Science

MAAAS

TARGETED T CELLS

Lipid nanoparticles deliver mRNA in vivo pp. 23 & 91

CONTENTS

20

CONTENTS

RESEARCH

ON THE COVER

Lipid nanoparticles (spheres) use antibodies against the CD5 protein (antibodies, blue; CD5, red) to specifically target T cells while inside the bloodstream and tissues. The nanoparticles deliver mRNA to T cells, reprogramming them in vivo to transiently express chimeric antigen receptors (gold) that target pathogenic cardiac fibroblasts. In vivo–engineered T cells mitigate cardiac fibrosis and improve heart function in a mouse model. See pages 23 and 91. *Illustration: C. Bickel/*Science

Diverse neuronal types form functional circuits in the rodent primary somatosensory cortex

ILLUSTRATION: NICOLLE FULLER AT SAYO STUDIO

Looking ahead, looking back

The first editorial of the year is generally an invitation to look forward. Often, we announce new initiatives at the *Science* family of journals or changes to our policies. This year, I want to look forward in a different way—by looking back. *Science* has a history that includes shame in addition to accomplishment. In 2021, we began to explore and acknowledge some of that regretful past, and we'll continue this examination in 2022.

In May of last year, *Science* recognized the 150th anniversary of Charles Darwin's *Descent of Man* with an outstanding review on the science and an editorial by Agustín Fuentes, who called attention to harmful views about race and gender that Darwin expressed. Some critics said we were trying to "cancel Darwin" or that we should have understood that he was a "man of his time." We reject both criticisms. We were simply providing a more complete description of one of the most important figures in science and noting the harm that his views produced for so many women and people of color.

In September, Sudip Parikh [the chief executive office of the American Association for the Advancement of Science (AAAS), the publisher of *Science*], John Slattery (a theologian and historian at AAAS), and I wrote a blog post about the role that the journal and AAAS played in promoting eugenics. In the 1920s, *Science* published articles by well-known American eugenicists, including Charles Davenport, Henry Osborn, and Leonard Darwin (son of Charles Darwin). Osborn went on to become the president of the AAAS. It is distressful to read their repugnant ideas in our pages; the harm done by the practice of eugenics is incalculable.

A year ago, David Christianson, a professor of chemistry at the University of Pennsylvania, wrote to me about a horrific event chronicled in *Science*. In the 1960s, researchers at Harvard Medical School, Massachusetts General Hospital, and Boston University School of Medicine carried out experiments designed to demonstrate the physiological effects of nuclear fallout. The test subjects were developmentally disabled children at Wrentham State School (now called the Wrentham Developmental Center) in Massachusetts—a state-run medical facility for the treatment of psychiatric and developmental disorders. Disabled children were given sodium iodide to determine the amount needed to suppress the uptake of radioactive iodine-131. The study was published in *Science* in 1962, and the children were demeaningly described as "mentally defective." Like Darwin's views and those of the eugenicists, the study reflects a willfulness to consider some individuals as less than human.

The study hit close to home for Christianson. His twin sister, Karen, was a patient at Wrentham State School at that time. She died shortly after she was admitted and after the experiments were done. Christianson says he will never know whether his sister was in the experiment or died because of it, because the experiments were apparently conducted without parental consent. The uncertainty only adds to the harm inflicted by such studies on marginalized communities. I contacted the provost of Harvard University to see if the university had uncovered any new information about this incident, but he was unable to find anything.

I am asked frequently what *Science* plans to do about other studies it has published that were discriminatory and even violated human rights. Is the best approach to look over all papers published since 1880 through today's lenses of systemic racism, sexism, and other prejudices, and then retract or put a notice on those that fall under this category? The practical aspects of such an effort and determining whether that is the best course forward will require more listening. But either way, it's time for honest discussions about such studies and shining a spotlight on the shameful views and actions of the scientific community's predecessors. *Science* is not afraid to point out its role in supporting malicious science—it is history that should not be forgotten and can guide us in working with the community to confront shortcomings, past and present, in our pages and across the scientific enterprise.

Despite the pain surrounding this topic, I believe there is reason to be hopeful. We owe it to the Christiansons and others to ensure a future that eliminates atrocities in the practice of science.

—H. Holden Thorp

> "...it's time for honest discussions about... the shameful views and actions of the scientific community's predecessors."

H. Holden Thorp
Editor-in-Chief, *Science* journals. hthorp@aaas.org; @hholdenthorp

NEWS

IN BRIEF | *Edited by* **Jeffrey Brainard**

A final view of the folded-up James Webb Space Telescope was snapped by the launch rocket's upper stage.

ASTROPHYSICS

Webb telescope embarks on new era in astronomy

The $10 billion James Webb Space Telescope—NASA's most expensive science mission ever—is already more than halfway to its new post 1.5 million kilometers from Earth. Launched on 25 December 2021 from French Guiana, the 6.5-meter telescope is considered the successor to the Hubble Space Telescope; it will gather infrared light to help scientists study the universe's first galaxies and the atmospheres of alien planets. NASA said the launch and two subsequent trajectory maneuvers were both precise and thrifty: Webb will have enough propellant left to maintain its orbit at L2, a gravitational balance point beyond the Moon, for "significantly" more than 10 years. As *Science* went to press, operators had unfurled Webb's kite-shaped, tennis court–size sunshield and tightened its membranes—two of the many mechanical deployments that have kept astronomers on edge. Later this week, operators will move the primary and secondary mirrors into their final positions. After Webb arrives at L2 in late January, it will take another 5 months to chill its instruments and align the 18 hexagonal segments of the main mirror before taking a first snapshot.

Plasma shows COVID-19 promise

THERAPEUTICS | The antibody-rich plasma of recovered patients can avert hospitalization if given early in the course of a COVID-19 infection, a clinical trial has found. The treatment already has an emergency use authorization from the U.S. Food and Drug Administration (FDA) for certain severe cases. But last month, the World Health Organization recommended against the use of convalescent plasma in SARS-CoV-2–infected people with "nonsevere" illness, citing 16 previous studies that found the treatment didn't prevent death or reduce the need for mechanical ventilation. In a study of 1181 U.S. patients within 8 days of symptom onset, posted as a preprint on 21 December 2021 on the medRxiv server, 2.9% of those who received convalescent plasma were hospitalized, versus 6.3% of those who got a placebo. The trial's investigators propose that health authorities allow use of convalescent plasma for outpatients and in settings where monoclonal antibodies or antiviral pills are not available.

