also on morphodynamic parameters, including molecular thresholds for fate decisions, organ growth rate, delay for meristems to start organ production, and organ production rate, which are independently regulated. Plant inflorescence architecture thus emerges from the complex interaction between the floral gene-regulatory network (GRN) and morphodynamic parameters. This is illustrated here by the *lfy* and *ap1 cal* mutants that have the same GRN outputs (Fig. 3C) but markedly different architectures (*6*, *27*). To study how this interaction operates in *Arabidopsis*, we integrated the SALT GRN in a 3D plant computational model implemented as an L-system (see the supplementary materials, modeling methods).

A multiscale model generates *Arabidopsis* cauliflower structures

The 3D model is made of the four types of organs that shape plant above-ground architecture: meristems, internodes, leaves, and flowers (Fig. 4A and see the supplementary materials). Each meristem's identity (vegetative, inflorescence, and floral) is determined by the GRN steady state, computed at each time step as a function of the meristem's previous state and external factors (auxin and F). The GRN model is implemented as single-compartment ordinary differential equations (see the supplementary materials, modeling methods). We assume that the GRN dynamics are faster than growth and reach steady state within a time step. A set of growth rules defines meristem production: A vegetative meristem produces a compressed stem (non-elongated internodes) with rosette leaves and dormant axillary meristems; an inflorescence meristem produces an elongating internode with either a cauline leaf and a new axillary shoot meristem in the leaf axil or a lateral flower meristem; and a floral meristem produces an internode terminating with a flower meristem devoid of bracts (leaf-like organs subtending flowers) because they are repressed by LFY (*6*). Each newly generated axillary meristem begins with maximal auxin level (*22*), SAX/LFY/AP1/CAL values inherited from the parent meristem, together with a fraction of the parent TFL1 value because in the real plant, this non–cell-autonomous protein is present in the primordia region (*30*). To match the WT plant architecture, indeterminate meristems at orders >2 (Fig. 4A) were kept quiescent, a likely effect of apical dominance (the inhibition of lateral meristem outgrowth) (fig. S3A). The model also contains rules describing organ growth dynamics (internode and leaf elongation, flower growth, organ production rate, and growth initiation delay). Simulated plants start with a single vegetative SAM and repeatedly produce new organs according to the GRN, the morphodynamic rules, and an input value of F.

By adjusting the GRN and morphodynamic parameters within a range of plausible values (see the supplementary materials), we successfully calibrated the model to produce realistic architectures for WT and *lfy* plants (movies S1 and S2), as well as for the *tfl1* mutant (Fig. 4, B to D) and a nonflowering phenotype for the *sax* mutant. However, our simulations could not generate a realistic *ap1*

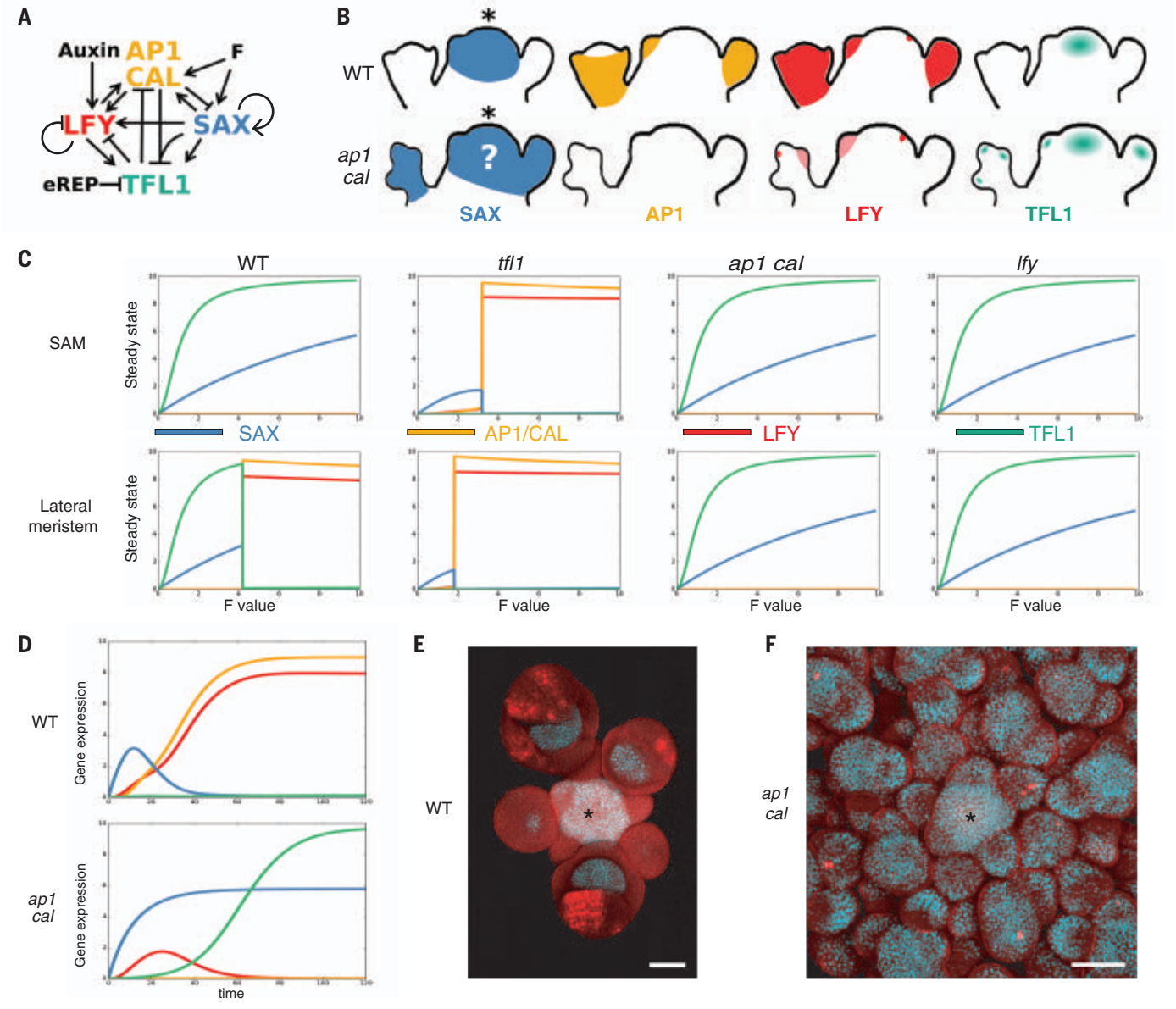


Fig. 3. SALT GRN model and experimental validation. (**A**) SALT GRN network structure. (**B**) Known expression patterns of *SAX*, *AP1/CAL*, *LFY*, and *TFL1* in the SAM and lateral primordia of the WT and the *ap1 cal* mutant. The question mark indicates a predicted expression pattern of the model. (**C**) WT, *tfl1*, *ap1 cal*, and *lfy* steady states of the model at different F values in the SAM (low auxin) and in lateral meristems (high auxin). The genetic identity predicted for WT and all mutant meristems correspond to the experimentally observed phenotypes. (**D**) Temporal simulation of gene expression in lateral primordia with high F values. (**E** and **F**) Expression of the SOC1:GFP (white/light blue signal) reporter construct in WT (E) and *ap1-7 cal-1* mutant (F) inflorescences. Asterisks mark the SAM. Scale bars, 50 μm.

cal mutant growing without bract or cauline leaves and displaying high-order meristems (fig. S3, A and B), suggesting that the cauliflower phenotype involves additional regulations. We reasoned that laterally produced *ap1 cal* inflorescence meristems are different from those produced in other genotypes because according to our GRN, they have been transiently exposed to LFY expression (Fig. 3D). Several pieces of evidence suggest that this transient LFY expression, already known to repress bracts (*6*), could also contribute to high-order meristem release. First, the outgrowth of otherwise inhibited axillary meristems in the rosette is stimulated by ectopic expression of LFY (or an LFY allele) (*31*, *32*). Second, it was established that the *lfy ap1 cal* triple mutant does not form cauliflowers (*6*), and we found that in this mutant, the number of high-order meristems was significantly reduced compared with *ap1 cal* (fig. S3, D to H), thus supporting our hypothesis.

We abstracted this critical molecular pathway by introducing in the model a factor X up-regulated when LFY exceeds a minimal threshold level. Up-regulated factor X releases high-order meristem growth and suppresses the bract. This was sufficient to unlock the recursive growth of lateral meristems and to generate the *ap1 cal* curd structure that arises from the transient but irreversible exposure of meristems to the floral signal without any alteration of WT growth dynamics (Fig. 4, E and H, and movie S3). Overall, our work shows that the *ap1 cal* and *lfy* architectures are different (Fig. 3C) because the molecular histories of their inflorescence meristems are different, thereby revealing the existence of a developmental hysteresis.

Growth dynamics define cauliflower and Romanesco curd structures

Our work in *Arabidopsis* offers a conceptual framework to explain how inflorescence architecture emerges from coupling a floral GRN to morphodynamic parameters. We wondered whether modifications affecting components of this framework could also explain the architecture of the cauliflowers that arose during domestication, namely the edible *B. oleracea (Bo)* var. *botrytis (Bob)* and its Romanesco variant. Whether similar genetic defects as in *Arabidopsis* are responsible for curd development in *B. oleracea* is still debated (*4*, *5*). To further investigate this point, we analyzed RNA-sequencing

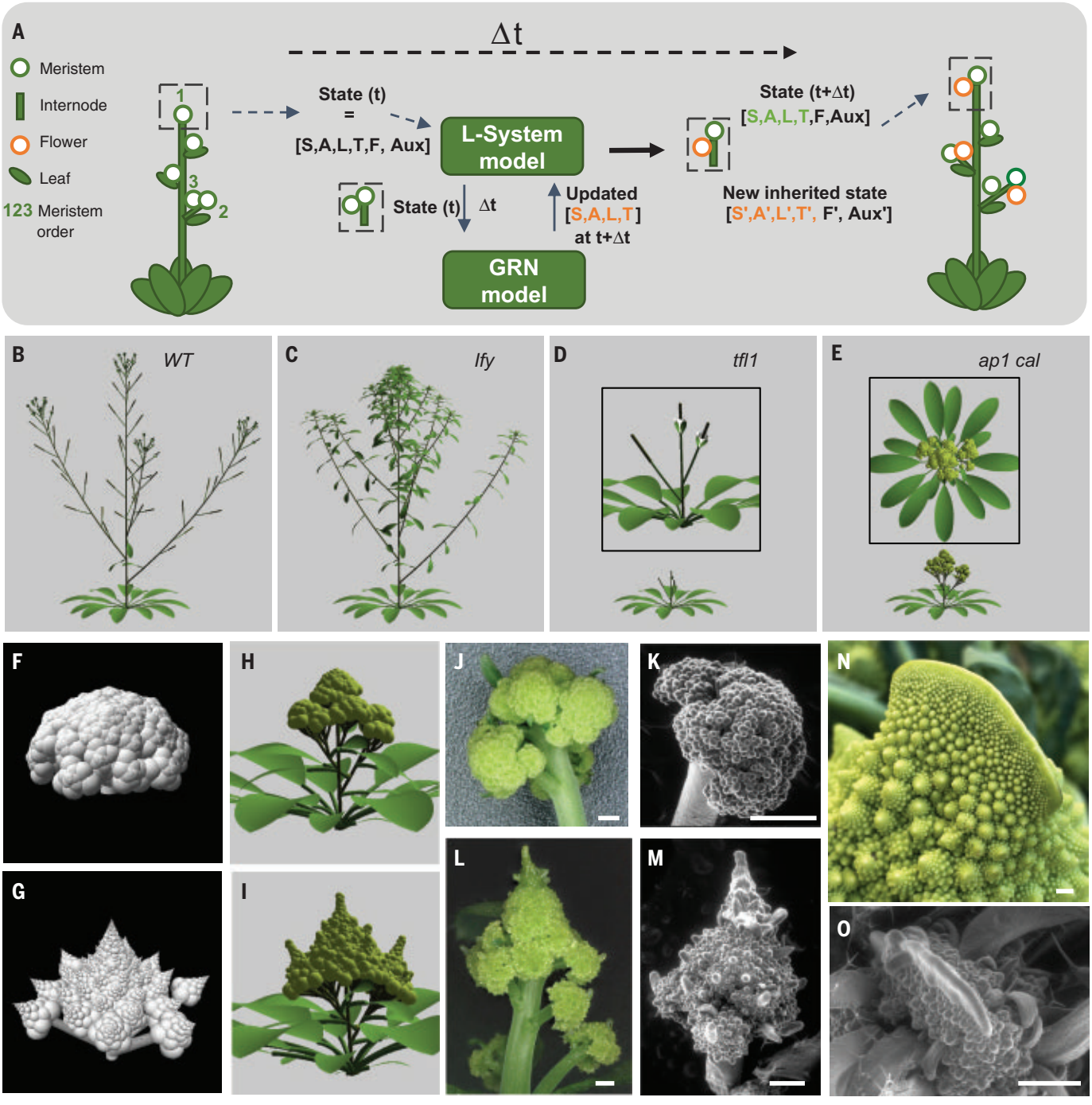


Fig. 4. Simulation and assessment of a GRN-based plant development model. (**A**) Schematic representation of the multiscale model of *Arabidopsis* development. Each meristem state is composed of signal levels (auxin, F) and a GRN steady state. At time *t*, the plant is made up of a collection of organs (left). At time *t*+Δ*t* (right), the model updates the signal levels and GRN state in each meristem. The steady state defines the identity of the meristems (vegetative, inflorescence, or flower) used to compute meristem lateral productions. Green numbers indicate meristem order. (**B** to **E**) Plant morphologies obtained in the WT (B), *lfy* (C), *tfl1* (D), and *ap1 cal* (E) simulations. (**F** to **I**) Simulated morphologies with constant [(F) and (H)] or increased [(G) and (I)] meristem production rate in a simplified model [(F) and (G)] and in the *Arabidopsis* model [(H) and (I)]. (**J** to **O**) Light micrographs [(J), (L), and (N)] and SEM [(K), (M), and (O)] of cauliflower structures in *Arabidopsis ap1 cal* [(J) and (K)], *Arabidopsis ap1 cal clv3* [(L), (M), and (O)], and Romanesco (N). The uninduced *AP1:GR* transgene is present in plants (J) to (M). Scale bars, 500 μm.

data of *Bob* curds. We confirmed the previously identified mutation in the *BobCAL* gene (fig. S4A) (*4*, *5*, *7*) and observed that the two *AP1* paralogs, *BobAP1-a* and *BobAP1-c*, are expressed at much lower levels than in cabbage (*Bo* var. *capitata)* inflorescences (fig. S4B). These functional proteins are induced with a delay only when the cauliflower elongates and starts forming normal flowers (*3*, *33*). Comparing cauliflower and cabbage sequences, we identified differences in binding sites for candidate regulators of *BoAP1* that could account for their delayed activation (fig. S4D). The combination of BoCAL inactivation and *BobAP1-a/c* expression delay (heterochrony caused by *cis* or *trans* mutations) thus likely participates in *Bob* curd development. Similar to *Arabidopsis ap1 cal*, cauliflowers have meristems of higher maximal order ($n \geq 7$) than cabbages ($n = 3$ to 4) (fig. S5). Nevertheless, the development of single massive cauliflower curds is not the exact equivalent of the *Arabidopsis* mutant (*3*, *5*) and involves additional multifactorial alterations of morphodynamics parameters (such as reduction of internode elongation and increase in branch diameter).

The conical shapes appearing in Romanesco spirals at all scales (Fig. 1F) represent an additional geometric variation obtained through domestication that seems to be associated with a change in morphodynamic parameters. Indeed, several such parameters remain constant during cauliflower development but vary in Romanesco (*34*): (i) the plastochron, the time between two successive meristem productions, (ii) the number of visual spirals originating from a given meristem, (iii) the time (measured in number of plastochrons) needed before a lateral primordium starts producing its own primordia (or lateral production onset delay), and (iv) the size of the meristems. Whether some of these parameters are causal to the Romanesco phenotype remains unclear, but phyllotaxis studies (*1*, *35*, *36*) indicate that the first three parameters are linked to the meristem size: An augmentation of the size of the meristem central zone should decrease the plastochron, which in turn increases the number of spirals, and the lateral production onset delay. We thus hypothesized that passing from a constant to a decreasing plastochron in meristems could change cauliflower into Romanesco morphologies. We first tested this in silico using a simplified, purely geometric model of curd growth that is independent from the *Arabidopsis* GRN and specific growth dynamics (see the supplementary materials). A decreasing plastochron was sufficient to produce Romanesco shapes (Fig. 4G), whereas constant values of this parameter produce cauliflower morphologies (Fig. 4F).

We then introduced the same change in the more complex, GRN-based *Arabidopsis* cauliflower architectural model while keeping its organ growth dynamics as calibrated on the WT. Although not as complete as in the purely geometric model, the curd changed toward a "Romanesco-like" morphology with typical conical curd shapes (Fig. 4, H and I). We then tested this hypothesis experimentally in *Arabidopsis* by altering the size of the meristem directly. We achieved this by introducing a mutation in the *CLAVATA3* (*CLV3*) gene that controls meristem homeostasis and induces an increase of the meristem central zone during growth (*37*, *38*). As predicted by our analysis, introduction of a *clv3* mutation in *ap1 cal Arabidopsis* mutant modified the curd shape, which lost its round morphology and acquired a more conical shape, with similar structures at different scales, features recognized as hallmarks of Romanesco curds (*39*) (Fig. 4, L and M). Two additional pieces of evidence support the hypothesis that meristem homeostasis is perturbed in Romanesco curds: (i) they occasionally show fasciation, a feature typical of meristem enlargement also observed in *clv3* or *ap1 cal clv3* mutants (Fig. 4, N and O) (*37*); and (ii) the expression of *CLV3* (and possibly two other genes acting in the same pathway) (*38*) is lower in Romanesco curds than in cauliflowers (fig. S6). Altogether, these observations establish that meristem size regulates the final curd morphology through control of plastochron value.

These results reveal how fractal patterns can be generated through growth and developmental networks that alter identities and meristem dynamics. Our data, GRN, and growth models now clarify the molecular and morphological changes over time by which meristems gain different identities to form the highly diverse and fascinating array of plant architectures found throughout nature and crops.

REFERENCES AND NOTES

1. C. Godin, C. Golé, S. Douady, *Development* **147**, dev165878 (2020).
2. C. F. Quiros, M. W. Farnham, "The genetics of *Brassica oleracea*," in *Genetics and Genomics of the Brassicaceae*, R. Schmidt, I. Bancroft, Eds. (Springer, 2011), pp. 261–289.
3. D. V. Duclos, T. Björkman, *J. Exp. Bot.* **59**, 421–433 (2008).
4. L. B. Smith, G. J. King, *Mol. Breed.* **6**, 603–613 (2000).
5. N. Guo *et al*., *BMC Biol.* **19**, 93 (2021).
6. J. L. Bowman, J. Alvarez, D. Weigel, E. M. Meyerowitz, D. R. Smyth, *Development* **119**, 721–743 (1993).
7. S. A. Kempin, B. Savidge, M. F. Yanofsky, *Science* **267**, 522–525 (1995).
8. G. Denay, H. Chahtane, G. Tichtinsky, F. Parcy, *Curr. Opin. Plant Biol.* **35**, 15–22 (2017).
9. A. Pajoro *et al*., *J. Exp. Bot.* **65**, 4731–4745 (2014).
10. B. Thomson, F. Wellmer, *Curr. Top. Dev. Biol.* **131**, 185–210 (2019).
11. K. E. Jaeger, N. Pullen, S. Lamzin, R. J. Morris, P. A. Wigge, *Plant Cell* **25**, 820–833 (2013).
12. C. Espinosa-Soto, P. Padilla-Longoria, E. R. Alvarez-Buylla, *Plant Cell* **16**, 2923–2939 (2004).
13. F. Leal Valentim *et al*., *PLOS ONE* **10**, e0116973 (2015).
14. P. Prusinkiewicz, Y. Erasmus, B. Lane, L. D. Harder, E. Coen, *Science* **316**, 1452–1456 (2007).
15. K. Goslin *et al*., *Plant Physiol.* **174**, 1097–1109 (2017).
16. C. Ferrándiz, Q. Gu, R. Martienssen, M. F. Yanofsky, *Development* **127**, 725–734 (2000).
17. D. Bradley, O. Ratcliffe, C. Vincent, R. Carpenter, E. Coen, *Science* **275**, 80–83 (1997).
18. X. Hou *et al*., *Nat. Commun.* **5**, 4601 (2014).
19. S. K. Yoo *et al*., *Plant Physiol.* **139**, 770–778 (2005).
20. A. Serrano-Mislata *et al*., *Development* **143**, 3315–3327 (2016).
21. R. V. Pérez-Ruiz *et al*., *Mol. Plant* **8**, 796–813 (2015).
22. D. Reinhardt *et al*., *Nature* **426**, 255–260 (2003).
23. P. A. Wigge, *Curr. Biol.* **21**, R374–R378 (2011).
24. J. Putterill, E. Varkonyi-Gasic, *Curr. Opin. Plant Biol.* **33**, 77–82 (2016).
25. C. Liu *et al*., *Dev. Cell* **24**, 612–622 (2013).
26. O. J. Ratcliffe *et al*., *Development* **125**, 1609–1615 (1998).
27. D. Weigel, J. Alvarez, D. R. Smyth, M. F. Yanofsky, E. M. Meyerowitz, *Cell* **69**, 843–859 (1992).
28. S. D. Michaels *et al*., *Plant J.* **33**, 867–874 (2003).
29. Y. Zhu *et al*., *Nat. Commun.* **11**, 5118 (2020).
30. L. Conti, D. Bradley, *Plant Cell* **19**, 767–778 (2007).
31. H. Chahtane *et al*., *Plant J.* **74**, 678–689 (2013).
32. D. Weigel, O. Nilsson, *Nature* **377**, 495–500 (1995).
33. X. Sun *et al*., *Environ. Exp. Bot.* **155**, 742–750 (2018).
34. M. Kieffer, M. P. Fuller, A. J. Jellings, *Planta* **206**, 34–43 (1998).
35. S. Douady, Y. Couder, *J. Theor. Biol.* **178**, 255–273 (1996).
36. Y. Refahi *et al*., *eLife* **5**, e14093 (2016).
37. J. C. Fletcher, U. Brand, M. P. Running, R. Simon, E. M. Meyerowitz, *Science* **283**, 1911–1914 (1999).
38. M. Kitagawa, D. Jackson, *Annu. Rev. Plant Biol.* **70**, 269–291 (2019).
39. L. E. Watts, *Euphytica* **15**, 224–228 (1966).

ACKNOWLEDGMENTS

We thank A.-M. Chèvre, R. Immink, R. Simon, L. Ostergaard, and M. Benitez for advice; T. Vernoux, C. Zubieta, and H. Chahtane for proofreading and useful feedback on the manuscript; D. Tardy, E. Giraud, R. Dumas, and V. Martin (OBS, France) for providing cauliflower samples; L. Bousset Vaslin for images and branch counting; F. Boudon for help with L-Py; R. Immink (Wageningen, Netherlands), C. Ferrándiz (IBMCP; Spain), G. Coupland (MPIPZ, Germany), M. Ángel Blázquez (IBMCP, Spain), R. Amasino (UWM, USA), and the European *Arabidopsis* Stock Centre for providing seeds; V. Berger (CEA/DRF) for financing the Keyence microscope; and C. Lancelon-Pin (Plateau de microscopie électronique - ICMG. CERMAV-CNRS) for help with SEM experiments. **Funding:** This work was supported by the INRAE Caulimodel project (to F.P. and C.Go.); Inria Project Lab Morphogenetics (to C.Go., E.A., and F.P.); the ANR BBSRC Flower model project (to F.P. and C.Go.); the GRAL LabEX (ANR-10-LABX-49-01) within the framework of the CBH-EUR-GS (ANR-17-EURE-0003) (to F.P., G.T., M.L.M., and J.L.); the EU H2020 773875 ROMI project (to C.Go.); and the Spanish Ministerio de Ciencia Innovación and FEDER (grant no. PGC2018-099232-B-I00 to F.M.). **Author contributions:** C.Go. and F.P. conceived the study. C.Go., E.A., and E.F. performed the modeling. A.S.-M., C.Gi., D.B., F.M., F.P., G.T., M.M.K., M.L.M., and V.G. designed and performed the plant experiments. N.P. performed the confocal imaging experiment. J.L. analyzed the RNA-seq and genomic data. C.Go., F.P., and E.A. wrote the paper with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main paper or the supplementary materials. All plant materials are available upon request. Raw and processed RNA-seq data are available at GEO under accession no. GSE150627. All source codes to run the simulations are available as a supplementary archive file (descriptions of installation and execution are available as README.txt).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/192/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S3
Movies S1 to S3
References (*40*–*108*)
Code Archive File: Architecture-Model.zip
MDAR Reproducibility Checklist

15 January 2021; accepted 3 June 2021
10.1126/science.abg5999

MARTIAN GEOLOGY

Brine-driven destruction of clay minerals in Gale crater, Mars

T. F. Bristow[1]*, J. P. Grotzinger[2], E. B. Rampe[3], J. Cuadros[4], S. J. Chipera[5], G. W. Downs[6], C. M. Fedo[7], J. Frydenvang[8], A. C. McAdam[9], R. V. Morris[3], C. N. Achilles[9], D. F. Blake[1], N. Castle[5], P. Craig[5], D. J. Des Marais[1], R. T. Downs[6], R. M. Hazen[10], D. W. Ming[3], S. M. Morrison[10], M. T. Thorpe[11], A. H. Treiman[12], V. Tu[11], D. T. Vaniman[5], A. S. Yen[13], R. Gellert[14], P. R. Mahaffy[9], R. C. Wiens[15], A. B. Bryk[16], K. A. Bennett[17], V. K. Fox[2]†, R. E. Millken[18], A. A. Fraeman[13], A. R. Vasavada[13]

Mars' sedimentary rock record preserves information on geological (and potential astrobiological) processes that occurred on the planet billions of years ago. The *Curiosity* rover is exploring the lower reaches of Mount Sharp, in Gale crater on Mars. A traverse from Vera Rubin ridge to Glen Torridon has allowed *Curiosity* to examine a lateral transect of rock strata laid down in a martian lake ~3.5 billion years ago. We report spatial differences in the mineralogy of time-equivalent sedimentary rocks <400 meters apart. These differences indicate localized infiltration of silica-poor brines, generated during deposition of overlying magnesium sulfate–bearing strata. We propose that destabilization of silicate minerals driven by silica-poor brines (rarely observed on Earth) was widespread on ancient Mars, because sulfate deposits are globally distributed.

Smectites are a group of clay minerals characterized by high surface areas and negative electrostatic charge (*1*). These properties promote intimate associations with organic material during sedimentation, enhancing organic preservation in Earth's geological record (*2*). The increased temperature and pressure experienced by sediments during burial cause physical, chemical, and mineralogical changes—collectively termed burial diagenesis—that break down mineral-organic associations. A common burial diagenetic change observed on Earth involves the transformation of smectite into other minerals (typically illite and chlorite), a process that coincides with molecular reconfiguration and degradation of associated organic matter (*1*). On Mars, Noachian- to Hesperian-age [~4.1 to 3.2 billion years ago (Ga)] sedimentary rocks that contain smectites but lack evidence for these transformations could therefore preserve information on geological or astrobiological processes (*2*, *3*).

Since landing in Gale crater in 2012, the *Curiosity* rover has traversed ~25 km across the crater floor and the base of Aeolis Mons (informally called Mount Sharp), a 5.5-km-high mountain in the center of Gale. Rover measurements have revealed smectite-bearing ~3.5-Ga sedimentary rocks with vertical thicknesses of several hundred meters (Fig. 1). The smectites in these rocks are largely unaffected by burial diagenetic transformations observed in sedimentary basins on Earth and preserve organic compounds (*4*–*6*). Bulk geochemical indices and mineralogical trends in these rocks predominantly reflect depositional processes and conditions, as they coincide with sedimentary indicators of changing environments within the lake system that once occupied the crater floor ~3.5 Ga (*5*, *7*–*9*).

Despite the limited evidence for burial diagenesis of clay minerals, outcrops surveyed by *Curiosity* display the influence of reactions involving liquid water (aqueous alteration) that postdate sediment deposition (and are therefore referred to as diagenetic reactions). Networks of millimeter- to centimeter-scale, calcium sulfate–filled fractures and veins are frequently found penetrating host rock (*4*, *10*). In some cases, the passage of fluids has altered the mineralogy and geochemistry of host sediments, producing decimeter-scale "alteration halos" adjacent to fractures (*11*). Some rocks along *Curiosity*'s traverse experienced enhanced mobilization of trace elements and mineralogical reactions that influence the physical and spectral properties of rocks (*12*–*16*). X-ray amorphous materials have been detected in all samples analyzed with the rover's CheMin x-ray diffraction (XRD) instrument, with abundances of ~15 to 70 wt %. This material could be the product of multiple episodes of aqueous interaction with sediment (*11*, *13*, *17*). Radiometric dating of jarosite—a potassium-bearing, iron-sulfate mineral—provides evidence for intermittent aqueous alteration extending over a period of at least half a billion years following deposition (*18*).

The protracted history of diagenetic reactions in Gale crater has made it challenging to identify the sources and drivers of the post-depositional passage of fluids (*13*–*15*). Natural variations in the depositional mineralogy of sediments, a function of changing sediment sources, depositional setting, and lake conditions (*9*), make the task more difficult. We attempt to eliminate this extra complexity by examining the influence of diagenetic reactions on the mineralogy of a lateral transect of ancient lake mudstones, deposited in similar environments at the same time.

Study area

Glen Torridon (GT) is a shallow west-to-east elongated trough ~0.3 to 1 km wide, lying between the slopes of Aeolis Mons to the south and the more resistant rocks of Vera Rubin ridge (VRR) to the north (Fig. 2). Rocks exposed within GT exhibit deep near-infrared absorption features in orbital reflectance spectra, which are attributed to smectites (*19*, *20*). The clay minerals in GT are part of a sequence of remotely detected hydrous and hydroxylated minerals exposed on Aeolis Mons. These clays are overlain by rocks enriched in sulfates and iron oxides (*19*, *20*), a sequence that is widely observed on Mars and is thought to have formed during exhaustion of the planet's near-surface liquid water (*3*, *21*).

Two main geomorphological subunits in GT were identified from orbital imagery before *Curiosity* reached GT in January 2019: a smooth ridged unit to the north, skirting VRR, and a fractured unit to the south (Fig. 2). During *Curiosity*'s traverse, rover imagery showed that the smooth-ridged unit is covered by a mixture of pebbles and sand. Occasional outcrops expose thinly to thickly laminated mudstones, which are sometimes fractured and rubbly (the likely source of the pebbles) and sometimes exhibit alternating recessive and resistant laminations. The overlying fractured unit primarily consists of sandstones. The rock compositions and their physical features (which, when grouped together, define a distinguishable rock unit or facies) resemble underlying and laterally equivalent deposits (*8*, *22*, *23*), indicating the continued presence of a lake with intermittent episodes of deposition by rivers and of wind-blown sediment. Our analysis of rover imagery shows no evidence of faulting or a depositional break marking the boundary between VRR and GT.

[1]Eobiology Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA. [2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [3]Astromaterials Research and Exploration Science Division, NASA Johnson Space Center, Houston, TX 77058, USA. [4]Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK. [5]Planetary Science Institute, Tucson, AZ 85719, USA. [6]Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA. [7]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN 37996, USA. [8]Globe Institute, University of Copenhagen, Copenhagen, Denmark. [9]Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [10]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA. [11]Jacobs Technology–Jacobs JETS Contract, Astromaterials Research and Exploration Science Division, at NASA Johnson Space Center, Houston, TX 77058, USA. [12]Lunar and Planetary Institute, Universities Space Research Association, Houston, TX 77058, USA. [13]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [14]Department of Physics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. [15]Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [16]Department of Earth and Planetary Science, University of California Berkeley, Berkeley, CA 94720, USA. [17]U.S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ 86001, USA. [18]Department of Earth, Environmental Sciences and Planetary Sciences, Brown University, Providence, RI 02912, USA.
*Corresponding author. Email: thomas.f.bristow@nasa.gov
†Present address: Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, MN 55455, USA.

The flat-lying bedding of the rocks, and the geomorphology of the GT topographic trough, mean that strata exposed in GT are lateral equivalents of the Jura member exposed on the north face of VRR (Fig. 2) (*24*). This is confirmed by our identification of distinctive rocks of the "Flodigarry facies" that can be traced from GT to VRR (Fig. 2). Therefore, GT mudstones correspond to the Jura member of the Murray formation, and overlying sandstones make up the Knockfarril Hill member (Figs. 1 and 2).

Curiosity drilled a pair of rock samples, informally called Aberlady (AL) and Kilmarie (KM), from laminated mudstones of the Jura member of the Murray formation exposed in GT. A second pair of drilled samples from GT [Glen Etive (GE) and Glen Etive 2 (GE2)] were collected from the Knockfarril Hill member (Fig. 1 and 2). We analyzed the GT samples (*25*) and their VRR counterparts from the Jura member [Rock Hall and Highfield (*14*, *15*)] to compare the mineral and geochemical properties of sedimentary rocks that share the same depositional history.

X-ray diffraction and evolved gas analysis

Mineralogical analyses of GT samples, using the CheMin instrument (*25*), show that they contain many of the phases identified in other Murray formation sediments (Table 1 and Fig. 2) (*5*, *12*–*14*). GT samples contain clay minerals, plagioclase feldspar, calcium sulfate minerals (bassanite and anhydrite), and an x-ray amorphous component in abundances >5 wt %. Minor constituents (<5 wt %) include hematite, pyroxene, and in KM and GE2, siderite (iron carbonate) (Table 1 and Fig. 2). The presence of carbonate minerals has been inferred from CO_2 released by rock samples in evolved gas analysis (EGA) experiments performed by the Sample Analysis at Mars (SAM) instrument on *Curiosity*, before the rover's arrival at GT (*26*). However, carbonates have not previously been definitively observed in CheMin data, indicating abundances below the instruments' detection limit of ~1 wt %. In KM, we detect multiple diffraction peaks from a single carbonate phase (siderite) in the CheMin XRD data (fig. S1).

We characterize clay minerals in rocks from GT using the position and breadth of XRD peaks, coupled with H_2O-release temperatures of structurally bound hydroxyl groups (the process of dehydroxylation) derived from SAM EGA experiments (*25*). As in much of the Murray formation, low bulk potassium content (table S1) and a broad XRD peak at low angles (the 001 peak in Fig. 3), corresponding to the ~10-Å spacing of clay mineral layers (known as the basal reflection), indicate the presence of smectite with collapsed interlayers, expected under the low-humidity conditions inside CheMin (*4*, *5*, *27*). SAM EGA experiments on KM and GE2 reveal H_2O-evolution temperatures of ~400° to 600°C (Fig. 4), consistent with dehydroxylation of a Fe^{3+}-rich dioctahedral smectite (*28*). However, H_2O lost in this temperature range could also arise from x-ray amorphous materials (*28*). The position of an XRD peak related to the size of smectite crystals along the *b* axis (the 02*l* peak) is sensitive to the occupancy and types of cations within smectite octahedral sheets (*4*, *5*, *27*). 02*l* peak positions from GT samples (Fig. 3) are consistent with a Fe^{3+}-rich dioctahedral smectite. The unit cell lengths of GT smectite crystals along the *b* axis were calculated from XRD data (*25*) and range from 9.081 to 9.112 Å (table S2). We estimate an iron content of ~0.9 to 1.2 atoms per formula unit (table S2), using previously published empirical relationships (*29*), which is approximately half of the dioctahedral cation sites.

The XRD data for KM contain another low-angle peak, corresponding to an interplanar spacing of ~9.22 Å (Fig. 3), not previously observed in *Curiosity* drilled or scooped samples. This peak is in a region typically occupied by basal reflections of clay minerals (*1*). However, the width of the peak is equivalent to the angular resolution of the CheMin instrument (fig. S2) and is considerably sharper than the smectite basal reflection in the same sample (Fig. 3), as well as other smectite-bearing samples from Gale (*4*, *5*). After eliminating other candidate phases, including hydrated iron sulfates, phosphates, and zeolites (*25*), we interpret this peak as arising from a mixed-layer serpentine-talc (S-T). S-T is a clay mineral in which layers of talc and serpentine are interstratified with each other. We fitted the peak positions and relative intensities in the KM data with a one-dimensional XRD model of S-T, finding a best-fitting composition of mixed-layer Fe^{2+}-bearing talc containing ~6% serpentine (*25*). An EGA H_2O release at ~715°C—lower than observed in talc lacking Fe^{2+} substitution—supports this interpretation (Fig. 4). S-T constitutes ~10% of the total clay mineral component in KM (*25*). Muted peaks at 9.22 Å in the XRD data of other GT samples indicate that they contain S-T in smaller amounts (Fig. 3).

Origin of mixed-layer serpentine-talc

On Earth, talc forms in a variety of geological settings, as an aqueous alteration product of

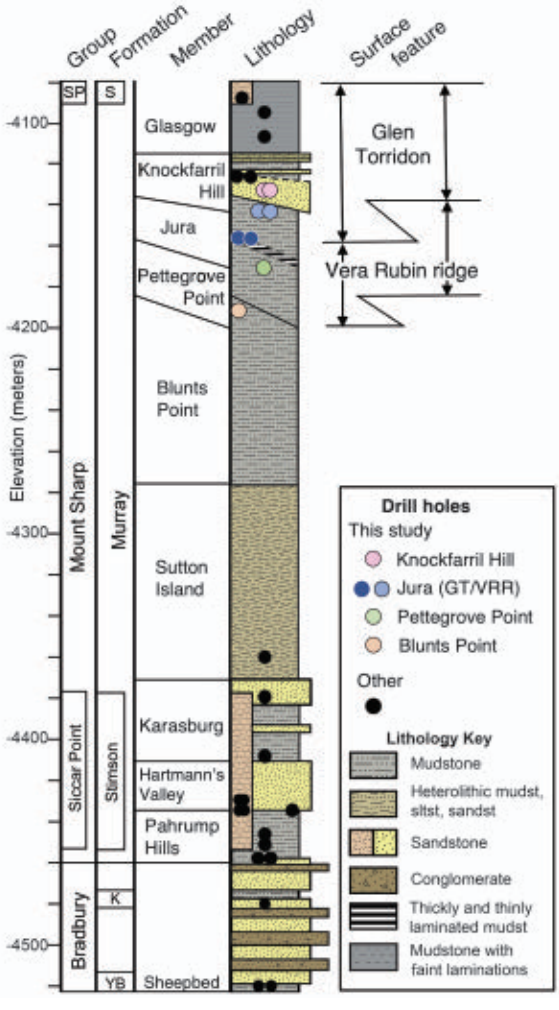


Fig. 1. Stratigraphic column of sedimentary rocks in Gale crater investigated by *Curiosity*. Comparisons of the mineralogy of four drill samples from the Jura member exposed at VRR and GT (blue circles; see also Fig. 2) reveal diagenetic processes. YB, Yellowknife Bay; K, Kimberley; SP, Siccar Point; S, Stimson; mudst, mudstone; sltst, siltstone; sandst, sandstone.

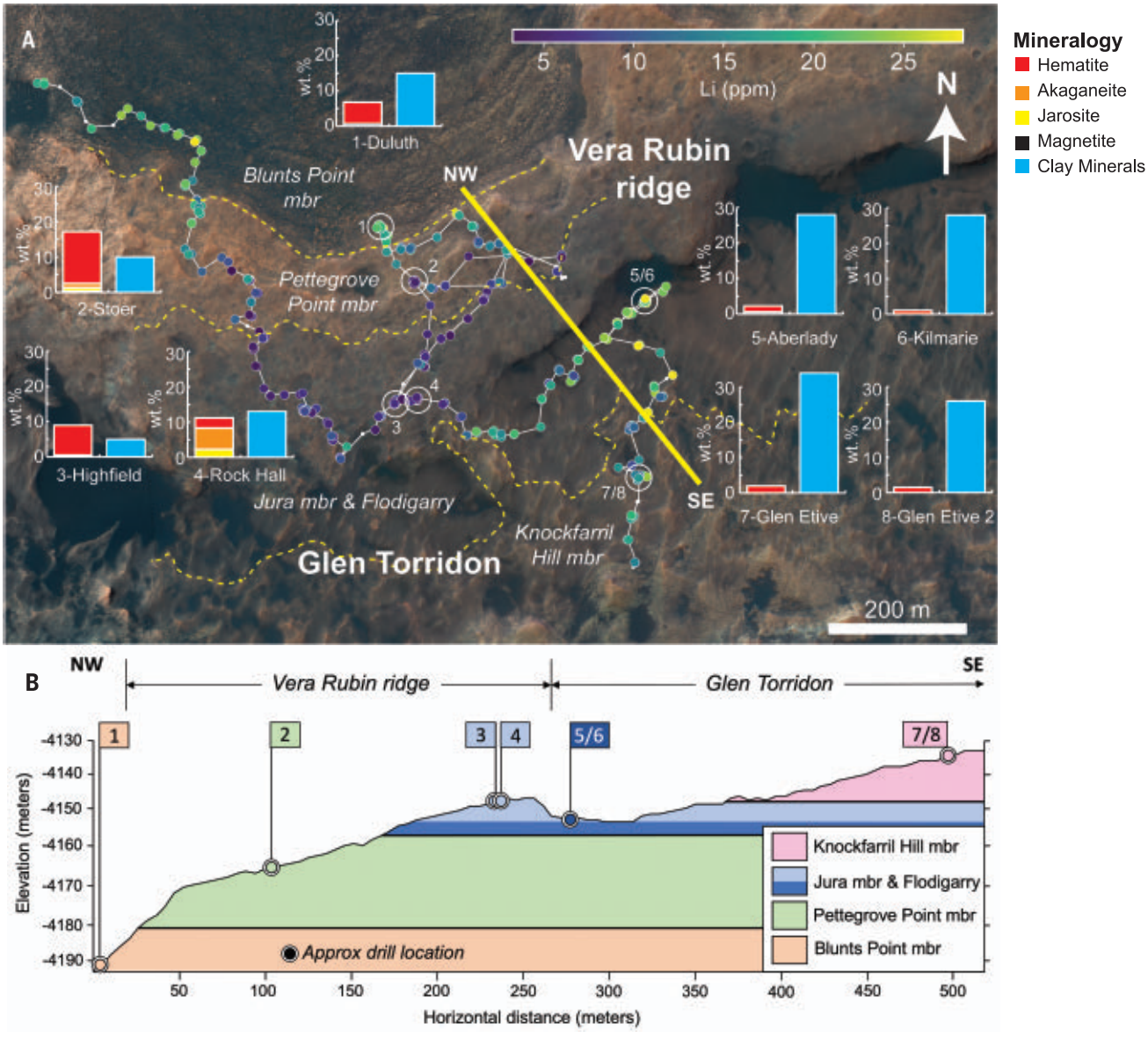


Fig. 2. Mineralogical, geochemical, stratigraphic, and geomorphologic overview of the areas of VRR and GT explored by *Curiosity*. (**A**) Mosaic of orbital images (in visible wavelengths) of VRR and GT (*40*) showing the rover path (white line), entering from the top left of the image, and progressing to GE in the southeast (SE). Drill locations along the path are indicated with white numbered circles. The numbers correspond to bar charts showing the abundances of clay minerals as well as iron oxide, oxyhydroxide, and sulfate minerals derived from CheMin data. ChemCam measurements of rock lithium abundances, a proxy for clay mineral content (*40*), are shown as small colored circles (see color bar). The straight yellow line shows the position of the cross section in (B). Dashed yellow lines trace the boundaries between stratigraphic units. (**B**) Stratigraphic cross section showing the surficial expression of Murray formation members (colored bands) and the approximate orthogonal projection of the drill sample locations marked as circles. The vertical exaggeration is 2.1. mbr, member.

magnesium-rich, silica-poor mafic and ultramafic volcanic rocks, as well as magnesium-rich sedimentary precursors (typically carbonates). Fe^{2+} substitution in talc (as inferred for KM) is usually characteristic of volcanic precursors (*30*). Talc can form directly from olivine and pyroxene in mafic and ultramafic volcanic rocks subjected to subsurface hydrothermal alteration, and in surficial weathering environments (*31*, *32*). From the chemical composition, mineralogy, and the degree to which sedimentary textures are preserved in GT, we infer that in situ S-T production within GT rocks is unlikely (see supplementary text). Instead, we suggest that S-T (a minor phase in GT rocks) was transported into Gale crater as sedimentary detritus, derived from erosion of older bedrock. As discussed below, our proposed origin of S-T is unlike that of smectites found in the same samples and in rocks from older stratigraphic levels, which likely formed close to the time of deposition of sedimentary strata (*4*, *5*, *27*).

Although we cannot exclude inputs of hydrothermal talc derived from basaltic rocks in Gale's catchment, any such sources would also have contributed chloritized aluminum-bearing phyllosilicates, zeolites, and amphiboles to GT rocks (*27*, *31*)—minerals that have not been identified in *Curiosity* data. Although mixed-layer S-T is rarely reported on Earth, it occurs as a retrograde product of metamorphosed ultramafic rocks (*33*) and forms during synthesis of serpentine minerals under experimental hydrothermal conditions (*34*). These results lead us to hypothesize that inputs of S-T to GT sediments were derived from ultramafic

lithologies in Gale's catchment transported to the crater floor by rivers.

Talc- and serpentine-bearing ultramafic rocks are products of serpentinization reactions, which generate free H_2 and methane (both greenhouse gases on Mars) and sequester inorganic carbon as carbonate minerals (*35*). The production rate of H_2 and hydrocarbons varies considerably, depending on mineral-reaction pathways (*36*). Local physicochemical conditions affect the potential astrobiological relevance of these environments (*37*). We can neither identify nor examine the source of S-T in Gale's catchment, so cannot assess its habitability. Nevertheless, our results indicate that serpentinization of ultramafic rocks took place in the vicinity of Gale prior to deposition of GT sediments at ~3.5 Ga.

Origin of smectite

The rover data on the composition, structural state, and abundance of clay minerals confirm orbital data that indicated Fe^{3+}-rich smectite is present along and adjacent to the rover traverse path through GT (*19*). The orbital detections were based on absorption features in reflectance spectra, including one arising from Fe^{3+}-OH vibrational modes centered at ~2.28 to 2.29 μm (*19*). The ~2.28-μm feature also exhibits a weak absorption at ~2.2 μm, which was assigned to Al-OH vibrations (*19*). The latter feature was originally interpreted as either (i) the presence of a second phase, such as montmorillonite or poorly crystalline kaolinite, or (ii) the presence of Al^{3+} in octahedral sites of a ferric smectite. The second interpretation is consistent with the *b* axis crystal cell dimensions of smectite that we derived from the XRD data.

Orbital spectral absorptions attributed to clay minerals were weaker and less spatially coherent in Murray formation strata traversed by *Curiosity* prior to arrival at GT (*5*). Nevertheless, previous samples analyzed by CheMin commonly contained smectite, with abundances of up to 28 wt % (*5*, *12*–*14*). Earlier in the mission, smectite was also found in the older Sheepbed mudstone of the Yellowknife Bay formation (*4*, *27*) (Fig. 1). The chemistry of smectite in Gale crater is variable: magnesium-rich trioctahedral varieties are found in the Karasburg member of the Murray formation rocks (Fig. 1), but smectites become increasingly rich in dioctahedral cations (Al^{3+} and Fe^{3+}) at higher stratigraphic levels, approaching VRR (*5*, *13*, *14*). Dioctahedral smectite is thought to be a product of oxidative chemical weathering in sediment source areas, during transport, or on the lake floor (*5*, *7*). This hypothesis is supported by reductions in the mafic mineral content of dioctahedral smectite-bearing sediments compared to underlying rock strata, sedimentological evidence of episodic lake desiccation, and bulk chemical signatures of weathering-related element mobilization (*5*, *7*).

The bulk mineralogy, geochemistry, and depositional environment of GT rocks are similar to those of the underlying members of the Murray formation, and therefore we propose a similar origin for the Fe^{3+}-rich smectite in GT. Variations in the iron, aluminum, and magnesium content of smectites are commonly observed in Earth lake systems—a consequence of variable lithologies and extent of weathering of catchment rocks, as well as processes that may modify the chemistry of clay minerals once they are deposited in the lake (*38*, *39*). The detection of S-T indicates that one possible cause for the higher abundance of iron in GT smectite compared with other parts of the Murray formation is a change in sediment provenance, with increasing smectite contributions from well-drained mafic or ultramafic rocks. An alternative is that clay minerals in GT sediments were not subject to processes capable of enriching smectite in magnesium and aluminum at the expense of iron, as documented in other parts of the Murray formation (*5*, *13*). The absence of trioctahedral smectites, and the lack of sedimentological evidence of desiccation seen in stratigraphically lower parts of the Murray formation, both support the latter hypothesis.

Diagenetic drivers

Textural, mineralogic, and geochemical observations from *Curiosity* have shown that VRR experienced enhanced alteration and cementation, making these rocks harder than other Murray formation exposures (*14*, *15*). Heterogeneities in the color (red, purple, gray) and spectral properties of VRR rocks are observed cross-cutting the primary bedding and are attributed to enhanced crystallization of hematite (*15*). Centimeter-scale iron-poor halos surround iron oxide–rich, dark-toned diagenetic features and crystal pseudomorphs resembling gypsum (*15*, *16*). Overall, the bulk chemistry of VRR is like that of the rest of the Murray formation, but diagenetic features and geochemical variability observed across the ridge indicate that elements, including iron, were locally remobilized (*15*, *16*, *40*). Several models have been proposed to explain the enhanced alteration of VRR (*15*). Most argue that VRR rocks were preferentially infiltrated by late diagenetic fluids, with flow focused along the contact between the Murray formation and a younger rock unit called the Stimson formation (Fig. 1).

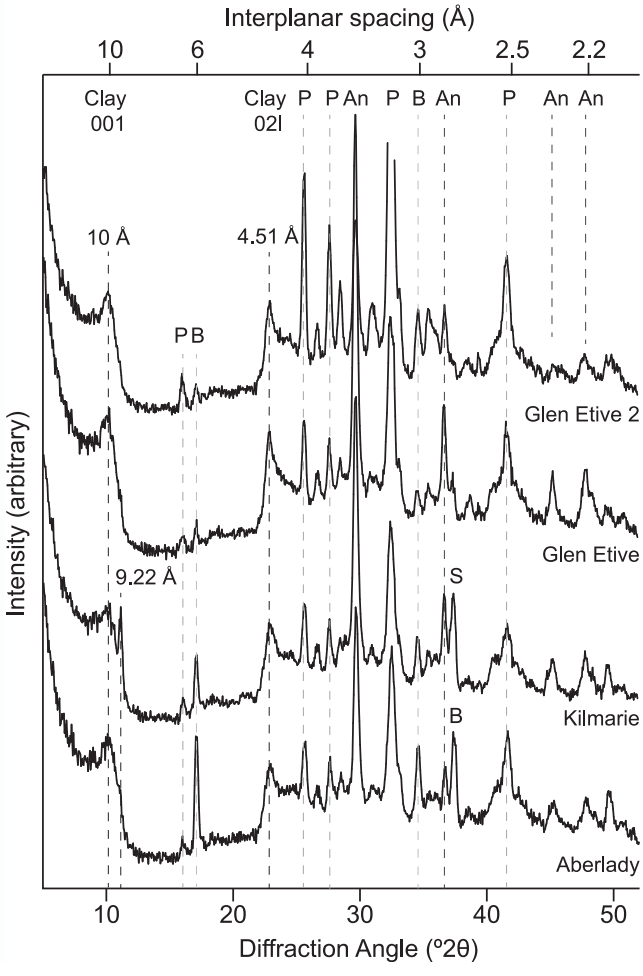


Fig. 3. XRD data from GT drill samples AL, KM, GE, and GE2. The peaks of clay minerals and other major mineral components are labeled and marked by dashed vertical lines. An, anhydrite; B, bassanite; P, plagioclase feldspar; S, siderite. KM data show a peak at ~11.1 °2θ (~9.22 Å), which we interpret as a mixed-layer serpentine-talc (*25*). Data have been arbitrarily shifted vertically for display. XRD data were collected by using Co Kα radiation (*25*).