Bullying complaints draw apology

WORKPLACE | A month before leaving her position as head of the U.K. Medical Research Council for a top job at an influential European science agency, stem cell biologist Fiona Watt issued a public apology to MRC colleagues who accused her of bullying behavior. A 15 December 2021 press release from UK Research and Innovation (UKRI), the body that includes MRC and six other funding agencies, says an independent investigator looked into complaints made against Watt in 2020. The agency did not disclose the findings or the exact nature of the incidents. But it says Watt has accepted the outcome and apologized to her accusers. "I was devastated to learn that my actions and behaviour had affected colleagues in a negative way," Watt said in the press release. UKRI Chief Executive Ottoline Leyser called it "profoundly upsetting that people have experienced this behaviour" and said she was grateful to the whistleblowers. Watt told *Science* she cannot comment before 10 January, when she leaves MRC to become director of the

ANIMAL HEALTH

Researchers probe bird flu epidemic hitting Europe, Israel

A growing epidemic of avian flu is afflicting wild birds and has caused major losses of domestic fowl in parts of Europe and Asia in recent weeks, with thousands of animals dead or culled to slow the spread. The virus, A(H5N1), had been detected in poultry in 16 European countries as of early December 2021, with most cases in Italy. In the United Kingdom, which is facing its worst outbreak ever, farmers have killed more than half a million fowl. In Israel, the deaths of more than 6000 cranes are a sign of "the most serious damage to wildlife in the history of the country," said Israeli Minister of Environmental Protection Tamar Zandberg last week. Researchers at Hebrew University are using GPS receivers attached to birds to assess the risk of the virus spreading in Israel's Hula Valley, where more than 30,000 common cranes (*Grus grus*) spend the winter. The virus, which causes highly pathogenic avian influenza, spreads easily among bird species, and migratory flocks carry it vast distances. It rarely jumps to humans, but can be lethal if it does. As of 13 December 2021, only China said it had detected human illnesses, reporting 13 cases to the World Health Organization.

Workers wearing protective gear retrieve dead cranes (background) in Israel's Hula Nature Reserve.

European Molecular Biology Organization, an appointment announced in July 2021.

Rx for COVID-19 falsehoods

PUBLIC OPINION | Some of the most widely trafficked misinformation about COVID-19 is finding fewer believers, a study suggests. A survey in November 2021 by researchers at the University of Pennsylvania found that 74% of 1800 U.S. adults correctly identified as false a statement that COVID-19 vaccines change people's DNA; two-thirds said a claim the vaccines cause infertility was probably or definitely false. For both, that's less than in earlier polls by the university's Annenberg Public Policy Center, and the researchers think accurate media reporting and statements by health authorities may explain the increases. The changes are welcome because the false statements have helped dissuade some people from receiving the vaccines, said Kathleen Hall Jamieson, the center's director.

Ancient head lice offer DNA source

ARCHAEOLOGY | For millennia, head lice have afflicted humans, including the elite, as shown by evidence from mummies. Now, a study reports that the ancient parasites offer archaeologists a valuable new tool. Studying the heads of mummies found in Argentina, an international team extracted human DNA and environmental DNA from the hard natural glue, or cement, that lice secrete to attach their eggs to human hair. The work, reported in the 28 December 2021 issue of *Molecular Biology and Evolution*, suggests the cement preserves DNA and offers an alternative to extracting DNA from ancient teeth and bones, which requires destructive sampling. Analysis of DNA from lice cement on the eight Argentine mummies, which are about 1500 years old, revealed details about migration patterns of people in that era. The researchers also detected the earliest known presence of the Merkel cell polyomavirus, which causes skin cancer.

USDA unveils GM food label

BIOTECHNOLOGY | The U.S. Department of Agriculture (USDA) has created a new consumer label for genetically engineered food, "bioengineered," that is intended to reduce confusion—but which advocates for transparency fault as flawed and deceptive. Previously, states were free to decide whether to require labeling of foods containing genetically modified ingredients, such as soybeans altered to resist herbicides. But a 2016 law directed USDA to create a national labeling standard. Environmental groups complain that the new rule, which took effect on 1 January, has loopholes. For example, the label is only required if bioengineered DNA is detectable, but it's often undetectable in oil and starch. And foods need not be labeled if the proportion of genetically modified ingredients is less than 5%.

Pandemic infants show delays

DEVELOPMENT | Babies born at two New York City hospitals in 2020 during the COVID-19 pandemic showed significant delays in development at 6 months of age compared with prepandemic babies, says a small study published this week. Compared with 62 infants born at one of the hospitals before the pandemic, the 227 pandemic-era infants scored significantly lower on measures of gross and fine motor skills, as reported by their mothers using a standardized assessment tool. The children also lagged on a measure of personal and social development, but not on measures of communication and problem

COVID-19 pills for the poor

Many health advocates celebrated after the U.S. Food and Drug Administration last month authorized the emergency use of two different oral treatments for early SARS-CoV-2 infections. They also lauded decisions by the pills' manufacturers, Pfizer and Merck & Co., to allow other drug companies to make low-cost versions accessible to the poorest countries via the Medicines Patent Pool (MPP). Set up in 2010, the nonprofit was modeled after a cross-licensing agreement to free airplane patents controlled largely by the Wright brothers. MPP's founder, attorney Ellen 't Hoen, notes that it has helped bring cheaper AIDS and hepatitis C drugs to many millions. 'T Hoen, who works at the nonprofit Medicines Law & Policy, spoke with *Science* about the COVID-19 deals. (A longer version of this interview is at https://scim.ag/COVIDpillpool.)

Q: Some prominent groups have criticized MPP's deals with Pfizer and Merck for not allowing generic manufacturing everywhere that could benefit. What do you think?
A: Those big, brand name NGOs [nongovernmental organizations] suffer a little bit from knee-jerk responses to things that aren't perfectly perfect. If you read the license agreements carefully, there are no barriers to [generic manufacturers] supplying drugs in countries where patents have not been filed or have not been granted—or where governments have decided to issue a compulsory license.

Q: Where has patent sharing yet to succeed with COVID-19?
A: Both Pfizer and Moderna have dug in their heels: They don't want to license their vaccines. What I'm hoping is that this experience Pfizer now has [with MPP] will lead it to take the next, and much more important, step to license its vaccine technology. And that would have to include a technology transfer package.

Q: The tech transfer is far more important for vaccines than drugs?
A: Indeed. It's the [manufacturing] know-how that needs to be transferred, and you don't find enough of that in the patent.

SCIENCE.ORG/NEWS
Read more news from *Science* online.

solving. The results were adjusted for factors including the infant's gestational age at birth and the mother's age and educational level. Lower educational attainment was found in children born during the influenza pandemic of 1918, note the researchers, who report the new findings this week in *JAMA Pediatrics*. The new findings, they write, "support the need for long-term monitoring of these children" to prevent outcomes seen in previous pandemics.

Fossil hunter Richard Leakey dies

SCIENTIFIC LIVES | Paleoanthropologist and conservationist Richard Leakey died at his home near Nairobi, Kenya, on 2 January at age 77. The son of famed paleoanthropologists Louis and Mary Leakey, Richard Leakey added to their legacy with numerous important finds in Kenya, where he was born. His team found fossils crucial to our current understanding of human evolution, including skulls of *Australopithecus boisei*, *Homo rudolfensis*, and the famed *H. erectus* skeleton, "Nariokotome boy." Several of his finds helped support the once-contentious notion that more than one type of ancient hominin had lived side by side. Leakey also worked for conservation, directed the Kenya Wildlife Service, and helped limit the poaching of elephants for ivory. He supported students and scientists, especially other Kenyans, studying their country's scientific riches.

CORRECTION
The news article "Omicron sparks a vaccine strategy debate" in the 24 December 2021 issue included an image described as a false-color electron micrograph of the coronavirus variant. Instead, it was an illustration based on an electron micrograph.
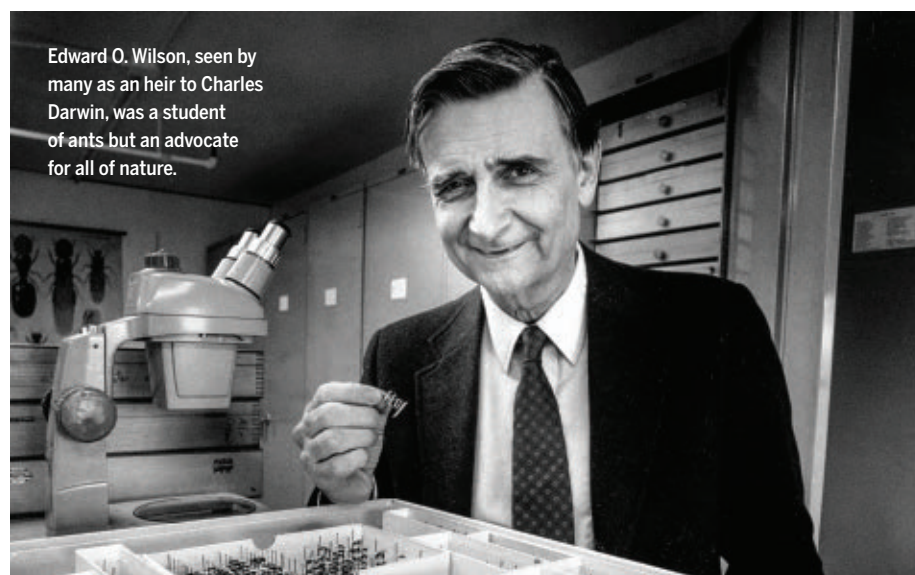
Edward O. Wilson, seen by many as an heir to Charles Darwin, was a student of ants but an advocate for all of nature.

BIOLOGY

Goodbye to two champions of biodiversity

December's end saw the deaths of two of the world's most influential naturalists and advocates for using science to protect the world's flora and fauna. Edward O. Wilson, 92, an ant researcher and evolutionary biologist at Harvard University, died on 26 December 2021. He introduced controversial theories about the evolutionary basis of animal societies, including human ones, and became a Pulitzer Prize–winning author of books about the natural world that inspired millions of readers. Thomas Lovejoy, a National Geographic Society conservationist who coined the term "biological diversity," drew attention to species loss in the Amazon, and helped rally the political will to protect them and their habitats, died at age 80 on Christmas Day. Both men were known for quirks of dress—Lovejoy for his bowties and Wilson for his field vest full of pockets—and both did research that drove home that smaller natural areas like those created by human activities tend to be more vulnerable to losing species. In large part through Wilson's writings, Lovejoy's "biological diversity" became "biodiversity" and a household word. The scientists "were really sort of bookends of the conservation movement," says Daniel Simberloff, an ecologist at the University of Tennessee, Knoxville, and a former Wilson student.

IN DEPTH

A temporary field hospital is constructed on the grounds of St. George's Hospital in London on 30 December 2021 to handle an expected surge in severe COVID-19 cases.

COVID-19

Omicron threat remains fuzzy as cases explode

Many countries break infection records; how much severe disease they will see is unclear

By **Kai Kupferschmidt** *and* **Gretchen Vogel**

As this year begins, the Omicron variant is smashing COVID-19 infection records across Europe, North America, Africa, and Australia. With massive numbers of people infected or in quarantine, tens of thousands of flights and trains have been canceled, and work and schools disrupted. Earlier in the pandemic, the frightful, near-vertical rise in cases would have triggered stringent lockdowns. Not this time: Many governments are banking on early indications that vaccines still protect against severe disease and that Omicron may be a gentler variant.

It's a risky bet, because scientists still can't predict Omicron's ultimate toll. Initial data from South Africa, Denmark, and the United Kingdom suggest it causes less severe disease, but they come with major caveats. And even less severe cases can strain hospitals already at the edge of their capabilities. In a worrying sign, COVID-19 hospitalization rates in the United States surged this week.

Virologists, molecular biologists, and epidemiologists are biting their nails and hoping cases will soon peak and begin to fall. They are also working at top speed to sort out the properties of the new variant. Here are some of the key questions they are trying to answer.

How does Omicron differ from previous variants?

Just 6 weeks after its discovery, it's clear the new strain behaves very differently from previous ones. Early studies by several groups hinted that Omicron replicates less well in lung cells than other variants, and two new studies have suggested a possible mechanism, based on how it enters human cells.

SARS-CoV-2 has two entry routes. After binding to ACE2, a receptor protein on the cell surface, virus particles can fuse with the cell directly when another human protein called TMPRSS2 cleaves the spike protein on the virus' surface. Alternatively, after binding to ACE2, the virus can be swallowed by the cell in a vesicle called an endosome. The virus then escapes into the cell's cytoplasm with the help of other spike-cleaving proteins called cathepsins.

Early lab studies showed TMPRSS2 cleaves Omicron's spike protein less efficiently than that of other variants, hampering the first, direct route. And in a preprint posted last week, virologist Joe Grove of the University of Glasgow and others showed that blocking TMPRSS2 with the chemical camostat inhibited pseudoviruses carrying the spike of the Alpha or Delta variants in cell culture, but not those carrying Omicron's. Conversely, blocking cathepsins with

a compound named E64d inhibited viruses carrying Omicron's spike, but not Alpha's or Delta's. Those data make "clear there's a very, very strong preference now for the endosomal route," Grove says. (Another new preprint, by virologist Thomas Peacock of Imperial College London and colleagues, had similar findings.)

Some researchers argue Omicron may spare the lungs and cause milder disease because TMPRSS2 is more common on cells in the lower airways, but there are few data to support that, Peacock says. Cells infected with Omicron may also be less likely to fuse with neighboring cells to form large cells called syncytia that might be a cause of severe disease.

"For now, the combined data seem to point in the direction of a less severe clinical picture, possibly explained by changes in how the viruses infect and which cell types they infect," says virologist Marion Koopmans of Erasmus Medical Center.

Could these changes in viral biology explain why Omicron spreads so fast?

Perhaps. It's possible they have shortened the time between exposure and onset of symptoms, which studies suggest is only 3 days for Omicron, down from about 4 days for Delta and more than 5 for previous variants. This is likely contributing to the steep rise in

COVID-19 case numbers. But other changes may play a role as well.

One is the variant's ability to get around immunity provided by previous infections and vaccines. A study of SARS-CoV-2 spread in nearly 12,000 households in Denmark, posted on 27 December 2021, provides some of the clearest evidence of Omicron's advantage. Economist Frederik Plesner Lyngse of the University of Copenhagen and the Danish Statens Serum Institute and his colleagues found that in households with a Delta outbreak, the unvaccinated were twice as likely to be infected by a household member as those who were fully vaccinated. In households struck by Omicron, unvaccinated and fully vaccinated people had roughly equal chances of catching the virus.

That doesn't mean COVID-19 shots don't work; other data clearly show they still prevent severe disease. And in the Danish study, a booster shot cut the risk of infection by Omicron in half. Being vaccinated also reduces an infected person's chance of infecting others, Lyngse notes: For both variants, an unvaccinated case was 41% more likely to infect another household member than a fully vaccinated one.