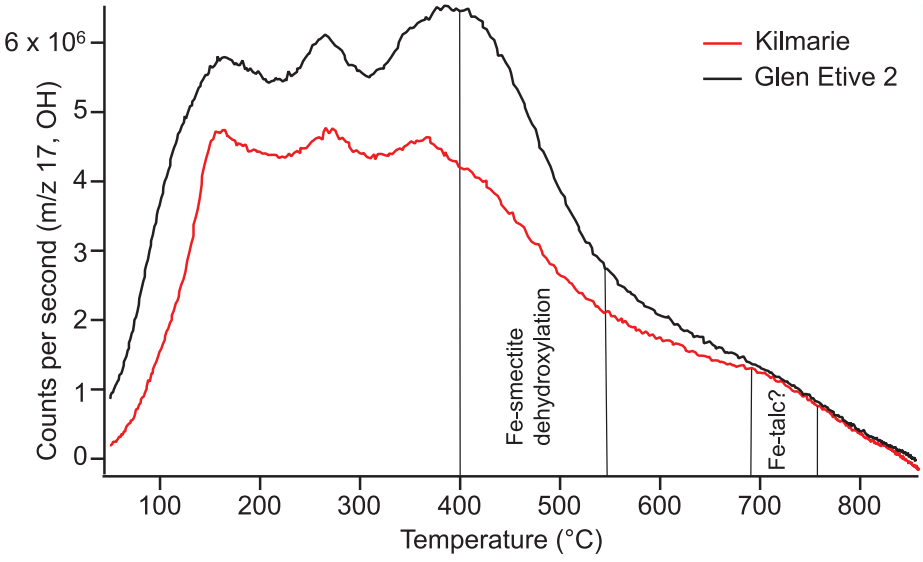


Fig. 4. Evolved H_2O traces from the SAM analysis of the KM (red line) and GE2 (black line) drill samples. H_2O, with a mass-to-charge ratio (*m/z*) of 18, saturated the quadrupole mass spectrometer detector during analyses; therefore, the signal from OH (*m/z* = 17) was used to study the temperature-dependent evolution of H_2O from KM and GE2.

Hypothesized fluid sources include (i) heated water migrating from the deeper parts of the crater or (ii) gravity-driven regional subsurface hydraulic systems with flow from Mars' southern highlands to the northern lowlands (*14*, *15*).

The types of minerals and their abundances in GT and VRR rocks sampled from the same stratigraphic interval provide evidence that the escarpment between GT and VRR marks a diagenetic front (*14*, *15*). GT drilled samples have clay mineral contents of 26 to 34 wt % (Table 1), compared to 5 to 13 wt % in VRR (*14*) (Fig. 2). Iron oxide and oxyhydroxide abundances show an opposite trend, with levels of 9 to 16 wt % in VRR samples (*14*) compared to ≤2 wt % in GT (Table 1 and Fig. 2). Unlike VRR samples, akaganeite, jarosite, or opaline silica were not detected in GT samples (*14*). Abundances of lithium, measured by the ChemCam instrument using laser-induced breakdown spectroscopy (*25*), provide a proxy for clay mineral content of rocks (*40*). Thousands of ChemCam data points collected across VRR and GT confirm that the drilled samples are representative of the differing clay mineral content of GT and VRR rocks (Fig. 2). These chemical and mineralogic differences cannot be explained by depositional processes, such as hydrodynamic sorting or variations in sedimentary sources, because of the stratigraphic equivalence of the units in immediately adjacent areas (Fig. 2).

Comparisons of GT and VRR mineralogy provide additional constraints on the diagenetic history of VRR. Our analysis and interpretation of GT clay minerals show that Fe^{3+}-rich smectites and Fe^{2+}-bearing S-T, originally deposited in lake sediments, were converted to iron oxides, oxyhydroxides, and opaline silica in VRR. A partially analogous reaction pathway is commonly observed in Earth weathering environments, where rainwater infiltrates soils, preferentially leaching silicon from smectite, forming iron oxides and aluminum-rich materials, such as allophane or kaolin group clay minerals (*32*). However, the lack of vertical elemental and mineralogical gradients shows that VRR has not experienced top-down weathering (*15*). Destruction of smectite during burial and diagenesis is rarely observed in Earth sedimentary basins, because of the strong influence of silica activity on the thermodynamic stability (fig. S3). Siliceous materials filling sedimentary basins typically keep fluids enriched in silica, which may lead to the precipitation of additional clay minerals as cements during burial and lithification of porous sedimentary rocks (*1*). The subsurface sources of late diagenetic fluids proposed to account for alteration at VRR (*14*, *15*) would also likely be high in silica through buffering with basalts and basaltic detritus (*41*) and thus cannot account for the destabilization of smectites. Another mechanism and source of fluids is required.

Brine-driven diagenesis

We propose that the conversion of Fe^{3+}-smectites and Fe^{2+}-substituted S-T into iron oxides and oxyhydroxides in VRR involved density-driven circulation of oxidizing, silica-poor brines (fig. S3) originating from the overlying sulfate-bearing unit (SBU) previously identified in orbital observations. The mode of deposition of the SBU cannot be determined from orbital data, and *Curiosity* is in the early stages of investigating the SBU. However, as indicated by the lateral extent of SBU exposures around Mount Sharp and the correspondence of orbital mineral variation to bedding, the SBU (or parts of it) could be a continuation of the lacustrine conditions within Gale, albeit with a shift in climatic and/or local hydrological conditions that led to the deposition of magnesium sulfate minerals (*19*, *20*).

Table 1. Mineralogical composition (wt %), with 1-σ errors, of drill samples from GT (Fig. 2). AL and KM are part of the Jura member. GE and GE2 come from the Knockfarril Hill member (Fig. 1). The detection limit for crystalline materials is ~1 wt %.

Mineral	Aberlady	Kilmarie	Glen Etive	Glen Etive 2
Andesine	10.8 ± 1.2	7.8 ± 0.9	11.4 ± 0.8	23.5 ± 1.1
Hematite	1.7 ± 0.4	0.9 ± 0.4	2.0 ± 0.7	1.5 ± 0.5
Magnetite	Trace	–	–	–
Ca sulfate	11.6 ± 0.6	9.4 ± 0.3	10.4 ± 0.4	5.0 ± 0.3
Sanidine	1.2 ± 0.5	Trace	1.4 ± 0.7	2.4 ± 0.6
Pyroxene	4.6 ± 1.7	3.2 ± 1.1	1.6 ± 0.5	4.2 ± 1.0
Quartz	Trace	Trace	Trace	Trace
Siderite	–	1.9 ± 0.2	–	Trace
Clay minerals	28 ± 5	28 ± 5	34 ± 6	26 ± 5
Amorphous	41 ± 20	48 ± 24	38 ± 19	37 ± 18

Orbital sulfate detections could also originate from sulfate-cemented clastic sedimentary rocks deposited in aqueous or eolian settings (*19*).

Irrespective of the mode of deposition, orbital observations of the ~400-m-thick SBU show that it is associated with spectral absorptions consistent with mono- and polyhydrated magnesium sulfates (*19*, *20*). These magnesium sulfates are highly soluble, and their precipitation requires intense evaporative concentration and the formation of dense brines (*42*). Combined with orbital data indicating possible conformable relationships with underlying perennial lake sediments, the thickness and areal extent of the SBU imply that density-driven brine infiltration likely occurred prior to the lithification of the Murray formation. This process is observed on Earth (*43*, *44*). For example, in the Tyrrell Basin of southeastern Australia, evaporative brines generated in salt lakes have infiltrated tens of meters into underlying freshwater lake sediments over the last ~30,000 years, in response to changing climate (*44*).

Sulfate-dominated brines that evolve in saline lake systems on Earth tend to maintain neutral to slightly acidic pH (*45*). This allows brines to become depleted in silica through precipitation of amorphous silica, clay minerals, and feldspars as brines are progressively concentrated (*46*, *47*). In the Lake Lewis Basin, central Australia, a climate-induced change from hydrologically open conditions to a closed salt-lake system resulted in the infiltration of brines into older, clay-mineral–rich, perennial lacustrine sediments (*46*). Brines beneath the dry lake are depleted in silica (<10 parts per million), resulting in the destabilization of detrital aluminosilicate components of older perennial deposits, including feldspar and smectite, and transformation into substantial portions (up to ~30 w %) of x-ray amorphous aluminosilicate materials (*46*). Therefore, we propose that the most concentrated brines generated during the deposition of the SBU could have also been depleted in silica and capable of destabilizing smectites and S-T. Exposure to oxidized brines, descending from the near-surface, provided an additional driver promoting the alteration of Fe^{2+}- substituted S-T.

Brine-driven diagenesis is predicted to be heterogeneous (*43*, *46*). On Earth, concentrated brines develop by progressive evaporation of fresher waters that migrate from marginal areas of sedimentary basins to the center. This process produces strong lateral gradients in aqueous geochemistry that follow chemical pathways governed by the initial water chemistry and precipitation of minerals during migration (*45*). Geochemical gradients may be preserved in the rock record as spatially distinct mineral facies (*43*). The position of mineral facies boundaries can vary stratigraphically, in response to the migration of the depocenter and changing hydrological conditions (*43*).

In Gale, the most concentrated brines would have been spatially restricted to topographic lows on the crater floor at any given time and migrated in response to variations in sediment supply and freshwater recharge. The interactions of brine with underlying perennial lacustrine sediments likely occurred before substantial burial, at temperatures of less than 50°C. Under these conditions, the nature and extent of mineralogical reactions would have been subject to kinetic controls and might not have reached equilibrium. For example, recrystallization of a secondary clay mineral after smectite dissolution may have been prevented by a combination of short residence of fluids and fast iron oxide and amorphous aluminosilicate precipitation. Reaction kinetics could have been governed by local variations in the geochemistry of interstitial fluids, as well as the composition and physical properties of sediments, such as permeability and surface area. Thus, the inherent heterogeneity of brine-driven diagenesis accounts for lateral differences in the conversion of clay minerals to iron oxide and oxyhydroxide in GT and VRR. This heterogeneity could also explain the localized mobilization and recrystallization of iron observed on VRR, through complexation of iron with highly concentrated SO_4^{2-} and possibly (based on the presence of the mineral akaganeite) Cl^--rich brines (*13*, *14*, *16*).

Other evidence for the influence of brines

The influence of descending brines on other intervals of the Murray formation appears to be more pervasive than previously documented. Heterogeneous, low-temperature interactions of descending brines provide a plausible explanation for the co-occurrence of calcium sulfates in multiple hydration states (gypsum, bassanite, and anhydrite) observed through much of the Murray formation (*10*, *14*). Halite has been detected directly by XRD in a drill sample, called Quela, from the Murray formation below VRR (*13*) but may be more widespread as shown by chlorine detections using ChemCam (*48*). Temperatures of SO_2 release in SAM EGA experiments indicate that x-ray amorphous magnesium and iron sulfates are present in many Murray formation drill samples (*10*, *26*, *28*). This is confirmed by positive linear correlations between the magnesium and sulfur content of bedrock, often in areas where concretionary or dendritic diagenetic features are observed (*9*). Patchy occurrences of polyhydrated magnesium sulfates have been detected in parts of GT (in an area adjacent to GE and GE2) and in some parts of the Murray formation stratigraphically beneath VRR, according to orbital spectral reflectance data (*20*).

Certain intervals of the Sutton Island and Blunts Point members of the Murray formation bedrock (Fig. 1) are particularly enriched in magnesium sulfates, as determined by ChemCam analysis (*49*). Although others have proposed that these magnesium sulfates precipitated at times when the lake dried out (*49*), they lack textural features associated with intense desiccation. Therefore, later enrichment by brines from overlying sediments cannot be ruled out (*49*). Magnesium and iron sulfates detected in other parts of the stratigraphy are unlikely to represent precipitation close to the time of deposition either, as sedimentological analysis shows that much of the Murray formation was deposited under perennial lake conditions (*8*, *9*, *23*). As revealed by the radiometric dating of the iron sulfate mineral jarosite (*18*), and magnesium sulfate-rich diagenetic concretions observed in the younger Stimson formation (Fig. 1) (*9*), brines appear to have continued to be intermittently active within Gale sediments after the lithification of the Murray formation as well.

Brine-driven diagenesis also helps to account for the ubiquity of x-ray amorphous materials in samples from Gale crater (*13*, *14*, *17*). Mass-balance calculations indicate that amorphous constituents include sulfates, iron, and aluminosilicate components, whose overall proportions vary from sample to sample (*13*, *17*). Comparisons with Lake Lewis sediments from Australia indicate that brines may have promoted the formation of x-ray amorphous materials at Gale through destabilization of clay minerals and possibly feldspar (*46*).

Implications for the martian sedimentary record

Tectonically driven subsidence and burial are the main drivers of diagenesis in sedimentary basins on Earth. Mars' sedimentary record has been preserved by the lack of tectonics, lower geothermal gradients, and timing of desiccation of the planet, which limited aqueous mineral reactions (*3*). The global, but temporally discrete, distribution of sulfate deposits on Mars is thought to record this desiccation process (*21*). Evidence for top-down diagenesis by infiltration of brines in Gale crater, combined with the global distribution of sulfates produced by Mars' shrinking hydrological budget, raise the possibility of more widespread alteration of older clay-mineral–bearing sediments through a mechanism that is rare on Earth. Brines can destabilize clay minerals and destroy intimate associations with organic molecules, potentially reducing the preservation capacity of the rock record. However, subsequent readsorption on sulfates may be an efficient check on organic decay (*50*). Although brine-driven destruction of clay minerals complicates geological interpretation, water released during this process would have acted as a negative feedback, slowing the pace of planetary desiccation and potentially extending

the period in which Mars had surface liquid water.

REFERENCES AND NOTES

1. C. E. Weaver, *Clays, Muds, and Shales* (Elsevier, 1989), p. 819.
2. R. E. Summons *et al*., *Astrobiology* **11**, 157–181 (2011).
3. J. P. Grotzinger, R. E. Milliken, The sedimentary rock record of Mars: Distribution, origins, and global stratigraphy, in *Sedimentary Geology of Mars*, J. P. Grotzinger, R. E. Milliken, Eds. (SEPM Special Publication, Tulsa, OK, 2012), vol. 102.
4. D. T. Vaniman *et al*., *Science* **343**, 1243480 (2014).
5. T. F. Bristow *et al*., *Sci. Adv.* **4**, eaar3330 (2018).
6. J. L. Eigenbrode *et al*., *Science* **360**, 1096–1101 (2018).
7. N. Mangold *et al*., *Icarus* **321**, 619–631 (2019).
8. J. P. Grotzinger *et al*., *Science* **350**, aac7575 (2015).
9. J. A. Hurowitz *et al*., *Science* **356**, eaah6849 (2017).
10. D. T. Vaniman *et al*., *Am. Mineral.* **103**, 1011–1020 (2018).
11. A. S. Yen *et al*., *Earth Planet. Sci. Lett.* **471**, 186–198 (2017).
12. E. B. Rampe *et al*., *Earth Planet. Sci. Lett.* **471**, 172–185 (2017).
13. C. N. Achilles *et al*., *J. Geophys. Res. Planets* **125**, JE006295 (2020).
14. E. B. Rampe *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006306 (2020).
15. A. A. Fraeman *et al*., *J. Geophys. Res. Planets* **125**, JE006527 (2020).
16. J. L'Haridon *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006299 (2020).
17. R. J. Smith *et al*., *J. Geophys. Res. Planets* **126**, e2020JE006782 (2021).
18. P. E. Martin *et al*., *J. Geophys. Res.* **122**, 2803–2818 (2017).
19. R. E. Milliken, J. P. Grotzinger, B. J. Thomson, *Geophys. Res. Lett.* **37**, L04201 (2010).
20. R. Y. Sheppard, R. E. Milliken, M. Parente, Y. Itoh, *J. Geophys. Res. Planets* **126**, e2020JE006372 (2021).
21. J. P. Bibring *et al*., *Science* **312**, 400–404 (2006).
22. K. M. Stack *et al*., *Sedimentology* **66**, 1768–1802 (2019).
23. L. A. Edgar *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006307 (2020).
24. N. T. Stein *et al*., *Geophys. Res. Planets* **125**, e2019JE006298 (2020).
25. Materials and methods are available as supplementary materials.
26. B. Sutter *et al*., *J. Geophys. Res.* **122**, 2574–2609 (2017).
27. T. F. Bristow *et al*., *Am. Mineral.* **100**, 824–836 (2015).
28. A. C. McAdam *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006309 (2020).
29. M. Heuser, P. Andrieux, S. Petit, H. Stanjek, *Clay Miner.* **48**, 97–103 (2013).
30. B. W. Evans, S. Guggenheim, *Rev. Mineral. Geochem.* **19**, 225–294 (1988).
31. H. Y. McSween Jr., T. C. Labotka, C. E. Viviano-Beck, *Meteorit. Planet. Sci.* **50**, 590–603 (2015).
32. M. J. Wilson, *Clay Miner.* **39**, 233–266 (2004).
33. D. R. Veblen, P. R. Buseck, *Science* **206**, 1398–1400 (1979).
34. P. C. Ryan, F. J. Huertas, L. N. Pincus, W. Painter, *Clays Clay Miner.* **67**, 488–506 (2019).
35. E. S. Amador, J. L. Bandfield, N. H. Thomas, *Icarus* **311**, 113–134 (2018).
36. F. Klein, W. Bach, T. M. McCollom, *Lithos* **178**, 55–69 (2013).
37. S. Q. Lang, W. J. Brazelton, *Philos. Trans. A Math. Phys. Eng. Sci.* **378**, 20180429 (2020).
38. L. L. Baker, O. K. Neill, *Am. Mineral.* **102**, 1632–1645 (2017).
39. J. F. Banfield, B. F. Jones, D. R. Veblen, *Geochim. Cosmochim. Acta* **55**, 2795–2810 (1991).
40. J. Frydenvang *et al*., *Geophys. Res. Planets* **125**, e2019JE006320 (2020).
41. S. M. McLennan, *Geology* **31**, 315–318 (2003).
42. N. J. Tosca, S. M. McLennan, M. P. Lamb, J. P. Grotzinger, *J. Geophys. Res.* **116** , E05004 (2011).
43. D. Larsen, *Geosphere* **4**, 612–639 (2008).
44. P. G. Macumber, *Chem. Geol.* **96**, 1–18 (1992).
45. D. M. Deocampo, B. F. Jones, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, 2014), vol. 7, pp. 437–469.
46. P. M. English, *Sediment. Geol.* **143**, 219–244 (2001).
47. J. Jankowski, G. Jacobson, *J. Hydrol. (Amst.)* **108**, 123–173 (1989).
48. N. H. Thomas *et al*., *Geophys. Res. Lett.* **46**, 10754–10763 (2019).
49. W. Rapin *et al*., *Nat. Geosci.* **12**, 889–895 (2019).
50. R. dos Santos, M. Patel, J. Cuadros, Z. Martins, *Icarus* **277**, 342–353 (2016).
51. T. Bristow, Data from: Brine driven destruction of clay minerals in Gale crater, Mars., Dryad, (2021); https://doi.org/10.5061/dryad.k3j9kd576.

ACKNOWLEDGMENTS

We thank R. Kleeberg for help developing the CheMin instrument profile model for BGMN and A. Derkowski for informative discussion. J. C. Corona kindly provided XRD data from Fe talc, and P. De Deckker shared knowledge of Australian lakes. We thank J. Bishop and two anonymous reviewers for helping improve the manuscript.
We acknowledge the support of the Jet Propulsion Lab engineering and management teams and MSL science team members who participated in tactical and strategic operations, without whom the data presented here could not have been collected. **Funding:** Some of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). A.A.F., A.C.M., and C.M.F. acknowledge funding by the MSL Participating Scientist Program, NASA solicitation NNH15ZDA001N. J.F. acknowledges support from the Carlsberg Foundation. J.P.G. received additional funding from the Simons Collaboration for the Origin of Life. **Author contributions:** T.F.B. wrote the manuscript, with corrections, discussions, and/or revised text from coauthors. J.C. led one-dimensional modeling of clay minerals. S.J.C. quantified the abundances of clay minerals and the x-ray amorphous phase from CheMin XRD data. G.W.D. led efforts to identify the 9.22-Å phase from XRD data. C.M.F. led efforts to establish the relationships between rock strata exposed at VRR and GT, capturing them in Figs. 1 and 2. J.F. analyzed ChemCam Li data and helped produce Fig. 2. A.C.M. led the interpretation of SAM EGA data and helped produce Fig. 4 and fig. S4. A.S.Y. compiled the bulk geochemical data from the Alpha Particle X-ray Spectrometer shown in table S1 and used to generate fig. S5. R.V.M. compiled the XRD data used to determine the angular resolution of the CheMin instrument (fig. S2). R.E.M. led comparisons of orbital and ground-based mineral data. C.N.A., T.F.B., D.W.M., S.M.M., E.B.R., M.T.T., V.T., and D.T.V. determined the abundances of crystalline phases in GT samples from XRD data. D.F.B., A.B.B., K.A.B., A.A.F., V.K.F., R.G., J.P.G., P.R.M., E.B.R., D.T.V., R.C.W., and A.R.V. designed the rover instruments and guided the mission. All authors performed operational roles in data collection. **Competing interests:** V.K.F. is also affiliated with the Department of Physics and Astronomy, Carleton College, Northfield, MN 55057, USA. **Data and materials availability:** All *Curiosity* data presented in this paper are archived in NASA's Planetary Data System; the URLs and file identifications are listed in table S5. The software written by the authors of this paper and files needed to replicate the analyses are publicly available at Dryad (*51*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/198/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S5
References (*52*–*79*)

24 January 2021; accepted 28 May 2021
10.1126/science.abg5449

REPORTS

GEOMORPHOLOGY

The life span of fault-crossing channels

Kelian Dascher-Cousineau*, Noah J. Finnegan, Emily E. Brodsky

Successive earthquakes can drive landscape evolution. However, the mechanism and pace with which landscapes respond remain poorly understood. Offset channels in the Carrizo Plain, California, capture the fluvial response to lateral slip on the San Andreas Fault on millennial time scales. We developed and tested a model that quantifies competition between fault slip, which elongates channels, and aggradation, which causes channel infilling and, ultimately, abandonment. Validation of this model supports a transport-limited fluvial response and implies that measurements derived from present-day channel geometry are sufficient to quantify the rate of bedload transport relative to slip rate. Extension of the model identifies the threshold for which persistent change in transport capacity, obliquity in slip, or advected topography results in reorganization of the drainage network.

The recognition in 1908 that ephemeral streams in the Carrizo Plain in southern California preserved a passive record of lateral fault offset transformed the study of active strike-slip faults (*1*–*5*). However, offset on these stream channels generally does not exceed tens to hundreds of meters. Flow eventually overtops channels, typically spilling straight across the fault and resetting the recorded offset (Fig. 1, A to C) (*2*, *3*, *6*). Variability in channel offsets suggests a reset mechanism that depends primarily on local channel configuration rather than regional climate or earthquake history. Slip rates on the order of centimeters per year on the San Andreas Fault imply that channels reset on millennial time scales, thereby limiting their utility as recorders of fault slip (*2*). We instead leveraged this tectonic-geomorphic interaction to better understand how the drainage network responds to perturbation. We developed a quantitative prediction for the channel life span as a function of fault slip rate, present-day channel geometry, and sediment transport capacity assuming transport-limited conditions. We validated our model by testing it against a LiDAR (light detection and ranging)–derived record of fault-crossing channels in the Carrizo Plain and explore implications for the long-term evolution of strike-slip fault landscapes.

Wallace (*2*) identified channel sedimentation as a control on the evolution of fault-crossing

Department of Earth and Planetary Sciences, University of California, Santa Cruz, CA, USA.
***Corresponding author. Email: kdascher@ucsc.edu**

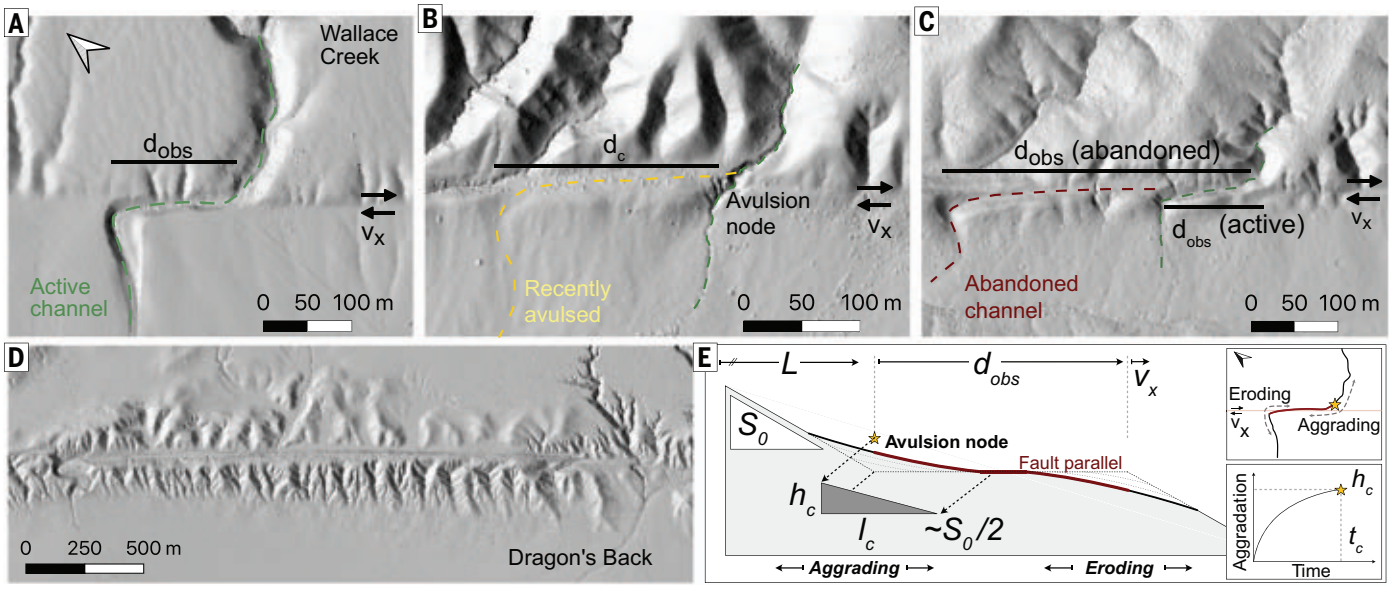


Fig. 1. The life span of fault-crossing channels. (**A** to **D**) Hillshading (*28*) illustrates, in progressive stages, (A) an offset channel, (B) a recently avulsed or reset channel at the critical offset d_c, and (C) a subsequently offset channel. (D) Large fault-parallel drainage along the advection path from the Dragon's Back. In this case, the penultimate avulsion caused a pulse of incision due to the resulting steeper gradient. Locations for (A) to (D) are shown in fig. S4. (**E**) Idealized elevation profile of an offset channel (shown in plan view in the upper inset), with a reach L and offset d_{obs}, undergoing horizontal fault slip v_x. Offset introduces a fault-parallel segment with near-horizontal slope (black dotted lines); the channel responds evolving with time (gray dotted lines) up until its current geometry (bold line). Avulsion occurs at time t_c if aggradation approaches the threshold height h_c (lower inset), which can be related to the upstream aggradation length scale l_c and the initial slope S_0 (gray triangle).

streams (*7*, *8*). He noted that offset increases channel length without changing the drop in elevation (Fig. 1A), thus reducing channel gradients in fault-crossing reaches. Where a channel enters a fault zone, the shallowing slope associated with a reduction in sediment transport causes deposition on the channel floor (*9*). Wherever the combined depth of aggrading sediment and water approaches the height of local topographic barriers, flow is susceptible to diversion in a sudden abandonment of the channel, known as an avulsion (*10*). Flow then incises a new channel in the steepest descent direction, typically orthogonal to the fault trace. In this way, avulsion resets the offset and orientation of fault-crossing channels (*2*, *7*) (Fig. 1B). The upstream bend locally incurs the most pronounced change in channel slope and therefore marks an avulsion node, a persistent point for subsequent avulsions to occur (Fig. 1C), much in the same way that the backwater length scale in lowland coastal rivers controls the location of geologically persistent flow deceleration, sedimentation, and therefore chronic avulsion in deltas (*11*). Trenching and geochronology along channels in the Carrizo Plain corroborate repeated cycles of aggradation, avulsion, and incision (*4*, *7*).

The importance of sedimentation in the life span of fault-crossing channels suggests that bedload transport, rather than the stream's ability to erode the bed, limits their long-term evolution. In a transport-limited channel, the rate of aggradation, v_a [L/T], implicitly determines the time required for a channel to avulse, t_c [T]:

$$h_c(t_c) = \int_0^{t_c} v_a(t)dt \qquad (1)$$

(*12*), where the avulsion threshold $h_c(t)$ is the net height of aggraded sediment required for the channel to overflow its container, which may be either a bank or a valley wall. Equation 1 evaluates the avulsion threshold at the time of avulsion, t_c; before this point in time, aggradation in active channels can never exceed the threshold. Throughout, the value of $h_c(t)$ may vary as a result of dip-slip along the fault, or because of advection of topographic features such as shutter ridges, depressions, or abandoned channel heads.

We first expand Eq. 1 with particular consideration of aggradation at the avulsion node and examine how the present-day geometry constrains and validates it, and then explore the influence of the avulsion threshold's time dependence. If we assume bedload transport capacity to be linearly related to channel slope, then the rate of erosion or aggradation [L/T] is proportional to a change in slope in the channel profile (*13*),

$$v_a = -\kappa\frac{\partial S}{\partial x} \qquad (2)$$

where x is the distance along the channel profile [L], S is the channel slope, and κ is a diffusivity coefficient [L^2/T]. The diffusivity coefficient represents the volumetric sediment transport capacity per unit channel width and slope. A channel conveying more sediment has a higher diffusivity coefficient and adjusts more rapidly to perturbations in the channel slope.

The channel elevation profile allows for useful approximations of the right side of Eq. 2 (Fig. 1E). For a narrow fault zone, offset introduces a fault-parallel segment with slope S_f, which is much smaller than the original slope S_0, such that $\Delta S = S_f - S_0 \approx -S_0$ approximates the slope change without aggradation. In response, a wedge of sediment near the avulsion node grows as a symmetric diffusive pulse. We define its horizontal half-width at the time of avulsion as l_c and thus approximate the channel response leading up to the avulsion as $-\kappa(\partial S/\partial x) \approx \kappa(S_0/l_c)$.

To relate Eq. 1 and Eq. 2, we cast l_c in terms of the initial channel slope S_0 and the avulsion threshold height h_c. After aggradation to the threshold level, the rise $S_0 l_c$ accommodates twice its original run, recovering half the original slope (Fig. 1E). This leads to the relationship $S_0/2 \approx h_c/l_c$ or, equivalently, $l_c \approx 2h_c/S_0$, an approximation that is consistent with measurements of millennial-scale slope changes along alluvial rivers induced by check-dams (*14*) and holds when channel response is rate-limiting

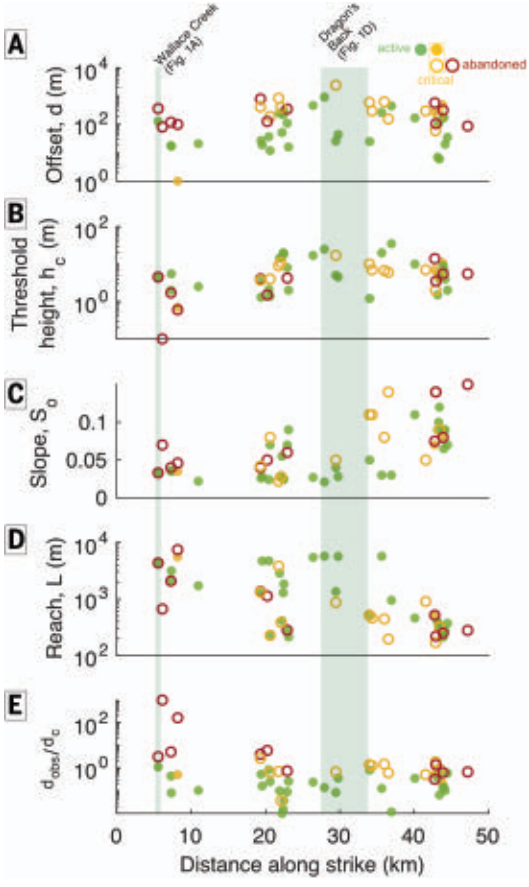


Fig. 2. Channel geometry along the San Andreas Fault in the Carrizo Plain. (**A** to **E**) Measurements of (A) offset d_{obs}, (B) initial avulsion threshold height h_c, (C) initial slope S_0, (D) channel reach L, and (E) normalized offset d_{obs}/d_c (table S1). Solid markers indicate active channels; open markers indicate abandoned channels. Yellow markers indicate channels that have evidence for incipient or recent avulsions. No individual set of measurements [(A) to (D)] appears sufficient to separate active and abandoned channels. Offsets in (A) and a slip rate of 3.5 cm/year are consistent with a channel response over millennial time scales. The uncharacteristically large offsets in (A) collapse near unity in (E) when normalized by the critical offset.

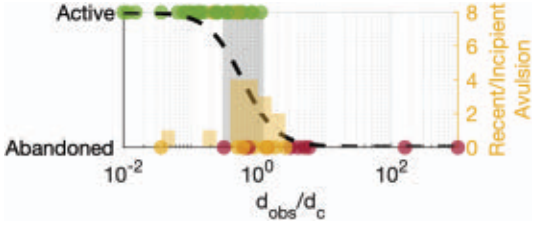


Fig. 3. Channel classification. The dashed line is a logistic regression through active and abandoned channels (collapsing Fig. 2E along the *x* axis) with respect to the normalized offset. Separation of the data, with a class boundary near unity, indicates consistency with the model framework. The gray box indicates the 99% confidence interval on the class boundary using 10,000 bootstrap samples. Near-critical channels (incipient or recent avulsion in yellow) are summarized in the histogram as an additional qualitative test.

(*15*). Finally, by combining Eq. 1 and Eq. 2 and approximating the integral in Eq. 1 as a product, we obtain a characteristic avulsion time scale,

$$t_c \approx \frac{4h_c^2}{\kappa S_0^2} \quad (3)$$

A full semi-analytical solution to Eqs. 1 and 2 with discrete slip events (*15*) (fig. S1) corroborates the approximate framework and suggests a response independent of whether slip is continuous or sequenced in individual earthquakes. Provided a long-term fault slip rate (v_x), a channel becomes unstable and susceptible to avulsion when the channel offset approaches a critical offset $d_c = v_x t_c$. In this way, through a known slip rate, the channel offset is an independent chronometer of the channel response. We can thus relate the advective and fluvial history of offset channels at the time of avulsion to determine a critical offset,

$$d_c \approx \frac{4h_c^2 v_x}{\kappa S_0^2} \quad (4)$$

Equation 4 is the first key result of the model. It defines the offset that resets fault-crossing channels and depends entirely on measurable quantities.

We tested this relationship on a collection of active and abandoned fault-crossing channels in the Carrizo Plain. Although it is rare to estimate d_c directly, Eq. 4 predicts observed offsets (d_{obs}) that are less than d_c on active channels and more than d_c on abandoned channels. We measured values of total offset (d_{obs}), avulsion threshold height (h_c), and initial slope (S_0) interpreted from LiDAR data (e.g., Fig. 2, A to C) for a set of 55 active and abandoned channels in the Carrizo Plain (*15*) (table S1 and figs. S6 to S64). We avoided channels where misalignment across the fault appears to result from the deflection of flow rather than progressive fault offset (e.g., downstream from an alluvial fan). We estimated the diffusivity coefficient using $\kappa \approx 0.1Lr$, where L is the corresponding catchment length (Fig. 2D) and r is the mean annual regional rainfall (*13*). We expect that the main source of error in our approach is this approximation of κ and its constancy through time (*15*). We recorded partial overspilling on an active channel or minimal offset on the active tributary of an abandoned channel, as these respectively indicate an incipient or a recent avulsion. In these cases, we assumed that the measured offset approximates the critical offset, d_c.

We compared measured offsets to corresponding predictions of d_c (Fig. 2E). A logistic regression through the log-transformed data finds that the expression $d_{obs}/d_c = 0.6^{+0.5}_{-0.3}$ (99% confidence interval on 10,000 bootstrap samples) best separates active and abandoned channels, indicating that the model prediction corresponds to observed critical offsets within a factor of 2 (Fig. 3). If the timing of avulsions were independent of the diffusive channel response, the ratio d_{obs}/d_c would not separate active and abandoned channels ($P = 0.0007$, following a Wald test) and has no reason to approach unity. Instead, the scaling analysis predicts the life span of fault-crossing channels well within model uncertainty, particularly in the parameterization of the diffusivity coefficient (*15*).

Under the assumptions of the model, any one parameter in Eq. 4 is constrained from the others. Posing Eq. 4 as an equality and isolating for the slip rate, the channel diffusivity coefficient (*15*) and channel geometry (table S1) on active and abandoned channels bracket a slip rate, respectively under- and overestimating it. A logistic regression then yields a $2.1^{+1.7}_{-1.0}$ cm/year slip rate in the Carrizo Plain (99% confidence interval on 10,000 bootstrap samples), compatible with geologic and geodetic estimates of 3.5 cm/year (*4*, *16*). With a known fault slip rate, a similar approach calibrates the diffusivity (and hence the sediment transport capacity) of channels. The critical offset encodes channel response best in arid environments where transport capacity is low relative to slip rate, which perhaps is a distinguishing feature of the Carrizo Plain (*15*). Where

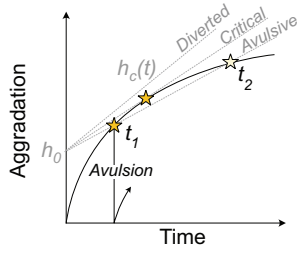


Fig. 4. Channel aggradation and relative vertical motion along the fault. Aggradation at the avulsion node is shown as a function of time for a channel with an avulsion threshold $h_c(t)$ that grows with time. Times t_1 and t_2 represent the two solutions to the quadratic form of the avulsion time scale; however, only time t_1 is physically realized, as the system resets once avulsion occurs. In detail, aggradation exhibits jagged high-frequency fluctuations representing cut-in-fill behavior but should be well approximated as a smooth diffusive curve on millennial time scales.

transport rates are high, the evolution of channels after individual earthquakes on centennial time scales may better capture the geomorphic process (fig. S3). Bedrock channels where sediment transport is not limiting may be better modeled as detachment-limited kinematic waves of erosion (*17*–*19*) with fault damage promoting headward incision along the fault (*2*, *20*).

Up to this point in our analysis, we used the avulsion threshold with respect to the present-day channel geometry, independent of consideration for its path and continued evolution. Expanding $h_c(t) = h_0 + \lambda v_x t + \cdots$, where h_0 reflects initial channel height or incision immediately after an avulsion, and λ is the dimensionless ratio of apparent vertical to horizontal motion in the fault plane at the avulsion node. Topographic gradients translated along strike and obliquity in slip (the ratio of strike-slip to dip-slip motion) contribute to λ. For the San Andreas Fault, obliquity in slip rarely exceeds 10% (*21*). For sharp topographic elements advected in front of the avulsion node, λ may be much higher (e.g., an angle of repose slope is approximately 35°, corresponding to a 70% gradient) over relatively short distances. The growth of the avulsion threshold, if sufficiently high, may outpace channel aggradation (Fig. 4). This results in persistent fault-parallel drainage, analogous to a river diverted by active uplift (*22*). In such cases the penultimate avulsion may not occur at the avulsion node, where aggradation is highest, but rather downstream, where uplift wanes (e.g., figs. S40 and S41). Unlike other fluvial environments susceptible to frequent avulsion, such as deltas or fans (*10*), aggradation in offset channels is effectively unbounded when there is sustained growth of $h_c(t)$. Sedimentation can therefore far exceed flow depths.

Introducing the time-dependent avulsion threshold into Eq. 3 yields a limit, λ_c, in which fluctuation in h_c outpaces aggradation,

$$\lambda_c \approx \frac{\kappa S_0^2}{16 v_x h_0} \quad (5)$$

If $\lambda > \lambda_c$, channels are diverted along strike (*15*) (Fig. 4). For the channels we studied, λ_c (estimated from Eq. 5) is typically on the order of a few percent and is therefore sensitive to near-fault topography, which may contribute to the ubiquity of fault-parallel drainage on the San Andreas Fault. Where near-fault relief is high, the timing of avulsion is modulated by the valley spacing or abandoned channels that are advected along the fault (*19*, *23*). In the Carrizo Plain, subdued near-fault topography and nearly horizontal slip implies that the ratio of vertical to horizontal motion rarely exceeds the critical obliquity λ_c and results in abundant cross-fault drainage. Along the Dragon's Back, however, $\lambda \approx 6\%$ along a kilometer of the fault (*24*) is consistent with uncharacteristically long channel offset and suppressed cross-fault drainage (Fig. 1D and figs. S40 and S41). The case of $\lambda < 0$ (downdropping) results in an unstable channel geometry, which is likely to develop alluvial fans where channels reset in high-flow events (*5*, *25*). For large rivers, the numerator κS_0^2 of Eq. 5 is roughly two orders of magnitude larger than the ephemeral channels of the Carrizo Plain (*26*), which suggests that established fault-crossing rivers are less susceptible to diversion. Therefore, major fault-parallel drainages along strike-slip faults may preferentially develop from the coalescence of smaller tributaries.

The geometry of offset channels is the product of an interactive fluvial-tectonic system. In the Carrizo Plain, channels offset by the San Andreas Fault provide an opportunity to characterize the long-term evolution of channels under a simple forcing history. In these reaches, the life span of fault-crossing channels is well described by a transport-limited response culminating in avulsions. The configuration of fault-crossing channels, readily obtained from topography data, quantifiably constrains the relative pace of the tectonic and fluvial systems. Our analysis implies that the geomorphology of strike-slip fault zones exists in a careful balance where subtle changes to the system can toggle drainage from fault-perpendicular to fault-parallel valleys.

REFERENCES AND NOTES

1. A. C. Lawson, *The California Earthquake of April 18, 1906* (Carnegie Institution of Washington, 1908).
2. R. E. Wallace, "Notes on stream channels offset by the San Andreas fault, southern Coast Ranges, California" (1968).
3. K. E. Sieh, *Bull. Seismol. Soc. Am.* **68**, 155–159 (1978).
4. K. E. Sieh, R. H. Jahns, *Geol. Soc. Am. Bull.* **95**, 883–896 (1984).
5. O. Zielke, J. R. Arrowsmith, L. Grant Ludwig, S. O. Akçiz, *Science* **327**, 1119–1122 (2010).
6. O. Zielke, Y. Klinger, J. R. Arrowsmith, *Tectonophysics* **638**, 43–62 (2015).
7. J. D. Sims, in *Proceedings of the Workshop on Paleoseismology* (1994), vol. 94, pp. 170–172.
8. S. Ouchi, *Geomorphology* **70**, 112–128 (2005).
9. L. B. Leopold, W. B. Bull, *Proc. Am. Philos. Soc.* **123**, 168–202 (1979).
10. R. Slingerland, N. D. Smith, *Annu. Rev. Earth Planet. Sci.* **32**, 257–285 (2004).
11. V. Ganti, A. J. Chadwick, H. J. Hassenruck-Gudipati, B. M. Fuller, M. P. Lamb, *Sci. Adv.* **2**, e1501768 (2016).
12. D. J. Jerolmack, D. Mohrig, *Geology* **35**, 463–466 (2007).
13. C. Paola, P. L. Heller, C. L. Angevine, *Basin Res.* **4**, 73–90 (1992).
14. L. B. Leopold, *Isr. J. Earth Sci.* **41**, 57–64 (1992).
15. See supplementary materials.
16. G. Schmalzle, T. Dixon, R. Malservisi, R. Govers, *J. Geophys. Res.* **111**, 5403 (2006).
17. A. R. Duvall, G. E. Tucker, *J. Geophys. Res. Earth Surf.* **120**, 2016–2026 (2016).
18. H. J. Gray *et al.*, *Geology* **46**, 59–62 (2018).
19. N. G. Reitman *et al.*, *J. Geophys. Res. Solid Earth* **124**, 13427–13451 (2019).
20. S. G. Roy, G. E. Tucker, P. O. Koons, S. M. Smith, P. Upton, *J. Geophys. Res. Earth Surf.* **121**, 1911–1930 (2016).
21. R. E. Wallace, *The San Andreas Fault System, California* (US Government Printing Office, 1990).
22. N. F. Humphrey, S. K. Konrad, *Geology* **28**, 43 (2000).
23. S. A. Harbert, A. R. Duvall, G. E. Tucker, *Geophys. Res. Lett.* **45**, 611–683 (2018).
24. G. E. Hilley, J. R. Arrowsmith, *Calif. Geol.* **36**, 367–370 (2008).
25. L. B. Grant, K. Sieh, *J. Geophys. Res. Solid Earth* **99**, 6819–6841 (1994).
26. S. Castelltort, J. Van Den Driessche, *Sediment. Geol.* **157**, 3–13 (2003).
27. QGIS Geographic Information System; https://qgis.org/en/site/.

ACKNOWLEDGMENTS

We thank the members of the UCSC seismology laboratory along with A. M. Rodriguez Padilla, A. Duvall, C. Abrahams, D. Draper, and M. Oskin for providing thoughtful insight and discussion. **Funding:** Supported by NASA FINESST fellowship 80NSSC19K1300 and the Southern California Earthquake Center (contribution no. 11001). SCEC is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047. **Author contributions:** Conceptualization, K.D.C., N.J.F., E.E.B.; data curation, K.D.C.; formal analysis, K.D.C.; methodology, K.D.C., N.J.F., E.E.B.; investigation, K.D.C., N.J.F., E.E.B.; visualization, K.D.C.; funding acquisition, K.D.C., E.E.B.; supervision, N.J.F., E.E.B.; writing–original draft, K.D.C.; writing–review and editing, K.D.C., N.J.F., E.E.B. **Competing interests:** To the best of the authors' knowledge, we have no conflict of interest publishing this research. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The data collected in this paper are available in table S1. Figures were produced using QGIS (*27*) and MATLAB, with source code provided at https://doi.org/10.5281/zenodo.4766502.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/204/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S64
Table S1
References (*28*–*33*)

13 October 2020; accepted 18 May 2021
10.1126/science.abf2320

MECHANOCHEMISTRY

Flyby reaction trajectories: Chemical dynamics under extrinsic force

Yun Liu[1,2], Soren Holm[3,4,5], Jan Meisner[3,4,5], Yuan Jia[1,2], Qiong Wu[1,2], Toby J. Woods[2,6], Todd J. Martinez[3,4,5]*, Jeffrey S. Moore[1,2]*

Dynamic effects are an important determinant of chemical reactivity and selectivity, but the deliberate manipulation of atomic motions during a chemical transformation is not straightforward. Here, we demonstrate that extrinsic force exerted upon cyclobutanes by stretching pendant polymer chains influences product selectivity through force-imparted nonstatistical dynamic effects on the stepwise ring-opening reaction. The high product stereoselectivity is quantified by carbon-13 labeling and shown to depend on external force, reactant stereochemistry, and intermediate stability. Computational modeling and simulations show that, besides altering energy barriers, the mechanical force activates reactive intramolecular motions nonstatistically, setting up "flyby trajectories" that advance directly to product without isomerization excursions. A mechanistic model incorporating nonstatistical dynamic effects accounts for isomer-dependent mechanochemical stereoselectivity.

Explanations of chemical reactivity are conventionally supported by statistical theories in chemical dynamics. In this classical picture, Boltzmann statistics are assumed to govern energy distributions on the potential energy surface and the corresponding rate constants. Consequently, product distributions are dictated by the relative frequencies with which energy barriers are surmounted among the competing intrinsic reaction coordinates. Emerging examples are challenging these perceptions by showing that even the outcomes of textbook organic reactions are influenced by nonstatistical dynamic effects (NDEs) (*1*–*3*). That is, the momenta (directions and speed) of the atoms in reacting species lessen the influence of Boltzmann statistics and lead to product distributions that depart from the statistical predictions (*4*–*8*). In some enzymes, NDEs are shown to steer reaction flux toward desired channels (*9*); on the benchtop, external controls over competing reaction paths have largely relied on the more conventional approach, modifying potential energy surfaces to affect relative rates (*10*–*13*). Extrinsic force is also known to modify potential energy surface to access reactivity and selectivity differently from their zero-force counterparts (*14*, *15*), enabling thermally forbidden trajectories (*16*), shifting chemical equilibria (*17*, *18*), and stabilizing otherwise fleeting intermediates (*19*). The acceleration of bond reorganization by force to promote nonequilibrium energy distributions inspired us to postulate that the extrinsic force can impose and control NDEs to steer reaction trajectories away from Boltzmann statistics even on the already force-modified potential energy surface (FMPES). It was recently observed that nonstatistical chemical dynamics can occur in mechanochemical reactions (*20*–*22*). However, the extent to which these trajectories are nonstatistical on the FMPES, and in particular, whether the nonstatistical behaviors are tunable by the extrinsic force are not known.

To investigate the postulated role of force in inducing and controlling NDEs, we chose a cyclobutane ring-opening reaction that follows a stepwise mechanism (*23*–*25*), is mechanochemically active (*25*–*27*), and exhibits a distribution of stereoisomeric products (*20*, *27*, *28*). The product's stereochemistry is determined after passing over the rate-limiting barrier of the first bond scission; from there, competing isomerization channels branch from the initially formed diradical intermediate (Fig. 1A). The reaction either proceeds directly to complete the ring opening with the second bond scission, or the intermediate undergoes isomerization before the second bond scission transpires. Owing to the instability of the diradical intermediates (*24*), the stereo-defining events are not expected to be rate-limiting under force. Consequently, on this FMPES, force-controlled NDEs can govern pathway selection by outpacing thermal equilibration of the diradical intermediate at the branch point. The resulting product stereoselectivity serves as a sensitive reporter of force-controlled NDEs.

We hypothesized that momenta derived from the extrinsic force can propel trajectories to "fly by" the isomerization branch points and thereby achieve nonstatistical product distribution even on the FMPES. Such flyby trajectories traverse the energy landscape via excited vibrational energy levels of the diradical intermediate without undergoing relaxation (Fig. 1B). Atomic motions associated with flyby trajectories are thus accelerated beyond their thermal rates, altering selectivity that is intrinsic to the FMPES such that the resulting product stereochemistry directly relates to the stereochemistry of the reactant. In contrast, deepening the potential well at the branch point increases the likelihood of vibrational relaxation, restoring the statistical reaction outcome of the FMPES.