There's another possible explanation for Omicron's explosive spread that the household study would not be able to pick up, notes epidemiologist Bill Hanage of the Harvard T.H. Chan School of Public Health. If the virus is really better at replicating in the upper airways, it might be expelled more readily into the surrounding air, making it more likely to trigger superspreading events. And if it truly causes milder illness, even people who shed a lot of virus might have few symptoms, making them more likely to be out and about. If so, banning large gatherings and closing down nightlife and restaurants might be even more effective at slowing the spread of Omicron than that of previous variants, Hanage says.

Is Omicron sending fewer people to the hospital than previous variants?

Data from South Africa, where cases have already started to decline, suggest Omicron put one-third as many people in the hospital as the Delta variant did. But previous infections and the country's relatively young population may have helped keep severe cases low.

Initial hospital data from England and Denmark also suggest Omicron cases are less severe. But those countries have high vaccination rates, and there, too, Omicron has spread most quickly among younger adults. Severe cases may increase in the wake of holiday parties where people of all ages mixed. "If we see that Omicron is capable of causing severe disease

in older age groups … I think it could be much worse than most people are thinking about at the moment," Hanage says. The U.S. picture looks less hopeful: More than 100,000 people there were hospitalized with COVID-19 when *Science* went to press—up from 75,000 a week earlier.

Moreover, there are signs Omicron can trigger dehydration from fever, vomiting, and diarrhea and can exacerbate other health issues such as diabetes. Such cases need less intensive care, but can still overwhelm hospitals. "I'm hearing story after story after story of hospitals that are full, health care workers that are infected, that can't do their job," says Maria van Kerkhove, an epidemiologist at the World Health Organization (WHO).

How severe Omicron is in people who are "immunologically naïve"—neither vaccinated nor previously infected—remains an open question. They are now a minority in most countries, but they too could add to the hospital burden if they get seriously ill. Deaths in Ghana, Ivory Coast, and Madagascar, where vaccination rates are low, are increasing sharply as COVID-19 cases surge.

Is Omicron the last variant of concern?

Probably not. More variants with Omicron's ability to evade immunity and spread successfully are likely to emerge, says Aris Katzourakis, who studies virus evolution at the University of Oxford—and they may prove more virulent than Omicron. After all, even Omicron itself may not have evolved to cause less severe disease, Katzourakis notes. It may be milder than Alpha or Delta, but it branched off from the family tree before those variants emerged. "Is it milder than what it evolved from? I don't think we know the answer to that," Katzourakis says.

The next variant to make global headlines may already be circulating. WHO is tracking two "variants of interest" and three "variants under monitoring," as well as 30 sublineages of Delta, Van Kerkhove says. "There are many more Greek letters that we can potentially go through."

But Omicron itself may help tame whatever comes next. Vaccination and natural infections have exposed many millions of people to earlier versions of SARS-CoV's spike protein, training their immune system to respond to those variants. Infections with Omicron, whose spike looks different, will likely both strengthen and broaden immunity, Hanage says, hopefully making new variants less dangerous. "I suspect that immunity, post-Omicron, will be pretty broad," Hanage says. "But I don't want to bet on it." ∎

What the Lieber verdict means for the China Initiative

Jury finds Harvard University chemist guilty of lying about ties to China

By **Jeffrey Mervis**

Last month, the U.S. government won the first conviction of an academic scientist under its 3-year-old China Initiative, when a federal jury in Boston found Harvard University chemist Charles Lieber guilty of lying about his research ties to China.

Andrew Lelling, the former U.S. attorney who led the investigation that 2 years ago resulted in criminal charges against Lieber, thinks it could also be the last such conviction. One reason is that defense attorneys now might choose to seek a plea bargain for their clients rather than risk an adverse verdict. And Lelling thinks the government could become more selective in deciding which cases to prosecute, focusing on imminent threats to U.S. economic and national security rather than violations of rules that require federally funded researchers to disclose all sources of research support.

Critics of the initiative have long argued for a narrower focus. They say it too often criminalizes bookkeeping mistakes by otherwise blameless researchers. They also believe prosecutions like Lieber's unnecessarily harm U.S. innovation by scaring scientists away from engaging in important research collaborations.

"Yes, [the initiative has] been a deterrent, just like cutting off the hands of thieves," says Peter Zeidenberg, a lawyer with Arent Fox who has defended several scientists charged under the China Initiative. "But it's also been terrifying. And it is not justice, nor is it of value to society."

So far, the government has prosecuted some two dozen academic scientists it says abetted China's efforts to steal U.S.-funded technology by not disclosing their ties to research programs funded by the Chinese government. Its record has been mixed. In eight cases, academics have pleaded guilty

Charles Lieber leaves the Boston courthouse during his trial last month.

and received prison sentences of up to 37 months. But the government dropped its prosecution of seven other scientists. And in September 2021, in the first China Initiative case to go before a jury, a federal judge acquitted Anming Hu, a former mechanical engineering professor at the University of Tennessee, Knoxville, of charges that he lied to NASA about his work with the Beijing University of Technology. An additional five researchers are still awaiting trial.

Lieber was charged with lying to federal research agencies about his participation in a Chinese foreign talent recruitment program, failing to report income from the program on his tax returns, and failing to disclose a Chinese bank account used to pay him. On 21 December 2021, it took the jury less than 3 hours to find him guilty. He faces up to 5 years in prison and a fine of $250,000.

Neither supporters nor critics of the China Initiative were surprised by the jury's short deliberations. "The charges were very straightforward," says Margaret Lewis, a law professor at Seton Hall University. "The government had to prove intentionality, and it was able to present evidence that was quite striking."

Much of that evidence came from a video that prosecutors played in court of Lieber's 3-hour interrogation by FBI agents after his arrest on 28 January 2020. It included Lieber's observation that the agents had "damning" evidence that, despite his de-

nial, he had joined a Thousand Talents recruitment program. "I may have signed a document that was in Chinese that I shouldn't have signed," Lieber says. Lieber also acknowledged on the recording that the money he received from the Wuhan Institute of Technology should have been declared and that he had failed to do so. He offered an unflattering motive for the collaboration: "This is embarrassing. Every scientist wants a Nobel Prize."

Lelling says Lieber's denials about participating in the Thousand Talents Program showed "bad intent." But although Zeidenberg agrees that Lieber was evasive in answering FBI questions, he's not sure his words proved intent. "What was clear is that [Lieber] was a member of the Thousand Talents Program," Zeidenberg says. "But what is less clear to me is what questions [FBI] asked."

Lieber's attorney, Marc Mukasey, has already filed a motion asking U.S. District Judge Rya Zobel, who presided over the case, to acquit him. That is standard practice, but Lewis says it's hard to see the conviction being overturned.

The verdict is unlikely to change the political rhetoric surrounding the China Initiative. Although Lieber wasn't accused of improperly sharing the results of his federally funded research with Chinese scientists, Senator Rob Portman (R–OH), who has called for tougher sanctions against scientists who don't disclose ties to foreign

funders, described the case as "a clear example of China's continued attempts to steal our taxpayer-funded research." And Joseph Bonavolonta, Lelling's former colleague and head of FBI's Boston office, said Lieber's conviction "reinforces our commitment to protect our country's position as a global leader in research and innovation."

Lelling, now in private practice at Jones Day, a law firm, thinks the Department of Justice will now "raise the bar" for prosecuting similar cases. "What's to be gained by targeting additional investigators for breaches of research integrity?" he asks. "You're not going to be getting additional deterrence. That message has already been received."

In place of pursuing "technical fouls," Lelling's phrase for cases involving a scientist's failure to disclose research ties, he expects the department's national security division, under newly installed Matthew Olsen, to focus on alleged theft of intellectual property, trade secrets, and other forms of economic espionage.

Critics would rather see Congress and President Joe Biden's administration pursue broader reforms. They are calling for clearer and more consistent rules about what researchers need to disclose, a more sophisticated approach to handling trade disagreements with China, and better training on how academic science operates for law enforcement agents tasked with sniffing out alleged economic espionage.

Some of those changes are already underway and won't be affected by the Lieber verdict, Lewis says. "There's a chance now for a more overarching view of how to maintain important legal principles in the course of advancing U.S. interests," she says, one that encompasses several federal agencies. "Remember, the goal of the China Initiative was to improve our economic position, not to send scientists to jail."

The next step in Lieber's case is sentencing. The prosecution will ask for a sentence based on guidelines that take into account the nature of the crime as well as the defendant's criminal history, acceptance of culpability, and willingness to cooperate with the government. Lieber's attorney is likely to argue for leniency, citing the researcher's ongoing treatment for cancer, his exemplary scientific record, and his remorse.

"I think you'll hear [Lieber] say, 'I screwed up, and I shouldn't have done it,'" Lelling predicts. "And that matters to judges." Lelling's bottom line: "I think it is highly unlikely that the government will seek jail time for Lieber. And it is equally unlikely that he will get any." ∎

A soybean plantation has replaced a swath of Amazon rainforest near Santarém, Brazil.

REMOTE SENSING

Satellites document rapid expansion of cropland

Farms added 100 million hectares globally over 2 decades, threatening biodiversity and accelerating climate change

By **Gabriel Popkin**

Farmland is overtaking much of the planet. A new global map assembled from satellite imagery shows that over the past 2 decades, fields of corn, wheat, rice, and other crops have eaten up more than 1 million additional square kilometers of land—roughly twice the area of Spain.

"The inexorable march of the human footprint is just brutal," says study co-author Matt Hansen, a geographer at the University of Maryland (UMD), College Park. The food needs of a fast-growing population in Africa are driving some of the expansion. But the study also highlights how Earth's land is becoming, in essence, a unified global farm, with wealthier countries increasingly outsourcing crop production to poorer regions. Half of the new fields have replaced forests and other natural ecosystems that stored large amounts of carbon, threatening efforts to conserve biodiversity and avert catastrophic climate change.

To construct their map, the researchers used data from the U.S. Geological Survey's and NASA's Landsat program, which has launched a series of satellites that periodically image every spot on Earth with pixels covering roughly 30 square meters, or about the size of a baseball diamond. The team visited farms around the world and used high-resolution commercial satellite photos available from Google to train algorithms to distinguish croplands from natural grasslands and other types of land cover. The maps yield both wide-angle and close-up views. "You can get a global story; you can also tell the story of [a single nation such as] Cambodia," Hansen says.

At the global scale, the cropland footprint increased 9% over the study period, which covered 2000 to 2019, the team reported on 23 December 2021 in *Nature Food*. The increase is several times higher than the 2.6% growth in "arable land" over the same period calculated by the Food and Agriculture Organization of the United Nations.

South America led the world in relative cropland expansion, with the continent's farms growing by nearly 50% during the study period. That's thanks largely to a booming soybean industry supplying livestock farmers in China and elsewhere. Africa saw the largest absolute growth in the total area of new fields as it struggled to feed a fast-growing population. Forty percent of Africa's cropland was created in the past 2 decades, and the rate is accelerating. Farmland also swelled in several nations in South Asia and in North America's Great Plains.

Crops didn't gain ground everywhere. In the former Soviet Union, for example, farmers abandoned unproductive areas.

But overall, the analysis underscores that people are creating far more cropland than they are abandoning or restoring to forest or grasslands, says Tim Searchinger, a senior fellow at the World Resources Institute who was not involved with the work. He notes that farming likely gobbled up even more land than the analysis found, because it didn't tally areas converted to new livestock pastures and tree plantations.

The map is sharper and more up to date than many currently in use, says geographer Amy Molotoks of the Stockholm Environment Institute. She and her colleagues "are highly likely to use it in the future" to determine where farms have replaced forests or other natural ecosystems, she says.

Those shifts have worrisome implications for efforts to preserve biodiversity. The conversion of rainforests like the Amazon to agriculture often gets the headlines, but the study found new crop fields took a bigger bite out of less heralded biodiversity hot spots, such as dry forests and savannas. In South America, important dry ecosystems known as the Chaco and Cerrado took major hits, says study lead author Peter Potapov, also at UMD. "They will disappear completely very soon," he fears.

Such losses accelerate climate change, because carbon stored in trees and soil escapes to the atmosphere when land is cultivated. Land clearing currently causes roughly one-eighth of humans' total carbon emissions, researchers estimate.

The study did reveal some hopeful trends. Over the study period, the growth rate of plant biomass in croplands increased by 25%, and per capita crop area decreased by 10%, suggesting humanity is continuing to find ways to squeeze more food out of a given hectare.

Molotoks says that, by showing how consumer food choices drive environmental change, such analyses could help conservation advocates make the case for greener practices. In the United Kingdom, she notes, some consumers have shifted to plant-based diets after learning that farmers in South America are clearing large tracts of forest to grow soybeans for animal feed.

The study also highlights the need to boost cropland productivity in Africa, which has the world's lowest crop yields, meaning more land is needed to grow a given amount of food. "If the world wants to solve climate change," Searchinger says, "from a purely self-interested standpoint, it needs to support Africa in solving its land use challenge, and that includes much higher yield growth and food security." ∎

SCIENTIFIC COMMUNITY

Afghan scholars find a warm welcome in Rwanda

Five academics have resettled in Kigali with help from the U.S. National Academies

By **Richard Stone**

One night in September 2021, Fatema Samim's fate hung in the balance. A civil engineer from Herat, in western Afghanistan, she had been on the run from the Taliban for a month and was holed up with her husband and two young sons in Mazar-i-Sharif. Charter flights were whisking Afghans with U.S. ties to safety, but the Taliban had begun to crack down on this escape route, and that evening, its fighters fanned out to the city's guest houses to detain people attempting to flee.

"That night was very bad for us," says Samim, a former dean at Herat University who led a rainwater harvesting project run by the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM). But after NASEM staffers evacuated the family to Kabul—a perilous 12-hour drive past several Taliban checkpoints—their odyssey ended in an unexpected haven: Rwanda. They flew via Pakistan to Kigali, the capital, where Samim and four other Afghan engineers and hydrologists with NASEM links are now assistant professors at the University of Rwanda.