To test our hypothesis, we designed and synthesized a series of cyclobutane stereoisomers from the stereoselective [2σ + 2σ + 2π] cycloaddition (*29*) between quadricyclane and ^{13}C-labeled alkenes (Fig. 2A). The alkyl and ester substituents offer different diradical stability (table S4) to vary the depth of the potential well at the branch point. The stereochemistry of the cyclobutanes was confirmed by spatial proximity of protons (figs. S1 to S4) and x-ray crystal structures (Fig. 2B). These isomers are denoted as *anti* (**A**), *down* (**D**), and *up* (**U**) following the orientations of substituents. Using the constrained geometries simulate external force (CoGEF) calculation (figs. S9 to S16), an ab initio energy-extension modeling method, all designed cyclobutane stereoisomers were predicted to undergo mechanically activated retro-[2+2] cycloaddition (*30*). Cu(0)-mediated radical polymerization was carried out from the initiator-bearing cyclobutanes using methyl acrylate (Fig. 2C) to render linear polymers of controlled molecular mass (M_n): 40 to 140 kDa, $Đ$ < 1.2 (figs. S19 to S23). The well-controlled bidirectional polymerization allowed for statistically chain-centered placement of cyclobutanes in the polymer chain while retaining high stereochemical purity (figs. S23 to S27).

We experimentally evaluated the mechanochemistry by sonication-induced solvodynamic extension of polymer chains (fig. S28) (*31*), and the resulting solutions were characterized by ^{13}C nuclear magnetic resonance (NMR) spectroscopy. For the alkyl-substituted cyclobutanes, sonication resulted in new ^{13}C-labeled methylene ($^{13}CH_2$) and alkene methine (^{13}CH) peaks at δ = ~60 and ~120 parts per million (ppm), respectively (Fig. 3A and figs. S29 to S31). The coupling between the methylene and alkene peaks (table S5) suggests ring opening of cyclobutanes into alkenes. On the basis of the three distinctive alkene peaks, reference compounds (fig. S34), and calculated ^{13}C chemical shifts (table S6), we assigned the ring-opening product of the **A**-isomer to an (*E*,*Z*)-bisalkene, the **D**-isomer to an (*E*,*E*)-bisalkene, and the **U**-isomer to a mixture of (*Z*,*Z*)- and (*E*,*Z*)-

[1]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [3]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [4]The PULSE Institute, Stanford University, Stanford, CA 94305, USA. [5]SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [6]3M Materials Chemistry Laboratory, School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
*Corresponding author. Email: todd.martinez@stanford.edu (T.J.M.); jsmoore@illinois.edu (J.S.M.)

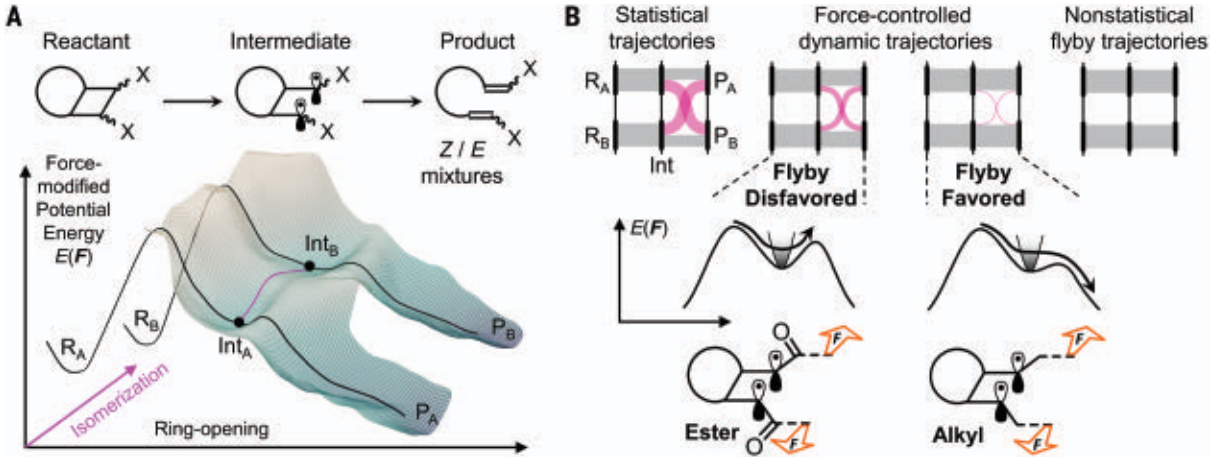


Fig. 1. Nonstatistical dynamic effects induced by extrinsic force. (**A**) A generic force-modified potential energy landscape for stepwise ring opening of isomeric cyclobutane reactants R_A and R_B. (**B**) The force-induced NDE hypothesis represented schematically by alluvial flow diagrams. Colored flows: gray, reaction trajectories that ring-open without undergoing isomerization; magenta, reaction trajectories that isomerize and then ring-open.

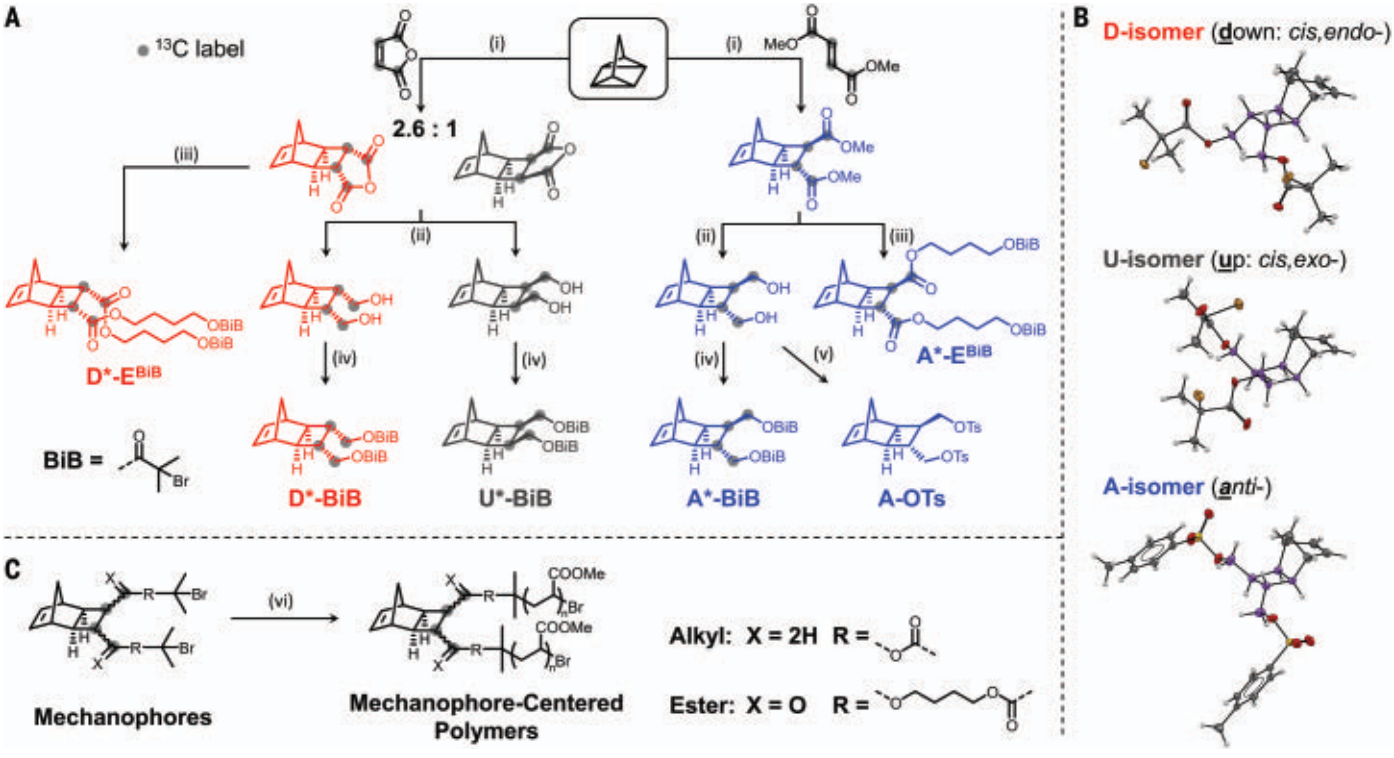


Fig. 2. ^{13}C-labeled cyclobutane mechanophores. (**A**) Synthesis, (**B**) stereochemical validation, and (**C**) polymer synthesis. Reaction conditions: (i) PhMe, 110°C. (ii) $LiAlH_4$, THF, 0°C. (iii) KOH or LiOH; then 4-hydroxybutyl 2-bromoisobutyrate, EDCI•HCl, DMAP, THF. (iv) BiB-Br, Et_3N, THF. (v) TsCl, Et_3N, THF (a nonlabeled reactant was used). (vi) Methyl acrylate, Cu(0), Me_6TREN, DMSO. THF, tetrahydrofuran; EDCI, *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide; DMAP, *N*,*N*-dimethylaminopyridine; Ts, 4-4-methylsulfonyl (tosyl); Me_6TREN, tris[2-(dimethylamino)ethyl]amine; DMSO, dimethylsulfoxide. Color coding in (B): carbon, gray; oxygen; red; bromine, brown; sulfur, yellow; hydrogen, white; cyclobutane ring and its α-CH_2 carbon, violet. All thermal ellipsoids are shown at 50% probability.

bisalkenes in an 8:1 ratio. The observation of (*E*,*Z*) by-product from the **U**-isomer suggests that the two radicals behave independently after generation, confirming the stepwise mechanism. Despite being stepwise, the ring opening of the alkyl cyclobutanes achieved high stereoselectivity. This conservation of stereochemical information suggests that the force-imparted momentum and/or force-modified energy landscape is effective in steering the reaction trajectories away from excursions at downhill branch points (Fig. 1B).

We then turned to the ester-substituted cyclobutanes, which stabilize the diradical intermediates more than the alkyl derivatives. Under

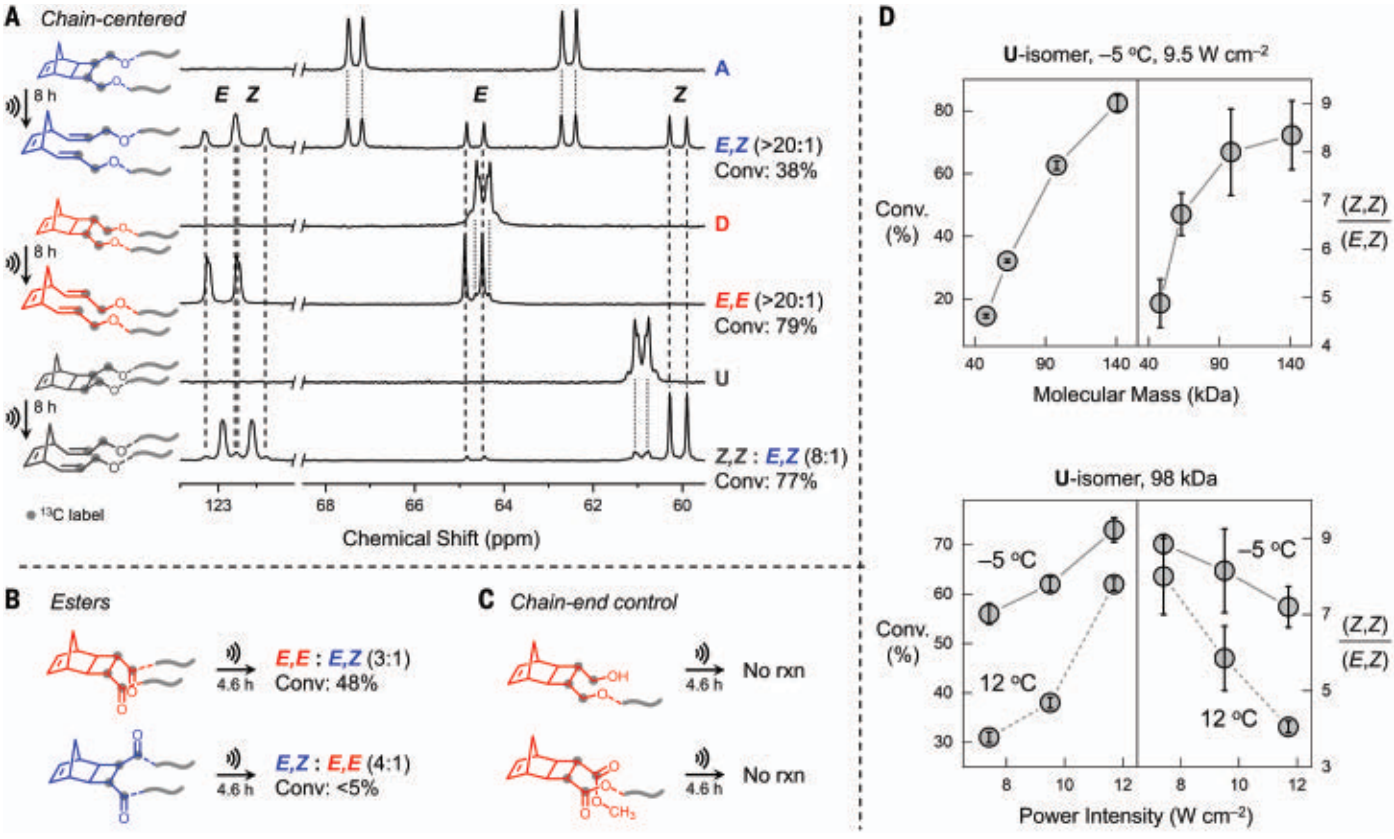


Fig. 3. Mechanochemical activation of cyclobutanes. (**A**) Stacked ^{13}C NMR spectra of chain-centered alkyl cyclobutanes in a ~100-kDa polymer before and after sonication (20 kHz, 9.5 W/cm^{-2}, −5°C, THF). (**B**) Mechanochemical activation of chain-centered ester variants. (**C**) Attempted mechanochemical activation of chain-end control cyclobutanes. (**D**) Impact of experimental variables on the conversion and product selectivity for the alkyl **U**-isomer. Errors are presented in 2σ from four measurements of relative integrations by using the two pairs of ^{13}C NMR peaks.

the same sonication conditions, mechanochemical production of α,β-conjugated diesters was observed for the **A**- and **D**-isomer (fig. S33); the **U**-isomer was not obtained because of isomerization during synthesis and was not experimentally studied. NMR analysis of carbon multiplicity (number of attached protons), spin-spin coupling, and alkene stereochemistry using the ^{13}C-labeled product peaks (fig. S35 and tables S5 and S7) revealed considerably lower stereoselectivity (Fig. 3B). The preferred alkene geometry remained the same as for the alkyl variants, but the selectivity fell from >20:1 to ~3:1. Loss of stereoselectivity arises from torsional rotations that occur via the diradical states, e.g., the **A**-isomer produces the (*E*,*E*)-bisalkene after the upward substituent rotates downward. The reduced stereoselectivity of ester variants suggests that force-driven dynamics compete less effectively with bond rotations when the there is a deeper energy well associated with the intermediate (Fig. 1B).

Mechanochemical activation was not observed in chain-end control polymers (Fig. 3C) in which nearly no force is exerted on the cyclobutanes. Nor did heating at 240°C result in any conversion (figs. S38 and S39), suggesting a pronounced rate acceleration by extrinsic force on high-barrier thermal reactions. We used relative integrations of the α-$^{13}CH_2$ carbon peaks to quantify the mechanochemical conversion of alkyl derivatives within a ~100-kDa polymer (figs. S43 to S46). The data were best-fitted with a competing first-order kinetics model to account for both activation and background polymer scission to obtain conversion rates (fig. S47). The **D**-isomer displayed the fastest conversion of 0.46 hour^{-1}, followed by the **U**-isomer (0.31 hour^{-1}) and then the **A**-isomer (0.14 hour^{-1}), consistent with the rank order of predicted rupture force by CoGEF (table S4) and literature precedents (*32*).

To alter the mechanochemical reaction outcomes, we explored a range of experimental variables (Fig. 3D) on the alkyl cyclobutanes. Our screening initially focused on the **U**-isomer as the test compound. Conversions were improved with increasing sonication-induced extensional force, by using longer polymers (fig. S52), lower solution temperatures, or higher ultrasound intensities (fig. S54) (*33*). The stereoselectivity of the **U**-isomer exhibited a positive dependence on force when molar masses and solution temperatures were changed, but a negative dependence on force-loading rate when ultrasound intensity was increased (eq. S1). Exploring the known effect of experimental variables, we altered the product ratios between the (*Z*,*Z*) and the (*E*,*Z*)-bisalkenes for the alkyl **U**-isomer from 4:1 to 10:1 (fig. S55). In comparison, the high stereoselectivities of the **A**-isomer and **D**-isomer persisted under all conditions (fig. S55), suggesting that force-imparted dynamics are isomer dependent.

We reasoned that the force-induced dynamics are bound by two limits. First, there is a thermal limit that reflects the statistical reaction outcome based on the FMPES; the imparted momenta largely dissipate by thermal relaxation. Second, there is a dynamic limit that reflects the experimentally achievable nonstatistical selectivity. Although experimentally we observed the nonstatistical outcome anticipated from our design, a theoretical investigation is necessary to quantify the dynamic and thermal contributions.

To characterize the thermal limit, we computationally mapped out the force-modified minimum energy paths (MEPs) for alkyl cyclobutanes using acetyl-terminated model

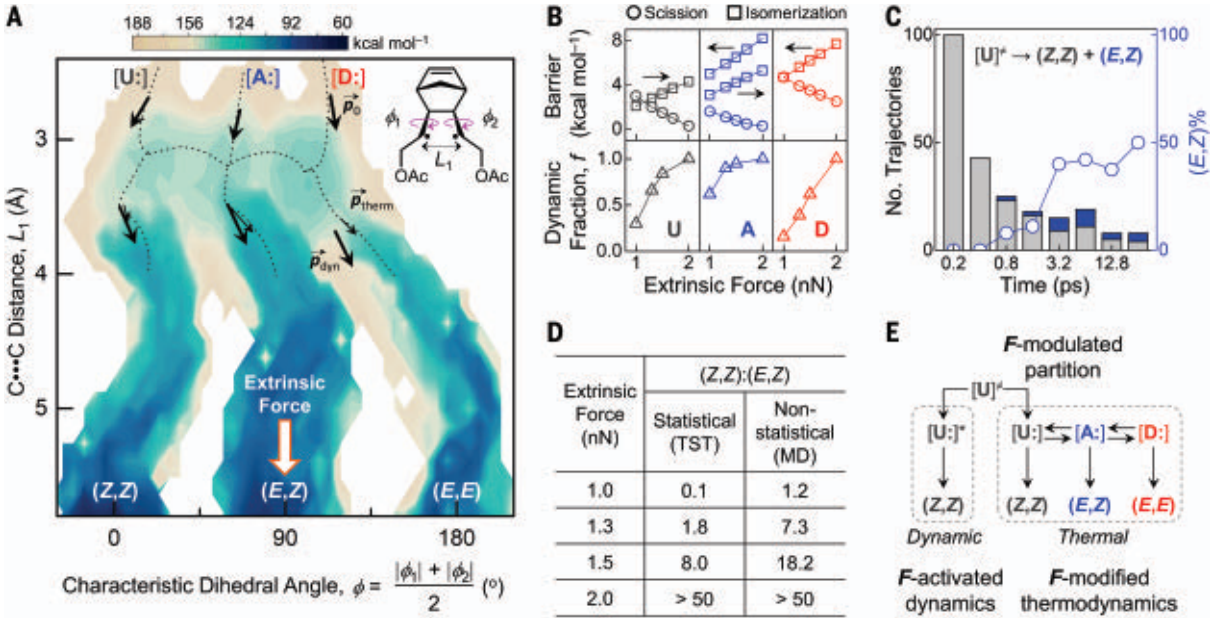


Fig. 4. Nonstatistical dynamics of cyclobutane ring opening under extrinsic forces. (**A**) FMPES constructed from AISMD (1.3 nN, 300 K) runs on model compounds; [**X:**] indicates the diradical intermediates. The extension coordinate (L_1) was chosen to represent bond scissions during ring opening, and the dihedral coordinate (ϕ, magenta) was chosen for isomerization. Dashed lines depict the MEPs. Colored arrows represent time- and ensemble-averaged reaction momenta over a 20-fs interval: thick black, entrance momenta $\vec{p}_0$ and dynamic exit momenta $\vec{p}_{dyn}$; thin black, thermalized exit momenta $\vec{p}_{therm}$. The orange, hollow arrow indicates that the direction of the extrinsic force is parallel to the ring-opening L_1 axis. (**B**) Top: Free energy barriers of the second C–C bond scission and diradical isomerization. Arrows indicate directions of isomerization along the **U**-**A**-**D** path. Bottom: Force-dependent partition coefficient (*f*) of dynamic trajectories. (**C**) Histogram of MD trajectories for **U** at 1.3 nN. $[\mathbf{U}]^{\ddagger}$ indicates the first-bond scission transition state. (**D**) Statistical (TST) and nonstatistical (MD) product selectivity under different extrinsic forces. (**E**) Hybrid NDE model exemplified by the ring opening of **U**. [**U:**]* indicates vibrationally excited diradical intermediates of the **U**-isomer.

compounds (Fig. 4A). When the pulling force was varied from 1 to 2 nN, energy barriers of the second scissions were lowered by 2 to 3 kcal mol^{-1} and those of isomerizations were raised by 3 to 4 kcal mol^{-1} (Fig. 4B). The mode selection is consistent with the coupling between force and reactive molecular extension during ring opening, whereas force-resistive contraction was seen during isomerization. Among the three isomers, **U** had the lowest isomerization barriers that became disfavored beyond 1.3 nN. At this cross-over point, we constructed a full FMPES using 470 ab initio steered molecular dynamics (AISMD) trajectories initialized from the rate-limiting transition states of the three isomers (Fig. 4A). The result features three ring-opening paths running downhill to the bisalkene products. The diradical intermediates were characterized as shallow calderas separated by isomerization barriers. The intermediate from **U** is less stable than the intermediate from **D**, raising the isomerization probability of the intermediate from **U**.

We garnered ~1000 force-steered MD trajectories (table S12) to study the force-induced dynamic effects. Exemplified by **U**, we saw that a large reaction flux flies by the branch areas at early simulation times (<0.5 ps) with a strong preference for the direct ring-opening channel and conservation of reactant stereochemistry (Fig. 4C, figs. S79 to S81, and movie S5). Over time, the force-imparted reaction momenta relax and isomerization ensues in conformity with the FMPES; this temporal distribution of product formation and selectivity are hallmarks of NDEs (*34*, *35*). The partition of trajectories into dynamic and thermal regimes resulted in bimodal rates of decay for the population of intermediates from the branch points (fig. S82). The fraction of dynamic trajectories (*f*) increased with increasing force for all three isomers but in an isomer-dependent manner: **A** > **U** > **D** (Fig. 4B); at 1.3 nN, dynamic trajectories make up 90, 66, and 39% of the total trajectories for **A**, **U**, and **D**, respectively. This dependence mirrors the rank order of their second scission barrier heights, supporting the prediction that a deeper potential well at the branch point reduces the number of dynamic trajectories.

To quantify the dynamic effects, we extracted atomic reaction momenta ($\vec{p}$) from the MD trajectories (Fig. 4A and table S13). An ensemble-averaged momentum characterizes the mean velocity and direction of the trajectories on the FMPES. In the dynamic regime (t < 0.5 ps), the momenta ($\vec{p}_{dyn}$) of trajectories exiting the branch areas deviate noticeably from the MEPs, showing a tendency to align with the extrinsic force. In contrast, the exit momenta in the thermal regime ($\vec{p}_{therm}$; t > 1 ps) closely adhered to the MEPs. The projected momenta on the extension coordinate (L_1, Fig. 4A) were fully conserved through the branch points, as shown by comparing the maximal momenta coming into and out of the areas; $\vec{p}_0$ and $\vec{p}_{dyn}$, respectively. The minimal loss in velocity along L_1 suggests little to no equilibration of the diradical intermediate in the dynamic regime, a signature of flyby trajectories. The force-derived momenta imparted to reactants on flyby trajectories were quantified by vector subtraction between dynamic and thermal exit momenta (fig. S91). We saw that the extrinsic force accelerates the atomic velocity of reactive molecular extension to 200% of their thermal values, reaching ~1.4 nm/fs (T_{eff} ~ 1000 K). In contrast, no consistent changes were observed for the dihedral coordinate between dynamic and thermal regimes, and their average angular velocity (8.5°/fs) is the same as in the competing isomerization trajectories. This distinction reveals that the extrinsic force selectively activates the chemical dynamics of reactive molecular extensions (Fig. 1B) over isomerization, leading to nonstatistically enhanced product stereoselectivity (Fig. 4D and fig. S95).

At 1.3 nN, the NDEs improve product selectivity by fourfold for **U** in comparison to the statistical outcome. The apparent enhancement on selectivity is much smaller (<10%) for **A** and **D**, because of **A** being a kinetic trap and an already preferred product channel on the FMPES and **D** having a relatively low dynamic contribution (Fig. 4B).

The isomer-dependent dynamic responses result from a varying degree of openness in the exit reaction channels along the dihedral coordinate, which is further determined by reactant stereochemistry (fig. S92). The 120° ridge on the FMPES blocks the majority of dynamic exit coming from **D**, leaving behind a narrow and risen gorge ($\Delta\phi$ = 3°) for its thermal exit (Fig. 4A). In contrast, the exit channel for **A** is wider ($\Delta\phi$ = 45°), whereas for **U** there is a modest constriction ($\Delta\phi$ = 35°). We used an additive model to relate dynamic population (f) and the relative configuration of stereocenters on the cyclobutane rings. Mathematically, we found that the stereochemistry along the pulling direction plays a primary role in promoting dynamic trajectories, with trans configuration much better than cis; the stereochemistry orthogonal to the force is secondary, with cis better than trans. This rule allowed us to rank the tendency to undergo dynamic ring opening for all possible cyclobutane stereoisomers (fig. S93).

The force-induced NDE is expressed as a linear hybridization of dynamic and thermal extremes with a force-dependent partition (Fig. 4E). The dynamic regime is stereospecific by definition, and the selectivity of product isomers in the thermal regime was assumed to be statistical. This statistical (transition state theory, TST) stereoselectivity (S_{TST}) was obtained from equilibrium product distributions calculated by solving the system of rate equations of the activation-isomerization reaction network (Fig. 4E, right box, and fig. S94). The rate constant for each rate equation was calculated by TST (table S14). The force-dependent nonstatistical stereoselectivity (S_{NDE}), defined as the ratio between direct stereoisomeric product and all isomerization products, is predictable from the statistical stereoselectivity, S_{TST}, and the dynamic fraction, f:

$$S_{NDE} = \frac{S_{TST} + f}{1 - f} \quad (1)$$

When the dynamic fraction increases from 0 to 1, the S_{NDE} goes from its thermal value, S_{TST} (f = 0), to being a stereospecific transformation in the limit when $f \to 1$ (table S15). This model reproduces the observed dynamic selectivity without the need to run full long MD simulations.

The hybrid NDE framework reveals that there is not necessarily a simple connection between the intermediate's stability and the dynamic selectivity of mechanochemical reactions. Although the stereoconvergent ring opening of the ester-substituted cyclobutane has been observed and attributed to its long-living intermediate (*28*), our study suggests that a long-lived intermediate need not be destined for lower stereoselectivity. Indeed, we have improved the product stereoselectivity of the ester derivatives compared to the reported literature values. Similarly, the alkyl **D**-isomer displayed high stereoselectivity under all experimental conditions despite its longest diradical lifetime as shown by radical quenching experiments with 4-hydroxyl TEMPO (table S17). Rather, stereoselectivity can be affected dynamically for a short-lived intermediate: for example, by a neighboring product channel serving as a dynamic sink to draw the trespassing trajectories into an irreversible path, similar to the formation of (*E*,*Z*)-bisalkene from the **U**-isomer.

We envision that external force is an experimental method for controlling nonstatistical dynamics and steering reaction trajectories away from constraints imposed by potential energy surfaces. The hybrid framework will offer ability to probe and understand the dynamic effects of mechanochemical phenomena in polymeric materials. The ability to control dynamics and stereochemistry of mechanochemical reactions may be helpful for programming properties of next-generation polymeric materials whose life cycles will go beyond the single-use paradigm.

REFERENCES AND NOTES

1. B. K. Carpenter, *J. Am. Chem. Soc.* **117**, 6336–6344 (1995).
2. S. C. Ammal, H. Yamataka, M. Aida, M. Dupuis, *Science* **299**, 1555–1557 (2003).
3. Y. Nieves-Quinones, D. A. Singleton, *J. Am. Chem. Soc.* **138**, 15167–15176 (2016).
4. C. Doubleday, C. P. Suhrada, K. N. Houk, *J. Am. Chem. Soc.* **128**, 90–94 (2006).
5. F. F. Crim, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12654–12661 (2008).
6. J. Rehbein, B. K. Carpenter, *Phys. Chem. Chem. Phys.* **13**, 20906–20922 (2011).
7. J. Rehbein, B. Wulff, *Tetrahedron Lett.* **56**, 6931–6943 (2015).
8. S. R. Hare, D. J. Tantillo, *Pure Appl. Chem.* **89**, 679–698 (2017).
9. Y. J. Hong, D. J. Tantillo, *Nat. Chem.* **6**, 104–111 (2014).
10. J. Wang, B. L. Feringa, *Science* **331**, 1429–1432 (2011).
11. C. J. Brown, F. D. Toste, R. G. Bergman, K. N. Raymond, *Chem. Rev.* **115**, 3012–3035 (2015).
12. A. C. Aragonès *et al.*, *Nature* **531**, 88–91 (2016).
13. W. Mtangi *et al.*, *J. Am. Chem. Soc.* **139**, 2794–2798 (2017).
14. M. T. Ong, J. Leiding, H. Tao, A. M. Virshup, T. J. Martínez, *J. Am. Chem. Soc.* **131**, 6377–6379 (2009).
15. P. Dopieralski *et al.*, *Nat. Chem.* **5**, 685–691 (2013).
16. C. R. Hickenboth *et al.*, *Nature* **446**, 423–427 (2007).
17. A. E. M. Beedle *et al.*, *Nat. Commun.* **9**, 3155 (2018).
18. M. E. McFadden, M. J. Robb, *J. Am. Chem. Soc.* **141**, 11388–11392 (2019).
19. J. M. Lenhardt *et al.*, *Science* **329**, 1057–1060 (2010).
20. Z. Chen *et al.*, *Nat. Chem.* **12**, 302–309 (2020).
21. R. Nixon, G. De Bo, *Nat. Chem.* **12**, 826–831 (2020).
22. M. Wollenhaupt, C. Schran, M. Krupička, D. Marx, *ChemPhysChem* **19**, 837–847 (2018).
23. E. Fischer, R. Gleiter, *Tetrahedron Lett.* **26**, 5289–5290 (1985).
24. S. Pedersen, J. L. Herek, A. H. Zewail, *Science* **266**, 1359–1364 (1994).
25. M. J. Kryger *et al.*, *J. Am. Chem. Soc.* **132**, 4558–4559 (2010).
26. H. M. Klukovich, Z. S. Kean, S. T. Iacono, S. L. Craig, *J. Am. Chem. Soc.* **133**, 17882–17888 (2011).
27. Z. Chen *et al.*, *Science* **357**, 475–479 (2017).
28. Z. S. Kean, Z. Niu, G. B. Hewage, A. L. Rheingold, S. L. Craig, *J. Am. Chem. Soc.* **135**, 13598–13604 (2013).
29. I. Tabushi, K. Yamamura, Z. Yoshida, *J. Am. Chem. Soc.* **94**, 787–792 (1972).
30. I. M. Klein, C. C. Husic, D. P. Kovács, N. J. Choquette, M. J. Robb, *J. Am. Chem. Soc.* **142**, 16364–16381 (2020).
31. P. A. May, J. S. Moore, *Chem. Soc. Rev.* **42**, 7497–7506 (2013).
32. M. J. Kryger, A. M. Munaretto, J. S. Moore, *J. Am. Chem. Soc.* **133**, 18992–18998 (2011).
33. J. M. Lenhardt, A. L. Black Ramirez, B. Lee, T. B. Kouznetsova, S. L. Craig, *Macromolecules* **48**, 6396–6403 (2015).
34. B. K. Carpenter, *Chem. Rev.* **113**, 7265–7286 (2013).
35. D. R. Glowacki, S. P. Marsden, M. J. Pilling, *J. Am. Chem. Soc.* **131**, 13896–13897 (2009).

ACKNOWLEDGMENTS

Y.L. thanks J. Fu and X. Zhu for the helpful discussions. We thank O. Davydovich and D. Loudermilk for figure designs, and the SCS NMR Lab for its technical support. **Funding:** This work was supported by the National Science Foundation (NSF-CMMI-19-33932). T.J.M. and S.H. thank the Army Research Office (W911NF-15-1-0525) for financial support. J.M. thanks the Dr.-Leni-Schöninger foundation and the Deutsche Forschungsgemeinschaft (no. 419817859) for financial support. This work used the XStream computational resource supported by the National Science Foundation Major Research Instrumentation program (ACI-1429830). **Author contributions:** Y.L. conceived this work and performed synthesis and sonication experiments; S.H. performed dynamic simulations to construct the energy surfaces and reaction momenta; J.M. calculated the paths and the rate constants; Y.L., J.M., and S.H. interpreted the data and built the hybrid model; Y.J. assisted in the synthesis and sonication instrumentation; Q.W. assisted in polymer characterizations; T.W. performed x-ray structural analysis; T.J.M. and J.S.M. supervised the project; Y.L., T.J.M., and J.S.M. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under reference CCDC nos. 2051124, 2051125, and 2051126. All other data supporting the findings of this study are presented in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/208/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S100
Tables S1 to S17
References (*36*–*62*)
Movies S1 to S5
Structures

29 March 2021; accepted 3 June 2021
10.1126/science.abi7609

SUPERCONDUCTIVITY

Magnetic excitations in infinite-layer nickelates

H. Lu[1,3], M. Rossi[1], A. Nag[2], M. Osada[3], D. F. Li[1]†, K. Lee[3], B. Y. Wang[3], M. Garcia-Fernandez[2], S. Agrestini[2], Z. X. Shen[1,3], E. M. Been[1], B. Moritz[1], T. P. Devereaux[1,3,4], J. Zaanen[5], H. Y. Hwang[1,3], Ke-Jin Zhou[2]*, W. S. Lee[1]*

The discovery of superconductivity in infinite-layer nickelates brings us tantalizingly close to a material class that mirrors the cuprate superconductors. We measured the magnetic excitations in these nickelates using resonant inelastic x-ray scattering at the Ni L_3-edge. Undoped $NdNiO_2$ possesses a branch of dispersive excitations with a bandwidth of approximately 200 milli–electron volts, which is reminiscent of the spin wave of strongly coupled, antiferromagnetically aligned spins on a square lattice. The substantial damping of these modes indicates the importance of coupling to rare-earth itinerant electrons. Upon doping, the spectral weight and energy decrease slightly, whereas the modes become overdamped. Our results highlight the role of Mottness in infinite-layer nickelates.

The mechanism of unconventional, high-temperature superconductivity, as embodied in families of copper oxides, or cuprates, remains a highly controversial subject in condensed matter physics. However, the mechanism is only one of a number of interrelated unresolved questions, which include the origins of the strange metal phase and the implications of intertwined orders and push the limits of established mathematical theory. Soon after the discovery of cuprates, the late P. W. Anderson recognized that this strangeness can be traced back to the strong local electron repulsion (Hubbard U) and the peculiar properties of doped Mott insulators—Mottness—which remains challenging to renormalize to more conventional Fermi-liquid or Bardeen-Cooper-Schrieffer (BCS) physics (*1*, *2*). Although large-scale classical simulations of Hubbard-type models have acquired benchmarking status and will be used to validate the first generation of quantum computers (*3*), string-theoretical methods can help make the case that the strangeness of cuprate electrons originates in dense, many-body quantum entanglement (*4*).

Although many cuprate families exist, it has proven very hard to find other material classes, based on different transition metals, that exhibit similar Mott physics, let alone the quasi–two-dimensional structure and small quantum spins that are deemed to be essential. The landscape changed very recently with the discovery of superconductivity in doped monovalent infinite-layer nickelates (*5*, *6*). Their crystal structures are very similar to the cuprates, with NiO_2 planes separated by spacer layers that contain a minimal set of chemical elements, which is simpler than most of the cuprates. They were predicted (*7*–*17*) to be isoelectronic to the cuprates: monovalent Ni^+ characterized by the same $3d^9$ state as that of Cu^{2+} in the cuprates.

Are the nickelates really cuprate cousins in the most important way, Mottness? There exists considerable uncertainty regarding the local Coulomb repulsion energy U because the atomic d-orbitals of monovalent Ni^+ are more extended than those of divalent Cu^{2+}, a factor that can have a large influence on the magnitude of U. The magnetic structure can provide a constraint that allows for proper categorization. In a weakly interacting metal, the magnetic excitations spread over the full bandwidth ($W \sim 3$ eV), whereas in a spin-density-wave–like system ($U \ll W$), the excitations accumulate at low energies and exclusively near the ordering wave vectors. At higher energies (near the magnetic zone boundaries), these "paramagnons" should be completely overdamped, decaying in the metallic continuum (Landau damping). However, in a Mott insulator ($U \gtrsim W$), the metallic continuum is "pushed up" by U, and instead one is dealing with the excitations of a pure spin system, which is characterized by long-lived propagating spin waves that survive all the way up to the zone boundary. Furthermore, information about magnetic excitations can clarify the debated energy scale of the spin exchange coupling strength J in $NdNiO_2$; some theories suggest J to be one order of magnitude smaller than in cuprates owing to a large charge transfer energy (*11*–*14*), whereas some others argue differently (*15*–*17*).

To address these questions, we report here the measurements of the spin excitation spectra in infinite-layer nickelates by using resonant inelastic x-ray scattering (RIXS). We studied ~10-nm-thick films of $Nd_{1-x}Sr_xNiO_2$ grown on a $SrTiO_3$ substrate with a $SrTiO_3$ capping layer of a few unit cells to maintain and orient the nickelate crystalline structure (figs. S1 and S2) (*5*, *18*). The Ni L_3-edge x-ray absorption (XAS) spectrum of $NdNiO_2$ is shown in Fig. 1A, top, which exhibits a single peak that corresponds primarily to the $2p^63d^9 \rightarrow 2p^53d^{10}$ transition (*10*), as in cuprates. Upon doping with Sr, the XAS broadens (fig. S3), which is consistent with the recent report from an electron energy loss spectroscopy measurement with a scanning transmission electron microscope (*19*). For RIXS

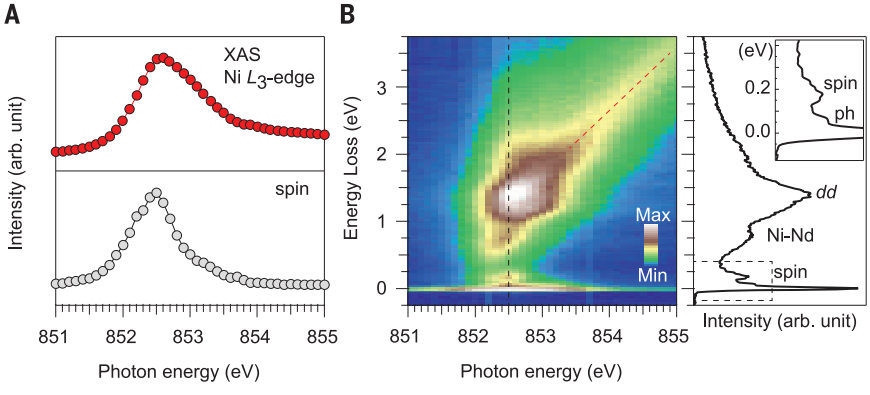


Fig. 1. XAS and RIXS map of elementary excitations in $NdNiO_2$. (**A**) (Top) Ni L_3-edge XAS measured with total electron yield at 20 K in normal incidence geometry. (Bottom) Resonant profile of spin-magnetic excitations obtained by integrating the RIXS intensity between 0.1 and 0.26 eV. (**B**) RIXS intensity map versus energy loss and incident photon energy across the Ni L_3-edge at 20 K. (Right) RIXS spectrum taken at a photon energy of 852.5 eV (black dashed line). The red dashed line indicates the fluorescence feature. (Inset) An enlarged view of the spectra within the dashed box. *dd*, Ni-Nd, spin, and ph denote orbital excitations within the Ni 3*d* orbitals, spectral feature of Ni 3*d* and Nd 5*d* orbital hybridization, magnetic excitation (~0.2 eV), and phonon (~0.07 eV), respectively.

[1]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory and Stanford University, Menlo Park, CA 94025, USA. [2]Diamond Light Source, Harwell Campus, Didcot OX11 0DE, UK. [3]Geballe Laboratory for Advanced Materials, Departments of Physics and Applied Physics, Stanford University, Stanford, CA 94305, USA. [4]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [5]Instituut-Lorentz for theoretical Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, Netherlands.
***Corresponding author. Email: kejin.zhou@diamond.ac.uk (K.-J.Z.); leews@stanford.edu (W.S.L.)** †Present address: Department of Physics, City University of Hong Kong, Kowloon, Hong Kong, China.

measurements, we adopted an experimental geometry well established to measure magnetic spin excitations in transition-metal oxides (*20*–*27*), including cuprates. A hierarchy of excitations can be resolved in a RIXS intensity map acquired by tuning the incident photon energy across the Ni L_3-edge (Fig. 1B): a fluorescence feature at 3 eV and above (Fig. 1B, red dashed line), dipole-forbidden *dd*-excitations from 1.2 to 3.0 eV, a peak at ~0.7 eV due to hybridization between Ni 3*d* and Nd 5*d* orbitals (*10*), and a low-energy feature at ~0.2 eV whose intensity is maximal near the peak of XAS (Fig. 1A, bottom) and is reminiscent of the magnetic excitations seen in RIXS maps of cuprates taken at the Cu L_3-edge (*20*–*22*). At lower energies, phonon excitations at ~0.07 eV can also be resolved (Fig. 1B).

To characterize the behavior of these low-energy excitations, we measured detailed momentum-resolved RIXS maps (*27*). As shown in Fig. 2, A and B, the excitations bear a strong resemblance to spin-1/2 antiferromagnet (AFM) magnons on a square lattice. Namely, they disperse strongly with maxima at (0.5, 0) and (0.25, 0.25), soften toward the conventional AFM ordering wave vector (0.5, 0.5), and exhibit spectral intensity suppression near (0.5, 0) (*27*). The magnetic excitations do not exhibit obvious dispersion along the *c* axis (Fig. 2C), indicating that they are quasi–two-dimensional. We fit the spectra to a damped harmonic oscillator (DHO) function $\chi''(q, \omega)$ (fig. S4) (*22*, *27*), given by

$$\chi''(q,\omega) = \frac{\gamma_q \omega}{\left(\omega^2 - \varepsilon_q^2\right)^2 + 4\gamma_q^2\omega^2} \quad (1)$$

where ε_q is the undamped mode energy and γ_q is the damping factor. The DHO function is equivalent to an antisymmetrized Lorentzian function (*21*) in the under-damped condition ($\gamma_q \ll \varepsilon_q$), but it provides a more physical description in strongly damped conditions (*22*). The fitted ε_q and γ_q along the three high-symmetry directions are shown in Fig. 3, except for the data near (0, 0), where the fitting is unreliable because the magnetic mode merges with the phonon and the tail of the elastic peak. The dispersion extracted from the DHO function is similar to that extracted by an anti-Lorentzian function, which essentially tracks the peak position of the spectrum (fig. S5). We also found a noticeable dispersion of ~50 meV along the AFM zone boundary [Figs. 2A and 3A, the (h, 0.5-h) direction], which is indicative of substantial exchange interactions beyond nearest-neighbor Ni (*28*). We fit the extracted dispersion to a linear spin wave form for the spin-1/2 square-lattice Heisenberg AFM (*29*), including nearest- and next-nearest-neighbor exchange couplings

$$H = J_1 \sum_{\langle i,j \rangle} S_i \cdot S_j + J_2 \sum_{\langle i,i' \rangle} S_i \cdot S_{i'} \quad (2)$$

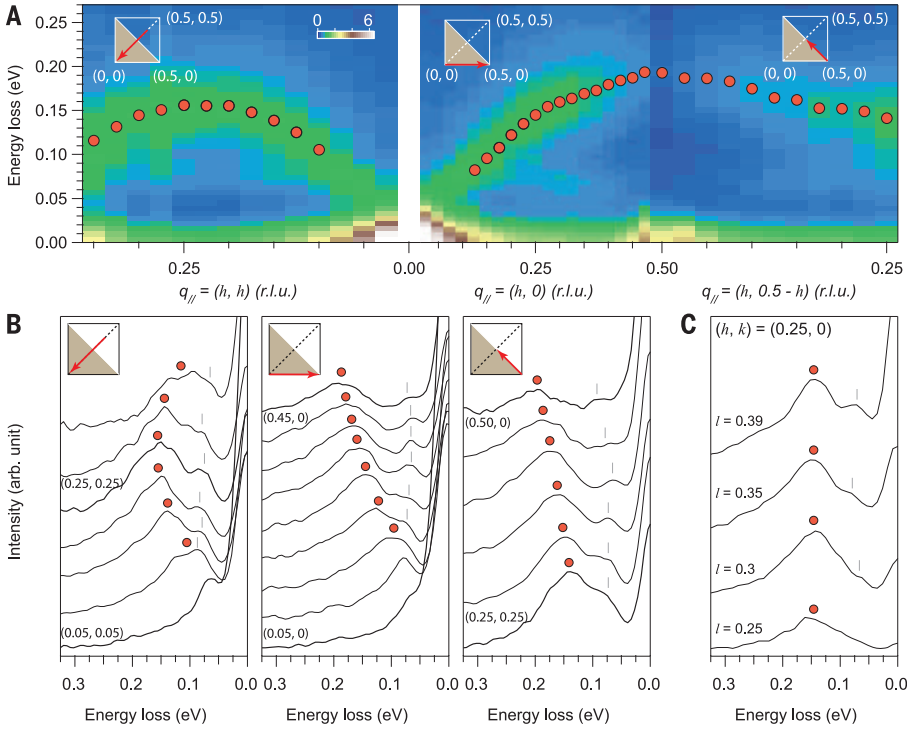


Fig. 2. Momentum-dependent RIXS intensity maps of $NdNiO_2$. (**A**) RIXS intensity maps versus energy loss and projected in-plane momentum transfer along three high-symmetry directions, as indicated with red arrows in the insets, which show a Brillouin zone with the first AFM zone shaded. Measurements were taken at 20 K. The red circles indicate peak positions of the magnetic excitation spectra. (**B**) Raw RIXS spectra at representative projected in-plane momentum transfers. The red circles indicate the peak positions of magnetic excitations, and the gray ticks indicate phonon excitations. (**C**) Raw RIXS spectra measured at a fixed in-plane momentum (0.25, 0) with different out-of-plane momentum *l*.

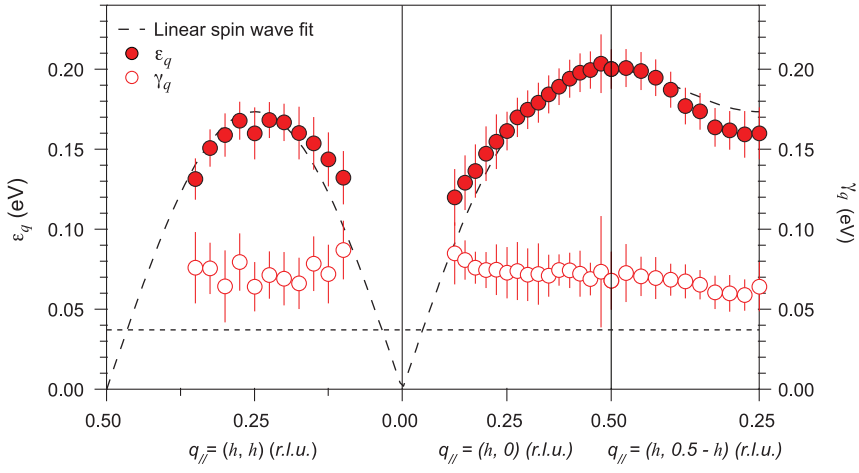


Fig. 3. Dispersion of magnetic excitations in $NdNiO_2$ and fit to the linear spin wave theory. A summary of fitted magnetic mode energy ε_q (solid red circles) and damping factor γ_q (open red circles) versus projected in-plane momentum transfer $q_{//}$ along high-symmetry directions at 20 K. The dashed curve is linear spin wave dispersion for a two-dimensional antiferromagnetic Heisenberg model fit to the data, with J_1 = 63.6 ± 3.3 meV and J_2 = −10.3 ± 2.3 meV. The energy resolution of our RIXS measurement is ~37 meV, as indicated with the horizontal dashed line. Error bars of ε_q were estimated by combining the uncertainty of zero-energy-loss position, high-energy background, and the standard deviation of the fits. Error bars of γ_q were estimated by combining the standard deviation of the fits and the uncertainty of high-energy background.