Rwanda's own dark past—a genocide in 1994 claimed more than 1 million lives and displaced 2 million—forged a bond. "What the Afghan scholars were experiencing struck a familiar chord," says chemist Valentine Uwamariya, Rwanda's minister of education. "We welcomed them warmly."

The U.S. Department of State had largely fumbled the evacuation of scholars and civil society actors after Kabul fell to the Taliban on 15 August 2021. Individual institutions scrambled to fill that void. Michigan State University, for instance, in late August managed to secure safe passage for 23 Afghan agricultural scientists and their families to Albania, where they are waiting for U.S. visas. Others who have managed to reach U.S. soil are stuck on military bases as their visas are processed.

NASEM sought a refuge for scholars funded through a competitive grants program the academies run for the U.S. Agency for International Development; they were likely to be in particular danger because of

their U.S. ties. "We were looking for someplace they wouldn't have to go through all the hurdles that it takes to get resettled in the U.S.," says National Academy of Sciences President Marcia McNutt. Rwanda leapt to mind: The country has long strived to raise its science game and, McNutt surmised, would sympathize with the Afghans. On 17 August, she wrote to Rwandan President Paul Kagame, who was eager to help. "We understood their plight," says Rwanda's

Fatema Samim (top) and Naeem Salarzai (bottom) have become assistant professors at the University of Rwanda.

U.S. ambassador, Mathilde Mukantabana, a former history professor. "Our country always opens our door to refugees."

"Never in my life did I imagine I'd end up in Rwanda," says Naeem Salarzai, a former director general for water management affairs in the Afghan government who led a 200-strong team that allocated the country's scarce water and managed transboundary agreements, including a fractious water treaty with Iran. When Kabul fell, Salarzai worried that Taliban sym-

pathizers in Iran, miffed by Afghanistan's assertive posture on water rights, might have shared his name with the Taliban. His participation in a NASEM workshop increased his jeopardy. "I was quite scared," he says.

Salarzai and his family moved from house to house in Kabul to evade Taliban pursuers. Three times in the latter half of August, NASEM directed them to head to the airport. The third time they almost reached the airport gate before giving up, fearing their two children might be crushed in the desperate crowd. Hours later, a suicide bomber killed scores of Afghan civilians and U.S. soldiers near the same gate. "It felt so horrible," he says.

Two weeks later, feigning a need for medical treatment in Pakistan, the family escaped to Islamabad and on to Kigali. "We found a place that is very green and very clean," Salarzai says. "Right from the first day we started loving it."

NASEM negotiated a 1-year contract for Salarzai and Samim and their three colleagues with the University of Rwanda. They are helping devise a curriculum for a new master's degree, supervising students, and reading theses, and will start to teach in February. Uwamariya hopes the Afghans will also use their "wealth of talent and experience" to advise a new government Water Resources Board. The five families—19 people altogether—cook together and go on weekend jaunts.

Still, the United States beckons. "Kigali is a good place, for a short time," says Samim, who intends to travel on once U.S. visas come through. Her younger sister is already in law school at the University of Notre Dame and working on visas for other family members now in Tajikistan and Iran. Salarzai hopes to join a brother whose family is living in California.

"I play out in my mind the long game here," McNutt says. "Will these scholars return to Afghanistan one day? That's what I hope against hope will happen." So does Salarzai: "I'm looking forward to the time when Afghanistan has the right kind of leaders and I can work for my own country again." ◾

ECOLOGY

Plant diversity is blowing in the wind

DNA in air reveals the plants below—and requires no tromping through bushes

By **Elizabeth Pennisi**

Inventorying the plants in a tract of woods or fields or searching for invasive species can take days of hot, hard work slogging through thorny brush and tick-infested grass. Now, researchers have shown that simply capturing and analyzing the DNA plants release into the air can work as well as putting boots on the ground—and in some cases even better.

"Airborne DNA could be a game changer in our ability to monitor and study biodiversity," says Kristine Bohmann, a molecular ecologist at the University of Copenhagen who was not involved with the work. The approach could help track how climate change is altering the makeup of plant communities, researchers say, and provide early warning of invading species.

Reported last month in *BMC Ecology and Evolution*, the work takes the study of environmental DNA (eDNA)—genetic material shed, defecated, coughed up, or otherwise released into the environment—into a new realm. Aquatic eDNA is now a proven tool for identifying fish and other marine and freshwater organisms, with no need to catch them. In soil, eDNA can reveal the presence of people and animals, current or ancient. "The time is right to look at another source," says Elizabeth Clare, a molecular ecologist at York University who has tracked animal species with airborne eDNA.

Plants already emit airborne tracers that are familiar to anyone with allergies: windborne pollen. The grains' distinctive shapes make it possible to identify unseen species simply by capturing their pollen. But pollen surveys have their limits. They can only detect plants whose pollen is spread by wind (other types depend on pollinating insects and other animals), require well-trained experts, and don't always produce species-specific identifications. Mark Johnson, a graduate student at Texas Tech University, wanted to know whether studying the eDNA that is shed into air not only as pollen but also in minuscule fragments of leaves or flowers would work better.

He and his colleagues developed better ways to collect plant eDNA in dust traps, and in 2019 they demonstrated that the filters capture DNA-bearing traces from all sorts of plants. "We could find species not flowering, not pollinating, or when they are not active like in the winter," Johnson says.

Now, he has shown how eDNA can inventory an entire plant community. He and his colleagues mounted dust traps in nine places across a well-studied short grass prairie owned by his university. They collected the dust every couple of weeks for 1 year, extracted the DNA, and sequenced a gene that varies among plant species, serving as a "DNA barcode" for identifying them. In the spring and again in the fall, his team also pulled on their boots and surveyed plants along 27 100-meter transects. They com-

Plants fill the air with their DNA, encased in the windblown pollen seen here, as well as in tiny particles of leaves and flowers.

pared the results of the two kinds of surveys.

The traditional surveys detected 80 species and the air eDNA study 91, the team reported. Both surveys uncovered the same 13 grass species, but the eDNA work found an additional 13. Among nonwoody flowering plants, both approaches yielded a total of 60 species, but each detected 20 or so that the other missed. eDNA was better at finding easily overlooked species with small flowers, such as weakleaf bur ragweed. But people were better at spotting plants too rare to release much eDNA, particularly when they had showy flowers, such as the chocolate daisy.

Airborne DNA also revealed tree of heaven, an invasive plant not detected by the survey. That's encouraging, says Loren Rieseberg, a plant evolutionary biologist at the University of British Columbia, Vancouver. "I think [airborne eDNA] will be especially useful for detecting invasive species before they become widespread and difficult to get rid of."

The technique recorded how the abundance of different species changed through time, including the rapid bloom and growth of the tansy mustard in early spring, which ground surveys missed. "This [report] might encourage more researchers to take up dust traps to complement" other kinds of surveys, particularly at long-term study sites, says Fabian Roger, an ecologist at ETH Zürich who was not involved with the work.

Along with plant DNA, the filters picked up DNA from fungi, and other researchers have captured insect, earthworm, and slug DNA from the air. "Potentially air DNA is incredibly diverse and representative of the full diversity of living organisms," Roger says. His own eDNA survey of insects in the wild detected just a fraction of the species known to be present, but he expects sensitivity to improve with a better understanding of how wind and other conditions affect DNA collection, and better technology.

Existing traps typically rely on natural air flow to carry in particles carrying eDNA, but the concentrations can be very low. More efficient filters, or traps with fans to suck in air, could work better. "You need a good system to trap the air," agrees Crystal Jaing, a molecular biologist at Lawrence Livermore National Laboratory who has been assessing high-altitude airborne microbes from specially equipped airplanes.

Joseph Craine, co-owner of Jonah Ventures, a company commercializing eDNA surveys, thinks the technology isn't ideal for large areas. "I can't see the application," he says. Finely tuned spectroscopic measurements from space are a better approach to surveying plants, he says.

But others point to how far aquatic eDNA studies have come in recent years and think air eDNA can do the same for monitoring terrestrial life. Roger says: "Air has the potential to be the 'water' over the land." ∎

Permafrost thaws have softened foundations and damaged buildings in Siberian cities such as Yakutsk.

CLIMATE CHANGE

Russia begins work on national permafrost monitoring system

Data could improve climate models and provide warnings for infrastructure vulnerable to thawing soil

By **Olga Dobrovidova**

With the Arctic warming up to four times faster than the global average, temperatures in the frozen soil, or permafrost, under northern Siberia have been rising, turning firm ground unstable and weakening foundations. On 29 May 2020, the thaw may have helped lead to a disaster, when a diesel fuel depot near the town of Norilsk collapsed and spilled more than 21,000 tons of fuel into a small river. The pollution turned the river rusty-red and ultimately reached the Arctic Ocean. The owner of the fuel depot, Norilsk Nickel, the world's largest miner of nickel and palladium, was fined almost $2 billion for the spill—the largest settlement for an environmental disaster in Russian history.

Now, the disaster has spurred the government to set up the first national system to monitor Russia's permafrost—the world's largest expanse of frozen soil, covering two-thirds of the nation. In October, President Vladimir Putin gave the go-ahead to a new $21 million system of 140 monitoring stations that could begin to deliver data as soon as 2023. Sensors placed in boreholes up to 30 meters deep will measure the temperature of permafrost at various depths, a critical parameter for tracking both the growing hazard thawing ground poses to infrastructure and the broader climate threat: that the thaw could release billions of tons of carbon into the atmosphere.

Researchers say the new nodes will bolster a patchy system of some 440 boreholes, run by research organizations and private companies like Norilsk Nickel, that had been in place since the collapse of the Soviet Union. "There was no national system at all, which was such a shame, and now we may get one," says Dmitry Streletskiy, a climatologist at George Washington University who co-leads the Circumpolar Active Layer Monitoring program, which has supported more than 250 sites globally and 75 in Russia since the 1990s with funding from the U.S. National Science Foundation. "I've been arguing for years that we need one."

The origin of the national monitoring system can be traced to a 2020 report authored by researchers from permafrost-focused institutes in Russia and commissioned by Alexander Kozlov, who was then minister for the development of the Russian Far East and Arctic. Kozlov later became the minister of natural resources and the environment, where he oversaw plans for the current system, which will be operated by Roshydromet, the state weather agency that operates under the aegis of the environment ministry.

Marat Sadurtdinov, who heads the Earth Cryosphere Institute in Tyumen and was a co-author of the report, is disappointed by the current plan. In the report, the researchers called for a larger $160 million system that would have tracked other parameters in addition to temperature, such as moisture or ice content, and would have been operated by an independent interagency entity. Sadurtdinov is concerned that the system will not gather enough information to provide warnings to the owners of vulnerable roads, pipelines, and buildings, and that Roshydromet's weather stations may not be the ideal locations for permafrost measurements. "We've talked about this everywhere, and the ministry knows it and is listening to us," Sadurtdinov says.

In a statement provided to *Science*, the environment ministry insists there is "no contradiction" between the commissioned report and its plans. It says Roshydromet was picked to lead the system in part because attaching the permafrost stations to the existing network of weather stations will minimize costs. Roshydromet has assigned the design and rollout of the system to the Arctic and Antarctic Research Institute (AARI), which currently has five permafrost monitoring sites in the Russian Arctic. Aleksandr Makarov, AARI director, says the first nodes could be deployed this summer.

The nodes would add to the sparse data sets used to calibrate climate models that predict changes in permafrost. Knowing how fast the permafrost is warming and thawing is critical to assessing the fate of the trillion or more tons of carbon in the frozen organic matter—more than the atmosphere now holds. Once it thaws, microbes could decompose the organic matter and release the carbon to the air—exacerbating global warming via the so-called "permafrost feedback." Researchers rely on models, says Oleg Anisimov, who heads climate change research at the State Hydrological Institute, because "you can't stick a sensor into every swamp to see what's going on with carbon emissions."

Adding sensors for carbon dioxide and methane would make the network even more powerful, says Guido Grosse, who studies permafrost at the Alfred Wegener Institute. But he agrees that even temperature data can help calibrate the modeling estimates for permafrost carbon fluxes, and he hopes researchers outside Russia will have access to the readings via the Global Terrestrial Network for Permafrost, which gathers data from existing boreholes. "It's important that the data becomes available to the scientific community at large—not 10 years from now but as soon as possible," he says. ∎

Olga Dobrovidova is a science journalist in Moscow.

FEATURES

THE PANDEMIC
WHISTLEBLOWER

Rick Bright raised the alarm about the Trump administration's response to COVID-19.
Now, he wants to build an alert system for future threats

By **Jon Cohen**

n May 2020, with anonymous callers vowing to kill him and similar threats mounting on social media, Rick Bright gave up his cellphone and went into hiding for more than a month. "If I heard tires rolling over the road in the middle of the night in the driveway where I was staying, it was panic," says the 55-year-old immunologist, who un-

til that month had been a powerful, if obscure, U.S. government public health official.

The threats began after Bright filed a whistleblower suit alleging he had been demoted from the top job at the Biomedical Advanced Research and Development Authority (BARDA) for protesting the govern-

ment's COVID-19 contracts and what he saw as its misguided, plodding response to the growing pandemic. He accused his bosses of trying to steer taxpayer dollars to firms run by "cronies" or "for political purposes." The "straw that broke the camel's back," the complaint stated, is that he publicly criticized hydroxychloroquine—the antimalaria drug then-President

Rick Bright has a lot of desks to fill at his new Pandemic Prevention Institute.

Donald Trump had touted as a coronavirus remedy—as useless.

The suit and his congressional testimony that soon followed catapulted Bright into the public eye. A *60 Minutes* story described him as "the highest ranking government scientist to charge the federal government's response to the coronavirus pandemic has been slow and chaotic." Trump took to Twitter to assail him, but many of Bright's peers in public health cheered him. "He was one of the very early people to tell the American people what was going on," says Nicole Lurie, who during former President Barack Obama's administration oversaw BARDA as assistant secretary for preparedness and response (ASPR) at the Department of Health and Human Services (HHS). "He had a lot of guts."