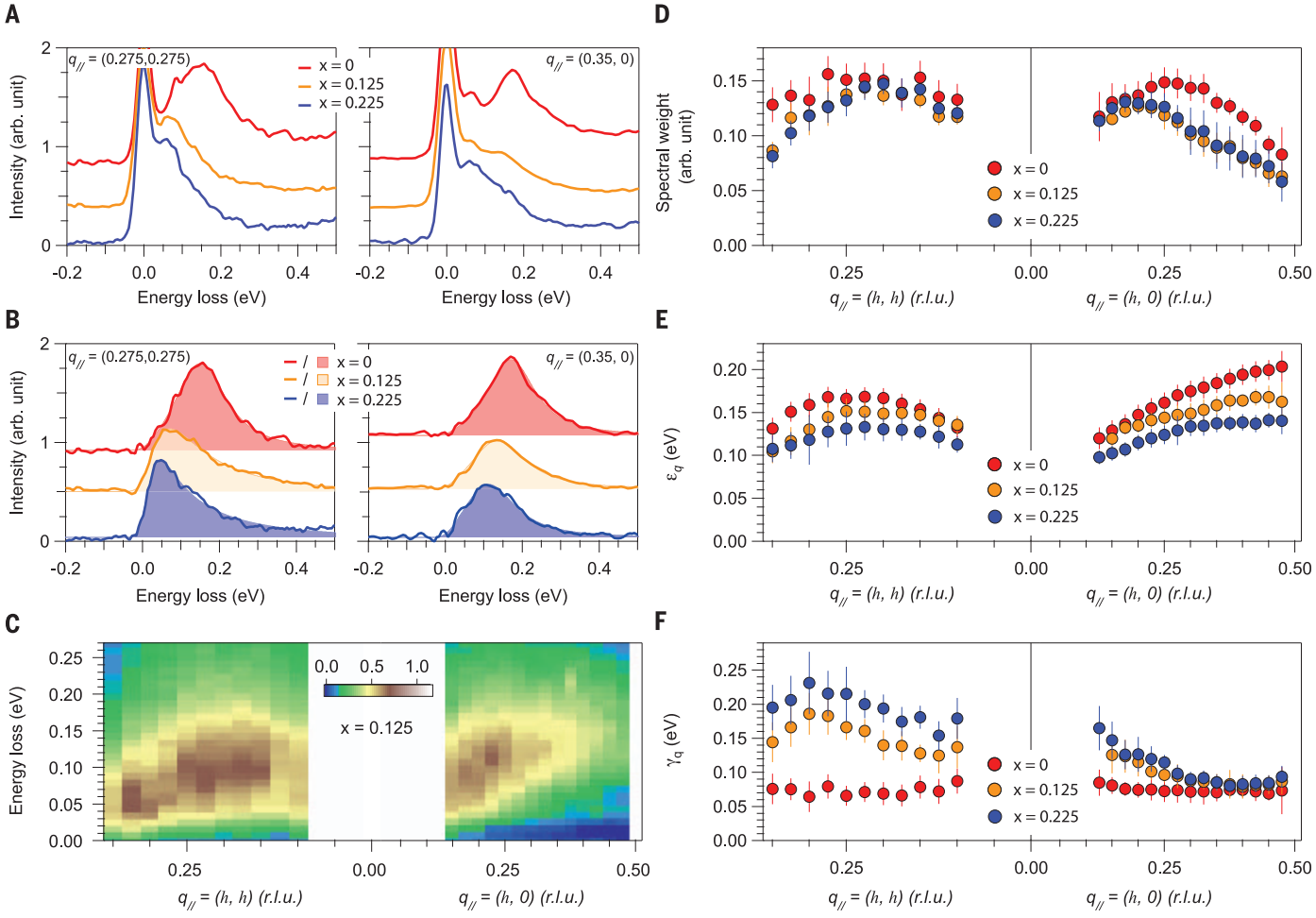


Fig. 4. RIXS spectra of $Nd_{1-x}Sr_xNiO_2$. (**A**) Raw spectra at representative momentum positions for x = 0, 0.125, and 0.225 at 20 K. (**B**) Magnetic spectra obtained by subtracting the elastic line, phonon, and background from the raw data shown in (A). The shaded areas indicate the associated DHO fitting. (**C**) Magnetic spectra map of $Nd_{1-x}Sr_xNiO_2$ (x = 0.125) along hh and h directions at 20 K. (**D** to **F**) Summary of (D) spectral weight, (E) mode energy ε_q, and (F) damping factor γ_q for the three doping concentrations. For x = 0.125 and 0.225, error bars of spectral weight and ε_q were estimated by changing the phonon intensity by ±30%, whereas those of γ_q are the estimate of the standard deviation to the fit and the uncertainty of the high-energy background. For x = 0, error bars of ε_q and γ_q are the same as those defined in Fig. 3, and error bars of the spectral weight were derived from the fittings.

where S_i, S_j, and $S_{i'}$ denote Heisenberg spins at site i, nearest-neighbor sites j, and the next-nearest-neighbor sites i', respectively (*27*). We found a substantial nearest-neighbor coupling J_1 = 63.6 ± 3.3 meV and a sizable next-nearest-neighbor coupling J_2 = −10.3 ± 2.3 meV.

The magnon bandwidth (~0.2 eV), which measures the strength of the exchange interaction, is quite comparable with that found in the parent cuprates (~0.3 to 0.4 eV) and notably higher than other nickelates (*24*, *25*). These results provide clarity to the debate in the theoretical literature on the strength of the exchange coupling (*11*–*17*). The theories, which predict a similar value of J_1, suggest that the effective onsite Coulomb interaction U falls in the range of 5 to 6 eV (*15*–*17*). However, we caution that the U value needs to be further refined with experimental values of electron dynamics parameters, which are not yet available. In addition, the magnitude of the next-nearest-neighbor coupling J_2 is intriguing. In the sense of the Zaanen-Sawatzky-Allen classification, one expects the spin exchange interaction to be rather short ranged because the nickelates are not charge-transfer compounds (*10*), in which the oxygen ligands serve as a major pathway for the super-exchange interaction (*30*). Thus, we speculate that this indicates a possible long-range RKKY metallic exchange mediated by the Nd 5*d* pockets. Although the negative J_2 (unfrustrated) should stabilize conventional Néel order, neutron powder diffraction studies indicated a lack of long-range magnetic order, despite development of magnetic correlations at low temperatures (*31*). However, the results of an absent magnetic order may be obscured by the quality of these metastable bulk materials, which is difficult to control. In addition, the heteroepitaxy stabilization by the $SrTiO_3$ substrate could also play a role in sculpting the magnetism in the film, such as through a ~0.4% compressive strain that is absent in the bulk. Our observations confirm the existence of AFM correlations but do not provide a definitive answer about the existence of long-range AFM order. A direct measurement of the magnetic spectrum at the putative AFM-ordering wave vector is needed, which is unfortunately inaccessible owing to insufficient momentum of the Ni *L*-edge photons. Further investigations that use different experimental probes on both film and bulk compounds are necessary to gain additional insight.

The other notable behavior revealed by the fits is in the spin wave damping factor γ_q (Fig. 3). The width of the magnetic peaks (~$2\gamma_q$ of ~130 meV) exceeds the value of our instrument resolution (~37 meV). This width is independent of the mode energy and momentum. The fingerprint of itinerant magnetism would be a damping that increases sharply as it moves away from the magnetic ordering wave vector [presumably at (0.5, 0.5)], and thanks to spin-wave interactions, the damping should be maximal at the magnetic zone boundary. The rather constant γ_q suggests that the magnons are instead coupled to a "heat bath," capable of dissipating only small momenta and energy. This may be consistent with the metallic Nd pockets and their small Fermi surfaces, as determined from either tight binding models (*10*, *15*) or larger experimental estimates (*5*). Incomplete chemical reduction and disorder that may suppress long-range antiferromagnetism also can play a role in damping these excitations.

We next discuss the doping evolution of the magnetic excitations in $Nd_{1-x}Sr_xNiO_2$ from x = 0 to 0.225, across the superconducting phase boundary (*32*). As shown in Fig. 4A, the magnetic excitations appear to soften with increasing doping concentration and overlap substantially with phonon excitations toward small momenta (fig. S6A). For those momentum positions, we assume the phonons to be doping independent and deduce magnetic spectra from our fitting analysis as shown in Fig. 4B (*27*), which contains information about the imaginary part of the dynamical spin susceptibility. As a function of momentum, the magnetic spectra are less dispersive (Fig. 4C and fig. S6B) than those in $NdNiO_2$. The magnetic spectral weight (Fig. 4D) decreases gently with doping, which is consistent with spin dilution as expected in a doped Mott insulator because some spins are replaced by holes. Consequently, the magnetic modes should soften, and the reduced lifetimes would indicate overdamped, relaxational dynamics. To extract the mode energy ε_q and damping γ_q, we fit the data to the DHO function (Eq. 1) (*27*). The superconducting compound acts as a doped Mott insulator, with ε_q mildly softened compared with the magnons in the parent compound with a similar dispersion (Fig. 4E). The most substantial change is in the γ_q along the *hh* direction, which increases greatly compared with the parent nickelate (Fig. 4F), causing substantial asymmetry in the magnetic spectrum (Fig. 4B). Because $\gamma_q \gtrsim \varepsilon_q$, these high-energy spin excitations in the superconductor are on the verge of becoming overdamped. The DHO fitting is consistent with a model-free spectral moment analysis, validating that our DHO analysis has captured the essential information of the magnetic spectrum (fig. S7) (*27*).

Compared with cuprates, which exhibit similar overdamped magnetic modes upon doping, the infinite-layer nickelates show subtle but important differences in the evolution of spectral weight and mode energies. In the cuprates, the spectral weight is found to be essentially unchanged, and the mode energy increases upon substantial doping (*22*). These effects seem, at first sight, counterintuitive but are rooted in the subtle interplay between magnetic interaction and longer-range hole dynamics, including the next-nearest neighbor hopping and coherent hopping involving multiple sites, such as the three-site exchange interaction (*33*, *34*). The doped nickelates exhibit mild softening and loss of spectral weight that conforms more closely to expectations of the simple *t-J* model (*35*). As charge-transfer insulators, the doped holes in cuprates reside mostly on oxygen, implying relatively extended hole–wave functions, which promote the nonlocal interactions that give rise to the interference effects. The nickelates appear to be more Mott-Hubbard–like (*10*, *15*), implying that hole–wave functions have more localized Ni *d*-like character, leading to a suppression of nonlocal interactions. With the added complexity caused by the presence of the Nd metallic pockets, the upcoming challenge is to find out how these microscopic differences affect collective phenomena in Mott systems. This should give further insight into the relationship between the microscopic ingredients determined with chemistry and the mechanism of superconductivity in these many-body–entangled quantum materials.

Note added in proof: During the review process, we became aware of RIXS results on triple-layer nickelates, finding similar values of J as in our study (*36*).

REFERENCES AND NOTES

1. P. W. Anderson, *The Theory of Superconductivity in the High-Tc Cuprates* (Princeton Univ. Press, 1997).
2. P. Phillips, *Ann. Phys.* **321**, 1634–1650 (2006).
3. J. Berges, *Nature* **569**, 339–340 (2019).
4. J. Zaanen, *SciPost Phys* **6**, 061 (2019).
5. D. Li *et al*., *Nature* **572**, 624–627 (2019).
6. S. Zeng *et al*., *Phys. Rev. Lett.* **125**, 147003 (2020).
7. V. I. Anisimov, D. Bukhvalov, T. M. Rice, *Phys. Rev. B* **59**, 7901–7906 (1999).
8. K. W. Lee, W. E. Pickett, *Phys. Rev. B* **70**, 165109 (2004).
9. A. S. Botana, M. R. Norman, *Phys. Rev. X* **10**, 011024 (2020).
10. M. Hepting *et al*., *Nat. Mater.* **19**, 381–385 (2020).
11. M. Jiang, M. Berciu, G. A. Sawatzky, *Phys. Rev. Lett.* **124**, 207004 (2020).
12. G.-M. Zhang, Y. Yang, F.-C. Zhang, *Phys. Rev. B* **101**, 020501 (2020).
13. L.-H. Hu, C. Wu, *Phys. Rev. Res.* **1**, 032046 (2019).
14. Z. Liu, Z. Ren, W. Zhu, Z. Wang, J. Yang, *npj Quantum Mater.* **5**, 31 (2020).
15. E. Been *et al*., *Phys. Rev. X* **11**, 011050 (2021).
16. V. M. Katukuri, N. A. Bogdanov, O. Weser, J. van den Brink, A. Alavi, *Phys. Rev. B* **102**, 241112 (2020).
17. R. Zhang *et al*., arXiv:2009.05816 [cond-mat.str-el] (2020).
18. K. Lee *et al*., *APL Mater.* **8**, 041107 (2020).
19. B. H. Goodge *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2007683118 (2021).
20. L. Braicovich *et al*., *Phys. Rev. Lett.* **102**, 167401 (2009).
21. M. Le Tacon *et al*., *Nat. Phys.* **7**, 725–730 (2011).
22. Y. Y. Peng *et al*., *Phys. Rev. B* **98**, 144507 (2018).
23. K. J. Zhou *et al*., *Nat. Commun.* **4**, 1470 (2013).
24. G. Fabbris *et al*., *Phys. Rev. Lett.* **118**, 156402 (2017).
25. Y. Lu *et al*., *Phys. Rev. X* **8**, 031014 (2018).
26. J. Kim *et al*., *Phys. Rev. Lett.* **108**, 177003 (2012).
27. Materials and methods are available as supplementary materials.
28. Y. Y. Peng *et al*., *Nat. Phys.* **13**, 1201–1206 (2017).
29. R. Coldea *et al*., *Phys. Rev. Lett.* **86**, 5377–5380 (2001).
30. Y. Wang, E. W. Huang, B. Moritz, T. P. Devereaux, *Phys. Rev. Lett.* **120**, 246401 (2018).
31. M. A. Hayward, M. J. Rosseinsky, *Solid State Sci.* **5**, 839–850 (2003).
32. D. Li *et al*., *Phys. Rev. Lett.* **125**, 027001 (2020).
33. C. J. Jia *et al*., *Nat. Commun.* **5**, 3314 (2014).
34. J. Bała, A. M. Oleś, J. Zaanen, *Phys. Rev. B* **52**, 4597–4606 (1995).
35. P. A. Lee, N. Nagaosa, X.-G. Wen, *Rev. Mod. Phys.* **78**, 17–85 (2006).
36. J. Q. Lin *et al*., *Phys. Rev. Lett.* **126**, 087001 (2021).
37. H. Lu *et al*., Replication data for "Magnetic excitations in infinite-layer nickelates". Harvard Dataverse, version 2 (2021).

ACKNOWLEDGMENTS

Funding: This work is supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under contract DE-AC02-76SF00515. We acknowledge the Gordon and Betty Moore Foundation's Emergent Phenomena in Quantum Systems Initiative through grant GBMF9072 for synthesis equipment. We acknowledge Diamond Light Source for providing the beam time at the I21-RIXS beamline under proposal NT25165. **Author contributions:** W.S.L. and K.-J.Z. conceived the research and designed the experiment. H.L., M.R., A.N., M.G.-F., S.A., K.-J.Z., and W.S.L. conducted the experiment at Diamond Light Source. H.L., M.R., A.N., K.-J.Z., and W.S.L. analyzed the data. M.O., D.F.L., K.L., B.Y.W., and H.Y.H. synthesized and characterized samples for the experiment. W.S.L., K.-J.Z., H.L., M.R., A.N., D.F.L., H.Y.H., E.M.B., B.M., Z.X.S., T.P.D., and J.Z. discussed and interpreted the results. H.L. and W.S.L. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data presented in this work are available online at Harvard Dataverse (*37*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/213/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*38*, *39*)

10 July 2020; resubmitted 8 September 2020
Accepted 21 May 2021
10.1126/science.abd7726

CATALYSIS

Stable and selective catalysts for propane dehydrogenation operating at thermodynamic limit

Ali Hussain Motagamwala[1,2]†, Rawan Almallahi[1,2], James Wortman[1,2], Valentina Omoze Igenegbai[1,2], Suljo Linic[1,2]*

Intentional ("on-purpose") propylene production through nonoxidative propane dehydrogenation (PDH) holds great promise for meeting the increasing global demand for propylene. For stable performance, traditional alumina-supported platinum-based catalysts require excess tin and feed dilution with hydrogen; however, this reduces per-pass propylene conversion and thus lowers catalyst productivity. We report that silica-supported platinum-tin (Pt_1Sn_1) nanoparticles (<2 nanometers in diameter) can operate as a PDH catalyst at thermodynamically limited conversion levels, with excellent stability and selectivity to propylene (>99%). Atomic mixing of Pt and Sn in the precursor is preserved upon reduction and during catalytic operation. The benign interaction of these nanoparticles with the silicon dioxide support does not lead to Pt-Sn segregation and formation of a tin oxide phase that can occur over traditional catalyst supports.

Shale gas production has led to a switch in the industrial feedstocks for propylene, from naphtha to propane (*1*). Global demand for propylene is >100 million metric tons per year, with ~4% annual growth (*2*). Oxidative propane dehydrogenation (OPDH) is an emerging technology in which thermodynamically driven overoxidation of propane to carbon oxides (CO_x) is a major drawback (*3*, *4*). For example, state-of-the-art hexagonal boron nitride OPDH catalysts achieve a relatively low per-pass propylene yield of <16% (*5*, *6*). Additionally, propylene separation from carbon oxides and other by-products of OPDH has further limited OPDH application at the commercial scale (*7*).

Nonoxidative propane dehydrogenation (PDH) produced 13.6 million metric tons of propylene in 2019, accounting for ~11% of global propylene production (*8*). Commercial PDH processes utilize either chromium (Cr)–based (Catofin process) or platinum (Pt)–based catalysts [Oleflex and steam-activated reforming (STAR) processes] (*9*). This endothermic reaction operates at a relatively high temperature (550° to 700°C) and atmospheric pressure to increase the equilibrium conversion (*10*). The main drawbacks with Cr-based catalysts are relatively low selectivity to propylene and high catalyst deactivation rate due to extensive deposition of solid carbon (coke) on the catalyst surface. Rapid catalyst deactivation requires regeneration of the Catofin catalyst every 12 min (*11*).

Relative to the Cr-based catalysts, Pt-based catalysts in which Pt nanoparticles (NPs) supported on Al_2O_3 and modified by the addition of a post-transition metal such as tin (Sn) exhibit higher selectivity. Performance depends critically on forming intermetallic PtSn NPs that improve propylene selectivity and enhance stability by reducing the rate of hydrogenolysis and coke formation (*12*). However, under PDH reaction conditions, PtSn intermetallic NPs on Al_2O_3 separate into Pt and Sn atoms and form segregated SnO_x and Pt phases on Al_2O_3 supports (*9*), and this phase-segregated catalyst deactivates rapidly through coking.

To alleviate this issue, large quantities of Sn are used in the catalyst preparation (the nominal atomic loading is Sn:Pt between 4 and 5), which increases the yield of intermetallic PtSn NPs (*12*). In addition, studies have shown that adding hydrogen to the reactant feed reduces the rate of catalyst deactivation (*13*). Although this strategy is implemented commercially (*10*), addition of hydrogen in the feed lowers the thermodynamic limit on per-pass propane conversion (fig. S1) and decreases productivity (*7*). Even with these augmentations, to combat deactivation, commercial Pt-based catalysts are operated at low per-pass propane conversion, and the catalyst is continuously regenerated with a moving-bed adiabatic reactor. Thus, catalysts are needed that can achieve higher per-pass propane conversion with high propylene selectivity and elevated catalyst stability compared with the current processes (*1*).

We developed a PDH catalyst that operates at thermodynamically limited conversion levels without the addition of hydrogen to the feed and exhibits excellent stability and selectivity to propylene (>99%). The catalyst is composed of silica-supported Pt_1Sn_1 NPs, synthesized and pretreated to ensure that the mixing of Pt and Sn is preserved during catalytic operation. We synthesized the catalyst by mixing chloroplatinic acid (H_2PtCl_6) and tin(II) chloride ($SnCl_2$) in 0.1 M hydrochloric acid solution to form a heterometallic Pt-Sn coordination complex. We used this solution to impregnate the SiO_2 support and reduce the catalyst to obtain very small Pt-Sn NPs. The principle behind catalyst synthesis is to begin with a heterometallic Pt-Sn coordination complex in which Pt and Sn are in immediate contact, and this interaction is retained during the reduction and PDH reaction. This catalyst has three main features: (i) The PtSn NPs form an atomically mixed Pt-Sn precursor. (ii) Upon reduction, very small (less than ~2 nm diameter) intermetallic PtSn NPs form. (iii) A benign interaction of these NPs with the SiO_2 support avoids segregation of Pt and Sn and formation of a SnO_x phase. We compared this Pt_1Sn_1/SiO_2 catalyst to a commercial mimic Pt-Sn/γ-Al_2O_3 [the commercially used nominal loading is Pt:Sn = 1:5 (*9*)] as well as a Pt/γ-Al_2O_3 catalyst. Catalytic performance was evaluated in a packed-bed reactor at 580°C under flowing propane, hydrogen, and an inert gas. Detailed descriptions of the catalyst synthesis, catalyst characterization, reactor operation, and analytical methods are available in the supporting information. Hock and coworkers used a similar approach to synthesize isolated Fe atoms on SiO_2 for PDH by grafting and reducing the Fe(oCp)$_2$ complex [bis(2,4-dimethyl-1,3-pentadienide) iron(II)] onto a silica support, however, poor propylene selectivity (~14%) was obtained on this catalyst (*14*).

Data in Fig. 1, A to C, show the performance of the Pt_1Sn_1/SiO_2 catalyst (1 wt % Pt and 0.6 wt % Sn) for propane dehydrogenation at 580°C, 16 vol % propane (with the remainder helium) and at a weight hourly space velocity (WHSV) of 4.7 hours^{-1}. Under these reaction conditions, the thermodynamic limit on propane conversion is 66.5% (fig. S2). The Pt_1Sn_1/SiO_2 catalyst showed near-thermodynamic conversion of >66% (Fig. 1A, red circles) and >99% selectivity to propylene (Fig. 1B, red circles). The Pt/γ-Al_2O_3 (0.5 wt % Pt) catalyst exhibited relatively poor performance, with only ~15% initial propane conversion (Fig. 1A, purple circles) and lower propylene selectivity (Fig. 1B, purple circles). Similarly, the commercial mimic Pt-Sn/γ-Al_2O_3 catalysts showed modest activity, with initial propane conversion of 22% (Fig. 1A, black triangles) and comparatively lower propylene selectivity (Fig. 1B, black triangles). Note that the active phase in these commercial mimics is Pt_3Sn, with the excess Sn present as the SnO_x layers on the alumina support (*12*).

Addition of hydrogen to the feed stream improved the activity and stability of the Pt/γ-Al_2O_3 and the commercial mimic Pt-Sn/γ-Al_2O_3 catalysts (Fig. 1, A and B, blue squares and green triangles, respectively), albeit to levels much lower than those achieved by Pt_1Sn_1/SiO_2. Addition of hydrogen did not compromise the performance of Pt_1Sn_1/SiO_2 (Fig. 1, A and B, purple diamonds). Data in Fig. 1C show the propylene yield (product of the selectivity and conversion) for the tested catalysts, demonstrating

[1]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA. [2]Catalysis Science and Technology Institute, University of Michigan, Ann Arbor, MI, USA.
*Corresponding author. Email: linic@umich.edu
†Present address: Shell Chemical Company, Shell Technology Center Houston, 3333 Highway 6 South, Houston, TX 77082, USA.

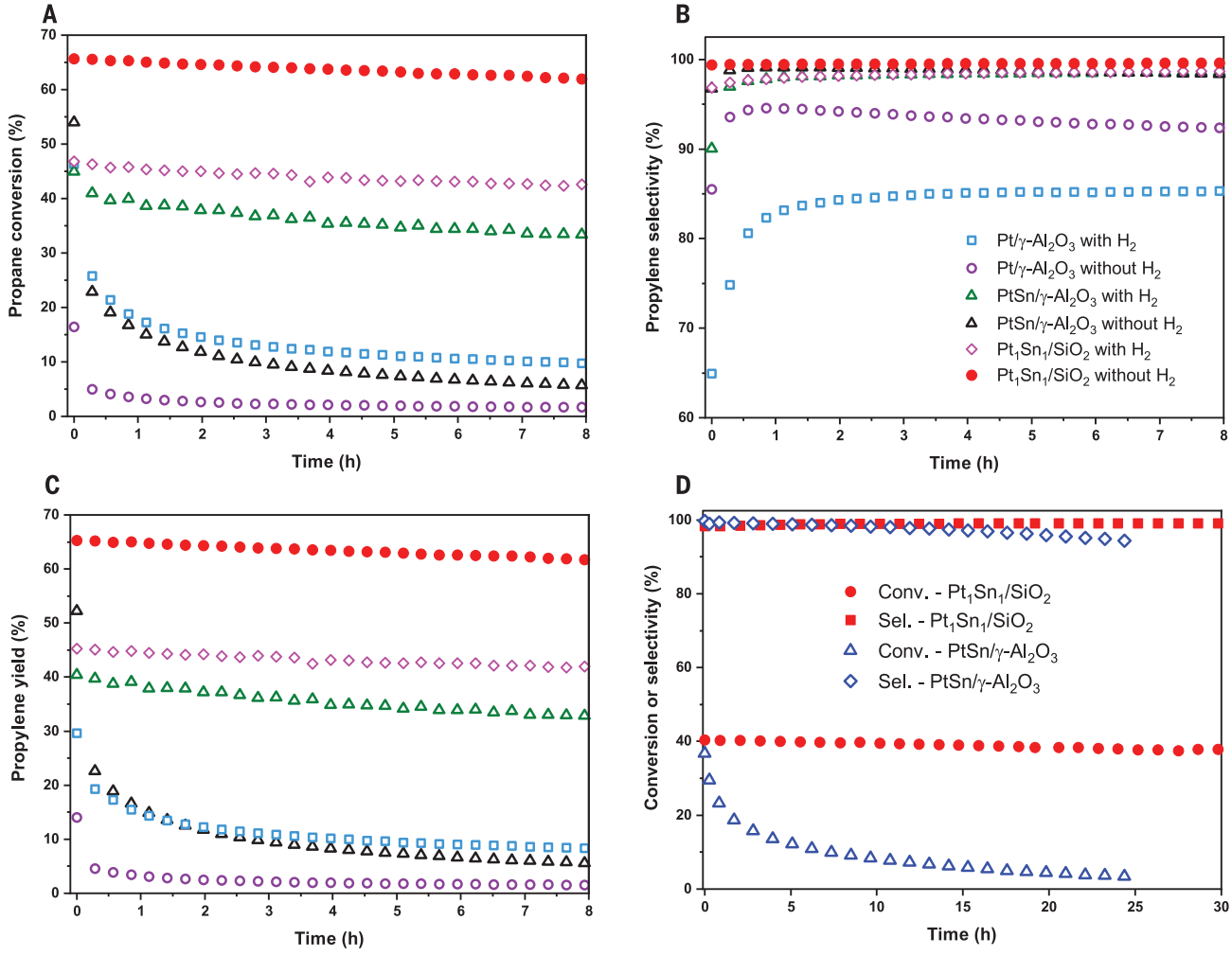


Fig. 1. Catalyst performance in propane dehydrogenation. (**A**) Propane conversion, (**B**) propylene selectivity, and (**C**) propylene yield as function of time on stream for supported PtSn catalysts with and without hydrogen added to the feed. Reaction temperature (T) = 580°C, propane partial pressure ($P_{C_3H_8}$) = 0.16, and WHSV = 4.7 hours^{-1}. (**D**) Performance of Pt_1Sn_1/SiO_2 and PtSn/γ-Al_2O_3 (commercial mimic) for PDH with an undiluted propane stream. T = 580°C, $P_{C_3H_8}$ = 1.0, and WHSV = 4.7 hours^{-1}. For Sn-promoted catalysts on an Al_2O_3 support, the atomic ratio of Sn:Pt is 5; on a SiO_2 support, the atomic ratio is 1.

the superior performance of the Pt_1Sn_1/SiO_2 catalyst. Increasing the Sn:Pt atomic ratio on SiO_2-supported catalysts to the levels used in the commercial Al_2O_3-supported catalysts (Pt:Sn = 1:5) led to poor performance (fig. S3A) (*15*). Similarly, decreasing the Sn:Pt atomic ratio on Pt-Sn/γ-Al_2O_3 to Pt_1Sn_1 led to rapid catalyst deactivation (fig. S3B).

The data in Fig. 1, A to C, were obtained with 16 vol % propane in helium. Propane streams are diluted in commercial operations to increase the limiting equilibrium conversion, which is higher at lower propane partial pressure (fig. S2), and to limit catalyst deactivation rates (*7*, *16*, *17*). This approach requires additional downstream separation units and limits process efficiency. In an undiluted propane stream, the Pt_1Sn_1/SiO_2 catalyst reached the thermodynamic conversion limit of 40% (Fig. 1D and fig. S2), with >99% selectivity to propylene. Even under these harsh, high-conversion and undiluted propane stream conditions, the Pt_1Sn_1/SiO_2 catalyst was stable and operated at the thermodynamic conversion limit and >99% selectivity for at least 30 hours on stream, whereas the PtSn/γ-Al_2O_3 (commercial mimic) deactivated rapidly under these conditions (Fig. 1D).

We also compared the performance of the Pt_1Sn_1/SiO_2 catalyst to various catalysts previously tested in propane dehydrogenation (*18*–*53*). Data in Fig. 2A and table S1 show the comparison of the initial (fresh catalyst) selectivity-conversion results reported for different catalysts. The slanted lines in the figure quantify the proximity of the measured conversion-selectivity data points to the highest possible conversion-selectivity lines for different conditions at which these catalysts were tested. The highest possible conversion-selectivity line has the value of 100% when the catalyst operates at equilibrium conversion and 100% propylene selectivity. By this metric, the Pt_1Sn_1/SiO_2 catalyst outperformed other catalysts, reaching very close to the highest possible conversion-selectivity line (100% selectivity to propylene at equilibrium conversion), whereas other tested catalysts were mostly <80% of the highest possible performance. We emphasize that on Pt_1Sn_1/SiO_2, the highest possible yield can also be achieved with undiluted propane feed (point labeled as 1 in Fig. 2A), which has not been demonstrated previously. For example, $PtSn/Al_2O_3$ nanosheets (*45*) reach near-equilibrium conversion (point labeled as 42 in Fig. 2A); however, the propane is diluted with hydrogen (C_3H_8:H_2 = 0.8) and an inert gas, leading to low propane partial pressure.

The overall productivity of a PDH catalyst is affected by the conversion-selectivity data as well as the catalyst stability and the inherent reaction rates. We developed a figure of merit that combined these two metrics and quantified the relative productivities of different catalysts. To quantify the stability of the catalysts, the

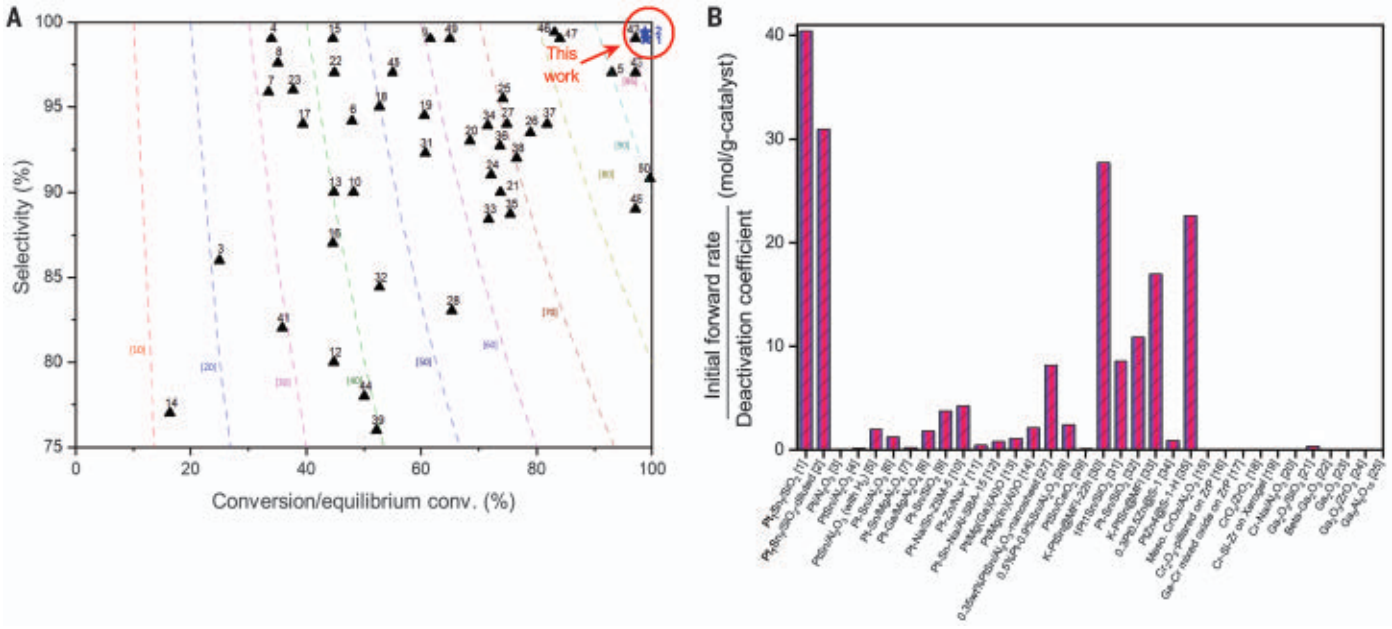


Fig. 2. Comparison of Pt_1Sn_1/SiO_2 with previously reported PDH catalysts. (**A**) Conversion-selectivity plots for different PDH catalysts. Numbers correspond to row numbers in table S1. (**B**) Catalyst productivity (product of the initial reaction rate and the inverse of the deactivation coefficient) for various PDH catalysts studied in the literature. Numbers in brackets correspond to row numbers in table S2. Two data points from this work are for two different conditions (with and without dilution in the feed).

propylene yield data, reported as a function of time for different catalysts, were analyzed assuming first-order deactivation kinetics, as is usually done for catalytic reactions on supported metal NPs (*9*). This analysis allowed us to calculate the catalyst deactivation coefficient, k_d (see supplementary text for derivation), for different catalysts. Data in table S2 show the calculated values of the deactivation coefficient.

To quantify the inherent kinetic rates for the same catalysts, the initial forward rates of propylene formation were obtained by analyzing the data reported in the literature. Most data are reported for an integral reactor design (i.e., high propane conversion), so we performed an integral reactor analysis to obtain an analytical expression for the initial reaction rates (see supplementary text for derivation). This analytical expression was used to calculate the initial forward rate of propylene production for various catalysts (table S2). The product of the initial forward reaction rate and the inverse of the deactivation coefficient (proportional to catalyst stability) provided a measure of the overall catalyst productivity. Data in Fig. 2B show the productivities of different catalysts operated under a range of conditions. The data show that the Pt_1Sn_1/SiO_2 catalyst reported here demonstrated the highest productivity for propylene production of all tested catalysts. For example, Pt_1Sn_1/SiO_2 in the absence of hydrogen was more productive than the commercial mimic Pt-Sn/γ-Al_2O_3 by two orders of magnitude.

The overall catalyst performance of the Pt_1Sn_1/SiO_2 catalysts was also studied at different WHSVs (i.e., including those away from the equilibrium conversion). Catalyst productivity remained high, between 28 and 40 moles per gram of catalyst, for different WHSVs (fig. S4 and table S2).

The data above show that the Pt_1Sn_1/SiO_2 catalysts exhibited very high productivity and stability compared with other catalysts. Because the dehydrogenation reaction is operated under carbon-rich reducing conditions, it is inevitable that, over time, catalyst deactivation due to deposition of solid carbon would compromise the performance of the catalysts. To investigate whether this Pt_1Sn_1/SiO_2 catalyst can be regenerated, we intentionally increased the rate of catalyst deactivation by increasing the propane flow rate (i.e., propane WHSV) and diluting the propane stream (which increases the propane equilibrium conversion). These conditions push the catalysts away from the equilibrium conversion and lead to a more rapid, forced carbon-induced deactivation. The data in fig. S5 show that, under these conditions, the propane conversion goes from ~41 to 39% in 120 min. Under identical conditions, the other tested catalysts deactivate very rapidly to very low conversion, as illustrated for the commercial mimic. A transmission electron microscopy (TEM) image of the spent catalyst, after 120 min of PDH reaction, is shown in fig. S6. The image shows that the Pt-Sn NPs are well dispersed, with an average particle diameter of ~1.6 ± 0.67 nm, which is consistent with the particle size in fresh samples. Raman spectroscopy (fig. S7) on the spent catalyst confirmed that the slight loss in catalyst activity was due to coking. We attempted to regenerate the Pt_1Sn_1/SiO_2 catalyst by oxidation, using 1% O_2 as a mild oxidant at 500°C followed by reduction at 600°C (see supplementary materials for details of the regeneration procedure). The regeneration studies showed that Pt_1Sn_1/SiO_2 catalysts can be partially regenerated. For example, we measured ~97% of the initial activity and >99% of the initial selectivity after three PDH/regeneration cycles (fig. S5).

We hypothesized that the performance and stability of the Pt_1Sn_1/SiO_2 catalyst was derived from favorable interactions among the support (SiO_2), promoter (Sn), and active metal (Pt), which allowed for a high degree of atomic mixing of Pt and Sn within very small intermetallic PtSn NPs. Mixing of Pt and Sn atoms in NPs can be compromised when these materials are introduced on other oxide supports such as Al_2O_3, where Pt and Sn atoms segregate, leading to the formation of Pt NPs and an Sn-oxide phase (*54*). We studied the structure of the catalyst and its precursors by using a combination of x-ray diffraction (XRD), ultraviolet-visible (UV-Vis) spectroscopy, x-ray photoelectron spectroscopy (XPS), diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS), TEM, and x-ray absorption spectroscopy [extended x-ray absorption fine structure (EXAFS) and x-ray absorption near-edge structure (XANES)].

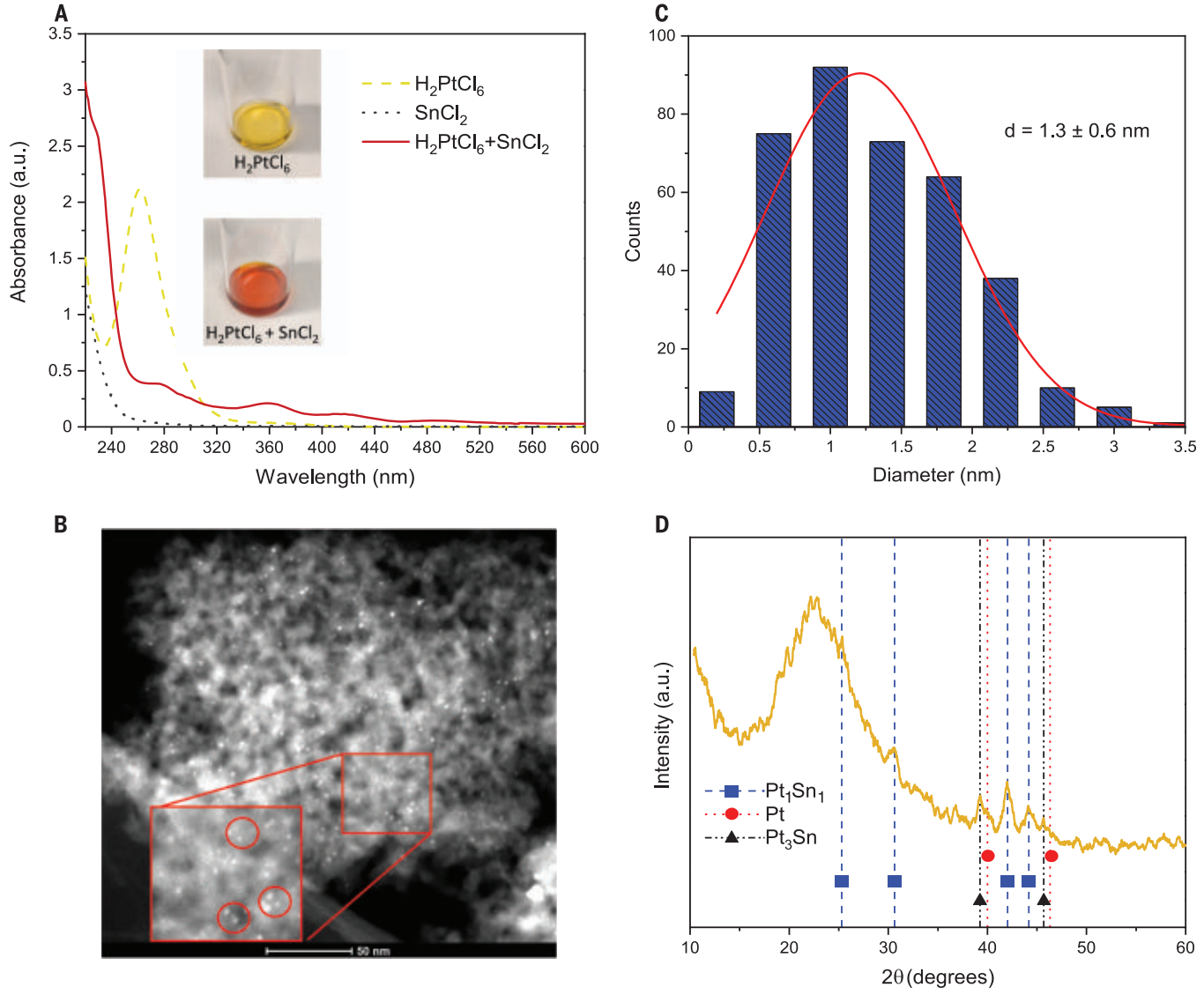


Fig. 3. Pt_1Sn_1/SiO_2 catalysts characterization. (**A**) UV-Vis absorption spectrum of precursor solutions. (Inset) Photographs of chloroplatinic acid solution (top) and a mixture of chloroplatinic acid and tin(II) chloride (bottom). a.u., arbitrary units. (**B**) Bright-field TEM image of Pt-Sn NPs formed on SiO_2. (Inset) Enlarged area of the image identifying small Pt-Sn NPs. (**C**) Particle size distribution of PtSn NPs shown in (B). (**D**) X-ray diffraction pattern for Pt_1Sn_1/SiO_2. The peak positions for Pt, PtSn, and Pt_3Sn were obtained from the International Centre for Diffraction Data. 2θ, diffraction angle.

Data in Fig. 3A show that the process of mixing chloroplatinic acid (H_2PtCl_6) and tin(II) chloride ($SnCl_2$) in 0.1 M hydrochloric acid solution formed a heterometallic Pt-Sn coordination complex. The UV-Vis absorption spectrum of H_2PtCl_6 is characterized by a prominent absorption peak at ~260 nm; $SnCl_2$ exhibits no absorption in the visible range. The combined precursor solution shows several additional UV-Vis absorption peaks at 279, 363, 420, and 488 nm that are characteristic of a heterometallic Pt-Sn complex (*55*), indicating the mixing of Sn and Pt atoms before reduction and growth of NPs on the support.

A bright-field TEM image of Pt-Sn NPs formed on SiO_2 upon the reduction of the Pt-Sn complex shows high metal dispersion on the support, with an average NP size of ~1.3 ± 0.6 nm (Fig. 3B). The NP size distribution (Fig. 3C) shows that most NPs were <2 nm. The XRD spectra in Fig. 3D show that the formed NPs were essentially alloys of Pt and Sn. This finding is supported by the presence of the XRD peaks corresponding to Pt_1Sn_1 (41.8° and 44.1°) and Pt_3Sn_1 (39.0° and 45.3°) stoichiometries.

To establish that the mixing of Sn and Pt atoms in PtSn NPs extended to their surface, we used DRIFTS to measure the infrared absorption spectra of carbon monoxide (CO), as its adsorption on Pt surfaces has been well characterized. The CO-DRIFTS spectra for Pt/SiO_2 and Pt_1Sn_1/SiO_2 (Fig. 4A) revealed linear (on top) and bridge-bonded CO on Pt/SiO_2 with infrared adsorption peaks at 2073 and 1821 cm^{-1}, respectively (*56*, *57*). On Pt_1Sn_1/SiO_2, the bridge-bonded CO was not present, and the linearly bonded peak had much lower intensity. Because CO does not adsorb at the bridge sites between Sn and Pt, the data suggest that Sn broke the ensembles of Pt atoms on the NP surfaces and created a checkerboard Pt-Sn surface structure (*58*). The decreased intensity of the linearly bonded CO adsorption peak was consistent with reduced surface coverage of CO on the PtSn compared with Pt (*59*), further demonstrating the mixing of Pt and Sn in Pt_1Sn_1/SiO_2. Introduction of Sn to the catalyst also slightly shifted the vibrational frequency of the linearly bonded CO band to 2078 cm^{-1}, indicating the perturbation of the electronic structure of surface Pt atoms by neighboring Sn atoms (*56*).

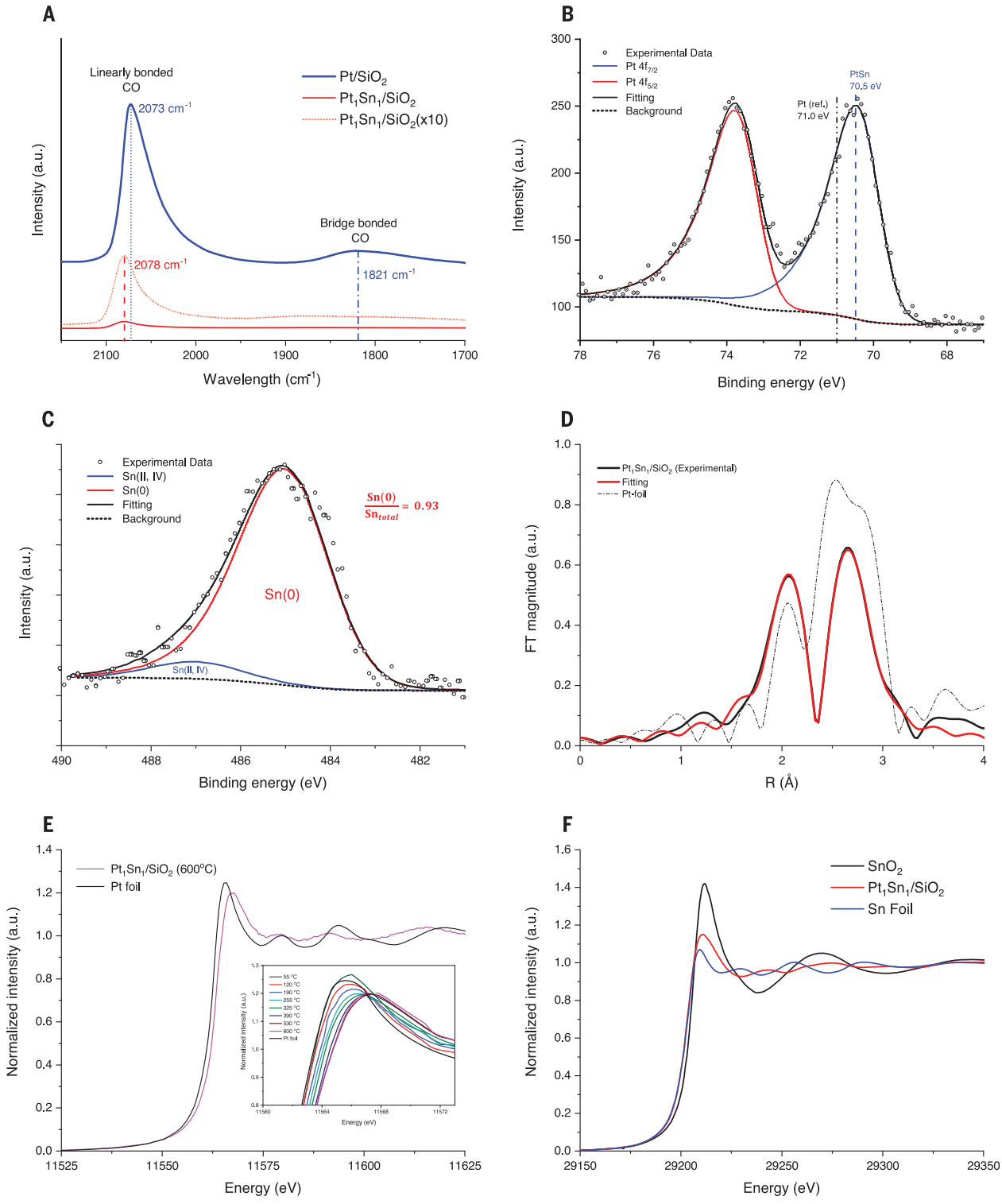


Fig. 4. Characterization of Pt_1Sn_1/SiO_2 catalysts. (**A**) DRIFTS spectra of adsorbed CO on Pt_1Sn_1/SiO_2. (**B**) Pt 4f photoemission spectra of reduced Pt_1Sn_1/SiO_2. (**C**) Sn 3d photoemission spectra of reduced Pt_1Sn_1/SiO_2. (**D**) EXAFS magnitude of the Fourier transform (FT) of Pt_1Sn_1/SiO_2 measured at room temperature in He after reduction in H_2 at 373 K. Coordination numbers for Pt-Sn and Pt–Pt are 4.2 ± 0.4 and 3.9 ± 0.9, respectively, and bond distances for Pt-Sn and Pt-Pt are 2.72 ± 0.01 and 2.75 ± 0.02, respectively (table S3). R, radial distance. (**E**) Pt L_{III}-edge XANES spectra of Pt_1Sn_1/SiO_2 reduced at 600°C and of Pt foil. (Inset) Pt L_{III}-edge XANES spectra of Pt_1Sn_1/SiO_2 collected at different temperatures while the catalyst was reduced from 55° to 600°C. (**F**) Sn K-edge XANES spectra of Pt_1Sn_1/SiO_2, SnO_2, and Sn foil.

To further confirm the presence of PtSn surface alloys, XPS of the freshly reduced Pt_1Sn_1/SiO_2 catalyst showed that the atomic ratio of Pt and Sn was ~1, consistent with the nominal loading (fig. S8). Figure 4B shows the XPS spectra associated with the Pt $4f_{7/2}$ electronic state, demonstrating that the Pt atoms were in the metallic Pt(0) state. Additionally, the Pt $4f_{7/2}$ peak shifted toward lower binding energy (70.5 eV) relative to a monometallic Pt reference (71.0 eV), in agreement with prior reports that the formation of Pt–Sn caused a slight electron transfer from Sn to Pt that increased the local electron density on Pt and local Coulombic repulsion, thereby lowering the electron binding energy (*60*–*62*). Furthermore, the Sn $3d_{5/2}$ XPS spectra in Fig. 4C show that most Sn atoms were metallic Sn, and a small fraction of Sn existed as SnO or SnO_2, further confirming the formation of PtSn intermetallic NPs (*63*, *64*).

The real-space Pt L_{III} EXAFS spectra for the Pt_1Sn_1/SiO_2 catalyst (Fig. 4D) are substantially different from the spectra of monometallic Pt and indicate the mixing of Pt and Sn atoms. The best-fit spectrum for the Pt_1Sn_1/SiO_2 catalyst is characterized by the presence of the Pt-Pt and Pt-Sn scattering paths, consistent with the formation of PtSn intermetallic phases. The parameters used to obtain the best-fit spectra are provided in the Fig. 4D caption.

When assigning EXAFS spectra to a geometric structure, it is instructive to compare the measured bond distances to those corresponding to thermodynamically stable phases of a material. For Pt_3Sn, both Pt-Pt and Pt-Sn bond distances are 2.83 Å. By contrast, the Pt-Pt bond distance in PtSn is 2.72 Å, whereas the Pt-Sn bond distance is 2.73 Å (*63*, *64*). The Pt-Pt and Pt-Sn bond distances experimentally measured for the Pt_1Sn_1/SiO_2 catalyst (2.75 and 2.71 Å, respectively) suggest that the catalyst was likely a mix of Pt_3Sn and PtSn, with a higher concentration of the PtSn phase, as no peaks were observed at 2.83 Å.

Data in Fig. 4, E and F, show the XANES spectra associated with the Pt L_{III}-edge and Sn K-edge, respectively, of the Pt_1Sn_1/SiO_2 catalyst and relevant controls. The Pt L_{III}-edge XANES spectra in Fig. 4E are consistent with the analysis of the EXAFS data, indicating that the mixing of Pt and Sn took place in the reduced Pt_1Sn_1/SiO_2 catalyst. Indeed, the formation of a stable Pt-Sn phase was complete by 390°C—above this temperature, no further changes were observed (Fig. 4E, inset). A slight increase in the Pt L_{III}-edge energy and a decrease in white-line intensity and its broadening observed for Pt_1Sn_1/SiO_2 compared with pure Pt could be attributed to the formation of the Pt-Sn phase, consistent with XPS results that showed slight electron transfer from Sn to Pt. Likewise, the slight increase in Sn K-edge energy and the increase in the white-line intensity relative to Sn foil in (Fig. 4F) indicated that Sn existed in both metallic and oxide forms on the catalyst surface.

REFERENCES AND NOTES

1. National Academies of Sciences, Engineering, and Medicine, *The Changing Landscape of Hydrocarbon Feedstocks for Chemical Production: Implications for Catalysis: Proceedings of a Workshop* (The National Academies Press, 2016); https://doi.org/10.17226/23555.
2. *Focus Catal.* **2019**, 5 (2019).
3. B. Frank, A. Dinse, O. Ovsitser, E. V. Kondratenko, R. Schomäcker, *Appl. Catal. A Gen.* **323**, 66–76 (2007).
4. P. Chaturbedy, M. Ahamed, M. Eswaramoorthy, *ACS Omega* **3**, 369–374 (2018).
5. J. T. Grant *et al*., *Science* **354**, 1570–1573 (2016).
6. Z. Zhang, E. Jimenez-Izal, I. Hermans, A. N. Alexandrova, *J. Phys. Chem. Lett.* **10**, 20–25 (2019).
7. A. Agarwal, D. Sengupta, M. El-Halwagi, *ACS Sustain. Chem. Eng.* **6**, 2407–2421 (2018).
8. R. Rabenhorst, "On Purpose – What's Driving New Propane Dehydrogenation Projects in North America?" Nexant (2019); www.nexant.com/resources/purpose-what-s-driving-new-propane-dehydrogenation-projects-north-america.
9. J. J. H. B. Sattler, J. Ruiz-Martinez, E. Santillan-Jimenez, B. M. Weckhuysen, *Chem. Rev.* **114**, 10613–10653 (2014).
10. W. Spieker, G. J. Nedohin, in *Handbook of Petroleum Refining Processes*, R. A. Meyers, Ed. (McGraw-Hill, ed. 4, 2016), pp. 173–182.
11. M. M. Bhasin, J. H. McCain, B. V. Vora, T. Imai, P. R. Pujadó, *Appl. Catal. A Gen.* **221**, 397–419 (2001).
12. A. Iglesias-Juez *et al*., *J. Catal.* **276**, 268–279 (2010).
13. B. Frank, T. P. Cotter, M. E. Schuster, R. Schlögl, A. Trunschke, *Chemistry* **19**, 16938–16945 (2013).
14. B. Hu *et al*., *ACS Catal.* **5**, 3494–3503 (2015).
15. Low interaction between the support and metal NP leads to the formation of PtSn NPs wherein the effect of Sn dominates, leading to poor activity but high selectivity and stability (fig. S3).
16. B. V. Vora, P. R. Pujado, in *Encyclopedia of Chemical Processing*, S. Lee, Ed. (Taylor & Francis, 2006), pp. 379–397.
17. B. V. Vora, *Top. Catal.* **55**, 1297–1308 (2012).
18. O. A. Bariås, A. Holmen, E. A. Blekkan, *J. Catal.* **158**, 1–12 (1996).
19. J. Salmones, J. A. Wang, J. A. Galicia, G. Aguilar-Rios, *J. Mol. Catal. Chem.* **184**, 203–213 (2002).
20. Y. Zhou, S. M. Davis, "Low-Pressure Dehydrogenation of Light Paraffins," US Patent 5,214,227 (1993).
21. N. Kaylor, R. J. Davis, *J. Catal.* **367**, 181–193 (2018).
22. Y. Zhang, Y. Zhou, L. Huang, M. Xue, S. Zhang, *Ind. Eng. Chem. Res.* **50**, 7896–7902 (2011).
23. P. L. De Cola, R. Gläser, J. Weitkamp, *Appl. Catal. A Gen.* **306**, 85–97 (2006).
24. Y. Duan, Y. Zhou, Y. Zhang, X. Sheng, M. Xue, *Catal. Lett.* **141**, 120–127 (2011).
25. G. Siddiqi, P. Sun, V. Galvita, A. T. Bell, *J. Catal.* **274**, 200–206 (2010).
26. P. Sun, G. Siddiqi, W. C. Vining, M. Chi, A. T. Bell, *J. Catal.* **282**, 165–174 (2011).
27. D. Shee, A. Sayari, *Appl. Catal. A Gen.* **389**, 155–164 (2010).
28. F. J. Pérez-Reina, E. Rodríguez-Castellón, A. Jiménez-López, *Langmuir* **15**, 8421–8428 (1999).
29. M. Alcántara-Rodríguez, E. Rodríguez-Castellón, A. Jiménez-López, *Langmuir* **15**, 1115–1120 (1999).
30. X. Zhang, Y. Yue, Z. Gao, *Catal. Lett.* **83**, 19–25 (2002).
31. K. L. Fujdala, D. Tilley, *J. Catal.* **218**, 123–134 (2003).
32. J. J. H. B. Sattler *et al*., *Chem. Commun.* **49**, 1518–1520 (2013).
33. M. Saito, S. Watanabe, I. Takahara, M. Inaba, K. Murata, *Catal. Lett.* **89**, 213–217 (2003).
34. B. Zheng, W. Hua, Y. Yue, Z. Gao, *J. Catal.* **232**, 143–151 (2005).
35. P. Michorczyk, J. Ogonowski, *Appl. Catal. A Gen.* **251**, 425–433 (2003).
36. B. Xu, B. Zheng, W. Hua, Y. Yue, Z. Gao, *J. Catal.* **239**, 470–477 (2006).
37. M. Chen *et al*., *J. Catal.* **256**, 293–300 (2008).
38. H. Zhu *et al*., *J. Catal.* **320**, 52–62 (2014).
39. Y. Zhang *et al*., *J. Mol. Catal. Chem.* **381**, 138–147 (2014).
40. Y. Dai *et al*., *J. Catal.* **381**, 482–492 (2020).
41. X. Liu *et al*., *Chem. Eng. J.* **247**, 183–192 (2014).
42. N. Prakash, M. H. Lee, S. Yoon, K. D. Jung, *Catal. Today* **293–294**, 33–41 (2017).
43. J. Li *et al*., *J. Catal.* **352**, 361–370 (2017).
44. Y. Zhang *et al*., *Catal. Commun.* **7**, 860–866 (2006).
45. L. Shi *et al*., *Angew. Chem. Int. Ed.* **54**, 13994–13998 (2015).
46. E. J. Jang, J. Lee, H. Y. Jeong, J. H. Kwak, *Appl. Catal. A Gen.* **572**, 1–8 (2019).
47. H. Xiong *et al*., *Angew. Chem. Int. Ed.* **56**, 8986–8991 (2017).
48. L. Liu *et al*., *Nat. Catal.* **3**, 628–638 (2020).
49. L. Deng *et al*., *ChemCatChem* **6**, 2680–2691 (2014).
50. L. Deng, T. Shishido, K. Teramura, T. Tanaka, *Catal. Today* **232**, 33–39 (2014).
51. L. Liu *et al*., *Nat. Mater.* **18**, 866–873 (2019).
52. Y. Wang, Z. P. Hu, X. Lv, L. Chen, Z. Y. Yuan, *J. Catal.* **385**, 61–69 (2020).
53. Q. Sun *et al*., *Angew. Chem. Int. Ed.* **59**, 19450–19459 (2020).
54. H. N. Pham, J. J. H. B. Sattler, B. M. Weckhuysen, A. K. Datye, *ACS Catal.* **6**, 2257–2264 (2016).
55. D. Rodríguez, J. Sánchez, G. Arteaga, *J. Mol. Catal. Chem.* **228**, 309–317 (2005).
56. K. Balakrishnan, J. Schwank, *J. Catal.* **138**, 491–499 (1992).
57. S. G. Podkolzin, J. Shen, J. J. De Pablo, J. A. Dumesic, *J. Phys. Chem. B* **104**, 4169–4180 (2000).
58. Q. Wang, D. Tichit, F. Meunier, H. Guesmi, *J. Phys. Chem. C* **124**, 9979–9989 (2020).
59. A. Moscu, Y. Schuurman, L. Veyre, C. Thieuleux, F. Meunier, *Chem. Commun.* **50**, 8590–8592 (2014).
60. J. Shen, J. M. Hill, R. M. Watwe, B. E. Spiewak, J. A. Dumesic, *J. Phys. Chem. B* **103**, 3923–3934 (1999).
61. J. N. Schwämmlein, P. A. L. Torres, H. A. Gasteiger, H. A. El-Sayed, *Sci. Rep.* **10**, 59 (2020).
62. J. A. Rodriguez, T. Jirsak, S. Chaturvedi, J. Hrbek, *J. Am. Chem. Soc.* **120**, 11149–11157 (1998).
63. J. M. Ramallo-López *et al*., *J. Phys. Chem. B* **107**, 11441–11451 (2003).
64. G. J. Siri *et al*., *Appl. Catal. A Gen.* **278**, 239–249 (2005).

ACKNOWLEDGMENTS

We thank R. J. Meyer, S. F. Liu, and J. R. Johnson at ExxonMobil for their help in obtaining EXAFS data and fruitful discussion on reaction data and catalyst characterization. We thank T. Ma and H. Sun at University of Michigan for their help in collecting TEM data. **Funding:** This material is based on work supported by RAPID under award DE-EE0007888. The materials characterization part of the work was supported by the US DOE Office of Basic Energy Sciences, Division of Chemical Sciences (DE-SC0021008). The authors acknowledge the financial support of the University of Michigan College of Engineering and technical support from the Michigan Center for Materials Characterization. R.A. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant DGE 1256260. **Author contributions:** S.L. and A.H.M. conceived the project and designed the experiments. A.H.M., R.A., J.W., and V.O.I. designed and constructed the reactor used to obtain reaction data. A.H.M. and R.A. performed the propane dehydrogenation experiments. R.A., A.H.M., J.W., and V.O.I. performed catalyst characterization. All authors analyzed the reaction data and catalyst characterization data. S.L., A.H.M., and R.A. wrote the paper. All authors edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. Additional data related to this paper may be requested from the authors.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/217/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S3
References

27 January 2021; accepted 28 May 2021
10.1126/science.abg7894

PHYSIOLOGY

Skeletal muscle thermogenesis enables aquatic life in the smallest marine mammal

Traver Wright[1,2]*, Randall W. Davis[3], Heidi C. Pearson[4,5], Michael Murray[6], Melinda Sheffield-Moore[1,2]

Basal metabolic rate generally scales with body mass in mammals, and variation from predicted levels indicates adaptive metabolic remodeling. As a thermogenic adaptation for living in cool water, sea otters have a basal metabolic rate approximately three times that of the predicted rate; however, the tissue-level source of this hypermetabolism is unknown. Because skeletal muscle is a major determinant of whole-body metabolism, we characterized respiratory capacity and thermogenic leak in sea otter muscle. Compared with that of previously sampled mammals, thermogenic muscle leak capacity was elevated and could account for sea otter hypermetabolism. Muscle respiratory capacity was modestly elevated and reached adult levels in neonates. Premature metabolic development and high leak rate indicate that sea otter muscle metabolism is regulated by thermogenic demand and is the source of basal hypermetabolism.

Sea otters (*Enhydra lutris*) are the smallest of all marine mammals. Their small size, combined with the high thermal conductivity of water (23 times that of air) and the cold water temperatures of their North Pacific habitat (0° to 15°C), imposes a thermoregulatory challenge to maintain a core body temperature of 37°C. Unlike other marine mammals that have subcutaneous blubber for thermal insulation, sea otters rely on air trapped in their dense fur (the highest hair density of any mammal) for insulation (*1*). However, this is not adequate to offset heat loss, so an increase in metabolic heat production is required to maintain a stable core body temperature. As a result, sea otters have a basal metabolic rate (BMR) approximately three times that predicted for a eutherian mammal of similar size (*1*, *2*).

In endothermic mammals, BMR includes both the metabolism needed to produce energy for basic physiological functions and that necessary to maintain a constant core body temperature. Although BMR generally scales with body mass (i.e., smaller mammals have a greater mass-specific metabolic rate), deviation from the mass-predicted rate can indicate ecologically-driven adaptive metabolic remodeling (e.g., increased heat production in response to low ambient temperature) (*3*). Polar and small-bodied marine mammals are particularly vulnerable to heat loss and require increased heat production to maintain body temperature (*2*, *4*, *5*). This thermogenic hypermetabolism in thermally challenged polar mammals appears to be a result of reduced tissue-level mitochondrial efficiency, which may explain the deviation from predicted metabolic scaling (*6*).

Whole-body BMR emerges as an aggregate of individual tissue metabolism (*7*). Because of its large relative mass and high metabolic capacity, skeletal muscle is an important determinant of BMR and endothermy in mammals (*8*). As a result, skeletal muscle is the primary target of adaptive metabolic remodeling in animals, and deviation from the predicted BMR in endotherms primarily reflects changes to muscle mass and activity (*9*). Within muscle, heat production occurs either as a by-product of contractile work [adenosine triphosphate (ATP)–driven functional contraction or shivering] or directly via nonshivering thermogenesis from mitochondrial proton leak and other "futile" ion cycling (*8*). To better understand thermogenic hypermetabolism in cold-adapted mammals, we characterized sea otter muscle metabolism. Given the elevated thermogenic BMR of sea otters and the adaptive metabolic capacity of muscle, we hypothesized that sea otter skeletal muscle would have a high capacity for thermogenic mitochondrial proton leak, which may explain their elevated BMR. Using sea otters as an extreme example of thermogenic hypermetabolism, we explored how tissue-level metabolic plasticity enables homeotherms to inhabit thermally challenging cold environments.

At the cellular level, aerobic metabolism occurs within the mitochondria where oxygen is consumed in the final step of the electron-transfer system. Energy derived primarily from the breakdown of lipids and carbohydrates is used to pump protons across the inner mitochondrial membrane and establish a proton gradient. The energy released by allowing this proton gradient to dissipate back across the membrane can be either coupled to ATP generation to store energy within the cell, or uncoupled via proton leak, which produces metabolic heat without producing ATP to power functional work (*10*).

Using high-resolution respirometry, we measured respiratory capacity in cranial tibial skeletal muscle samples taken from both northern and southern sea otters encompassing a broad range of body mass (1.4 to 44.5 kg) and representing age classes from neonate to adult (table S1). Steady-state oxygen flux was measured at five induced respiratory states (detailed in table S2), including the three primary states of LEAK (native mitochondrial proton leak), OXPHOS (native proton leak in addition to oxidative phosphorylation), and ETS (maximal flux through the electron-transport system), which reflects mitochondrial density (*11*) [see supplementary materials for detailed methods and Texas Data Repository for DatLab respirometry data files (*12*)].

We found that OXPHOS and ETS respiratory capacities in sea otter skeletal muscle are modestly elevated compared with values measured in the muscle of other active mammals but do not reach the extreme levels observed in some elite performance animals such as Alaskan husky Iditarod dogs (Table 1 and table S1). The 15.7% additional ETS capacity beyond OXPHOS results in a flux control

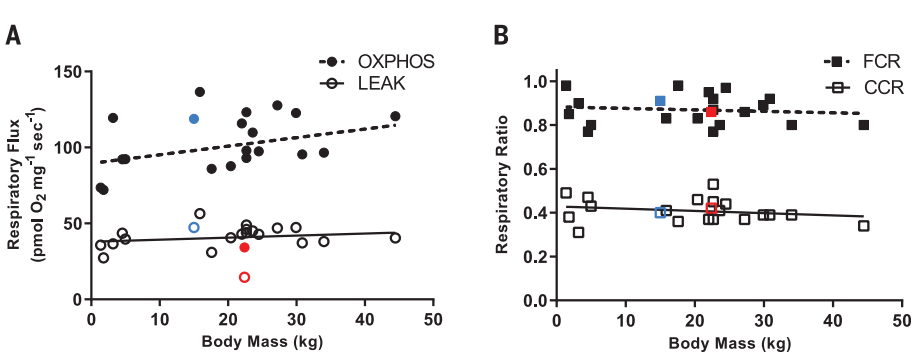


Fig. 1. Sea otter skeletal muscle respiratory flux. (**A**) In individuals ranging from 1.4 to 44.5 kg, body mass was not predictive of OXPHOS or LEAK respiratory capacity. (**B**) Body mass was not predictive of CCR (LEAK/OXPHOS) or FCR (OXPHOS/ETS). Respiratory capacity in a rehabilitated, captive-raised individual (infancy to adulthood) (blue) was indistinguishable from that of wild conspecifics but was drastically reduced in one emaciated, stranded adult (red).

[1]Department of Health and Kinesiology, Texas A&M University, College Station, TX, USA. [2]Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX, USA. [3]Department of Marine Biology, Texas A&M University at Galveston, Galveston, TX, USA. [4]Department of Natural Sciences, University of Alaska Southeast, Juneau, AK, USA. [5]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Juneau, AK, USA. [6]Monterey Bay Aquarium, Monterey, CA, USA.
*Corresponding author. Email: traywright@gmail.com

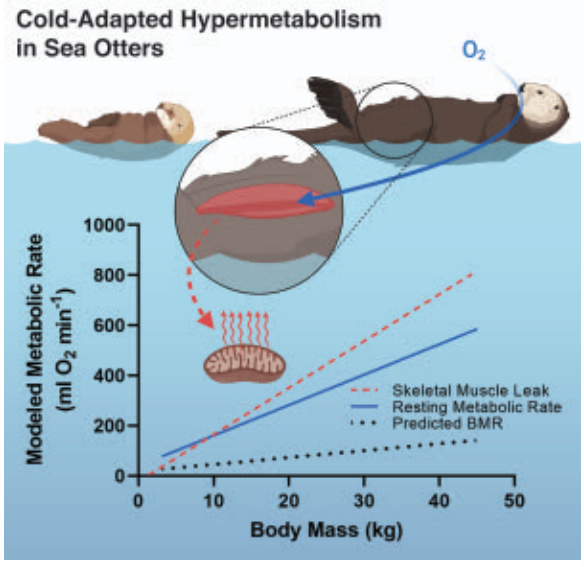


Fig. 2. Whole-body estimates of sea otter metabolism scaled to body mass. Estimated scaled whole-body metabolic rates include predicted [BMR; predicted on the basis of scaled rate for eutherian mammals (*3*)], sea otter resting metabolic rate, and whole-body skeletal muscle leak capacity.

ratio (FCR; OXPHOS/ETS) of 0.87 (Table 1 and table S1), indicating that phosphorylation capacity limits maximum OXPHOS respiration (*13*). The LEAK respiratory capacity in sea otter skeletal muscle resulted in a coupling control ratio (CCR; LEAK/OXPHOS) of 0.41 (Table 1 and table S1), indicating that leak respiratory capacity accounts for up to 41% of OXPHOS capacity. This mitochondrial LEAK capacity is elevated compared with that of other mammals (Table 1); absolute LEAK capacity and CCR are, respectively, two and four times that of comparably sized Alaskan husky Iditarod dogs. These values are also higher than comparable measures reported for any known mammal, excluding extremely small mice species with body mass ranging from 5 to 22 g (*14*). Increased thermogenic mitochondrial leak is predicted for polar animals relying on metabolic heat production to maintain body temperature (*6*) and is described here in sea otter skeletal muscle.

Because of the scaling of surface area to body mass, smaller neonatal mammals are particularly susceptible to heat loss (*4*, *15*). In addition, metabolically active skeletal muscle constitutes a smaller relative portion of total body mass in younger, less physically developed individuals (*16*), and their immature skeletal muscle tends to be less thermogenic (*17*). Sea otters are altricial; born relatively inactive, they depend on their mothers for feeding and grooming and become progressively more active as they mature during the first 3 months of life (*18*, *19*). Given the broad ranges of body mass and age classes included in our data, we determined how sea otter skeletal muscle respiratory capacity changed throughout ontogenetic development.

Respiratory capacity was surprisingly consistent across all age classes, and Pearson correlation indicated that body mass was not correlated with respiratory measures of LEAK (r = 0.172; P = 0.457; Fig. 1A), OXPHOS (r = 0.274; P = 0.229; Fig. 1A), or ETS (r = 0.294; P = 0.197). Body mass was likewise not predictive of FCR (r = −0.112; P = 0.628) or CCR (r = −0.220; P = 0.339) (Fig. 1B). In altricial mammals, skeletal muscle is underdeveloped at birth and is delayed in developing adult measures of mechanical function (*20*) as well as metabolic and thermogenic capacity (*21*). However, sea otter muscle metabolic capacity appears fully developed in even the smallest neonates, although other measures of muscle maturity, including muscle mass and myoglobin concentration, do not reach adult levels until animals are ~2 years of age (*16*). The consistent skeletal muscle mitochondrial leak capacity and CCR across a broad range of body mass and age classes indicate that mitochondrial leak does not constitute a greater portion of total OXPHOS capacity in smaller, less mature individuals as might be predicted.

Samples collected from two animals in unusual circumstances provided additional insights. Included in our study was one captive-raised, rehabilitated juvenile. This individual was housed using a flow-through system of water from its natural habitat, ensuring an endemic temperature range. Despite reduced functional demand on skeletal muscle for foraging, diving, and swimming, this sea otter was metabolically indistinguishable from its wild counterparts (Fig. 1, A and B; blue). This is consistent with a skeletal muscle metabolic capacity that is modulated by thermogenic demand and not by contractile muscle load. Other thermally challenged marine mammals, including Weddell seals (*Leptonychotes weddellii*) (*22*) and northern elephant seals (*Mirounga angustirostris*) (*23*), also demonstrate similarly high neonatal muscle metabolic capacity (*5*).

One stranded, emaciated adult was an exception to the consistent high metabolic capacity seen in this study. Owing to its extensive dental wear, this individual was estimated to be >12 years of age; it was dehydrated and emaciated at the time of stranding, which suggests that it may have had difficulty feeding. This geriatric sea otter demonstrated a universally profound reduction in skeletal muscle metabolic capacity, with OXPHOS and LEAK respiratory capacity only 34% and 36% of the average value, respectively (Fig. 1A, red). Despite the metabolic disturbance, LEAK, OXPHOS, and ETS respiratory capacities were uniformly reduced and respiratory ratios of FCR and CCR remained constant (Fig. 1B, red). Although the cause of the emaciation and stranding cannot be deduced, metabolic factors may have affected thermogenic capacity and contributed to stranding.

To estimate the potential metabolic contribution of skeletal muscle to sea otter resting metabolic rate and thermogenesis throughout life, we combined our measures of skeletal muscle leak capacity with published measures of scaled sea otter muscle mass to determine whole-body skeletal muscle leak capacity (Fig. 2; see supplementary methods for calculations). In sea otters with body mass greater than ~9 kg, skeletal muscle leak capacity exceeds whole-body resting oxygen consumption. However, at less than ~9 kg, skeletal muscle leak is not independently adequate to fully account for the resting metabolic rate owing to the lower relative muscle mass. For smaller animals to maintain body temperature, maximal use of thermogenic leak from skeletal muscle and other tissues may be critical. Smaller animals may also require additional muscle thermogenesis via either shivering or sarcoplasmic reticulum calcium leak to use a greater portion of skeletal muscle respiratory capacity (fig. S1). Notably, neonatal pups have dense natal fur and spend most of their time resting on the mother's abdomen, so there is little heat loss to water until the age of 1 month (*24*).

Thermogenic leak metabolic capacity in sea otter muscle is elevated across all age and size classes and, if extrapolated to whole-body muscle mass, is adequate to explain the elevated resting metabolic rate of adult sea otters. However, smaller, less mature individuals with lower relative muscle mass require either non-leak muscle thermogenesis (e.g., shivering, active grooming) or other metabolically active tissues to account for their elevated resting metabolism. Although mass-specific LEAK respiratory capacity does not change throughout ontogenetic development in sea otter skeletal muscle (Fig. 1A), skeletal muscle leak capacity constitutes a greater portion of resting metabolic rate in larger animals owing to larger relative muscle mass (Fig. 2).

Table 1. Skeletal muscle respiratory capacity in various mammals. Respiratory flux (picomoles O_2 per second per milligram) includes similarly derived measures of LEAK, OXPHOS, and ETS. Leak respiration was determined using substrates without adenylates. In the current study, LEAK was induced with both mitochondrial complex I and complex II substrates, whereas others included only complex I substrate. Respiratory ratios of CCR (LEAK/OXPHOS) and FCR (OXPHOS/ETS) were reported directly in the cited study or calculated on the basis of reported respiratory measures. Respiratory measures from the current study represent the combined average of all age classes from both northern and southern subspecies. Measures for American quarter horses were estimated from published graphs. For Alaskan huskies, humans, mice, and rats, CCR and FCR were calculated from published respiratory measures.

Condition	LEAK	OXPHOS	ETS	CCR	FCR	Reference
Sea otter (cranial tibial)						
Average	40.6	100.6	116.4	0.408	0.87	Current study
Northern elephant seal (longissimus dorsi)						
Pre-molt females	13.0	100.5	116.5	0.13	0.87	(*28*)
American quarter horse (triceps brachii)						
Young	6	85	115	0.068	0.74	(*29*)
Old	6.5	80	116	0.075	0.68	(*29*)
Alaskan huskies (biceps femoris)						
Raced	14	245	253	0.057	0.97	(*30*)
Nonraced	24	237	254	0.101	0.93	(*30*)
Human, young healthy male (vastus lateralis)						
Untrained	10.7	54.5	64.0	0.196	0.85	(*31*)
Trained	21.8	75.6	104.4	0.288	0.72	(*31*)
Mouse, male BALB/c (spinotrapezius)						
Sham	33.9	63.9	69.7	0.531	0.92	(*32*)
Rat - male Long-Evans (soleus)						
Control	5.8	101.7	110.8	0.057	0.92	(*33*)

As a group, carnivorous marine mammals (e.g., whales, dolphins, and seals) have a 2.3-fold increase in BMR relative to the allometric prediction for eutherian mammals (*2*). This consistently elevated BMR is noteworthy given that this group evolved from different phylogenetic orders (Carnivora and Artiodactyla) to live in differing marine habitats. In comparison, the mean BMR for six other species of terrestrial and semiaquatic Mustelidae (mean body mass: 1.8 kg), including river otters, is 1.6 times greater than the allometric prediction (*3*), which is less than that of fully aquatic sea otters. Only shrews (*Sorex sp.*; mean body mass 5.6 g) have an elevated BMR (3.1 times higher than the allometric prediction) comparable to that of sea otters. Hence, the high BMR in sea otters is not exclusive, but it is the highest reported for any mammal with body mass >1 kg.

For homeothermic mammals, basal metabolism and tissue-level muscle metabolism are each shaped by both environmental acclimation and evolutionary adaptation (*25*–*27*). Although other tissues may also contribute to sea otter thermogenesis (including brown adipose tissue that is not well documented in sea otters), we show here that skeletal muscle leak capacity alone is adequate to account for adult sea otter hypermetabolism. Sea otter skeletal muscle metabolic capacity is indistinguishable between adults and neonates, but it is not clear if this development occurs in utero (precocial) or rapidly after birth (precocious). Premature development of skeletal muscle metabolic capacity combined with high thermogenic capacity for mitochondrial leak throughout life suggests that thermogenesis is a fundamental determinant in the development and lifelong regulation of sea otter muscle metabolism in response to the thermal demands of the marine environment. Some combination of adaptive and acclimative forces shapes sea otter skeletal muscle for thermogenic hypermetabolism.

The generally elevated BMR in most marine mammals indicates similar metabolic remodeling as an adaptation for aquatic life. Instead of the strict allometric relationship between BMR and body mass historically ascribed to mammals (as well as birds), our results indicate that metabolic plasticity may be fundamental for mammals to adapt to varied thermal environments for enhanced efficiency, performance, and fitness. Enhanced thermogenesis may have been a critical adaptation enabling the terrestrial ancestors of marine mammals to overcome the thermoregulatory challenges of aquatic life and inhabit the world's oceans.

REFERENCES AND NOTES

1. D. P. Costa, G. L. Kooyman, *Can. J. Zool.* **60**, 2761–2767 (1982).
2. R. W. Davis, *Marine Mammals: Adaptations for an Aquatic Life* (Springer Nature, 2019).
3. B. K. McNab, *Comp. Biochem. Physiol. A* **151**, 5–28 (2008).
4. A. S. Blix, J. B. Steen, *Physiol. Rev.* **59**, 285–304 (1979).
5. G. C. Whittow, *Mar. Mamm. Sci.* **3**, 220–241 (1987).
6. F. J. Ballesteros *et al.*, *Sci. Rep.* **8**, 1448 (2018).
7. Z. Wang, T. P. O'Connor, S. Heshka, S. B. Heymsfield, *J. Nutr.* **131**, 2967–2970 (2001).
8. M. Periasamy, J. L. Herrera, F. C. G. Reis, *Diabetes Metab. J.* **41**, 327–336 (2017).
9. B. K. McNab, *J. Exp. Biol.* **222**, jeb.205591 (2019).
10. R. A. Busiello, S. Savarese, A. Lombardi, *Front. Physiol.* **6**, 36 (2015).
11. D.-M. Votion, E. Gnaiger, H. Lemieux, A. Mouithys-Mickalad, D. Serteyn, *PLOS ONE* **7**, e34890 (2012).
12. T. Wright, "Sea Otter Skeletal Muscle High-Resolution Respirometry," Texas Data Repository, V1 (2021); https://doi.org/10.18738/T8/TQBMXQ.
13. E. Gnaiger, *Int. J. Biochem. Cell Biol.* **41**, 1837–1845 (2009).
14. M. Boël *et al.*, *J. Exp. Biol.* **223**, jeb.215558 (2020).
15. S. R. Noren, L. E. Pearson, J. Davis, S. J. Trumble, S. B. Kanatous, *Physiol. Biochem. Zool.* **81**, 868–879 (2008).
16. N. M. Thometz, M. J. Murray, T. M. Williams, *Physiol. Biochem. Zool.* **88**, 311–327 (2015).
17. M. Pant, N. C. Bal, M. Periasamy, *J. Exp. Biol.* **218**, 2321–2325 (2015).
18. N. M. Thometz, M. T. Tinker, M. M. Staedler, K. A. Mayer, T. M. Williams, *J. Exp. Biol.* **217**, 2053–2061 (2014).
19. M. Cortez, C. E. C. Goertz, V. A. Gill, R. W. Davis, *J. Exp. Mar. Biol. Ecol.* **481**, 81–91 (2016).
20. D. S. Gokhin, S. R. Ward, S. N. Bremner, R. L. Lieber, *J. Exp. Biol.* **211**, 837–843 (2008).
21. C. E. Robertson, G. B. McClelland, *J. Exp. Biol.* **222**, jeb.210963 (2019).
22. S. B. Kanatous *et al.*, *J. Exp. Biol.* **211**, 2559–2565 (2008).
23. A. J. Chicco *et al.*, *J. Exp. Biol.* **217**, 2947–2955 (2014).
24. M. Cortez, R. Wolt, F. Gelwick, S. K. Osterrieder, R. W. Davis, *J. Exp. Mar. Biol. Ecol.* **481**, 71–80 (2016).
25. N. J. Dawson, S. A. Lyons, D. A. Henry, G. R. Scott, *Acta Physiol.* **223**, e13030 (2018).
26. S. Mahalingam, Z. A. Cheviron, J. F. Storz, G. B. McClelland, G. R. Scott, *J. Physiol.* **598**, 5411–5426 (2020).
27. M. Konarzewski, A. Książek, *J. Comp. Physiol. B* **183**, 27–41 (2013).
28. T. J. Wright *et al.*, *Front. Physiol.* **11**, 564555 (2020).
29. C. Li, S. H. White, L. K. Warren, S. E. Wohlgemuth, *J. Appl. Physiol.* **121**, 299–311 (2016).
30. B. Miller *et al.*, *PLOS ONE* **12**, e0174874 (2017).
31. C. Porter, P. T. Reidy, N. Bhattarai, L. S. Sidossis, B. B. Rasmussen, *Med. Sci. Sports Exerc.* **47**, 1922–1931 (2015).
32. C. Porter *et al.*, *Burns* **42**, 112–122 (2016).
33. E. J. Stephenson *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **302**, E1541–E1549 (2012).

ACKNOWLEDGMENTS

The authors thank M. Gho, C. Pearson, K. Randolph, and S. Atkinson for their help in sample collection and logistical support. We would also like to thank T.M. Williams for assistance with manuscript preparation and helpful edits, M. Young for post collection processing and shipping of samples collected at the Monterey Bay Aquarium, and A. Filley and E. Parker for laboratory assistance. Figure 2 was created with BioRender.com. **Funding:** This research was funded by a Texas A&M University College of Education and Human Development Catapult award issued to T.W. (241398-18008) and by Texas A&M University College of Education and Human Development Startup funds issued to M.S.-M. (131505-51333). **Author contributions:** Conceptualization: T.W., R.W.D., and M.S.-M. Funding acquisition: T.W., R.W.D., and M.S.-M. Methodology: T.W., M.M., R.W.D., and M.S.-M. Resources: T.W., H.C.P., M.M., R.W.D., and M.S.-M. Investigation: T.W. Formal analysis: T.W. Visualization: T.W. Writing – original draft: T.W. and R.W.D. Writing – review & editing: H.C.P., M.M., and M.S.-M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** DatLab respirometry data files available at Texas Data Repository (*12*). All data and analysis are available in the manuscript and supplementary materials. Southern sea otter samples were collected at the Monterey Bay Aquarium as authorized by the Monterey Bay Aquarium IACUC (protocol 15-01) and USFWS (permit MA 032027-2). Northern sea otter samples were acquired under letter of authorization from USFWS.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/223/suppl/DC1
Materials and Methods
Fig. S1
Tables S1 and S2
References (*34*, *35*)

29 October 2020; accepted 25 May 2021
10.1126/science.abf4557

SENSORY EVOLUTION

Early origin of sweet perception in the songbird radiation

Yasuka Toda[1,2,3], Meng-Ching Ko[4], Qiaoyi Liang[4], Eliot T. Miller[5], Alejandro Rico-Guevara[6,7], Tomoya Nakagita[8], Ayano Sakakibara[9], Kana Uemura[9], Timothy Sackton[10], Takashi Hayakawa[11,12], Simon Yung Wa Sin[13,14], Yoshiro Ishimaru[2], Takumi Misaka[1], Pablo Oteiza[15], James Crall[14,16], Scott V. Edwards[14], William Buttemer[17,18], Shuichi Matsumura[9], Maude W. Baldwin[4,14]*

Early events in the evolutionary history of a clade can shape the sensory systems of descendant lineages. Although the avian ancestor may not have had a sweet receptor, the widespread incidence of nectar-feeding birds suggests multiple acquisitions of sugar detection. In this study, we identify a single early sensory shift of the umami receptor (the T1R1-T1R3 heterodimer) that conferred sweet-sensing abilities in songbirds, a large evolutionary radiation containing nearly half of all living birds. We demonstrate sugar responses across species with diverse diets, uncover critical sites underlying carbohydrate detection, and identify the molecular basis of sensory convergence between songbirds and nectar-specialist hummingbirds. This early shift shaped the sensory biology of an entire radiation, emphasizing the role of contingency and providing an example of the genetic basis of convergence in avian evolution.

Sensory systems evolve and adapt, allowing animals to perceive the species-specific cues relevant for survival. Sensory receptor modifications can have profound ecological consequences, affecting behaviors such as foraging (*1*) and mate choice (*2*) and even driving speciation (*3*). Evolution of novel sensory adaptations enables organisms to exploit extreme environments and new niches (*4*–*6*). Determining the timing of sensory changes is essential for understanding the underlying ecological causes and consequences of shifts in perception, as sensory differences may reflect not only adaptation to current lifestyles but also persistence of traits established earlier in the evolutionary history of a clade.

Taste is an important sense used to discriminate between nutrient-rich and toxic food items. Most basic taste categories, such as bitter (eliciting aversion) and umami (the appetitive taste of amino acids), are conserved in mammals and fish (*7*, *8*). An appetitive taste for sugars (conferred by the T1R2-T1R3 sweet receptor heterodimer) is widespread in mammals (*9*, *10*), but *T1R2* was lost early in bird evolution (*11*). Despite this loss, divergent lineages of birds (including hummingbirds, parrots, and honeyeaters) consume sugar-rich nectar and fruit. Hummingbirds, a large radiation of nectarivores, acquired the ability to detect sugars through modifications to the ancestral savory receptor heterodimer (T1R1-T1R3) after divergence from their close relatives, swifts (*11*). Whether the myriad other frugivorous and nectarivorous birds can taste sweet is currently unknown.

To understand the origins of avian sweet taste, we examined nectar consumption patterns across the phylogeny. Unexpectedly, we observed a marked enrichment of nectar-taking behavior in songbirds from a variety of dietary guilds [Fig. 1, fig. S1, and table S1; diet data from (*12*, *13*)]. Ancestral state reconstructions created with a hidden Markov model (*14*) suggest infrequent gains and losses of sweet taste but frequent transitions to and from nectar-feeding once sweet taste had been gained (fig. S2 and tables S3 and S4). We therefore wondered whether sweet taste was gained early in the songbird radiation and subsequently retained, even in species for which nectar is not a major dietary component.

To understand the mechanism underlying the possible gain of sweet taste in songbirds, we first conducted brief-access taste trials on New Holland honeyeaters, which are specialized flower visitors (movie S1). In our two-choice assay, honeyeaters exhibited a clear preference for sugars (Fig. 2A, fig S3, and table S5) over water controls. To assess whether a sugar preference also exists in non-nectarivorous songbirds, we performed brief-access sucrose tests with canaries, granivorous finch distantly related to honeyeaters. The canaries' responses to sucrose (Fig. 2B and table S5) suggest that the ability to taste sugar may have persisted in members of the songbird radiation, regardless of diet.

To examine whether songbirds, like hummingbirds, evolved a mechanism to taste sugars that involved changes to the savory receptor, we cloned and functionally profiled T1R1-T1R3 receptors from honeyeaters and canaries. We also tested receptors from representatives of different dietary guilds (Fig. 2C and fig. S4). We observed a strong response to carbohydrates in the savory receptors of the honeyeater, white-eye, and bulbul— species that consume large amounts of fruit and nectar (Fig. 2C and fig. S5). Surprisingly, receptors from the canary and great tit, two non-nectar specialists, also showed significant sugar responses (Fig. 2C and table S6).

Next, we cloned taste receptors from the barred antshrike (*Thamnophilus doliatus*) and the rusty-margined flycatcher (*Myiozetetes cayanensis*), members of the sister group of songbirds (suboscines). Receptors from both species exhibited strong responses to amino acids but did not respond to sugars (Fig. 2C), suggesting that the sugar response seen in songbirds evolved after these two passerine clades diverged. Receptors from the brown treecreeper (*Climacteris picumnus*), an early-diverging Australian songbird that is primarily insectivorous but occasionally takes nectar, exhibited a strong response to amino acids but also a small response to sugars (fig. S4), implying an early origin of sugar perception in songbirds.

To investigate whether songbird receptors employed a shared mechanism to respond to sugars, we examined responses of cross-species T1R1-T1R3 pairs. Our study of hummingbird receptors indicated that sugar detection required coordinated functioning of both members of the heterodimer (T1R1 and T1R3). We therefore hypothesized that if songbirds had evolved a response to sugars early in their evolutionary history, a response that was retained by later lineages rather than evolving multiple times independently, then cross-species heterodimers may still respond to sucrose. First, we examined mixed pairs of hummingbird and honeyeater receptors and observed responses to amino acids but not to sugars, confirming that receptor heterodimers

[1]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan. [2]Department of Agricultural Chemistry, School of Agriculture, Meiji University, Kawasaki, Kanagawa 214-8571, Japan. [3]Japan Society for the Promotion of Science, Tokyo 102-0083, Japan. [4]Evolution of Sensory Systems Research Group, Max Planck Institute for Ornithology, Seewiesen, Germany. [5]Macaulay Library, Cornell Lab of Ornithology, Ithaca, NY, USA. [6]Department of Biology, University of Washington, Seattle, WA 98105, USA. [7]Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98105, USA. [8]Proteo-Science Center, Ehime University, Matsuyama, Ehime 790-8577, Japan. [9]Faculty of Applied Biological Sciences, Gifu University, Gifu, 501-1193, Japan. [10]Informatics Group, Harvard University, Cambridge, MA, USA. [11]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan. [12]Japan Monkey Centre, Inuyama, Aichi 484-0081, Japan. [13]School of Biological Sciences, The University of Hong Kong, Pok Fu Lam Road, Hong Kong. [14]Department of Organismic and Evolutionary Biology and the Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA, USA. [15]Flow Sensing Research Group, Max Planck Institute for Ornithology, Seewiesen Germany. [16]Department of Entomology, University of Wisconsin-Madison, WI, USA. [17]Centre for Integrative Ecology, Deakin University, Geelong, Victoria, Australia. [18]School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, New South Wales, Australia.
*Corresponding author. Email: mbaldwin@orn.mpg.de

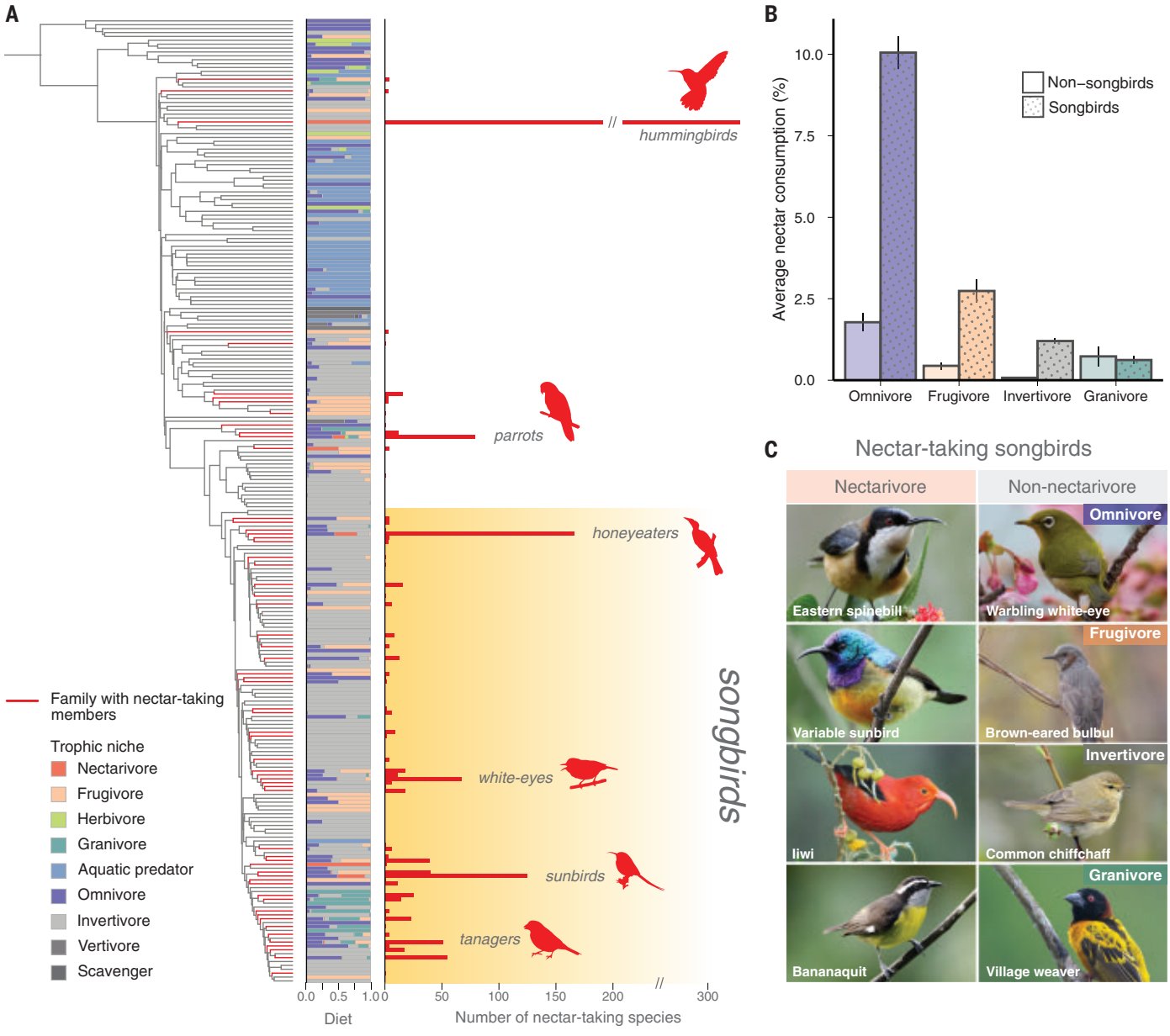


Fig. 1. Widespread nectar consumption across songbirds. (**A**) Songbirds (yellow shading) display extensive nectar-taking behavior. Proportion of species per trophic niche and number of nectar-taking species are plotted per family; see fig. S1. Red branches: >1% of species take nectar. (**B**) Average nectar consumption (percent of diet) from species occupying diverse trophic niches is higher in songbirds compared with other birds (see table S1 and fig. S1). Error bars indicate standard error of the mean. (**C**) Examples of specialized songbird nectarivores, as well as nectar-consuming songbird species from other trophic groups [diet classified in accordance with (*13*)]; (see table S2 for photo credits).

were still functional and suggesting a distinct mechanism for sugar detection in each radiation (Fig. 3, A and B, and fig. S6). We next tested mixed pairs between songbirds and close relatives. Responses to sugars were seen in honeyeater receptors coexpressed with corresponding receptors from the canary, great tit, bulbul, and white-eye, but not with those of the antshrike or lyrebird (Fig. 3C), even though these pairs displayed robust amino acid responses (fig. S6).

Next, we synthesized and tested ancestral receptors (see supplementary materials) to confirm that the receptors of early songbirds could respond to sugars. Receptors from the ancestor of honeyeaters and all other songbirds (Anc2), but not from the earlier ancestor (Anc1), responded strongly to carbohydrates (Fig. 4A). Changes in both T1R1 and T1R3 were necessary for this response (fig. S7). By creating chimeric receptors, we identified two critical regions involved in carbohydrate detection (Fig. 4B); responses were observed only when both regions were present (Fig. 4B and fig. S8). We further narrowed the region to seven residues in T1R1 and nine in T1R3, which together elicited a strong sucrose response when inserted in the background of Anc1 receptors (Fig. 4, B and C, and fig. S9). Surprisingly, most of the identified residues in the ligand-binding region are in T1R1 rather than in T1R3, as in hummingbirds, suggesting that convergence between these lineages occurs at the level of the tertiary structure

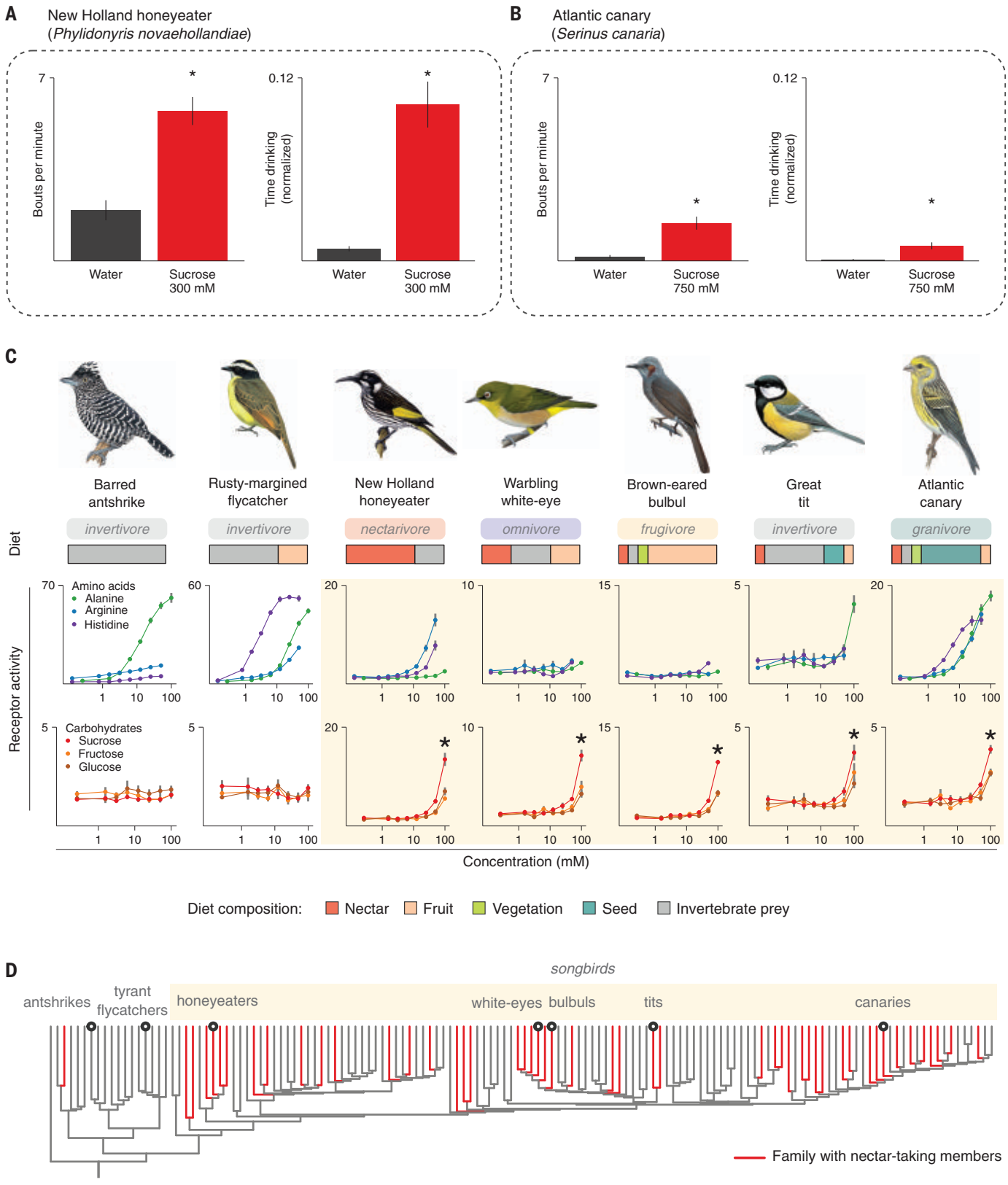


Fig. 2. Songbirds with diverse diets respond to sugars. Honeyeaters (**A**) (n = 10) and canaries (**B**) (n = 8) display higher drinking rates and spend more time per trial drinking sucrose than water (mean ± SE, *P < 0.001; table S5). (**C**) T1R1-T1R3 from five songbird species (yellow shading) responds to carbohydrates; receptors from both suboscines respond only to amino acids (n = 8 to 10, mean ± SE; *P < 0.01 shown for sucrose; for other stimuli, see table S6). Illustrations reproduced with permission of Lynx Edicions. (**D**) Phylogenetic position of tested songbirds (red: families with >1% nectar-taking species).

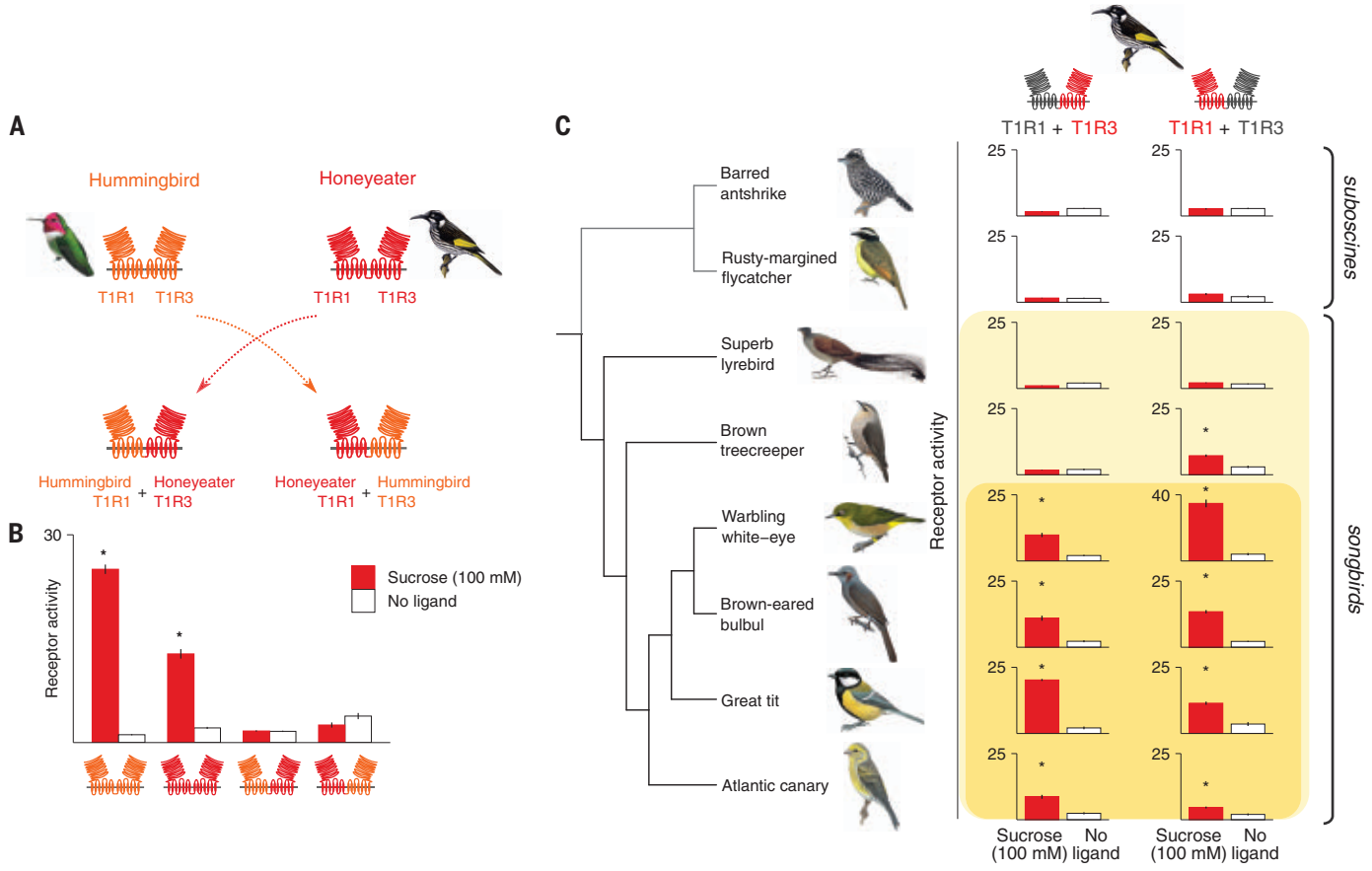


Fig. 3. Unique sensory shift evolves early in songbirds. (**A** and **B**) Honeyeater receptors coexpressed with the corresponding T1R from hummingbirds (mixed pairs) (A) reveal the lack of response to sugars (B) (n = 3 to 6, mean ± SE, *P < 0.01). (**C**) Mixed pairs between taste receptors from honeyeaters and other species (yellow shading, songbirds). (Left) Honeyeater T1R3; (right) honeyeater T1R1 (n = 6, mean ± SE, *P < 0.01). Responses from both pairs (dark yellow) suggest a shared mechanism that has evolved in songbirds. Bird illustrations in (A) and (C) reproduced with permission of Lynx Edicions.

of the protein rather than via identical residues. In T1R3, functionally important sites are located across the protein, including in the transmembrane domain (TM) and in the cysteine-rich domain (CRD), which connects the extracellular Venus flytrap (VFT) domain to the transmembrane domain (Fig. 4D and fig. S10). Notably, homology modeling revealed that T1R1 sites are located on the surface of the ligand-binding region facing T1R3, with most residues located in helices at the dimer interface involved in receptor activation (*15*) (Fig. 4, D and E, and fig. S11). Two songbird residues (Arg^{139} and Thr^{162}) are adjacent to the orthosteric binding pocket (Fig. 4E and fig. S11), and one (Thr^{162}) occurs at the same location as a critical site in the hummingbird binding pocket of T1R3 (Ile^{167}) (Fig. 4F). Thus, songbirds and hummingbirds both independently modified the ligand-binding region of the umami taste receptors. However, these changes occurred in alternate heterodimeric partners in each radiation (Fig. 4G) and, with the exception of a single identical site, involve distinct subsets of residues within this domain.

The changes in ancestral songbird taste receptors imply a complex set of modifications in regions involved in ligand binding, signal transmission, and receptor activation. These suggest coordinated changes, both between the different domains and between the two members of the heterodimer (consistent with some selection tests; fig. S13). Moreover, examining the responses of single residues also reveals extensive inter- (*16*) and intramolecular epistasis (fig. S14). Compared with shifts in visual pigment tuning, which can be caused by small numbers of substitutions and can therefore occur frequently across the phylogeny (*17*), the molecular basis of the acquisition of carbohydrate detection appears more complex; this complexity may lessen the likelihood that sweet taste has repeatedly evolved.

The ecology driving the initial acquisition of sugar sensing in songbirds is enigmatic. Many extant Australian birds rely heavily on additional and unique sources of sugar. By the Oligocene, Australia had become increasingly arid and was dominated by eucalypts (*18*), which can produce large quantities of a sugary exudate known as manna. Both manna and insect secretions (called lerp and honeydew) make up substantial components of the diets of many Australian songbirds (*19*). The divergence times of the ancestral nodes reconstructed here appear to predate the period of Australian aridification (*20*). However, because eucalypts, which originated in Gondwana, may have been present in Australia during this period (*21*), it is unclear whether sweet taste originated in birds with diets favoring nectar or fruit, or lerp and/or manna. Our results indicate that Australasian treecreepers, the sister group to the frugivorous bowerbirds (*20*), use a mechanism that differs from but is potentially related to that described here. This suggests a scenario in which initial, permissive changes (*22*) may have evolved earlier than the ancestral receptors that we characterize, facilitating later functional change (Fig. 4G).

Our results reveal an unexpected early event with widespread consequences for the diets and ecologies of later lineages. Although subsequent loss may also have occurred in some species, songbirds appear to have broadly

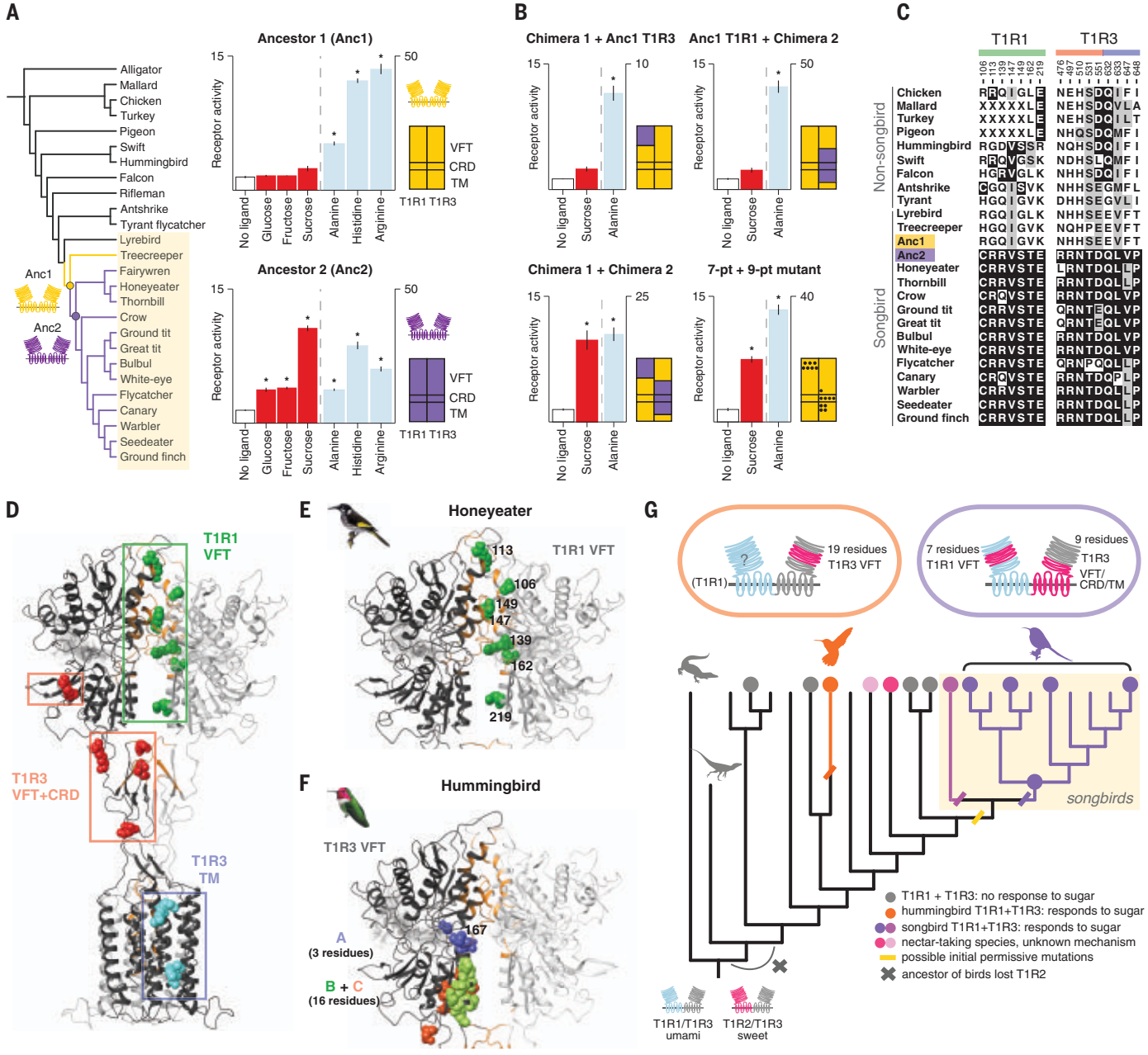


Fig. 4. Molecular basis of songbird sweet perception. (**A**) Anc2 (purple) but not Anc1 (yellow) T1R1-T1R3 responds to sugars (n = 6; mean ± SE; *P < 0.01). (**B**) Residues from two domains confer sugar responsiveness (n = 6; mean ± SE; *P < 0.01). (**C**) The 16 residues are largely conserved across songbirds. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**D**) Homology model of T1R1-T1R3 showing residues located across domains. (**E**) T1R1 VFT residues face T1R3 (yellow, within 4 Å of other T1Rs; binding pocket shaded gray). Bird illustrations in (D) and (E) reproduced with permission of Lynx Edicions. (**F**) Location of hummingbird VFT residues in T1R3 [in accordance with (*11*)]. (**G**) Model showing convergent evolution of sweet perception: Songbirds and hummingbirds independently recruit distinct regions of the ancestral savory receptor.

retained an ancestral sensory ability as they radiated out of Australia and colonized diverse habitats across the planet. This sensory shift enabled non-nectar specialists to opportunistically exploit novel or seasonally varying food sources, such as the nectar consumed by many insectivorous species during migration (*23*). The series of molecular changes that conferred the ability to sense sugars in the Australian ancestors of songbirds shaped the sensory *umwelt* and subsequent evolution of nearly half of the world's birds.

REFERENCES AND NOTES

1. T. O. Auer *et al.*, *Nature* **579**, 402–408 (2020).
2. Y. Grosjean *et al.*, *Nature* **478**, 236–240 (2011).
3. O. Seehausen *et al.*, *Nature* **455**, 620–626 (2008).
4. V. Matos-Cruz *et al.*, *Cell Rep.* **21**, 3329–3337 (2017).
5. O. Eigenbrod *et al.*, *Science* **364**, 852–859 (2019).
6. Z. Musilová *et al.*, *Science* **364**, 588–592 (2019).
7. G. Nelson *et al.*, *Nature* **416**, 199–202 (2002).
8. H. Oike *et al.*, *J. Neurosci.* **27**, 5584–5592 (2007).
9. G. Nelson *et al.*, *Cell* **106**, 381–390 (2001).
10. D. Glaser, *Pure Appl. Chem.* **74**, 1153–1158 (2002).
11. M. W. Baldwin *et al.*, *Science* **345**, 929–933 (2014).
12. H. Wilman *et al.*, *Ecology* **95**, 2027 (2014).
13. A. L. Pigot *et al.*, *Nat. Ecol. Evol.* **4**, 230–239 (2020).
14. J. D. Boyko, J. M. Beaulieu, *Methods Ecol. Evol.* **12**, 468–478 (2021).
15. A. Koehl *et al.*, *Nature* **566**, 79–84 (2019).
16. D. W. Anderson, A. N. McKeown, J. W. Thornton, *eLife* **4**, e07864 (2015).
17. A. Ödeen, O. Håstad, *BMC Evol. Biol.* **13**, 36 (2013).
18. G. H. Orians, A. V. Milewski, *Biol. Rev.* **82**, 393–423 (2007).
19. D. C. Paton, *Emu* **80**, 213–226 (1980).

20. C. H. Oliveros *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7916–7925 (2019).
21. A. H. Thornhill *et al.*, *Aust. Syst. Bot.* **32**, 29–48 (2019).
22. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, *Science* **317**, 1544–1548 (2007).
23. J. G. Cecere, F. Spina, S. Jenni-Eiermann, L. Boitani, *J. Ornithol.* **152**, 923–931 (2011).
24. Y. Toda *et al.*, Data from "Early origin of sweet perception in the songbird radiation," Dryad (2021); https://doi.org/10.5061/dryad.xksn02vfz.

ACKNOWLEDGMENTS

We thank those who supported these experiments; see supplementary materials for a full list. **Funding:** Funding for this work was obtained from the Putnam Expedition Grant (Harvard University) and the Max Planck Society to M.W.B. and the Japan Society for the Promotion of Science to Y.T. (18K14427 and 20H02941), Y.I. (#18KK0166), and T.M. (#16H04918, #19H02907).

Author contributions: Y.T., Q.L., M.-C.K., E.T.M., A.R.-G., T.N., T.S., T.H., S.S., Y.I., T.M., P.O., J.C., S.V.E., W.B., S.M., and M.W.B. designed the research. A.R.-G., W.B., and M.W.B. designed and conducted field experiments with honeyeaters. J.C., P.O., Q.L., and M.W.B. designed experiments with canaries. Q.L. conducted behavioral trials with canaries. Q.L., J.C., P.O., and M.W.B. analyzed behavioral data. E.T.M. performed hidden Markov modeling. E.T.M. and M.W.B compiled results of hidden rates analyses. M.-C.K., P.O., Y.T., T.N., and M.W.B. designed the figures. T.N. performed homology modeling. A.S. and K.U. cloned taste receptors from white-eyes and bulbuls under the supervision of S.M. Y.T. and M.W.B. cloned taste receptors from remaining species, created and tested chimeric receptors, and determined critical residues. Y.T. collected data for dose response curves and ancestral receptors. Y.T., M.-C.K., and M.W.B. analyzed cell assay experiments. T.S., M.-C.K., and M.W.B. performed ancestral reconstruction and selection tests. Y.I., T.M., W.B., S.M., and M.W.B. acquired funding. M.W.B. wrote the manuscript with input from all authors. **Competing interests:** the authors declare no competing interests. **Data and materials availability:** T1R sequences are accessioned in GenBank (accession numbers MZ220489 to MZ220511), and alignments and other data are available at Dryad (*24*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/226/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S7
References (*25*–*62*)
MDAR Reproducibility Checklist
Movie S1

10 November 2020; accepted 19 May 2021
10.1126/science.abf6505

ANTIVIRAL DEFENSE

An isoform of Dicer protects mammalian stem cells against multiple RNA viruses

Enzo Z. Poirier[1]*, Michael D. Buck[1], Probir Chakravarty[2], Joana Carvalho[3]†, Bruno Frederico[1], Ana Cardoso[1], Lyn Healy[4], Rachel Ulferts[5], Rupert Beale[5,6], Caetano Reis e Sousa[1]*

In mammals, early resistance to viruses relies on interferons, which protect differentiated cells but not stem cells from viral replication. Many other organisms rely instead on RNA interference (RNAi) mediated by a specialized Dicer protein that cleaves viral double-stranded RNA. Whether RNAi also contributes to mammalian antiviral immunity remains controversial. We identified an isoform of Dicer, named antiviral Dicer (aviD), that protects tissue stem cells from RNA viruses—including Zika virus and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—by dicing viral double-stranded RNA to orchestrate antiviral RNAi. Our work sheds light on the molecular regulation of antiviral RNAi in mammalian innate immunity, in which different cell-intrinsic antiviral pathways can be tailored to the differentiation status of cells.

Type I and III innate interferons (IFNs) are rapidly induced in mammalian cells in response to virus infection. These cytokines act in an autocrine and paracrine manner to promote the transcription of multiple interferon-stimulated genes (ISGs), which encode a variety of viral restriction, cellular arrest, and cell death factors (*1*). The inducible protection conferred by IFN receptor signaling is much more marked in differentiated cells than in embryonic and adult stem cells, which lack expression of components of the pathways that lead to IFN induction and IFN responsiveness (*2*–*4*). This may ensure that infected stem cells are spared the cytostatic and cytotoxic effects of IFN exposure.

[1]Immunobiology laboratory, Francis Crick Institute, London NW1 1AT, UK. [2]Bioinformatics and Biostatistics, Francis Crick Institute, London NW1 1AT, UK. [3]Experimental Histopathology, Francis Crick Institute, London NW1 1AT, UK. [4]Human Embryo and Stem Cell Unit, Francis Crick Institute, London NW1 1AT, UK. [5]Cell Biology of Infection Laboratory, Francis Crick Institute, London NW1 1AT, UK. [6]Division of Medicine, University College London, London WC1E 6BT, UK. ***Corresponding author. Email: enzo.poirier@crick.ac.uk (E.Z.P.); caetano@crick.ac.uk (C.R.S.)** †Present address: Histopathology Scientific Platform, Champalimaud Centre for the Unknown, 1400-038 Lisboa, Portugal.

The IFN system is absent from invertebrates and plants, which protect themselves from viral infection by means of RNA interference (RNAi) (*5*). Antiviral RNAi starts with the protein Dicer, which recognizes and cleaves double-stranded RNA (dsRNA) produced during RNA virus infection to generate small interfering RNAs (siRNAs). These guide the sequence-specific degradation of viral RNAs by a slicing-active Argonaute protein such as Argonaute 2 (Ago2), present in insects and mammals. Irrespective of infection, RNAi also has a distinct role in regulating cellular gene expression through microRNAs (miRNAs) produced by Dicer cleavage of precursor miRNAs (pre-miRNAs) (*5*). Recent work suggests that mammals, like invertebrates and plants, can co-opt RNAi for antiviral immunity (*5*). Examples of mammalian antiviral RNAi in vitro and in vivo have been reported for Nodamura virus, human enterovirus 71, Zika virus, and other RNA viruses (*6*–*14*), although other studies have argued against the existence of such a response (*15*–*18*). Part of the controversy may relate to the fact that IFN inhibits mammalian dsRNA-mediated RNAi and the latter may therefore only be relevant in cells that are hyporesponsive to IFNs, such as stem cells (*5*, *19*, *20*). Notably, stem cells can resist virus infection, which has been partly attributed to IFN-independent constitutive expression of restriction factors (*21*). Whether stem cells additionally possess specializations that favor antiviral RNAi remains unclear.

Plants and insects that use RNAi both as a means of regulating translation of cellular mRNAs and as an antiviral mechanism encode multiple Dicers, each dedicated to one arm of the pathway. In contrast, mammals possess a single *DICER* gene with one canonical protein product, which cleaves pre-miRNA but processes dsRNA poorly (*22*, *23*). Interestingly, a truncated form of Dicer with improved antiviral capacity can be produced from the *Dicer* gene in mice, but its expression is restricted to oocytes (*24*). By analogy, we hypothesized that antiviral RNAi in mammals may involve expression of an isoform of Dicer that processes dsRNA more efficiently than canonical Dicer. By performing a polymerase chain reaction (PCR) on total cDNA from mouse small intestine, we identified an alternatively spliced in-frame transcript of Dicer missing exons 7 and 8 (Fig. 1A). In silico translation of this transcript resulted in a truncated Dicer protein in which the central Hel2i domain of the N-terminal helicase segment is absent (Fig. 1A). For simplicity, hereafter we refer to canonical Dicer (which includes the sequences encoded by exons 7 and 8) as Dicer and its truncated form as antiviral Dicer (aviD). Using a reverse transcription quantitative PCR (RT-qPCR) assay that distinguishes *aviD* and *Dicer* mRNA, both isoforms could be detected in mouse cells, including neural stem cells, embryonic stem cells (ES cells), and a 3T3 cell line, as well as in organs from pre-weaning or adult mice (fig. S1, A to D). The *AVID* and *DICER* transcripts were also found in human ES cells, human induced pluripotent stem cells (iPSCs), and some human cell lines (fig. S1, E to H). In general, transcripts encoding aviD appeared to be less abundant than transcripts encoding full-length Dicer by at least a factor of 10 (fig. S1), explaining

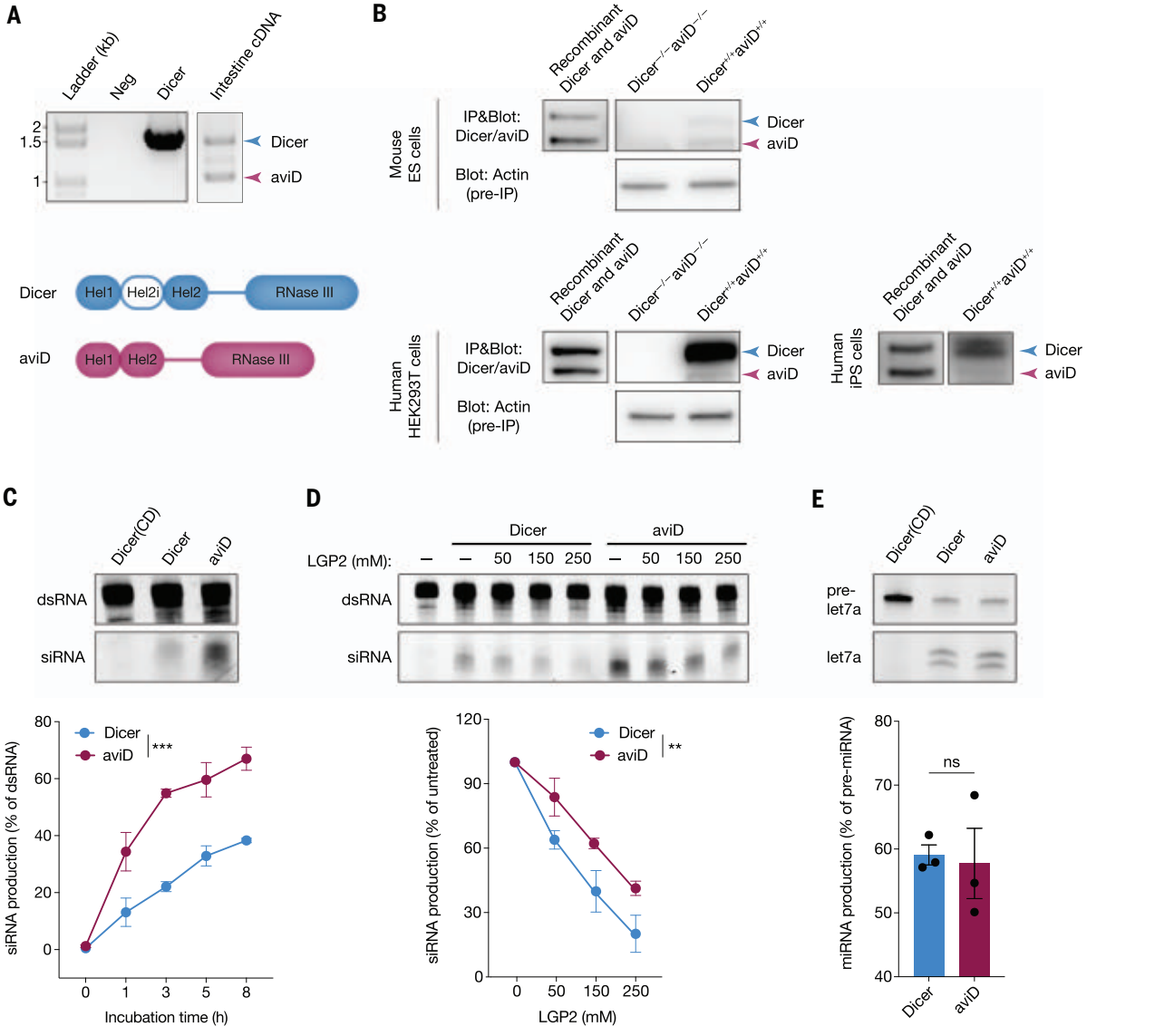


Fig. 1. aviD is an isoform of Dicer that efficiently cleaves dsRNA. (**A**) Dicer PCR amplicons using vehicle (Neg), a plasmid coding for Dicer, or mouse small intestine cDNA templates. In addition to a canonical product corresponding to full-length Dicer, an in-frame transcript missing exons 7 and 8 (nucleotides 705 to 1346 of the coding sequence) was detected. This corresponds to an isoform termed antiviral Dicer (aviD) lacking the Hel2i domain of the helicase (white). (**B**) Immunoblots from wild-type Dicer$^{+/+}$aviD$^{+/+}$ or Dicer$^{-/-}$aviD$^{-/-}$ mouse ES cells, HEK293T cells, or Dicer$^{+/+}$aviD$^{+/+}$ human iPSC lysates before (pre-IP) or after immunoprecipitation with a Dicer/aviD-specific antibody. Recombinant Flag-tagged Dicer and aviD were included as controls. (**C**) Recombinant Flag-tagged Dicer, Dicer catalytically deficient [Dicer(CD), used as a negative control], or aviD were incubated with synthetic Cy5-labeled dsRNA at 37°C for the indicated time. The reactions were resolved on a denaturing polyacrylamide gel and visualized by Cy5 in-gel fluorescence, and Dicer versus aviD cleavage was quantitated by densitometry. (**D**) Increasing concentrations of recombinant LGP2 were added to the in vitro dicing reaction as in (C) and incubated for 3 hours at 37°C. After densitometric quantitation, the siRNA amount was normalized to the amount of siRNA produced in a reaction without LGP2. (**E**) Immunopurified Flag-tagged Dicer, Dicer(CD), or aviD were incubated with *let-7a* pre-miRNA at 37°C for 20 min. The reactions were resolved on a denaturing polyacrylamide gel and visualized by Cy5 in-gel fluorescence, and Dicer versus aviD cleavage was quantitated by densitometry. Data in (C) to (E) are means ± SEM pooled from three independent experiments. $^{**}P < 0.01$, $^{***}P < 0.001$ [two-way analysis of variance (ANOVA) in (C) and (D), Mann-Whitney test in (E)]; ns, not significant.

in part why they are not easily found in public domain RNA-seq datasets. Nonetheless, they resulted in detectable protein, as immunoprecipitation using an antibody that dually recognizes Dicer and aviD demonstrated the presence of low levels of aviD protein in mouse ES cells, human iPSCs and human embryonic kidney (HEK) 293T cells (Fig. 1B) but, as expected, not in *Dicer* gene–deficient (Dicer$^{-/-}$ aviD$^{-/-}$) cells used as a negative control (*25*, *26*). aviD lacks part of the helicase domain, which negatively regulates Dicer's ability to process dsRNA (*22*, *23*). Consistent with that notion, recombinant aviD produced about twice as much siRNA from synthetic dsRNA as did recombinant Dicer in an in vitro dicing assay (Fig. 1C). In addition, aviD was more resistant to LGP2, an ISG product that inhibits dsRNA cleavage by Dicer and is partly responsible for IFN-mediated inhibition of antiviral RNAi in differentiated mammalian cells (*20*) (Fig. 1D). In contrast to dsRNA cleavage, both Dicer and aviD generated equivalent amounts of *let-7a* miRNA from pre-miRNA (Fig. 1E). These results suggest that loss of the Hel2i domain does not impair the ability of aviD to process miRNA precursors but confers enhanced capacity to

dice dsRNA into siRNAs, a hallmark of Dicers involved in antiviral RNAi.

To assess the ability of aviD to mediate antiviral RNAi, we complemented *Dicer* gene–deficient (Dicer$^{-/-}$aviD$^{-/-}$) HEK293T "NoDice" cells (*26*) by stable transfection with constructs encoding FLAG-tagged Dicer or aviD to generate sublines denoted as Dicer$^{+/+}$aviD$^{-/-}$ or Dicer$^{-/-}$aviD$^{+/+}$ 293T, respectively (figs. S2A and S3). By immunofluorescence, aviD and Dicer localization was predominantly cytoplasmic (fig. S3). Consistent with the in vitro pre-miRNA cleavage assays (Fig. 1E), the expression of either Dicer or aviD was sufficient to restore miRNA production to Dicer$^{-/-}$aviD$^{-/-}$ "NoDice" cells (fig. S2, B and C). We then infected Dicer (Dicer$^{+/+}$aviD$^{-/-}$)–expressing or aviD (Dicer$^{-/-}$aviD$^{+/+}$)–expressing cells with Sindbis virus (SINV) or with Zika virus (ZIKV), two RNA viruses that are targets of antiviral RNAi in insects. We did not include Dicer$^{-/-}$aviD$^{-/-}$ cells in these experiments to avoid confounding effects from loss of miRNA-regulated protein expression. Notably, cells expressing only aviD displayed lower production of SINV (Fig. 2A) and ZIKV (Fig. 2B) virus progeny than did cells that only expressed Dicer. We further tested doxycycline-inducible acute expression of the proteins in the same cells. aviD but not Dicer induction impaired SINV-GFP viral replication over time, as measured by accumulation of green fluorescent protein (GFP) (Fig. 2, C to F). This was not observed with a version of aviD that was catalytically deficient in dsRNA cleavage [aviD(CD)] (Fig. 2, C to F). These data demonstrate that aviD but not Dicer possesses antiviral function that is dependent on its catalytic domain, consistent with a role in RNAi.

Mammals encode four Ago proteins, all of which can mediate miRNA-driven gene silencing. However, only Ago2 possesses endonuclease activity to mediate target "slicing" in antiviral RNAi. Silencing Ago2 in Dicer$^{-/-}$aviD$^{+/+}$ cells (fig. S4A) rescued ZIKV particle production to levels similar to those in Dicer$^{+/+}$aviD$^{-/-}$ cells treated with control or *Ago2* siRNA (Fig. 2G). We also tested the effect of the B2 protein of Nodamura virus, a well-characterized viral suppressor of RNAi (VSR) that shields dsRNA from Dicer cleavage (*5*, *12*). SINV-GFP and SINV expressing B2 (SINV-B2) grew to similar levels in baby hamster kidney cells (fig. S4, B and C), Dicer$^{-/-}$aviD$^{-/-}$ cells (fig. S4D), and Dicer$^{+/+}$aviD$^{-/-}$ cells (fig. S4E). In contrast, infectious virion production in Dicer$^{-/-}$aviD$^{+/+}$ cells infected with SINV-B2 was greater than in the same cells infected with SINV-GFP by as much as a factor of 100 (fig. S4F). Finally, given the current human impact of the coronavirus pandemic, we further tested the ability to resist infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Dicer$^{+/+}$aviD$^{-/-}$ or Dicer$^{-/-}$aviD$^{+/+}$ cells that had been engineered

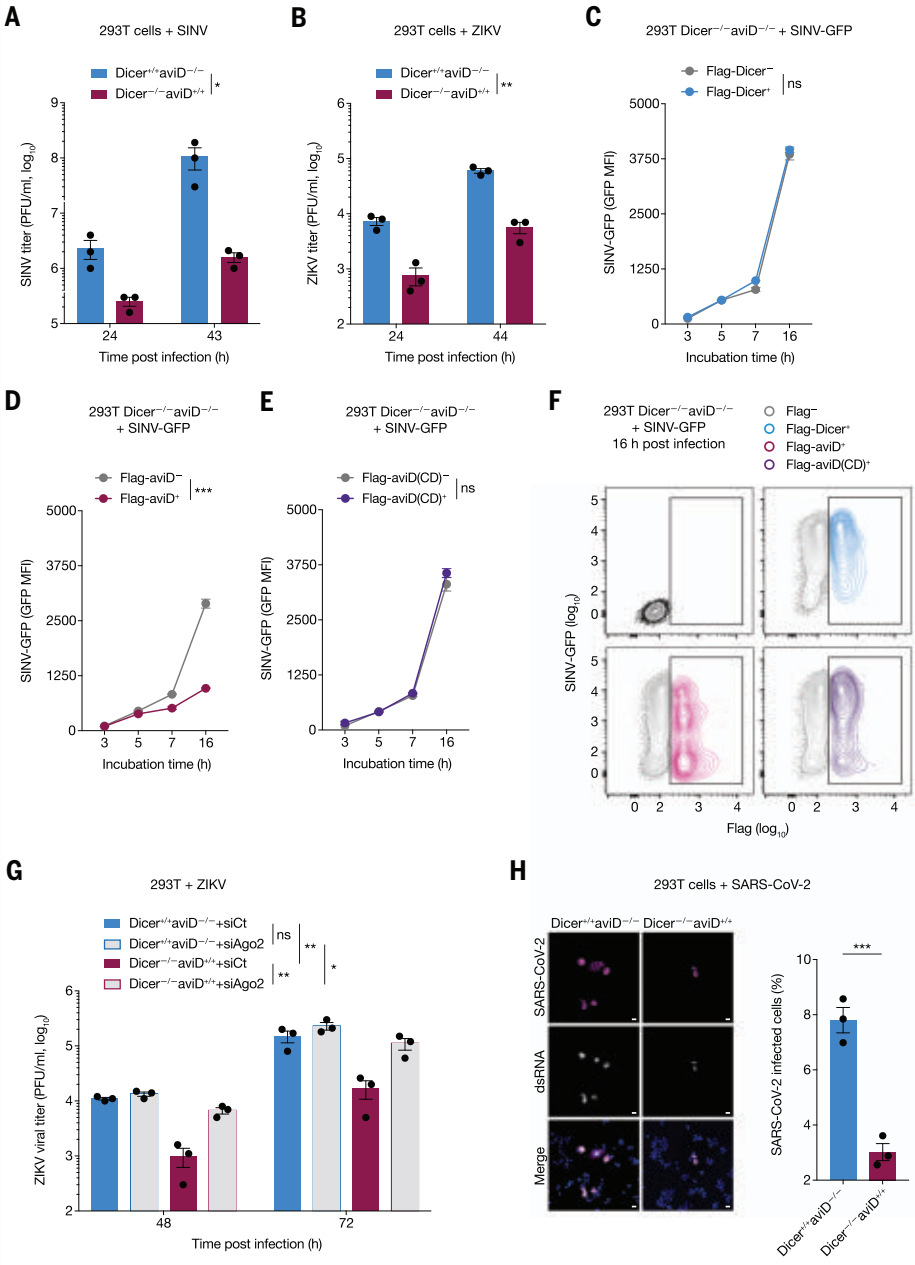


Fig. 2. aviD can mediate antiviral RNAi. (**A** and **B**) HEK293T Dicer$^{-/-}$aviD$^{-/-}$ cells complemented with Dicer (Dicer$^{+/+}$aviD$^{-/-}$) or aviD (Dicer$^{-/-}$aviD$^{+/+}$) were infected with SINV (A) or ZIKV (B) at MOI (multiplicity of infection) of 0.1. Supernatant was collected at the indicated time points and viral content determined by plaque assay. (**C** to **E**) HEK293T Dicer$^{-/-}$aviD$^{-/-}$ cells induced by doxycycline to express Flag-Dicer (C), aviD (D), or aviD(CD) (E) were infected with SINV-GFP. Flow cytometry was used to monitor the expression of Dicer/aviD (via anti-Flag staining) and SINV replication (via GFP fluorescence). (**F**) Representative contour plots from 16 hours after infection. Boxes represent Flag-positive cells defined on the basis of the uninduced controls (top left plot). (**G**) Dicer$^{+/+}$aviD$^{-/-}$or Dicer$^{-/-}$aviD$^{+/+}$ HEK293T cells were transfected with siRNA targeting Ago2 (siAgo2) or with control siRNA (siCt) and infected with ZIKV at MOI of 0.1. Supernatant was collected at the indicated time points and viral content was determined by plaque assay. (**H**) Immunofluorescence of Dicer$^{-/-}$aviD$^{+/+}$ Dicer$^{+/+}$aviD$^{-/-}$ HEK293T cells expressing ACE2 infected with SARS-CoV-2 and stained for SARS-CoV-2 N protein (magenta) and dsRNA (white). Scale bar, 20 μm. Graph shows percentage of infected cells. Data in all panels are from one of three independent experiments. Data are means ± SEM; $n = 3$ biological replicates. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ [unpaired t test (H), two-way ANOVA (other panels)].

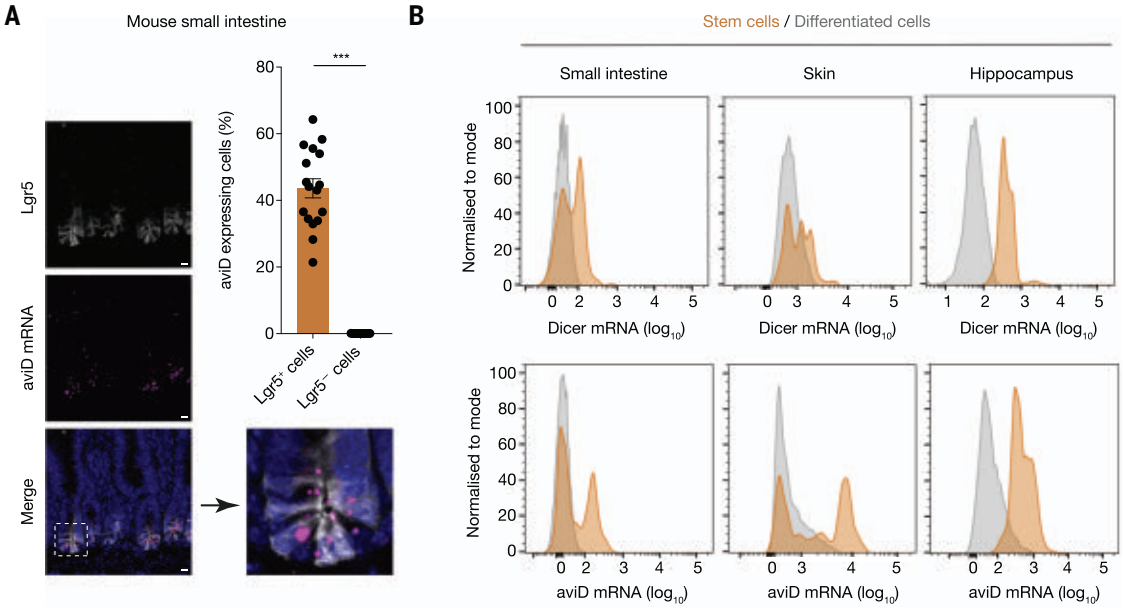


Fig. 3. *aviD* mRNA is enriched in tissue stem cells. (**A**) Small intestine from an Lgr5-GFP reporter mouse was fixed, sectioned, and probed for *aviD* mRNA (magenta) by fluorescence in situ hybridization. The *aviD* probe was designed to detect the exon-exon junction specific to *aviD* and cannot detect *Dicer* mRNA (fig. S5A). $Lgr5^+$ stem cells were identified with anti-GFP (white) and nuclei were visualized by DNA staining (Hoechst, blue). Scale bar, 30 μm. Percentages of stem ($Lgr5^+$) or differentiated ($Lgr5^-$) cells expressing aviD mRNA were determined on 17 images with at least three villi each. Data are means ± SEM. ***$P < 0.001$ (Mann-Whitney test). (**B**) *aviD* or *Dicer* mRNA was measured by PrimeFlow cytometry in stem ($Lgr5^+$) or differentiated ($Lgr5^-$) cells from small intestine or skin isolated from Lgr5-GFP reporter mice or in stem or differentiated cells from hippocampus distinguished by the presence or absence of *Sox2* mRNA, respectively.

to express the SARS-CoV-2 entry receptor ACE2 (angiotensin-converting enzyme 2) (fig. S2D). We observed a factor of 3 reduction of the number of infected $Dicer^{-/-}aviD^{+/+}$ cells relative to infected $Dicer^{+/+}aviD^{-/-}$ cells (Fig. 2H). Together, these data reveal that expression of aviD allows for an antiviral RNAi response that restricts replication of several RNA viruses. In contrast, replication of two DNA viruses, vaccinia virus and herpes simplex virus 1, was similar in $Dicer^{+/+}aviD^{-/-}$ and $Dicer^{-/-}aviD^{+/+}$ cells (fig. S4, G and H).

We examined the expression of aviD in mice. aviD transcripts could be detected by fluorescence in situ hybridization (fig. S5A) in the crypts of mouse small intestine, where they colocalized with Lgr5, a marker of intestinal stem cells, but were not found in differentiated cells along the villi (Fig. 3A). By PrimeFlow cytometry, validated using complemented $Dicer^{-/-}aviD^{-/-}$ HEK293T cells (fig. S5B), *aviD* mRNA was found to be predominantly expressed in a fraction of $Lgr5^+$ stem cells in the intestine, as well as in $Lgr5^+$ hair follicle stem cells of the skin and in $Sox2^+$ neural stem cells of the hippocampus (Fig. 3B). Consistent with the latter, *aviD* expression by RT-qPCR was found in cultured neural stem cells but, unlike *Dicer* mRNA, was lost when the cells were made to differentiate into astrocytes (fig. S5C). These data suggest that aviD is expressed preferentially by stem cells rather than differentiated cells within adult mouse tissues.

To assess the role of aviD in stem cells, we took advantage of an in vitro model of organ generation using ES cells. We complemented mouse $Dicer^{-/-}aviD^{-/-}$ ES cells (*25*) with either Dicer or aviD (fig. S6A). The interferon unresponsiveness of ES cells relies in part on their production of miR-673, which inhibits MAVS (mitochondrial antiviral signaling protein) to block coupling of RNA virus detection to IFN gene transcription (*2*). As a consequence, ES $Dicer^{-/-}aviD^{-/-}$ cells lacking miRNAs produce type I interferons and transcribe ISGs in response to viral stimulation, unlike their wild-type counterparts (*2*). We confirmed that introduction of either Dicer or aviD into $Dicer^{-/-}aviD^{-/-}$ ES cells restored miRNA production, including that of miR-673, and inhibited the induction of ISGs in response to cell stimulation with a viral RNA mimic (fig. S6, B to E). Complementation with Dicer or aviD additionally suppressed the constitutive activation of protein kinase R (PKR) that has been reported to result from *Dicer* loss (*27*) and inhibits growth (fig. S7A). Finally, we found that aviD could also mediate dsRNA-induced gene silencing in ES cells (fig. S7B) (*19*).

Brain organoids derived from ES cells recapitulate the overall organization of the adult brain (*28*), and $Sox2^+$ neural stem cells present in organoids derived from wild-type ($Dicer^{+/+}aviD^{+/+}$) ES cells expressed more aviD and Dicer transcripts than differentiated cells in the same tissue (fig. S8A). Both $Dicer^{-/-}aviD^{+/+}$ and $Dicer^{+/+}aviD^{-/-}$ ES cells generated organoids similar to those made by wild-type $Dicer^{+/+}aviD^{+/+}$ ES cells, including differentiated neuronal layers and astrocytes (fig. S8B). ZIKV infection of brain organoids preferentially targets $Sox2^+$ stem cells, resulting in slower organoid growth and increased stem cell demise by apoptosis, which recapitulates the microcephaly phenotype observed in humans (*29*). Uninfected organoids grew similarly irrespective of genotype (fig. S8C). In contrast, upon infection with ZIKV, $Dicer^{+/+}aviD^{-/-}$ organoids grew more slowly than $Dicer^{+/+}aviD^{+/+}$ and $Dicer^{-/-}aviD^{+/+}$ organoids (Fig. 4A) and produced more infectious viral particles (Fig. 4B). Thus, despite being expressed at low levels (Fig.1B), endogenous aviD in $Dicer^{+/+}aviD^{+/+}$ organoids can display antiviral activity equivalent to that of ectopically expressed aviD in $Dicer^{-/-}aviD^{+/+}$ organoids. Consistent with the notion that absence of aviD compromises stem cell resistance to viral infection, $Sox2^+$ stem cells in $Dicer^{+/+}aviD^{-/-}$ organoids displayed increased infection with ZIKV (Fig. 4C); accumulated more viral dsRNA, the substrate for aviD (Fig. 4D); and displayed decreased 5-ethynyl-2′-deoxyuridine (EdU) incorporation indicative of lower proliferation (Fig. 4E). ZIKV-derived small RNAs from infected organoids displayed canonical features of viral siRNAs, such as a predominant length of 22 nucleotides (nt) and a read-phasing consistent with the presence of 2-nt 3′ overhangs (fig. S9, A to D). $Dicer^{+/+}aviD^{-/-}$ organoids showed decreased accumulation of these viral siRNAs (fig. S9B), consistent with impaired ability to restrict ZIKV infection. SARS-CoV-2 can also display brain tropism and infect brain organoids (*30*). We engineered ES cells to express ACE2 (fig. S8D) and infected organoids with SARS-CoV-2. As for ZIKV infection, the absence of aviD in $Dicer^{+/+}aviD^{-/-}$ organoids correlated with an increase in the percentage of virally infected stem cells (Fig. 4F) as well as loss of viral siRNA production (fig. S9E). Taken together, these data indicate that aviD can

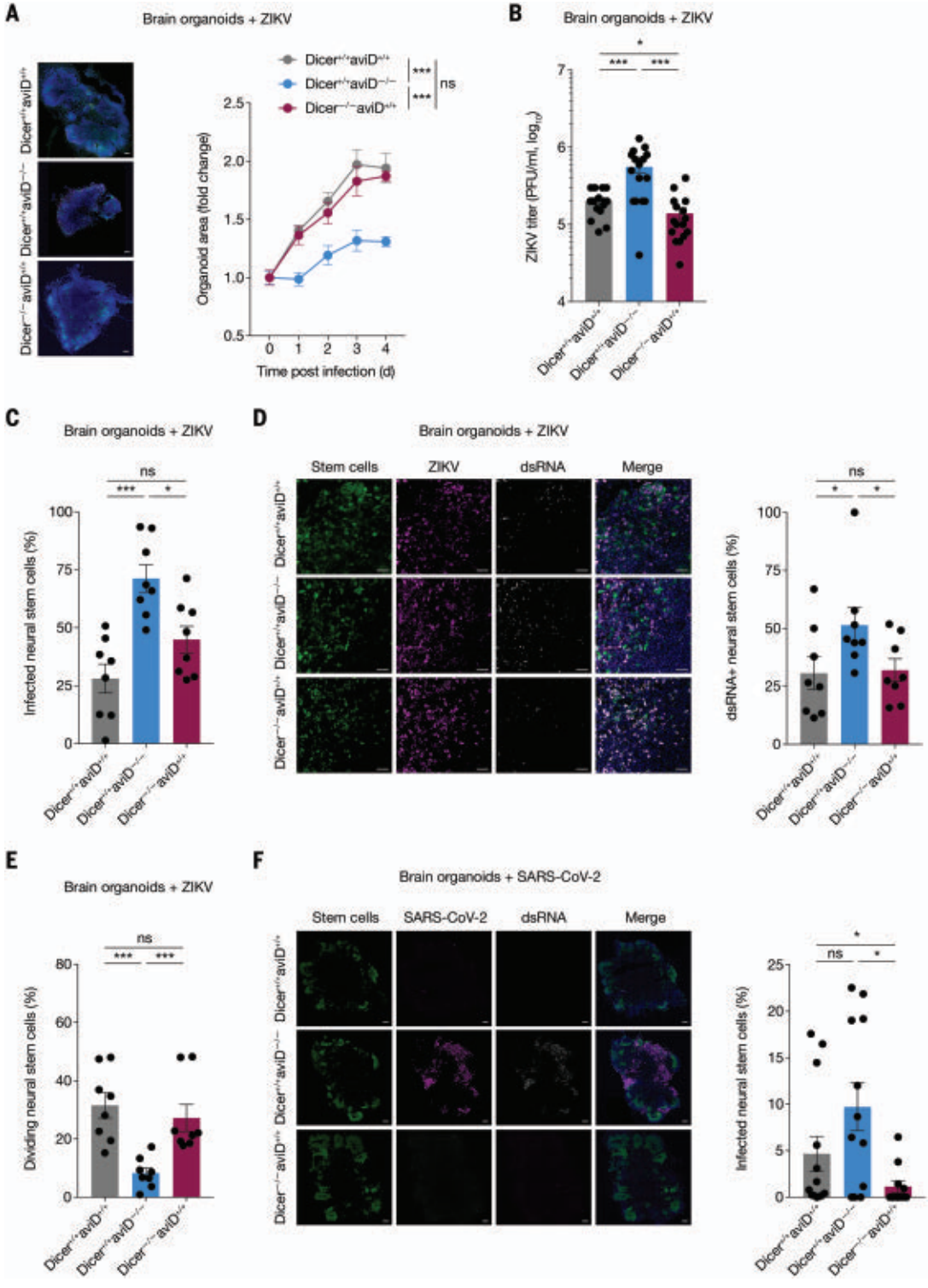


Fig. 4. aviD thwarts viral infection in stem cells. (**A**) Individual Dicer$^{+/+}$aviD$^{+/+}$, Dicer$^{+/+}$aviD$^{-/-}$, or Dicer$^{-/-}$aviD$^{+/+}$ brain organoids were infected with ZIKV, and organoid area was monitored for 4 days. Immunofluorescent stanining and confocal microscopy on organoid sections were carried out to identify stem cells by Sox2 expression (green) and infected cells by ZIKV glycoprotein expression (magenta). Scale bar, 100 μm. (**B**) Production of viral particles from ZIKV-infected organoids was determined by transferring individual organoids into fresh medium at day 3 after infection and collecting the supernatant 24 hours thereafter to determine viral content by plaque assay. In (A) and (B), n = 16 organoids per condition. (**C**) Percentage of ZIKV-infected stem cells was measured 4 days after infection by immunofluorescence on organoid sections. (**D**) dsRNA in infected stem cells was visualized by immunofluorescence on organoid sections after 4 days of infection. Images show dsRNA (gray) in ZIKV-infected (magenta) stem cells (green). Scale bar, 20 μm. (**E**) Stem cell division rate was measured by pulsing organoids with EdU at day 3 for 1 hour and chasing for 24 hours. Organoid sections were analyzed by immunofluorescence, with Sox2 staining to identify stem cells and EdU staining to mark cells in S phase at time of pulsing. In (C) to (E), n = 8 organoids per condition, 8 with highest fold change in area at day 4 for Dicer$^{+/+}$aviD$^{+/+}$ and Dicer$^{-/-}$aviD$^{+/+}$, 8 with lowest fold change in area for Dicer$^{+/+}$aviD$^{-/-}$. (**F**) Dicer$^{+/+}$aviD$^{+/+}$, Dicer$^{+/+}$aviD$^{-/-}$, or Dicer$^{-/-}$aviD$^{+/+}$ brain organoids expressing ACE2 were infected with SARS-CoV-2 for 48 hours. Percentage of infected stem cells (n = 11 organoids per condition) was determined by immunofluorescence on sections stained for the stem cell marker Sox2 (green) and for the SARS-CoV-2 N protein (magenta). Scale bar, 100 μm. Data are means ± SEM. *P < 0.05, ***P < 0.001 [two-way ANOVA in (A), Mann-Whitney test in (B) to (F)].

protect adult stem cells from ZIKV and SARS-CoV-2 virus infection by orchestrating an antiviral RNAi response.

Our results show that the *DICER* gene can generate an alternative transcript that encodes aviD, a truncated Dicer that helps protect mouse and human stem cells against RNA virus infection and compensates in part for stem cell hyporesponsiveness to innate IFNs. Our data reveal that mammals, like plants or insects, can produce at least two Dicer proteins, one of which is superior at initiating antiviral RNAi. Interestingly, aviD can also process pre-miRNAs and compensates for Dicer loss in miRNA generation when it is

the only isoform expressed in the cell; this would imply that aviD is not fully specialized for antiviral RNAi. Antiviral RNAi has been noted in some studies with differentiated cells, especially when using viruses deficient in VSRs (*8*–*10*, *12*–*14*). Whether such observations were due to aviD activity is unknown, as our data suggest that aviD is expressed only at low levels in differentiated cells. Why this should be the case is unclear. However, one element to consider is the interplay between antiviral RNAi and the IFN response (*5*, *19*, *20*). The action of aviD could deplete infected cells of viral dsRNA, thereby eliminating a key trigger of dsRNA-activated proteins of the IFN response pathway such as RIG-I, MDA5, PKR, or ribonuclease (RNase) L. This is less important for stem cells that are not reliant on the IFN pathway for antiviral resistance. Notably, aviD-mediated antiviral RNAi is not the only defense mechanism in stem cells and likely acts in concert with others conferred by IFN-independent expression of restriction factors encoded by ISGs (*21*). An aviD-specific knockout mouse will help to delineate the nonredundant contributions of these distinct strategies. Antiviral innate immunity in mammals is therefore a composite of pathways that are tailored to the differentiation status of the cell and that display complementarity as well as redundancy.

REFERENCES AND NOTES

1. D. Goubau, S. Deddouche, C. Reis e Sousa, *Immunity* **38**, 855–869 (2013).
2. J. Witteveldt, L. I. Knol, S. Macias, *eLife* **8**, e44171 (2019).
3. W. D'Angelo *et al.*, *J. Immunol.* **198**, 2147–2155 (2017).
4. X.-X. Hong, G. G. Carmichael, *J. Biol. Chem.* **288**, 16196–16205 (2013).
5. P. V. Maillard, A. G. van der Veen, E. Z. Poirier, C. Reis e Sousa, *EMBO J.* **38**, e100941 (2019).
6. P. V. Maillard *et al.*, *Science* **342**, 235–238 (2013).
7. Y. Li, J. Lu, Y. Han, X. Fan, S.-W. Ding, *Science* **342**, 231–234 (2013).
8. Y. Li *et al.*, *Nat. Microbiol.* **2**, 16250 (2016).
9. Y. Qiu *et al.*, *Immunity* **46**, 992–1004.e5 (2017).
10. Q. Qian *et al.*, *J. Virol.* **94**, e01233-19 (2019).
11. Y.-P. Xu *et al.*, *Cell Res.* **29**, 265–273 (2019).
12. Q. Han *et al.*, *mBio* **11**, e03278-19 (2020).
13. Y. Zhang *et al.*, *Emerg. Microbes Infect.* **9**, 1580–1589 (2020).
14. Y. Qiu *et al.*, *Sci. Adv.* **6**, eaax7989 (2020).
15. B. R. Cullen, S. Cherry, B. R. tenOever, *Cell Host Microbe* **14**, 374–378 (2013).
16. S. Schuster, L. E. Tholen, G. J. Overheul, F. J. M. van Kuppeveld, R. P. van Rij, *MSphere* **2**, e00333-17 (2017).
17. K. Tsai, D. G. Courtney, E. M. Kennedy, B. R. Cullen, *RNA* **24**, 1172–1182 (2018).
18. S. Schuster, G. J. Overheul, L. Bauer, F. J. M. van Kuppeveld, R. P. van Rij, *Sci. Rep.* **9**, 13752 (2019).
19. P. V. Maillard *et al.*, *EMBO J.* **35**, 2505–2518 (2016).
20. A. G. van der Veen *et al.*, *EMBO J.* **37**, 97479 (2018).
21. X. Wu *et al.*, *Cell* **172**, 423–438.e25 (2017).
22. E. Ma, I. J. MacRae, J. F. Kirsch, J. A. Doudna, *J. Mol. Biol.* **380**, 237–243 (2008).
23. E. M. Kennedy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6945–E6954 (2015).
24. M. Flemr *et al.*, *Cell* **155**, 807–816 (2013).
25. J. E. Babiarz, J. G. Ruby, Y. Wang, D. P. Bartel, R. Blelloch, *Genes Dev.* **22**, 2773–2785 (2008).
26. H. P. Bogerd, A. W. Whisnant, E. M. Kennedy, O. Flores, B. R. Cullen, *RNA* **20**, 923–937 (2014).
27. C. Gurung *et al.*, *J. Biol. Chem.* **296**, 100264 (2021).
28. M. A. Lancaster *et al.*, *Nature* **501**, 373–379 (2013).
29. P. P. Garcez *et al.*, *Science* **352**, 816–818 (2016).
30. C. Iadecola, J. Anrather, H. Kamel, *Cell* **183**, 16–27.e1 (2020).

ACKNOWLEDGMENTS

We thank P. Maillard, G. Kassiotis, A. Wack, and members of the Immunobiology laboratory for useful discussions; A. Baulies Domenech for sharing her protocol for small intestine dissociation; K. Ng for his help with ACE2 staining by flow cytometry; the Crick Flow Cytometry facility for technical assistance; the Crick Advanced Sequencing facility for generating the small RNA libraries; A. Acha for help with the BioAnalyzer; I. Dalla Rosa for advice regarding vaccinia virus infection; L. Frangeul for help with analyzing small RNA libraries; J. Canton for help with microscopy analysis; M. Way for vaccinia virus; M. Vignuzzi for the ZIKV plasmid; S. Macias for Dicer$^{-/-}$aviD$^{-/-}$ ES cells; and B. Cullen for NoDice cells. **Funding:** Supported by the Francis Crick Institute [which receives core funding from Cancer Research UK (FC001136), the UK Medical Research Council (FC001136), and the Wellcome Trust (FC001136)], ERC Advanced Investigator grant AdG 268670, Wellcome Investigator Award WT106973MA, and a prize from the Louis-Jeantet Foundation. E.Z.P. and M.D.B. are supported by EMBO Long-Term Fellowships ALTF 536-2108 and ALTF 1096-2018 and Marie Skłodowska-Curie Individual Fellowships 832511 and 837951. **Author contributions:** E.Z.P. and C.R.S. designed experiments and analyzed data; E.Z.P. conducted experiments with the assistance of M.D.B., A.C., B.F., and L.H.; J.C. performed the BaseScope in situ hybridization experiment; P.C. analyzed the small RNA sequencing data; R.U. and R.B. provided SARS-CoV-2 reagents; E.Z.P., M.D.B., and C.R.S. wrote the manuscript; C.R.S. supervised the project. **Competing interests:** C.R.S. has an additional appointment as professor in the Faculty of Medicine at Imperial College London and owns stock options and/or is a paid consultant for Bicara Therapeutics, Montis Biosciences, Oncurious NV, Bicycle Therapeutics, and Sosei Heptares, all unrelated to this work. The remaining authors declare no competing interests. **Data and materials availability:** RNA-seq data used for fig. S9 have been deposited in GenBank under accession number GSE173946. All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/231/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*31*–*43*)

18 December 2020; accepted 19 May 2021
10.1126/science.abg2264

CORONAVIRUS

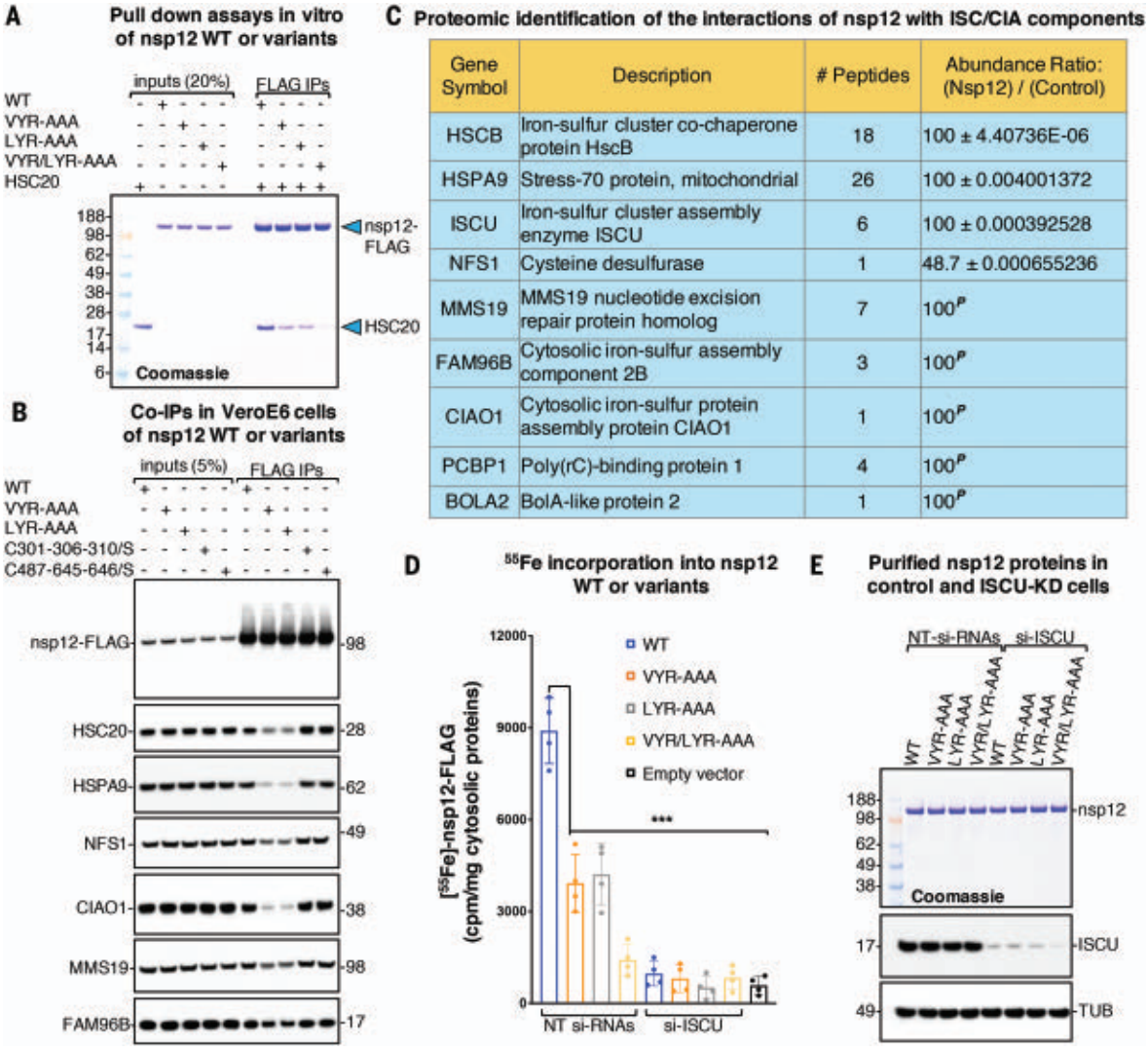
Fe-S cofactors in the SARS-CoV-2 RNA-dependent RNA polymerase are potential antiviral targets

Nunziata Maio[1], Bernard A. P. Lafont[2], Debangsu Sil[3], Yan Li[4], J. Martin Bollinger Jr.[3,5], Carsten Krebs[3,5], Theodore C. Pierson[6], W. Marston Linehan[7], Tracey A. Rouault[1]*

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causal agent of COVID-19, uses an RNA-dependent RNA polymerase (RdRp) for the replication of its genome and the transcription of its genes. We found that the catalytic subunit of the RdRp, nsp12, ligates two iron-sulfur metal cofactors in sites that were modeled as zinc centers in the available cryo–electron microscopy structures of the RdRp complex. These metal binding sites are essential for replication and for interaction with the viral helicase. Oxidation of the clusters by the stable nitroxide TEMPOL caused their disassembly, potently inhibited the RdRp, and blocked SARS-CoV-2 replication in cell culture. These iron-sulfur clusters thus serve as cofactors for the SARS-CoV-2 RdRp and are targets for therapy of COVID-19.

The novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic known as COVID-19 (*1*–*3*), which can be prevented by vaccines but for which antiviral treatments are much needed. Coronaviruses employ a multisubunit machinery for replication and transcription. A set of nonstructural proteins (nsps) produced as cleavage products of the ORF1a and ORF1ab polyproteins (*4*) assemble to facilitate viral replication and transcription. The core component of this complex is the catalytic subunit (nsp12) of an RNA-dependent RNA polymerase (RdRp) (*5*), which catalyzes the synthesis of viral RNA and thus plays a central role in the replication and transcription cycle of SARS-CoV-2, with the assistance of nsp7 and nsp8 as accessory factors (*6*, *7*). Structures of the RdRp (nsp12-nsp7-nsp8 complex) alone and in complex with the helicase have been determined by cryo–electron microscopy (cryo-EM) (*8*–*11*); in all of these structures, the RdRp of SARS-CoV-2 was proposed to contain zinc ions ligated in the same locations as those observed in SARS-CoV (*7*) in highly conserved metal binding motifs

[1]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [2]SARS-CoV-2 Virology Core, Laboratory of Viral Diseases, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [3]Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA. [4]Proteomics Core Facility, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA. [5]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA. [6]Laboratory of Viral Diseases, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [7]Urologic Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA.
***Corresponding author. Email: rouault@mail.nih.gov**



Gene Symbol	Description	# Peptides	Abundance Ratio: (Nsp12) / (Control)
HSCB	Iron-sulfur cluster co-chaperone protein HscB	18	100 ± 4.40736E-06
HSPA9	Stress-70 protein, mitochondrial	26	100 ± 0.004001372
ISCU	Iron-sulfur cluster assembly enzyme ISCU	6	100 ± 0.000392528
NFS1	Cysteine desulfurase	1	48.7 ± 0.000655236
MMS19	MMS19 nucleotide excision repair protein homolog	7	100^P
FAM96B	Cytosolic iron-sulfur assembly component 2B	3	100^P
CIAO1	Cytosolic iron-sulfur protein assembly protein CIAO1	1	100^P
PCBP1	Poly(rC)-binding protein 1	4	100^P
BOLA2	BolA-like protein 2	1	100^P

Fig. 1. Fe-S cluster incorporation into nsp12 occurs through its interactions with components of the Fe-S biogenesis machinery. (**A**) Representative Coomassie blue staining of pull-down assays performed with purified proteins. Purified nsp12-FLAG (0.25 μg) or the variants wherein either or both LYR motifs were replaced by alanines (VYR-AAA, LYR-AAA, and VYR/LYR-AAA, respectively) were combined with 0.25 μg of HSC20, as indicated. Immunoprecipitations (IPs) were performed with anti-FLAG antibody to immunocapture nsp12 proteins. The presence of HSC20 (i.e., HSCB) in the eluates after IPs of nsp12 proteins was analyzed by SDS–polyacrylamide gel electrophoresis and Coomassie staining. Aliquots corresponding to 20% of the inputs were run on the gel for comparison (n = 5 biological replicates). (**B**) Eluates after IPs of nsp12 WT or variants recombinantly expressed in Vero E6 cells, as indicated, were probed with antibodies against FLAG to verify the efficiency of IP and against components of the Fe-S cluster (HSC20, HSPA9, and NFS1) and of the cytoplasmic Fe-S (CIA) assembly machinery (CIAO1, MMS19, and FAM96B) (n = 6). (**C**) Mass spectrometry identification of affinity purified interacting partners of nsp12 that are components of the Fe-S cluster biogenesis pathway (see data S1 for a complete list). The protein ratios were calculated as reported in the methods (n = 6). The maximum allowed fold change value was set to 100. In the instances (marked with a superscript P) in which the interacting partner was detected in the nsp12-only samples and not in the negative controls, the nsp12/control ratios were set to 100 and reported without P values. (**D**) Levels of radiolabeled iron (^{55}Fe) incorporated into nsp12 WT or the variants in control cells transfected with nontargeting siRNAs (NT siRNAs) and in cells transfected with siRNAs directed against the main scaffold protein ISCU (si-ISCU). Levels of iron stochastically associated with the beads in lysates from cells transfected with the backbone plasmid (empty-vector, p3XFLAG-CMV-14) are also reported (accounting for 587 ± 292.62 cpm/mg of cytosolic proteins) and were not subtracted from measurements of radiolabeled iron incorporated into nsp12 WT or variants in the chart (n = 4). Significance was determined by two-way analysis of variance (ANOVA) and Sidak's multiple comparisons test. Mean ± 95% confidence interval (CI). ***P < 0.001. (**E**) Representative Coomassie staining showing levels of nsp12 WT or variants in control and ISCU-depleted cells that were quantified in (D) for their iron content. Immunoblots to ISCU, showing the efficiency of its silencing (knock down), and to α-tubulin (TUB), used as a loading control, are also shown.

composed of H295-C301-C306-C310 and C487-H642-C645-C646 (fig. S1). These zinc ions have been proposed to serve a structural role in maintaining the integrity of the RdRp architecture (*7*–*11*) (see supplementary text in the supplementary materials). Zinc has long been known to be capable of replacing endogenous iron-sulfur (Fe-S) metal cofactors during standard aerobic purification of proteins (*12*–*15*), because Fe-S clusters are inherently susceptible to destabilization and degradation by oxidants, including oxygen, superoxide (O_2^-), and nitric oxide (*16*). Notably, Fe-S clusters, inorganic cofactors often associated with biological redox reactions (*17*, *18*), have been

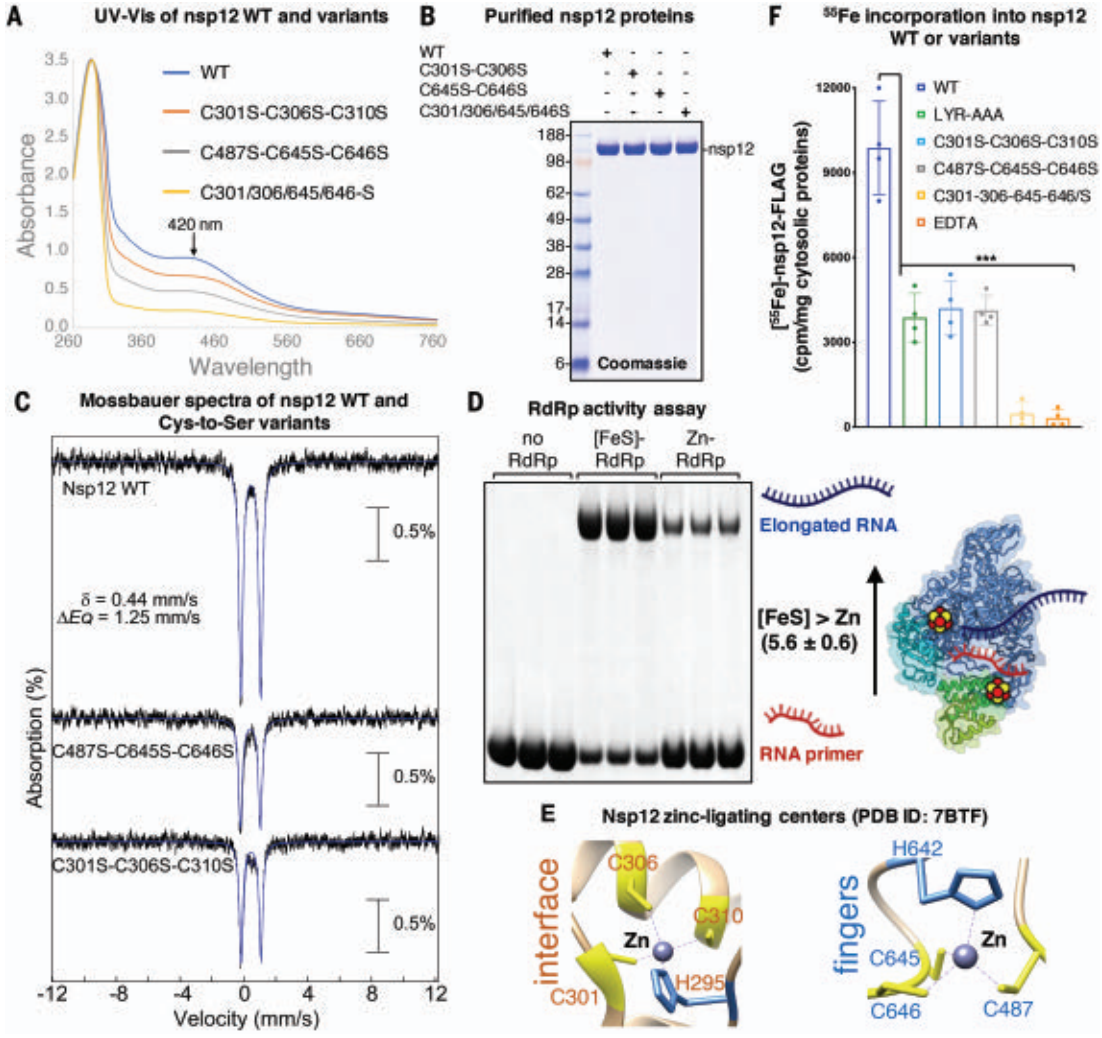


Fig. 2. Evidence for ligation of two Fe-S metal cofactors by nsp12. (**A**) UV-vis spectra of nsp12 WT or variants of the cysteine residues in the two metal ligating centers. (**B**) Representative Coomassie blue staining of purified nsp12 WT or variants analyzed in (A). (**C**) Mössbauer spectra of nsp12 WT and variants exhibited the parameters typical of [Fe_4S_4] clusters. For each of the two nsp12 Cys-to-Ser variants, ~95% of iron was still associated with a quadrupole doublet that matched parameters of WT nsp12. (**D**) RNA polymerase activity of anoxically purified RdRp ([Fe-S]-RdRp at 1 μM) and aerobically purified RdRp reconstituted with zinc and containing two zinc ions per protomer (Zn-RdRp at 1 μM) (n = 4). (**E**) Conserved zinc-binding motifs in SARS-CoV-2 nsp12 [Protein Data Bank (PDB) ID 7BTF] (*8*) rendered in the ribbon representation. H295-C301-C306-C310 ligate zinc at the interface between the NiRAN and the catalytic domain, whereas the C487-H642-C645-C646 residues ligate zinc in the catalytic domain. (**F**) Levels of radiolabeled iron (^{55}Fe) incorporated into nsp12 WT or variants, as indicated. ^{55}Fe content of nsp12 treated with the chelator EDTA is also reported to provide a control for the complete loss of ^{55}Fe in the protein (n = 4). Significance was determined by two-way ANOVA and Sidak's multiple comparisons test. Mean ± 95% CI. ***P < 0.001.

identified in numerous proteins involved in DNA and RNA metabolism, where they play a variety of critical functional roles (*12*, *13*, *19*–*26*).

Having recently demonstrated that we are able to predict the presence of Fe-S cofactors in candidate proteins based on the identification of specific amino acid sequence motifs (*27*), we analyzed the primary sequences of SARS-CoV-2 proteins to investigate whether any might incorporate Fe-S clusters. We identified two highly conserved LYR (leucine-arginine-tyrosine)–like motifs (fig. S2A) in nsp12 that have been previously characterized as potential binding sites for the cochaperone HSC20 (also known as HSCB) of the Fe-S biogenesis machinery (*27*–*30*), which facilitates Fe-S cluster transfer from the main scaffold protein, ISCU (iron-sulfur cluster assembly scaffold), to recipient proteins (fig. S2B). To assess whether the LYR-like motifs were involved in direct binding of nsp12 to HSC20, we incubated full-length SARS-CoV-2 nsp12 wild type (WT) or variants wherein either or both LYR motifs were replaced by alanines (A) (fig. S2C) with purified HSC20. Nsp12 WT bound HSC20, indicating that the RdRp subunit interacts directly with the cochaperone (Fig. 1A). Substitution of either of the two LYR motifs with alanines decreased the amount of bound HSC20 (Fig. 1A), which was even more profoundly diminished by loss of both motifs in nsp12$^{VYR/LYR-AAA}$ (Fig. 1A). Coimmunoprecipitation (co-IP) experiments in Vero E6 cells and mass spectrometry analysis confirmed that nsp12 transiently interacted with HSC20 and with components of the de novo Fe-S cluster (the chaperone HSPA9, the cysteine desulfurase NFS1, and the main scaffold ISCU) and cytoplasmic Fe-S (CIA) biogenesis (CIAO1, MMS19, and FAM96B) machineries (Fig. 1, B and C; fig. S2D; and data S1), suggesting that these interactions may be required for Fe-S cluster acquisition by nsp12. To investigate whether nsp12 coordinated an Fe-S cluster, we quantified ^{55}Fe incorporation into the protein expressed in cells transfected with either a pool

of nontargeting small interfering RNAs (NT siRNAs) or with siRNAs against the initial Fe-S biogenesis scaffold, ISCU. In control cells (NT siRNAs), nsp12 WT bound radiolabeled iron (8312 ± 775 cpm/mg of cytosolic proteins) (Fig. 1, D and E), whereas nsp12 that lacked the LYR motifs did not interact with HSC20 and bound significantly less iron (250 ± 92 cpm/mg of cytosolic proteins) (Fig. 1, D and E). Nsp12 expressed in cells silenced for ISCU (si-ISCU) failed to incorporate iron (Fig. 1, D and E). Taken together, these results demonstrate that nsp12 binds iron, likely in the form of an Fe-S cluster. Nsp12 expressed in Expi293F mammalian cells and purified anoxically exhibited a shoulder at ~420 nm in its ultraviolet–visible (UV-vis) absorption spectrum (Fig. 2, A and B, and fig. S3, A and B), suggesting that it harbored one or more Fe-S clusters (*31*, *32*). To determine the type and stoichiometry of the Fe-S cluster(s), a ^{57}Fe-enriched nsp12-FLAG sample was analyzed by Mössbauer spectroscopy (Fig. 2C). The 4.2-K Mössbauer spectrum collected in a 53-mT magnetic field applied parallel to the direction of gamma radiation (Fig. 2C) shows the presence of a single quadrupole doublet with parameters typical of $[Fe_4S_4]^{2+}$ clusters [isomer shift (δ) of 0.44 mm/s and quadrupole splitting parameter (ΔE_Q) of 1.25 mm/s, blue line] (*33*). Wild-type nsp12 bound 7.5 ± 0.35 iron atoms per monomer, and we thus interpret the Mössbauer spectrum as two $[Fe_4\text{-}S_4]^{2+}$ clusters. The X-band electron paramagnetic resonance (EPR) spectrum, recorded at 20 K, showed no signal (fig. S3C), ruling out the presence of Fe-S clusters with a half-integer spin ground state. However, upon reduction with dithionite, EPR signal characteristics of $[Fe_4S_4]^{+}$ clusters were observed (fig. S3D) (*34*). Notably, the nsp12-nsp7-nsp8 complex anoxically purified with the Fe-S cluster(s) showed markedly increased binding to the template and RNA primer (fig. S4) and increased polymerase activity relative to the aerobically purified complex that contained two zinc ions per protomer (Fig. 2D and fig. S4).

The available cryo-EM structures of the RdRp complex have assigned two chelated zinc ions in the highly conserved metal binding motifs of nsp12 composed of H295-C301-C306-C310 at the interface between the NiRAN (nidovirus RdRp-associated nucleotidyltransferase) domain and the catalytic domain and of C487-H642-C645-C646 in the fingers of the catalytic domain (*7*–*11*) (Fig. 2E and fig. S1; see supplementary text). By replacing selected cysteines with serines and characterizing the variant nsp12 proteins, we tested the hypothesis that the two $[Fe_4\text{-}S_4]$ clusters are coordinated by these motifs. The two variants lacking any one of the set of three Cys residues of either the interfacial motif ($nsp12^{C301S\text{-}C306S\text{-}C310S}$) or the catalytic domain ($nsp12^{C487S\text{-}C645S\text{-}C646S}$) (replaced by Ser) contained 3.8 ± 0.2 and 3.67 ± 0.3 Fe per nsp12 protomer, respectively, and exhibited approximately half of the absorbance at 420 nm (Fig. 2, A and B, and fig. S3, A and B) and half of the ^{55}Fe radiolabel seen for the WT nsp12 (Fig. 2F). The 4.2-K/53-mT Mössbauer spectra of these two variants revealed that ~95% of Fe is associated with the quadrupole doublet with the same parameters deduced from the spectrum of WT nsp12, thus revealing the presence of one $[Fe_4S_4]^{2+}$ cluster in the unmodified binding site (Fig. 2C). The 20-K X-band EPR spectra of the variants after they were treated with sodium dithionite are also consistent with the presence of one $[Fe_4S_4]^{2+}$ cluster (fig. S3D). A variant lacking a total of four cysteines from both motifs ($nsp12^{C301S\text{-}C306S\text{-}C645S\text{-}C646S}$) did not bind Fe and had no absorbance at 420 nm, consistent with

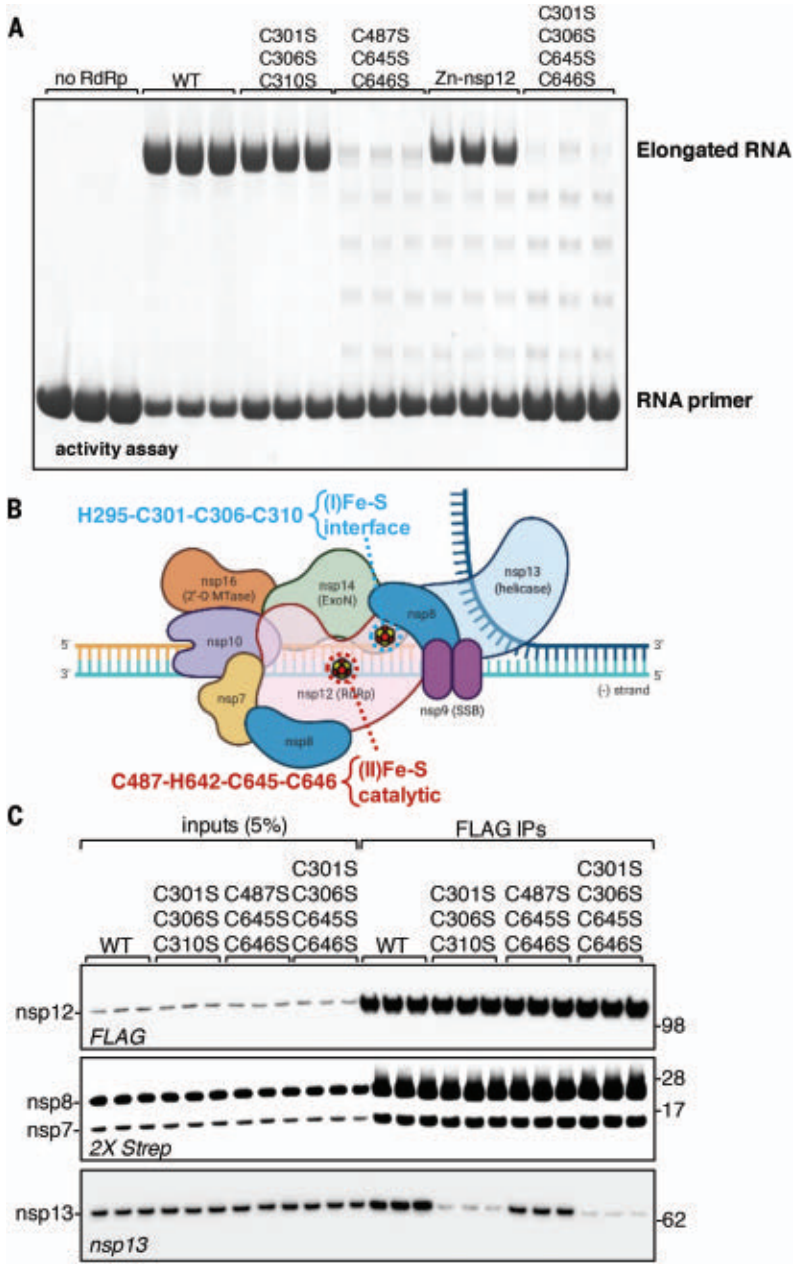


Fig. 3. Fe-S cluster sites in nsp12 are important for activity and interactions with nsp13. (**A**) RNA polymerase activity of anoxically purified RdRp (all lanes except Zn-nsp12) (RdRp at 1 μM) and of aerobically purified and Zn-reconstituted RdRp containing two zinc ions per protomer (three technical replicates are shown; *n* = 4). (**B**) Schematic of the complex required for coronaviral replication (*10*), in which the two Fe-S clusters and their coordination spheres are highlighted. ExoN, exoribonuclease; SSB, single-stranded DNA-binding protein; 2′-O MTase, 2′-O methyltransferase. (**C**) Co-IP of nsp12 WT or variants recombinantly expressed in Vero E6 cells cotransfected with helicase nsp13 and accessory factors nsp7 and nsp8 (Strep II tagged) probed with antibodies against FLAG, Strep II, or nsp13 (three technical replicates are shown; *n* = 4).

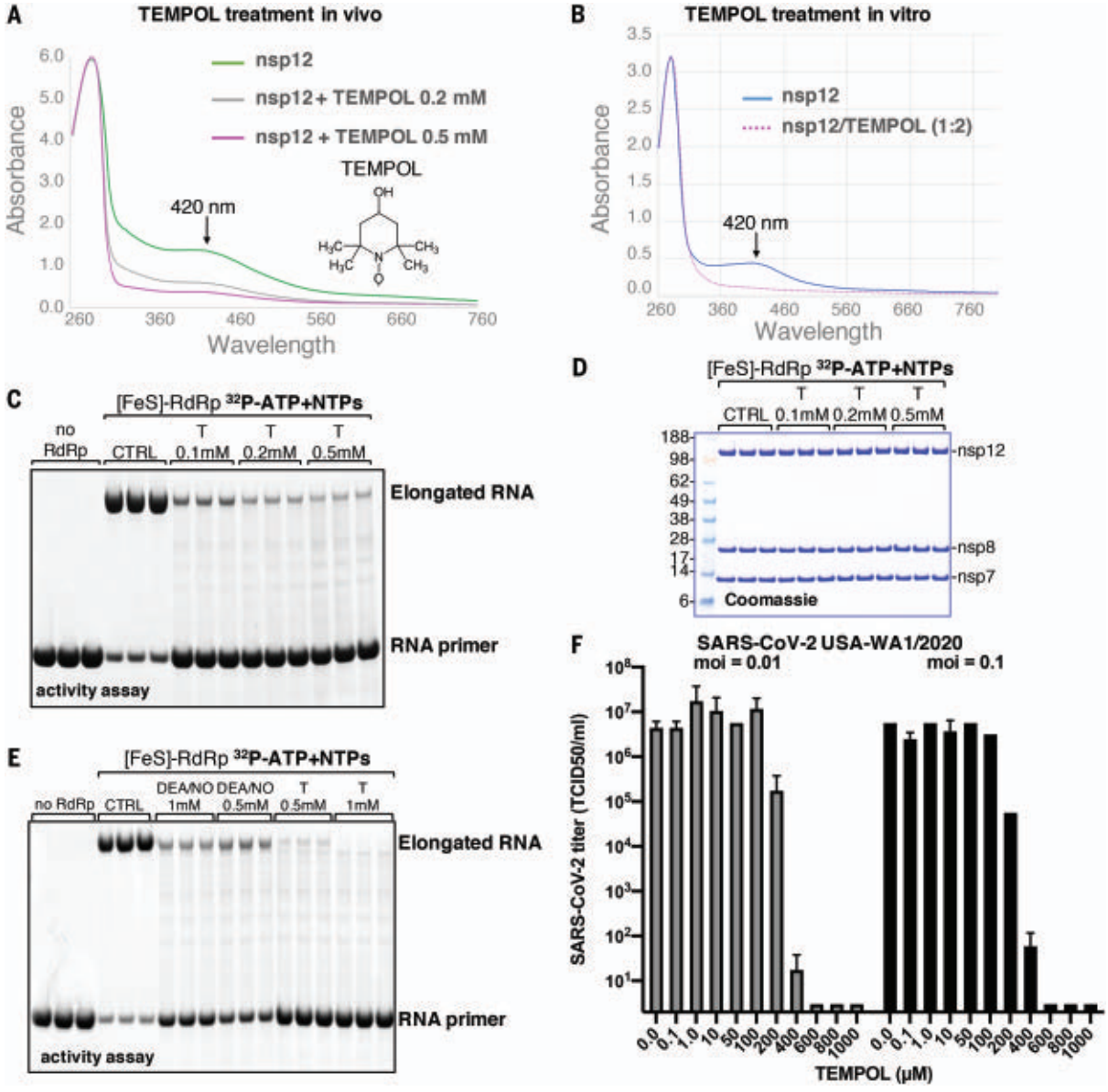


Fig. 4. The stable nitroxide TEMPOL potently inhibited the RdRp by causing disassembly of its Fe-S clusters and blocked viral replication in cell culture models of SARS-CoV-2 infection. (**A**) UV-vis spectra of nsp12 anoxically purified from Expi293F control cells and from cells treated with TEMPOL. (**B**) UV-vis spectra of purified nsp12 and of purified nsp12 incubated with TEMPOL (1:2 ratio nsp12:TEMPOL) for 10 min. (**C**) RNA polymerase activity of the RdRp complexes anoxically purified from control and TEMPOL-treated (T) Expi293F cells. (**D**) Representative Coomassie staining of the RdRp complexes analyzed for activity in (C). (**E**) RNA polymerase assay of the RdRp complexes (at 1 μM) anoxically purified from control or DEA/NO- or TEMPOL-treated Vero E6 cells, as indicated (n = 4). (**F**) Titer of infectious virus produced at 48 hours measured by $TCID_{50}$ (median tissue culture infectious dose) assay in Vero E6 cells infected with SARS-CoV-2 at a multiplicity of infection (moi) of 0.1 or 0.01 (n = 3).

the notion that both [Fe_4S_4] cluster binding sites had been eliminated (Fig. 2, A and B, and fig. S3, A and B). The two [Fe_4S_4]$^{2+}$ clusters incorporated in a mammalian overexpression system are thus ligated by cysteine residues located in the two zinc-binding sites identified in the cryo-EM structures.

We next aimed to characterize the role of the two Fe-S clusters in the RdRp. Functional studies revealed that the [Fe_4S_4] cluster in the catalytic domain of nsp12 is required for the RNA polymerase activity of the nsp12-nsp7-nsp8 complex (Fig. 3, A and B), in addition to presumably maintaining structure. In fact, the absence of the cysteine ligands in the catalytic domain in the nsp12$^{C487S-C645S-C646S}$ variant caused a more profound decrease in the polymerase activity than was observed in the zinc complex (Fig. 3A), suggesting that Zn, by coordinating the same cysteine residues, can partially fulfill the structural role of the Fe-S cluster, preserve the architecture of the fingers subdomain, and maintain some polymerase activity, which is strictly associated with the palm of the catalytic domain (*8*, *11*). Fe-S enzymes involved in DNA and RNA metabolism have often been mischaracterized as zinc-containing proteins, as Fe-S clusters readily undergo oxidative degradation during standard aerobic purification procedures of proteins, allowing zinc to coordinate the same cysteine residues. Moreover, zinc-containing enzymes have been shown to retain activity in vitro on short templates (*14*, *35*), which previously supported the conclusion that zinc was the physiological cofactor of these enzymes. Fe-S clusters in nucleic acid

metabolism enzymes have been thought to participate directly not in catalysis but in modulating binding of the enzyme to the template and/or to other components of the replication complex (*26*, *36*, *37*), as well as in increasing processivity and enabling repair through a proposed charge-transfer mechanism (*38*, *39*). Consistent with the notion that zinc is likely not the physiological cofactor in several viral replicases that have so far been crystallized with chelated zinc ions, supplementation with zinc has been reported to inhibit replication in several cell culture models of viral infection (*40*–*42*). Loss of the [Fe_4-S_4] cluster ligated by H295-C301-C306-C310, which is located at the interface between the NiRAN and the catalytic domain of nsp12, had minimal effect on the RNA polymerase activity (Fig. 3, A and B). However, loss of this cluster profoundly diminished the interaction with the helicase nsp13 (Fig. 3, B and C), which is an essential component of the replication complex.

We attempted to exploit the sensitivity of Fe-S clusters to oxidative degradation (*43*) to prevent coronavirus replication in cell culture models. Previous studies have shown that a stable nitroxide, TEMPOL (4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl), was beneficial in two different animal models of human conditions through its ability to oxidize and disassemble the Fe-S cluster of cytosolic aconitase (IRP1), thereby converting it into the iron responsive element (IRE)–binding apo-form (*44*–*46*). RdRp isolated from Expi293F cells that had been treated with TEMPOL (Fig. 4A) had diminished absorbance at 420 nm relative to the complex isolated from untreated cells, indicative of loss of the Fe-S clusters of nsp12. Likewise, treatment with TEMPOL of the Fe-S cluster–containing protein in vitro caused loss of absorbance in the same region (Fig. 4B). Either treatment resulted in loss of polymerase activity (Fig. 4, C to E). The TEMPOL treatment of cells did not impact the activities of several mitochondrial Fe-S enzymes, including the respiratory complexes and mitochondrial aconitase (ACO2), and the cytosolic Fe-S enzyme dihydropyrimidine dehydrogenase (DPYD) (figs. S5 and S6, A to F), nor did it cause any cytotoxicity at doses up to 5 mM (fig. S6G). TEMPOL treatment also did not affect the interactions of nsp12 with the components of the Fe-S and CIA biogenesis machinery from which nsp12 acquires its Fe-S clusters (fig. S7). We thus infer that TEMPOL directly reacts with Fe-S clusters in RdRp, leading to their degradation.

In support of this mechanism of action, diethylamine nonoate (DEA/NO), a nitric oxide donor that readily reacts with Fe-S clusters to form dinitrosyl complexes with diminished absorbance (*47*, *48*), also inhibited the RdRp (Fig. 4E and fig. S8), although less effectively than TEMPOL. We found that TEMPOL was both a more potent RdRp inhibitor (fig. S9) and synergized with remdesivir (RDV) (fig. S10), a nucleoside analog that has been used to target the replication of SARS-CoV-2 (*49*). RDV was notably less effective against the Fe-S–RdRp than the zinc-RdRp (fig. S11).

Having demonstrated a strong inhibitory effect of TEMPOL on the activity of the RdRp of SARS-CoV-2, we asked whether TEMPOL might exhibit antiviral activity against live virus replication. Vero E6 cells were infected with the SARS-CoV-2 USA-WA1/2020 isolate in the presence of increasing concentrations of TEMPOL (range: 0.1 to 1 mM). TEMPOL exhibited strong antiviral activity at concentrations above 0.2 mM. Viral titers were reduced by more than 5 $\log_{10}$ in the presence of 0.4 mM TEMPOL, which is reported to have a 50% cytotoxic concentration (CC50) greater than 100 mM (*50*). Our studies present a molecular basis for pursuing TEMPOL—with its low cytotoxicity and known access to tissues relevant for COVID-19 infection (*51*, *52*)—and other related stable nitroxides as potential SARS-CoV-2 therapies during active viral infection.

REFERENCES AND NOTES

1. J. F. Chan *et al*., *Lancet* **395**, 514–523 (2020).
2. N. Chen *et al*., *Lancet* **395**, 507–513 (2020).
3. F. Wu *et al*., *Nature* **579**, 265–269 (2020).
4. P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
5. D. G. Ahn, J. K. Choi, D. R. Taylor, J. W. Oh, *Arch. Virol.* **157**, 2095–2104 (2012).
6. L. Subissi *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3900–E3909 (2014).
7. R. N. Kirchdoerfer, A. B. Ward, *Nat. Commun.* **10**, 2342 (2019).
8. Y. Gao *et al*., *Science* **368**, 779–782 (2020).
9. H. S. Hillen *et al*., *Nature* **584**, 154–156 (2020).
10. J. Chen *et al*., *Cell* **182**, 1560–1573.e13 (2020).
11. W. Yin *et al*., *Science* **368**, 1499–1504 (2020).
12. G. D. Shimberg *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4700–4705 (2016).
13. D. J. Netz *et al*., *Nat. Chem. Biol.* **8**, 125–132 (2011).
14. G. D. Shimberg, J. D. Pritts, S. L. J. Michel, *Methods Enzymol.* **599**, 101–137 (2018).
15. A. R. Conlan *et al*., *J. Mol. Biol.* **392**, 143–153 (2009).
16. J. A. Imlay, *Mol. Microbiol.* **59**, 1073–1082 (2006).
17. N. Maio, T. A. Rouault, *Trends Biochem. Sci.* **45**, 411–426 (2020).
18. N. Maio, T. A. Rouault, in *Biochemistry, Biosynthesis and Human Diseases*, vol. 2 of *Iron-Sulfur Clusters in Chemistry and Biology*, T. A. Rouault, Ed. (De Gruyter, ed. 2, 2017).
19. S. Klinge, J. Hirst, J. D. Maman, T. Krude, L. Pellegrini, *Nat. Struct. Mol. Biol.* **14**, 875–877 (2007).
20. B. E. Weiner *et al*., *J. Biol. Chem.* **282**, 33444–33451 (2007).
21. S. Vaithiyalingam, E. M. Warren, B. F. Eichman, W. J. Chazin, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13684–13689 (2010).
22. L. Sauguet, S. Klinge, R. L. Perera, J. D. Maman, L. Pellegrini, *PLOS ONE* **5**, e10083 (2010).
23. V. B. Agarkar, N. D. Babayeva, Y. I. Pavlov, T. H. Tahirov, *Cell Cycle* **10**, 926–931 (2011).
24. A. G. Baranovskiy *et al*., *J. Biol. Chem.* **290**, 5635–5646 (2015).
25. M. L. Kilkenny, G. De Piccoli, R. L. Perera, K. Labib, L. Pellegrini, *J. Biol. Chem.* **287**, 23740–23747 (2012).
26. M. Girbig *et al*., *Nat. Struct. Mol. Biol.* **28**, 210–219 (2021).
27. G. Liu *et al*., *Nat. Commun.* **11**, 6310 (2020).
28. N. Maio *et al*., *Cell Metab.* **19**, 445–457 (2014).
29. N. Maio, T. A. Rouault, *Biochim. Biophys. Acta* **1853**, 1493–1512 (2015).
30. K. S. Kim, N. Maio, A. Singh, T. A. Rouault, *Hum. Mol. Genet.* **27**, 837–852 (2018).
31. E. C. Duin *et al*., *Biochemistry* **36**, 11811–11820 (1997).
32. S. Arragain *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7355–7360 (2017).
33. M. E. Pandelia, N. D. Lanz, S. J. Booker, C. Krebs, *Biochim. Biophys. Acta* **1853**, 1395–1405 (2015).
34. W. R. Hagen, *Biomolecular EPR Spectroscopy* (CRC Press, 2020).
35. J. K. Barton, R. M. B. Silva, E. O'Brien, *Annu. Rev. Biochem.* **88**, 163–190 (2019).
36. D. Martin *et al*., *FASEB J.* **27**, 1074–1083 (2013).
37. J. Ter Beek *et al*., *Nucleic Acids Res.* **47**, 5712–5722 (2019).
38. E. M. Boon, A. L. Livingston, N. H. Chmiel, S. S. David, J. K. Barton, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12543–12547 (2003).
39. E. O'Brien *et al*., *Science* **355**, eaag1789 (2017).
40. A. J. te Velthuis *et al*., *PLOS Pathog.* **6**, e1001176 (2010).
41. N. Kaushik *et al*., *J. Virol.* **91**, e00754-17 (2017).
42. R. O. Suara, J. E. Crowe Jr., *Antimicrob. Agents Chemother.* **48**, 783–790 (2004).
43. T. A. Rouault, N. Maio, *Mol. Cell* **78**, 1–3 (2020).
44. M. C. Ghosh *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12028–12033 (2008).
45. M. C. Ghosh, D. L. Zhang, H. Ollivierre, M. A. Eckhaus, T. A. Rouault, *J. Clin. Invest.* **128**, 1317–1325 (2018).
46. G. Costain *et al*., *Brain* **142**, 1195–1202 (2019).
47. S. R. Jaffrey, N. A. Cohen, T. A. Rouault, R. D. Klausner, S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12994–12998 (1994).
48. E. M. Palmieri *et al*., *Nat. Commun.* **11**, 698 (2020).
49. M. Wang *et al*., *Cell Res.* **30**, 269–271 (2020).
50. L. B. Oliveira, F. S. Celes, C. N. Paiva, C. I. de Oliveira, *Front. Cell. Infect. Microbiol.* **9**, 237 (2019).
51. A. P. Cotrim *et al*., *Clin. Cancer Res.* **13**, 4928–4933 (2007).
52. Y. Wang *et al*., *Iran. J. Basic Med. Sci.* **21**, 1238–1244 (2018).
53. N. Maio *et al*., Supplementary data for "Fe-S cofactors in the SARS-CoV-2 RNA-dependent RNA polymerase are potential antiviral targets," MassIVE (2021); https://doi.org/10.25345/C5MV60.

ACKNOWLEDGMENTS

The authors thank S. Holland (NIAID) for insightful discussions and guidance, NIAID for access to live viral testing, A. Singh (NICHD) and B. Fubara (NINDS) for technical assistance, all members of the Rouault lab for feedback that greatly improved the quality of this work, and the Eunice Kennedy Shriver National Institute of Child Health and Human Development for support. **Funding:** This work was supported by the Intramural Research Program of the National Institutes of Health (T.A.R.); the Center for Cancer Research, National Cancer Institute (W.M.L.); the Division of Intramural Research, NIAID (T.C.P.); and award R35 GM-127079 from the National Institutes of Health (to C.K.). **Author contributions:** N.M., T.A.R., and W.M.L. conceived of the project. N.M. designed the project, wrote the manuscript, and designed and performed most of the experiments, except SARS-CoV-2 viral infection assays (B.A.P.L.), EPR and Mossbauer spectroscopies (D.S.), and mass spectrometry sample preparation and analysis (Y.L.). N.M., T.A.R., W.M.L., B.A.P.L., D.S., Y.L., J.M.B., T.C.P., and C.K. analyzed the data. T.A.R. supervised the study and wrote the manuscript. All authors revised the manuscript. **Competing interests:** On the basis of the implications of the discoveries reported here, N.M., T.A.R., and W.M.L. have filed a patent (application no. 63/193656). **Data and materials availability:** The mass spectrometry data have been deposited to MassIVE (*53*). All other data needed to evaluate the conclusions of the paper are present in the main text and supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/236/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (*54*–*66*)
MDAR Reproducibility Checklist
Data S1

20 March 2021; accepted 28 May 2021
Published online 3 June 2021
10.1126/science.abi5224

NEUROSCIENCE

Nonlocal spatiotemporal representation in the hippocampus of freely flying bats

Nicholas M. Dotson[1] and Michael M. Yartsev[1,2]*

Navigation occurs through a continuum of space and time. The hippocampus is known to encode the immediate position of moving animals. However, active navigation, especially at high speeds, may require representing navigational information beyond the present moment. Using wireless electrophysiological recordings in freely flying bats, we demonstrate that neural activity in area CA1 predominantly encodes nonlocal spatial information up to meters away from the bat's present position. This spatiotemporal representation extends both forward and backward in time, with an emphasis on future locations, and is found during both random exploration and goal-directed navigation. The representation of position thus extends along a continuum, with each moment containing information about past, present, and future, and may provide a key mechanism for navigating along self-selected and remembered paths.

Bats are renowned for their exceptional navigational abilities and are believed to use map-based navigation to forage in the wild (*1*–*4*). The neural circuitry for map-based representations includes the hippocampus, which encodes the immediate spatial position in terrestrial animals (*5*–*7*). Consistent with these studies, the basic components for representing the immediate three-dimensional (3D) spatial position in freely flying bats have been identified (*8*–*10*). However, is it enough for navigational information to be represented only at the present moment or should such representations extend from the past and into the future during active movement (*11*–*15*)? Several mechanisms have been identified in the rodent hippocampus for representing spatial locations extending beyond the immediate position. Theta-phase and theta-sequence coding have been shown to indicate the animal's past and future positions as well as possible future paths (*16*–*24*). Rate modulations based on past or future events have also been observed in a subset of place cells (*25*–*27*). However, neurons providing nonlocal positional information are still primarily active when the animal is within the cell's firing field. Intriguingly, a subset of place cells has been found to be most informative when the neural spiking activity is assigned to the animal's position a short period of time (~120 ms) into the future (*28*), suggesting the existence of an additional, nonlocal neural mechanism based on rate-shifted coding. However, because of the movement speed of rodents in small enclosures, an ~120-ms shift corresponds to a position directly under the rat's nose (*29*), therefore still representing relatively local position. Here, we investigated rate-shifted coding mechanisms in flying bats, which provide an advantageous model system for studying nonlocal positional coding because of their high-speed (*30*), ballistic motion through space, which could facilitate identifying neural representations of locations that are spatially distant but still temporally close.

Typically, place coding is thought to be carried out by place cells that encode the immediate spatial position (Fig. 1A, top). We posited that the neural code for 3D spatial navigation in a fast-flying animal may include additional neural populations encoding places that the animal does not currently occupy (nonlocal coding) but that are within the flight path (Fig. 1A, bottom). In this hypothetical spatiotemporal code, a population of neurons is simultaneously active at each spatial position, but only some optimally encode the immediate position, whereas others optimally encode past or future positions that are meters away, leading to a representation of the full flight path in any given moment (Fig. 1A, bottom). This hypothesis predicts that neurons representing distant positions along the path may not exhibit any spatial tuning at all when the neural activity is considered with respect to the bat's immediate position. For these neurons, the prediction is that spatially selective fields may only emerge when the appropriate lag is applied between the positional and neural information; in other words, their spatial selectivity will be divorced from the present and imbedded in the past or in the future.

We initially performed wireless neurophysiological recordings from the dorsal CA1 region of four Egyptian fruit bats exploring the full 3D volume of a large flight room (Fig. 1B). In this random exploration experiment, a total of 532 well-isolated single units were recorded, of which 304 were sufficiently active during flight and included in subsequent analyses (see the supplementary materials and methods). Bats were encouraged to fly throughout the room to ensure dense spatial coverage (Fig. 1B). The 3D position of the bats was tracked at millimeter spatial resolution (see the supplementary materials and methods).

We used a time-shifting procedure to assess the neuron's positional coding and to determine whether it is higher at past, present, or future spatial positions (*28*). We calculated the spatial information after time shifting the spiking activity along flight trajectories (−1000 to +1000 ms in 100-ms steps; see the supplementary materials and methods; fig. S1). In agreement with previous studies (*8*), 27% (81/304 neurons; $P < 0.05$, shuffle test with Bonferroni correction) of the neurons were spatially selective at zero lag between spatial position and spikes (τ_0) (Fig. 1C and fig. S2). Overall, 46% (140/304 neurons, $P < 0.05$, shuffle test with Bonferroni correction) of the neurons were spatially selective at one or more time shifts. For most hippocampal neurons, spatial information was highest at nonzero time shifts, with most being optimally informative about future positions (Fig. 1, D and E, and fig. S3). Figure 1D shows three examples of neurons with no significant spatial selectivity at zero lag (τ_0) and high spatial selectivity at a future lag. If the hippocampal code were primarily related to the present, then we would expect to see the maximum spatial information concentrated around τ_0 (Fig. 1E, red line, and fig. S4). Instead, we saw a continuum, with most neurons being maximally informative at future lags while maintaining a similar amount of spatial information across lags (Fig. 1F). For a large fraction of spatially selective neurons, the spatial information at τ_0 was negligible (42%; 59/140 selective neurons), indicating that these neurons would go undetected as being spatially selective using standard procedures, and are thus truly nonlocal. Further, the spatiotemporal firing fields were distributed throughout the entire room and maintained consistent volumes and interfield distances across lags (Fig. 1, G and H), with typically one (46%) or two (29%) fields per cell (89% had three or fewer fields). Thus, time-lagged firing fields appear to be equivalent to nonlagged fields in terms of their spatial information, spatial distribution, and total volume. Furthermore, we found that spatiotemporal fields were largely stable throughout the session (fig. S5), and the optimal lag time was not related to the mean speed of the bat (fig. S6). A spatial information analysis using distance lags that matched the temporal lags (based on the mean flight speed) produced similar results (fig. S7), and a cross-correlation analysis indicated that the timing relationship of correlated firing is not related to the optimal lag order (fig. S8). Last, nonlocal heading tuning was prominent as well (54%, 142/304), with a high degree of overlap with spatially informative neurons (76%, 107/140), although the distribution of optimal heading tuning was shifted toward the past (fig. S9).

[1]Department of Bioengineering, University of California, Berkeley, CA 94720, USA. [2]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: myartsev@berkeley.edu

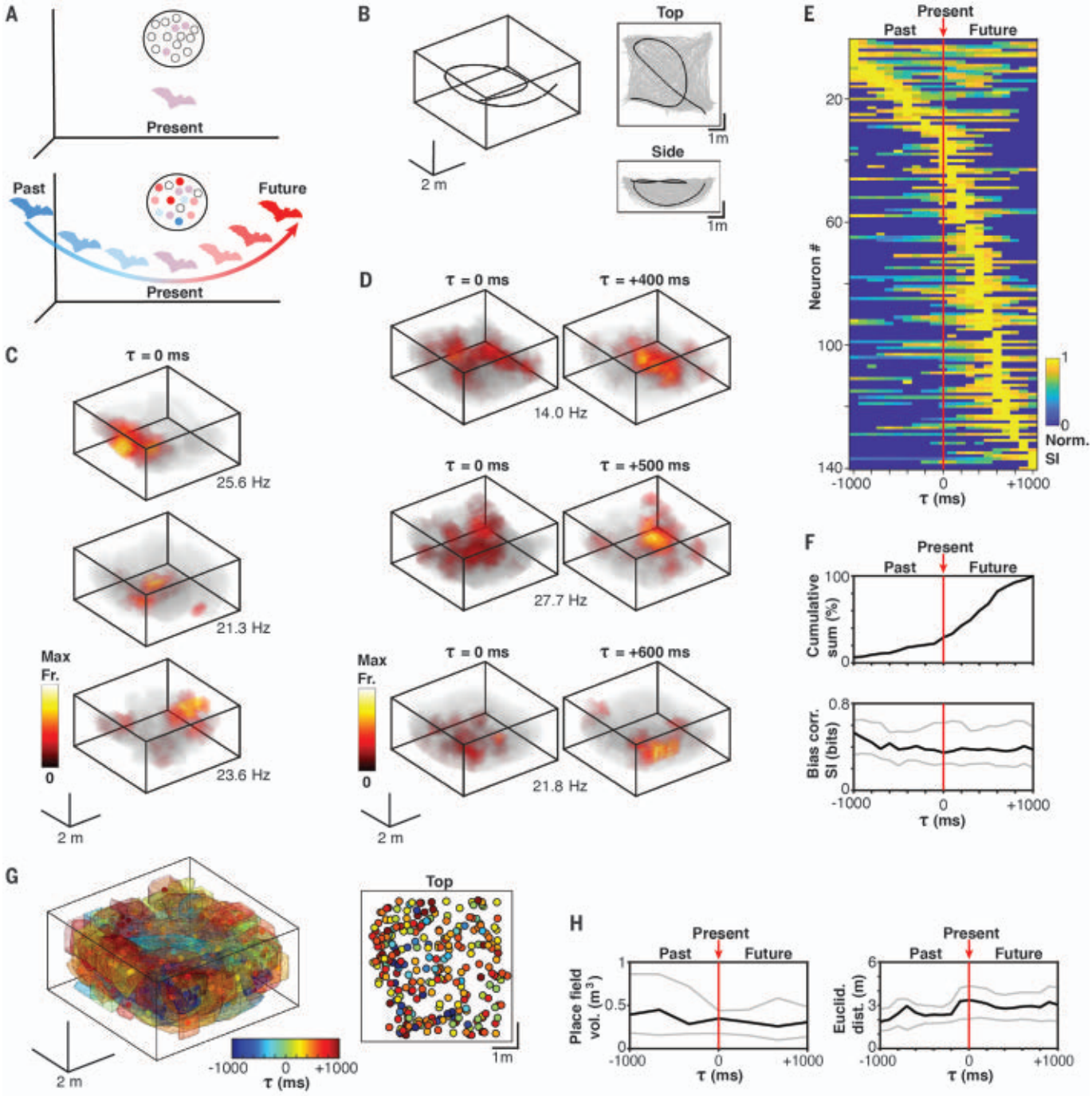


Fig. 1. CA1 neurons in flying bats encode nonlocal positions, primarily in the future. (**A**) Illustration of nonlocal coding hypothesis. Top: A subset of classical place cells (light purple circles) with place fields overlapping with the immediate spatial position (local). Bottom: Neurons with nonlocal fields encoding past (blue) and future (red) spatial positions. (**B**) Example of a single flight (left) and the spatial coverage (gray lines) shown from the top and the side for an entire recording session. Black trace corresponds to the example flight trajectory. (**C**) 3D color-coded firing rate maps for three examples of 3D place cells. Bias-corrected spatial information values for the three cells is as follows: top, 0.81 bits; middle, 0.61 bits; and bottom, 0.61 bits. The peak firing rate is indicated. (**D**) Same as (C) but for three example neurons exhibiting higher 3D spatial information for future positions. Spatial information values for the example neurons are as follows: top, lag zero = 0.10 bits [not significant (n.s.)] and optimal lag = 0.28 bits; middle: lag zero = 0.11 bits (n.s) and optimal lag = 0.30 bits; and bottom, lag zero = 0.18 bits (n.s.) and optimal lag = 0.58 bits. (**E**) Normalized spatial information across time lags for each neuron with significant spatial information at one or more time lags (–1000 to 1000 ms in 100-ms steps). Rank order is based on the time lag with maximal spatial information. (**F**) Cumulative sum of the time lags with maximal spatial information for each neuron (top) and summary of the bias-corrected spatial information values at each lag (one value per neuron) (bottom). The 25th (bottom gray line), 50th (middle black line), and 75th (top gray line) percentiles are shown. (**G**) Left: Convex hull representation of all significant spatiotemporal field locations across all recorded neurons. Right: top-down view showing location of spatiotemporal field centers. Color code indicates the time lag of maximal spatial information. (**H**) Field volume across lags (left) and Euclidean distance between all field centers at each lag (right). The 25th (bottom gray line), 50th (middle black line), and 75th (top gray line) percentiles are shown.

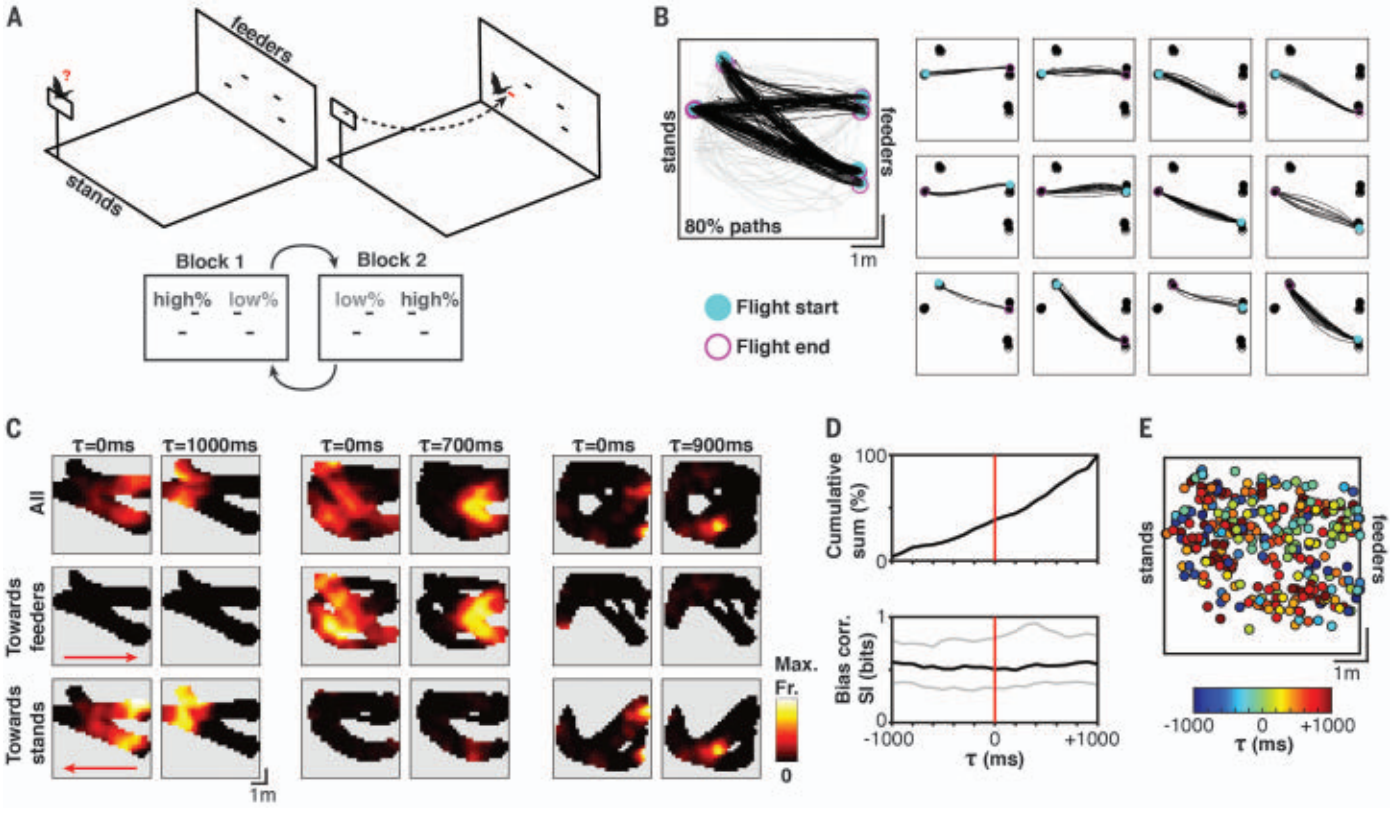


Fig. 2. Nonlocal spatiotemporal coding is present during goal-directed navigation. (**A**) Illustration of the foraging task (not drawn to scale; room dimensions: 5.6 m × 5.2 m × 2.5 m). Bats choose between four different feeders. Feeders fed at predetermined probabilities that switched during the session (see the supplementary materials and methods). (**B**) Left: all the flights on repeatable paths (black lines) are shown for a single recording session. Non-path flights are shown in gray. Right: break out of individual flight paths. Each subplot represents a distinct flight path (see the supplementary materials and methods). (**C**) Rate maps for three examples neurons with maximum spatial information at future time shifts. Top: all flight data. Middle: parts of flights with heading angles toward the feeders. Bottom: parts of flights with heading angles toward the stands. Maximum firing rate and bias-corrected spatial information values are as follows: left, max fr. = 30.98 Hz, lag zero = 0.65 bits, optimal lag = 1.08 bits; middle, max fr. = 56.11 Hz, lag zero = 0.49 bits, optimal lag = 0.93 bits; and right, max fr. = 9.76 Hz, lag zero = 0 bits (n.s.), optimal lag = 0.37 bits. (**D**) Cumulative sum of peak spatial information (top) and bias-corrected spatial information across lags. The 25th (bottom gray line), 50th (middle black line), and 75th (top gray line) percentiles are shown. (**E**) Top-down view showing the locations of spatiotemporal field centers. Color code indicates time lag of maximal spatial information.

What role can a representation of space that is divorced from the present moment play in a navigating animal? Egyptian fruit bats navigate landscapes spanning hundreds of kilometers to forage for food, flying in highly reproducible paths to and from foraging sites (*1*). Members of this species also make novel shortcuts between distant foraging sites (*3, 4*). This suggests that during goal-directed navigation, these bats maintain a detailed spatial memory of both their environment and of paths taken within it. To understand how this nonlocal code may be used for path-based, goal-directed navigation, we engaged animals in a freely paced foraging task in an automated environment [Fig. 2A; (*31*); see the supplementary materials and methods]. Four reward feeders positioned along one wall of the room fed at different probabilities, with one side feeding at higher probabilities than the other (e.g., 70 and 30%). The probabilities switched after a set number of feeds to induce exploration to all feeders and to keep the animals attentive to the task (see the supplementary materials and methods).

We recorded neural activity from three bats engaged in the goal-directed foraging task. One bat performed the task alone and the other two performed the task at the same time (204 of 281 total cells were sufficiently active during behavior; see the supplementary materials and methods). There were no differences in the results between the individual bat and the pair, nor were there any indications of social representations of the other bat under the conditions of this experiment (fig. S10). We therefore combined the data from all three bats in subsequent analyses. Bats formed common paths, and typically only visited the feeders and a few places in the room to rest between feedings (Fig. 2B). Individual bats developed distinct movement patterns (fig. S11), underscoring the self-selected nature and reproducibility of their flight paths during goal-directed navigation (fig. S11B).

To determine whether nonlocal positions are encoded during foraging, we performed the same time-shifting analysis as in the random exploration experiments, focusing on the flights occurring along reproducible paths. Because flight paths were constrained to narrow portions of the room, leading to a heterogeneous sampling of the z axis with respect to the x and y axes, we constrained the analysis to 2D, as was done previously [(*32, 33*); similar results were found in 3D; fig. S12]. Similar to the random exploration experiment (Fig. 1), neural activity predominantly encoded nonlocal positions, was again shifted toward the future, and maintained a consistent level of spatial information for all temporal lags (Fig. 2, C and D, and fig. S12). Results were robust to multiple shuffle tests designed to account for the flight pattern structure (fig. S13; see the supplementary materials and methods). We found a very high percentage of neurons with significant spatial information (90%, 183/204

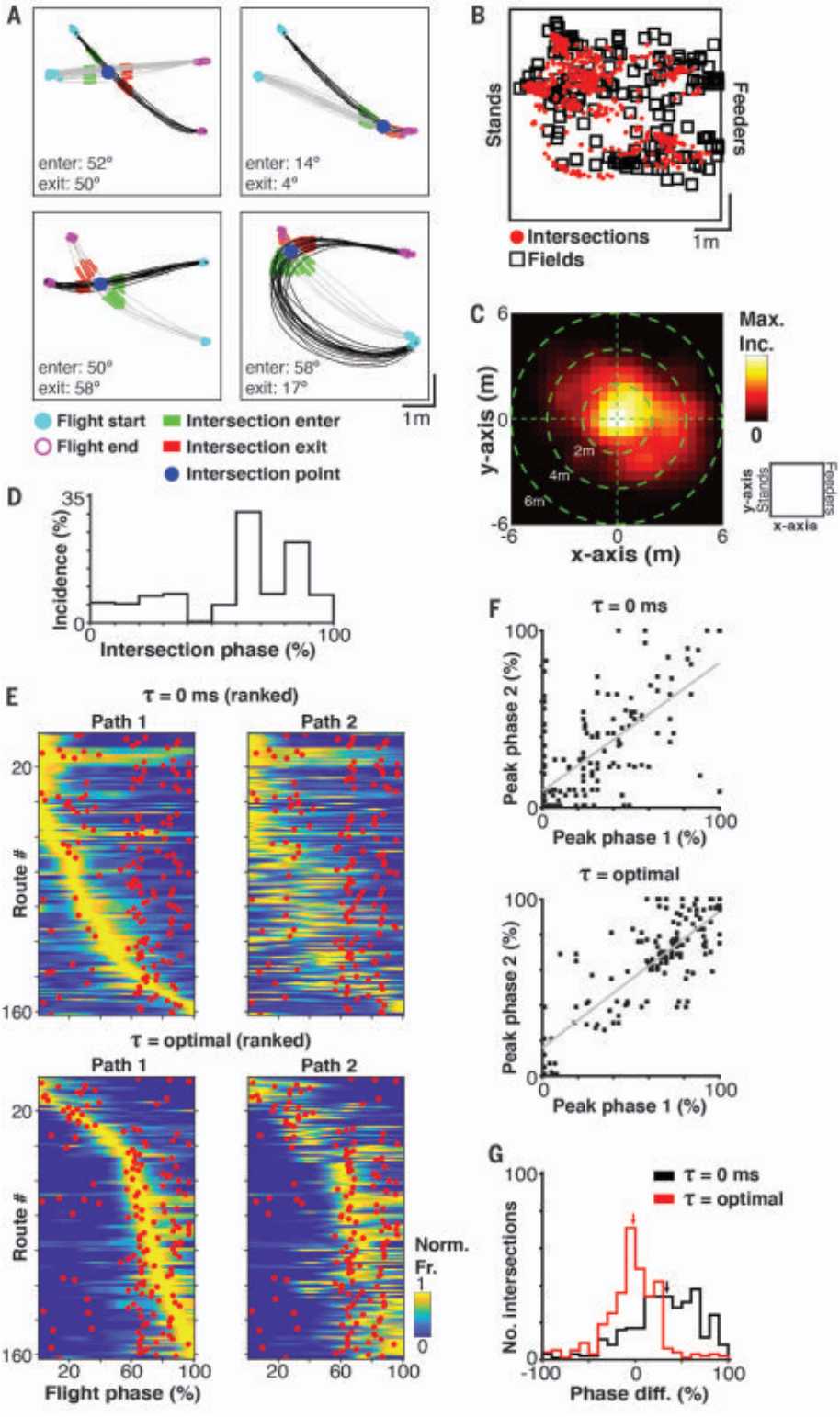


Fig. 3. Flight path intersections are aligned with nonlocal spatiotemporal fields. (**A**) Examples of intersecting flight paths. Each panel shows the flights for two paths (one path in gray and one in black). Blue dot indicates the intersection point. Green (before intersection) and red (after intersection) highlights indicate the portion of the flights used to calculate the intersection angles. Intersection angles (enter and exit) are indicated for each panel. (**B**) Locations of spatiotemporal fields (black squares; largest field is shown if multiple fields exist per neuron) and intersections (red dots). Note the tight correspondence between the location of firing fields and intersections. (**C**) 2D heat map of the differences in *x* axis (stand-feeder axis) and *y* axis positions between spatiotemporal fields and intersection locations (normalized to maximum incidence). The schematic on the right indicates the *x* axis and *y* axis reference frame of the room. (**D**) Distribution of the intersection phases. For each intersection, there are two phases, one for each path. (**E**) Firing rate tuning profiles for each pair of paths at lag zero (top) and the optimal lag (bottom). Rank is ordered by the peak firing rates in path #1. Red dots indicate the intersection phase. (**F**) Scatter plots for the peak phase of path 1 and path 2 at lag zero (top; $r = 0.69$, $P < 0.01$) and the optimal lag (bottom; $r = 0.81$, $P < 0.01$). Least-squares fit lines are shown in gray. (**G**) Histograms showing the difference between the peak firing rate phase and the phase of the intersection at lag zero and the optimal lag. Small black and red arrows indicate the median difference for lag zero and the optimal lag, respectively (lag zero median = 34.03%, optimal lag median = −2.27%).

neurons) compared with the random exploration experiment, which may be attributed to the increased attentional demand and highly structured behavioral patterns exhibited during the task (*34*). This high degree of nonlocal spatiotemporal information was exemplified by the ability to decode past, present, and future positions on individual flight trajectories, with higher performance for decoders using the optimal time lag of each neuron (fig. S14; see the supplementary materials and methods). Furthermore, neurons appeared directionally tuned to either the feeders or the stand directions, suggesting that the task environment had a polarizing effect on the neural representations (Fig. 2C). Indeed, 85% (155/183) of spatially informative neurons and 81% (165/204) of total neurons were significantly tuned for heading at one or more time lags (see the supplementary materials and methods). The resulting spatiotemporal fields were distributed

throughout the portion of the room used by the bats during foraging (Fig. 2E), with typically one (54%) or two (24%) fields per cell (91% had three or fewer fields).

The results presented thus far suggest that CA1 neurons encode nonlocal navigational information during flight but they do not indicate what might determine the locations of the spatiotemporal fields. We reasoned that spatiotemporal field organization might be linked to the self-organized behavioral structure that bats exhibit during goal-directed navigation. Because we found that neurons were tuned to flight paths headed in the same direction (Fig. 2C), we posited that neurons might also be tuned to other spatial commonalities between paths, specifically the places where paths intersect. To address this hypothesis, we first determined whether there was any overlap between path intersections and the spatiotemporal field locations during the goal-directed foraging task (Fig. 2). We identified the intersections between all pairs of flight paths (Fig. 3A and fig. S15; 10.8 mean ± 8.6 SD path intersections per neuron, for a total of 1931 intersections). We only considered intersections where paths were headed in the same direction (enter and exit angles < |90°|) and only intersected once (see the supplementary materials and methods). Intersections based on this criterion were found for 97% (178/183) of neurons with spatiotemporal fields. We found a high degree of overlap between the nonlocal spatiotemporal fields and intersection locations (Fig. 3B), supporting the hypothesis that temporally shifted firing fields were centered at locations where flight paths intersected. We computed the distance between intersection locations and spatiotemporal field locations and found that 72% (129/178) of the neurons had an intersection within ≤2 m from the neuron's spatiotemporal field. Moreover, 50% (973/1931) of all the intersections were <2 m from their respective neuron's spatiotemporal field (Fig. 3C; less dispersed than would be expected by chance, $P < 0.01$, randomization test; see the supplementary materials and methods). On the basis of the strong directional orientation, it is apparent that neurons are only active along a subset of paths, which could explain why a portion of intersections are far from spatiotemporal fields. Thus, there is a strong correlation between the location of a large subset of intersections and spatiotemporal fields.

Next, we sought to determine whether spatiotemporal fields could result from the patterns of neural activity along intersecting paths. If the neural activity on each path were temporally advanced or delayed with respect to the intersection, then this could result in a spatiotemporal field at the intersection. For each neuron, we identified pairs of paths with significant spatial information (see the supplementary materials and methods) and an intersection <2 m from the spatiotemporal field of the neuron. We identified 162 path pairs from 29% (53/183) of the spatially informative neurons. We then investigated whether the peak firing rates are aligned to path intersection points when the optimal lag is taken into consideration. We first scaled all the flights from 0 to 100% (flight phase) and then identified the phase of intersections (Fig. 3D) and the peak firing rate phases at lag zero and at the optimal lag (optimal lag is the lag with maximal spatial information for that neuron). Figure 3E shows the intersection phases (red dots) overlaid on the firing rate profiles for each pair of paths at lag zero (top) and the optimal lag (bottom). The peak firing rates for both paths, even though most have little overlap in space, became aligned with the intersection phase at the optimal lag. Furthermore, there was a strong positive correlation between the peak firing rate phases for all pairs of intersecting paths, which increased when accounting for the optimal lag (Fig. 3F). Indeed, simply accounting for the optimal lag substantially shifted the distribution of differences between the peak firing rate phase and the intersection phase toward zero (Fig. 3G; $P <$ 0.01 rank sum test). These results demonstrate a precise alignment between self-organized navigation patterns and neural dynamics, suggesting that spatiotemporal firing fields may form around intersections.

When we recorded from bats flying freely at high speeds during either random exploration or in a goal-directed manner, we found that the neural activity in bat CA1 robustly encodes nonlocal navigational information. Classical place tuning appears to be simply part of a larger continuum representing the bat's past, present, and future locations. Much of this information would go undetected if past or future positions were not taken into consideration. These findings also complement reports demonstrating that cells with no identified firing fields, i.e., non–place cells, can contribute to the neural code for space at the population level (*35*, *36*), albeit in a different manner. Last, these findings, although functionally similar, are mechanistically distinct from theta-phase and theta-sequence coding and reveal another complementary mechanism by which positional information that extends beyond the animal's present position can be represented in the hippocampal formation. The diversity of temporal and rate-based coding schemes by which nonlocal positional information can be represented raises important questions for future studies to consider about how simultaneous representations of past, present, and future positions in the hippocampus can be read out effectively by downstream regions.

Our results further indicate that nonlocal representations are anchored near both future and past intersections of self-selected flight paths, which can occur at all phases of flight. This differs from a vectorial tuning to a specific endpoint goal location (*32*), yet the two could serve as important complements to each other during navigation. These results are also consistent with studies showing evidence for path-invariant representations of spatial positions (*29*) such as intersections and goal locations. Previous investigations into rate-shifted coding mechanisms in rats showed that neurons optimally encoded a position immediately in front of the animal (*28*, *29*). This could have been because of the slower speeds of movement exhibited by rodents in small experimental environments compared with the high speeds of bat flight, which necessitates planning and rapidly predicting positions far into the future. Combined, these results reveal a positional representation in flying bats that extends along a continuum of space and time and could support a representation of remembered paths.

Spatial coding has been observed in the hippocampus (or analogous structures) across a wide variety of species that have evolved to navigate in very different environments, whether underwater, on the ground, or in the sky (*5*, *8*, *37*–*39*). The way in which a given species negotiates its environment may necessitate weighing the relevance of past and future positions differently. For example, monkeys jumping between tree branches or humans driving a car or skiing downhill at high speeds may require a higher weight placed on future locations where the positional information ahead could be more important for survival than the present position. This notion is consistent with more recent theories highlighting the function of the hippocampus as a flexible predictive map (*40*). Furthermore, different species may also weigh the importance of specific locations in their environment differently. An Egyptian fruit bat, which naturally forms highly reproducible paths (*1*) along which it traverses repeatedly at high speeds, may benefit from a nonlocal representation of specific locations along flight paths (e.g., intersection points), yet this may not be the case for an animal exploring the environment in a different manner. Our results further highlight the importance of studying neural circuits during spontaneously emerging behavioral patterns across a diversity of species. Examining neural activity from an ethological perspective that carefully considers how a specific animal moves and operates could help to better inform us of the underlying neural computations that generate behavior (*41*, *42*).

REFERENCES AND NOTES

1. A. Tsoar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, E718–E724 (2011).
2. M. Geva-Sagiv, L. Las, Y. Yovel, N. Ulanovsky, *Nat. Rev. Neurosci.* **16**, 94–108 (2015).
3. L. Harten, A. Katz, A. Goldshtein, M. Handel, Y. Yovel, *Science* **369**, 194–197 (2020).
4. S. Toledo *et al.*, *Science* **369**, 188–193 (2020).

5. J. O'Keefe, J. Dostrovsky, *Brain Res.* **34**, 171–175 (1971).
6. E. I. Moser, M. B. Moser, B. L. McNaughton, *Nat. Neurosci.* **20**, 1448–1464 (2017).
7. R. M. Grieves *et al*., *Nat. Commun.* **11**, 789 (2020).
8. M. M. Yartsev, N. Ulanovsky, *Science* **340**, 367–372 (2013).
9. M. Geva-Sagiv, S. Romani, L. Las, N. Ulanovsky, *Nat. Neurosci.* **19**, 952–958 (2016).
10. M. J. Wohlgemuth, C. Yu, C. F. Moss, *Front. Cell. Neurosci.* **12**, 270 (2018).
11. Nonlocal positional coding has also been observed in awake, stationary animals before and after spatial exploration and at decision points (*12*–*15*).
12. D. J. Foster, M. A. Wilson, *Nature* **440**, 680–683 (2006).
13. K. Diba, G. Buzsáki, *Nat. Neurosci.* **10**, 1241–1242 (2007).
14. A. Johnson, A. D. Redish, *J. Neurosci.* **27**, 12176–12189 (2007).
15. A. D. Redish, *Nat. Rev. Neurosci.* **17**, 147–159 (2016).
16. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317–330 (1993).
17. W. E. Skaggs, B. L. McNaughton, M. A. Wilson, C. A. Barnes, *Hippocampus* **6**, 149–172 (1996).
18. G. Buzsáki, *Neuron* **33**, 325–340 (2002).
19. G. Buzsáki, *Hippocampus* **15**, 827–840 (2005).
20. R. L. Buckner, *Annu. Rev. Psychol.* **61**, 27–48, C1–C8 (2010).
21. A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, A. D. Redish, *Nat. Neurosci.* **15**, 1032–1039 (2012).
22. T. Eliav *et al*., *Cell* **175**, 1119–1130.e15 (2018).
23. M. Wang, D. J. Foster, B. E. Pfeiffer, *Science* **370**, 247–250 (2020).
24. K. Kay *et al*., *Cell* **180**, 552–567.e25 (2020).
25. E. R. Wood, P. A. Dudchenko, R. J. Robitsek, H. Eichenbaum, *Neuron* **27**, 623–633 (2000).
26. L. M. Frank, E. N. Brown, M. Wilson, *Neuron* **27**, 169–178 (2000).
27. J. Ferbinteanu, M. L. Shapiro, *Neuron* **40**, 1227–1239 (2003).
28. R. U. Muller, J. L. Kubie, *J. Neurosci.* **9**, 4101–4110 (1989).
29. F. P. Battaglia, G. R. Sutherland, B. L. McNaughton, *J. Neurosci.* **24**, 4541–4550 (2004).
30. The average flight speed of a bat is ~3 m/s, with peak speeds up to ~6 m/s under laboratory conditions.
31. D. Genzel, M. M. Yartsev, *J. Neurosci. Methods* **348**, 108970 (2021).
32. A. Sarel, A. Finkelstein, L. Las, N. Ulanovsky, *Science* **355**, 176–180 (2017).
33. D. B. Omer, S. R. Maimon, L. Las, N. Ulanovsky, *Science* **359**, 218–224 (2018).
34. C. G. Kentros, N. T. Agnihotri, S. Streater, R. D. Hawkins, E. R. Kandel, *Neuron* **42**, 283–295 (2004).
35. L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, W. Bialek, *Neuron* **96**, 1178–1191.e4 (2017).
36. F. Stefanini *et al*., *Neuron* **107**, 703–716.e4 (2020).
37. G. E. Hough, V. P. Bingman, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **190**, 1047–1062 (2004).
38. H. S. Courellis *et al*., *PLOS Biol.* **17**, e3000546 (2019).
39. E. Vinepinsky *et al*., *Sci. Rep.* **10**, 14762 (2020).
40. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, *Nat. Neurosci.* **20**, 1643–1653 (2017).
41. M. M. Yartsev, *Science* **358**, 466–469 (2017).
42. S. R. Datta, D. J. Anderson, K. Branson, P. Perona, A. Leifer, *Neuron* **104**, 11–24 (2019).

ACKNOWLEDGMENTS

We thank O. Tchernichovski for extensive discussion of the results, comments on the manuscript, and help with analysis; D. Foster for discussion of the results and manuscript; J. Widloski, K. Kay, and D. Foster for helpful discussions on a revised version of the manuscript; L. Kang, A. Forli, W. Liberti, T. Schmid, B. Styr, and M. Snyder for careful reading and comments on the manuscript; members of the Yartsev laboratory for discussion and comments; D. Genzel for assistance with the design of the experimental task and training protocols; Y. Minton and L. Loomis for experimental room maintenance and animal care; and G. Lawson and the staff of the Office of Laboratory Animal Care for support with animal husbandry and care. **Funding:** This work was supported by the New York Stem Cell Foundation (NYSCF-RNI40), the Air Force Office of Scientific Research (FA9550-17-1-0412), the Office of Naval Research (N00014-21-1-2063), the Packard Fellowship (2017-66825), the National Institute of Neurological Disorders and Stroke (R01NS118422-01), the Valle Foundation (VS-2020-34), and the Searle Scholars Program (SSP-2016-1412) to M.M.Y. **Author contributions:** N.M.D. and M.M.Y. designed the research, analyzed the data, and wrote the manuscript. N.M.D. performed the experiments. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6551/242/suppl/DC1
Materials and Methods
Figs. S1 to S15
References (*43*–*54*)
MDAR Reproducibility Checklist

11 December 2020; accepted 1 June 2021
10.1126/science.abg1278

YOU'D SMILE TOO IF YOU JUST ADVANCED YOUR CAREER.

Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets

THE UNIVERSITY OF CHICAGO

Assistant Instructional Professor in General Biology

Description - The Biological Sciences Collegiate Division at the University of Chicago is accepting applications for one or more positions as Assistant Instructional Professor (AIP) in general biology. This is a full-time teaching position with a start date of September 1, 2021. Appointment will be made at the rank of Assistant Instructional Professor for an initial term of two years with reappointment and progression possible following review.

Responsibilities include teaching six lecture and/or laboratory courses for biology majors and non-majors. Expected course topics include introductory cell biology, genetics, developmental biology, and molecular biology for biology majors and general introductory biology for non-majors. Additional duties include hiring, training, and supervising teaching assistants. Instructional Professors of all ranks are required to engage in regular professional development.

Qualifications - Applicants must have a PhD in the Biological Sciences in hand prior to start date and at least one year of teaching university level biology lecture and laboratory courses.

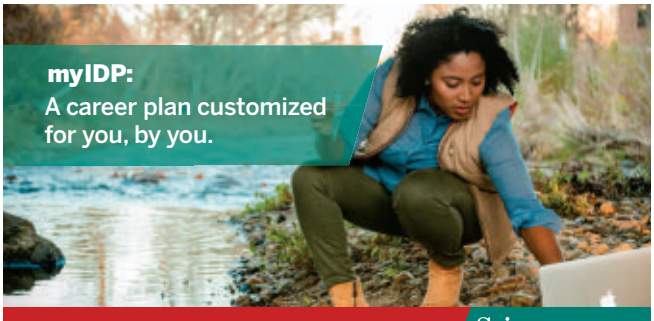
Application Instructions - Applicants must apply online at the University of Chicago's Interfolio website at **apply.interfolio.com/89633**. An application must include: 1) a cover letter; 2) a curriculum vitae; 3) a sample course syllabus; 4) course evaluations or evidence of past teaching performance; 5) the names and contact information for three references. Review of applications will begin on **July 28, 2021** and will continue until the position is filled or the search is closed.

This position will be part of the Service Employees International Union.

We seek a diverse pool of applicants who wish to join an academic community that places the highest value on rigorous inquiry and encourages diverse perspectives, experiences, groups of individuals, and ideas to inform and stimulate intellectual challenge, engagement, and exchange. The University's Statements on Diversity are at https://provost.uchicago.edu/statements-diversity.

The University of Chicago is an Affirmative Action/Equal Opportunity/Disabled/Veterans Employer and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender identity, national or ethnic origin, age, status as an individual with a disability, protected veteran status, genetic information, or other protected classes under the law. For additional information please see the University's Notice of Nondiscrimination.

Job seekers in need of a reasonable accommodation to complete the application process should call 773-702-1032 or email equalopportunity@uchicago.edu with their request.



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
University of Massachusetts Medical School UMASS
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

Universität Konstanz

The University of Konstanz is one of eleven Universities of Excellence in Germany. Since 2007 it has been successful in the German Excellence Initiative and its follow-up programme, the Excellence Strategy.

The *Zukunftskolleg* is an Institute for Advanced Study for early career researchers in the natural sciences, humanities and social sciences. It is an central component of the university's strategy „creative.together". The *Zukunftskolleg* promotes interdisciplinary discourse between distinguished researchers from Germany and abroad, and provides them with the resources to conduct their research independently and free from administrative constraints. The working language is English.

The *Zukunftskolleg* of the University of Konstanz offers:

Three 5-year Research Fellowships at the Zukunftskolleg

(Salary Scale TV-L 14)

to develop and implement individual research projects and build up their own research group.

Fellowships will begin on May 1, 2022.

Reference number 2021/152

Three 2-year Postdoctoral Fellowships at the Zukunftskolleg

(Salary Scale TV-L 13)

to develop and implement individual research projects. Conditionally on the submission of an external grant, the position can be extended for an additional year.

Fellowships will begin on April 1, 2022.

Reference number 2021/153

Zukunftskolleg

Applications, supporting materials and two letters of reference should be submitted in English **by 15 October 2021 11:00 am (Konstanz time) using the Online Application Platform** https://www.uni-konstanz.de/zukunftskolleg/online-application/.

Details concerning the application, required documents, and information about the Zukunftskolleg are available on our website: https://www.uni-konstanz.de/zukunftskolleg/ .

By **Manya Ruckhaus**

Standing my ground

When I saw the black bear emerge from the trees, I knew I shouldn't turn around and flee. So I held my ground in spite of my fear. It was a moment I'd been dreading for months. I'd been hired as an undergraduate field assistant to do geology work in an area with healthy populations of bears and cougars. I went into the summer not wanting to be alone in the field, fearing I'd encounter a potentially deadly animal. But bears and cougars turned out to be the least of my problems. I spent 6 weeks working alongside a male colleague who constantly belittled my abilities, leaving me mentally exhausted and questioning whether I belonged in the field. My encounter with the bear, in contrast, turned out to be empowering.

I grew up in the Rocky Mountains of Colorado and spent much of my childhood hiking, skiing, and otherwise exploring outdoors. I was fascinated by the mountains, streams, and canyons, so I decided to study geology at college. I loved my program and was excited when I had the opportunity to participate in summer field research. One summer, I spent nearly 3 months hiking every day and living in a rustic cabin. It was a great experience. My other summer in the field, however, was not.

I was hired to work alongside my supervisor and one other undergraduate field assistant. Before we went into the field, I overheard the student telling others I was a slow hiker and that he hoped I wouldn't show up. It wasn't the first time he had disparaged my physical or intellectual abilities. Our supervisor didn't know our history before hiring us for the project. I desperately wanted to tell him about it, but I needed the research experience and I did not want him to second-guess hiring me.

Once we were in the field, the other student never missed an opportunity to play the game of one-upmanship. He argued with me constantly. No matter what I said, he voiced an opposite position. The more it happened, the quieter I became. After a few weeks of misery, I could see that my supervisor was starting to have doubts about me, mistaking my silence for incompetence. Sensing his disappointment in me, I began to believe I wasn't cut out to be a scientist.

Although I felt deeply uncomfortable with our team dynamics, I didn't want to work alone. We were in a remote area with rugged terrain, and I worried one of us might get injured and need help. I was also uneasy about encountering the locals alone. One man we had crossed paths with earlier in the summer had muttered something about seeing a woman and joked that he thought my rock hammer was a weapon.

"I realized that if a wild animal feared me, I wasn't powerless after all."

But when our supervisor left a few weeks early, the other student suggested we split up to cover more ground. I protested, telling him I was uncomfortable working alone. But he countered that I was paranoid. I felt I had no choice but to agree.

A week later, I spotted the bear. At first I was terrified. But when it quickly ran away, simply because I was standing there, my feelings started to change. I realized that if a wild animal feared me, I wasn't powerless after all. I went back to camp with the confidence I needed to stand up to the real threat I faced that summer: the other student.

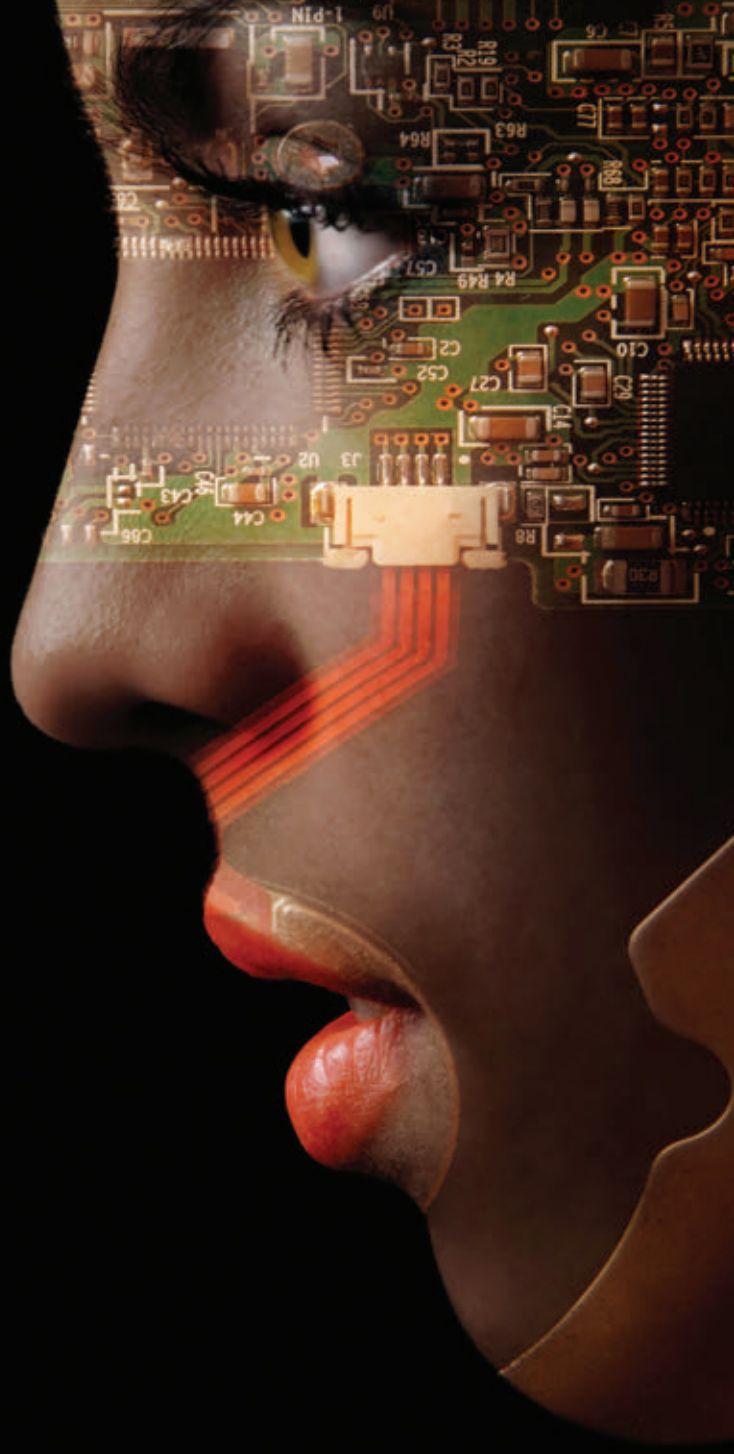
That evening, after he criticized the way I was setting up the camp table for dinner, I threw the table legs down and told him how disgusted I was with how he'd treated me. He didn't apologize, but I felt better after getting it off my chest. I realized that in the future I need to address problems head-on rather than internalizing them and letting them affect my self-confidence.

I am now in grad school, thanks in part to a supportive female mentor. I haven't faced any other dangerous beasts, but I assume I will someday. And the next time I do confront a bear, cougar, or menacing colleague, I won't turn around and run. ■

Manya Ruckhaus is a master's student at the University of Vermont. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Share **Your Robotics Research** with the **World**.



Shaping the future of robotics with high impact research!

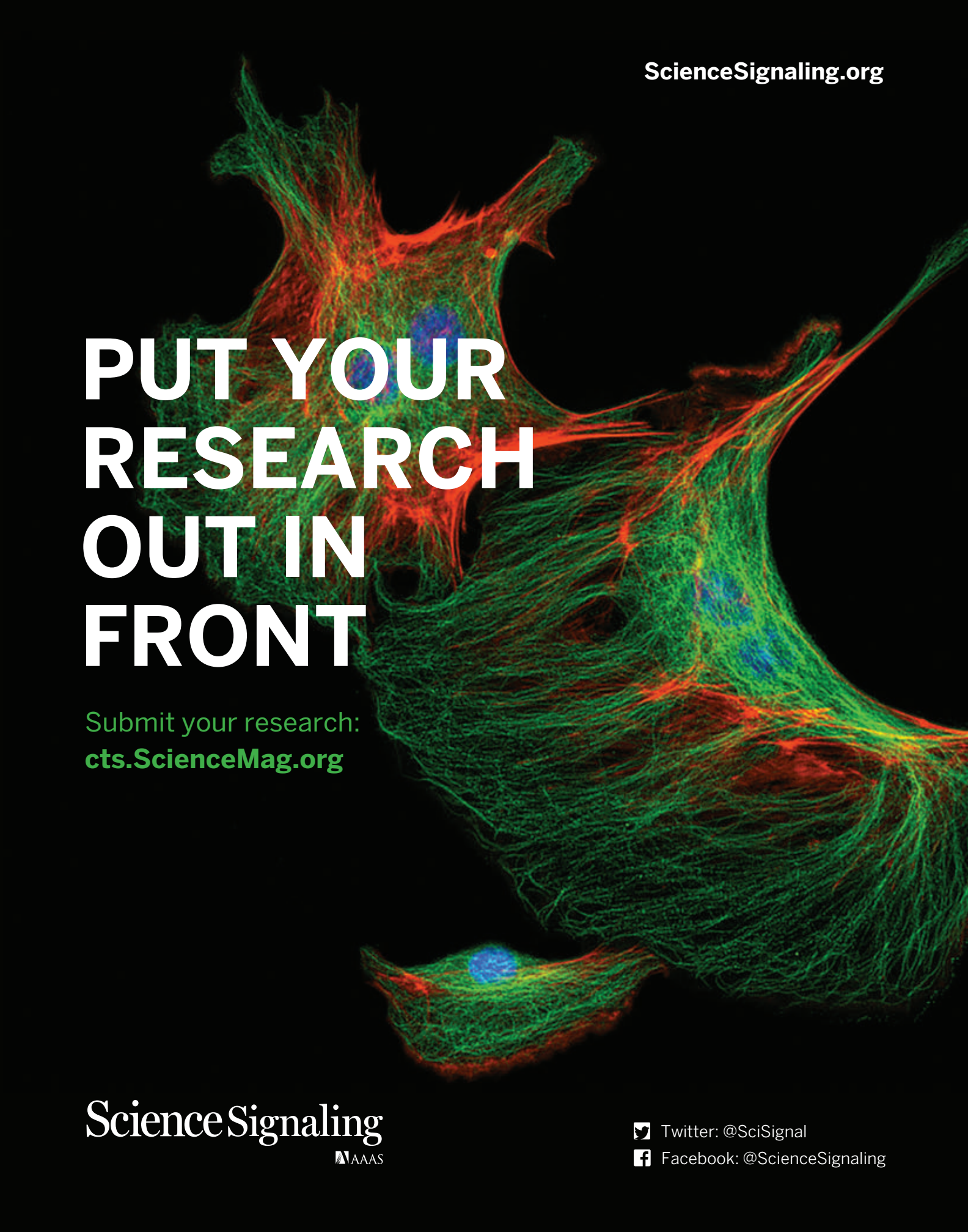
As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

Submit your research today. Learn more at: **robotics.sciencemag.org**

Twitter: @SciRobotics
Facebook: @ScienceRobotics

Science Robotics
AAAS

ScienceSignaling.org



PUT YOUR RESEARCH OUT IN FRONT

Submit your research:
cts.ScienceMag.org

Science Signaling
AAAS

Twitter: @SciSignal
Facebook: @ScienceSignaling