Bright's abrupt, tumultuous exit from BARDA, where he had for 4 years overseen a $1 billion–plus research budget aimed at protecting the country from pandemics and bioweapons, marked but one more dramatic chapter in a rough-and-tumble life. Now, in a bold gamble on his ability to make something from nothing, the Rockefeller Foundation has hired Bright to head a new bid to protect the world from future pandemics.

Rockefeller will give the Pandemic Prevention Institute (PPI) $150 million in seed money over the next 3 years to tap and quickly share pathogen surveillance data gathered by myriad sources. "We're setting out to build an environment for sharing data around the world at all levels—not just governments—that will allow us to make smarter decisions," Bright says.

"I'm wildly supportive that the Rockefeller's doing this," says Eric Lander, director of the White House's Office of Science and Technology Policy, which recently issued its own ambitious, multibillion-dollar prescription to better address pandemics. Rajiv Shah, Rockefeller's president, is certain Bright can turn PPI into a powerful force. "Rick is absolutely the best person on the planet to lead it," Shah says. Bruce Gellin, an epidemiologist who for 15 years led HHS's National Vaccine Program Office and is one of PPI's first 16 employees, says, "Rick is a 50-year-vision person. That's what he does."

But even some admirers wonder how Bright's new venture will stand out among efforts by governments, academia, industry, and the World Health Organization (WHO) that share PPI's elusive aspiration, including several new ones with similarly large backing. And his detractors charge that Bright can be arrogant. "His ego is big-ger than his managerial skills," says physician Robert Kadlec, Lurie's successor as ASPR under Trump and the main target of Bright's blistering whistleblower complaint.

Lurie sees it differently. Bright, she says, has no fear of speaking what he perceives as truth to power. "Even when it was unpopular, it was something he did, whether it was about programmatic stuff or individuals," she says. "If Rick didn't respect somebody, it was difficult for him to play along without saying something."

BRIGHT'S HARDSCRABBLE ROOTS help explain his willingness to speak out, Lurie says. "When you overcome a huge amount of adversity, it builds a new kind of self-

> ## "If Rick didn't respect somebody, it was difficult for him to play along without saying something."
>
> **Nicole Lurie**, Coalition for Epidemic Preparedness Innovations

confidence and resilience." But little else foreshadowed that Bright would become a leader in vaccine development and pandemic preparedness. Nearly 40 years ago, his high school in Hutchinson, Kansas, told him he could not attend his senior year because he had not received the vaccine against measles, mumps, and rubella. Turns out, he says, "my mother never vaccinated us for anything."

Bright, who bounced between eight foster homes after his family fell apart because of an abusive stepfather, stresses that his mother wasn't antivaccine. "We were a low-income family in a small town," says Bright, who has six siblings. "It was an educational thing."

He got vaccinated and finished high school—a notable feat given his background. "My choice in life growing up was to drop out of school early and take over the auto salvage business or work on the farm," he says. College, he adds, never would have occurred to him, but while in high school he worked 40 hours a week at a restaurant owned by a family that valued education and encouraged him to keep studying. He enrolled at the University of Kansas as an accounting major, his ticket out of Hutchinson. But he soon dropped out and moved to Kansas City, Missouri, with no concrete plans.

A help wanted ad in the paper led to a job at a dance studio, cold calling potential clients. During a coffee break, one of the studio's teachers—a "glamorous woman with a glamorous life"—asked him whether he had any interest in becoming an instructor. After 1 month of dance lessons, he had his first career. A wealthy student—an elderly woman who had lost a son his age the year before—took a shine to him and challenged him to restart his college education. She offered him $10,000 to quit his job, and he accepted. "I went back for a year and dropped out again, still lost."

Bright became seriously interested in science when a lab that did diagnostic testing hired him to write educational material. He decided to try college again, moving to Alabama with the dream of going to medical school. To be eligible for in-state tuition, he needed to live there for 1 year, so he took a job at Blockbuster Video, which helped him pay off $25,000 in credit card debt. He eventually enrolled at Auburn University, paying his way through school by washing glassware in a lab. Bright was 30 years old in 1997 when he graduated, magna cum laude, with a double major in biology and chemistry.

For his Ph.D., Bright attended Emory University, where he told his adviser, HIV vaccine researcher Harriet Robinson, he wanted to do something unique. He was thrilled when she let him focus on developing a vaccine against H5N1, a highly lethal bird flu strain that had jumped to humans. "I want to save lives, and I want to protect people," Bright says. He got his first taste of research with dangerous pathogens, running experiments in high-biosecurity labs at the nearby Centers for Disease Control and Prevention (CDC).

Bright returned to CDC after getting his Ph.D., studying the limitations of an entire class of influenza drugs and publishing first-author papers in *The Lancet* and *JAMA*. He soon turned to influenza vaccines again, first at an up-and-coming company called Novavax (which recently developed a COVID-19 vaccine) and then at PATH, a nonprofit that focuses on global health. There he helped a government-owned manufacturer in Vietnam produce shots for seasonal and potentially pandemic influenza viruses.

The PATH project received funding from BARDA, which lured Bright in 2010 to oversee its international projects, and he soon moved into the influenza and emerging diseases division. Ted Ross, a postdoc in Robinson's lab while Bright was doing his Ph.D. who now develops influenza vaccines at the University of Georgia, Athens, soon found himself seeking money from his old friend and collaborator—who sometimes turned him down. "He could listen to 100 proposals and not really flinch," Ross says. "He knew when people were kind of BS-ing him versus what was real."

During Bright's decade at BARDA, he oversaw billions of dollars of investments into countermeasures against potential chemical and nuclear threats, as well as drugs and vaccines for bioweapons and infectious diseases such as pandemic influenza, Middle East respiratory syndrome, Zika, and then COVID-19. He also helped shape the government's response to epidemics at home and abroad, represented the country at WHO, and regularly briefed Congress. As Bright documents in his whistleblower suit, he received "stellar performance appraisals" from no less than Kadlec.

BRIGHT SAYS TENSIONS with Kadlec predated the pandemic. The former Air Force officer, he asserts, believed BARDA should emphasize protecting against bioweapons over emerging infectious diseases. But COVID-19 made simmering bad blood boil. Bright and Kadlec battled about whether to fund specific masks, drugs, and vaccines to thwart SARS-CoV-2. Discord also grew after Bright visited the White House at the invitation of Peter Navarro, a trade adviser to Trump who early on advocated for more aggressive actions to stop the emerging virus. Bright says he lobbied for a crash program to make COVID-19 vaccines—a pandemic "Manhattan Project"—which Navarro spelled out in a memo on 9 February 2020 to the White House's Coronavirus Task Force. It took 2 months before HHS endorsed the concept, when it formed Operation Warp Speed.

Kadlec, Bright contends, was livid about the meetings with Navarro, which spurred the White House to push HHS to ramp up mask production, purchase potentially helpful drugs, and invest billions in vaccine development. "Kadlec was very uncomfortable with it," Bright says. "He actually could see that pressure was mounting. There were jokes in the hallway about Rick and his new friend, Peter."

Bright also convinced Congress that to better respond to the pandemic, BARDA needed a substantial infusion of funding that it could control—without ASPR's oversight. And then Bright shared with a reporter concerns about what he later called in his complaint HHS's "reckless and dangerous push" of hydroxychloroquine and its analogs as COVID-19 treatments. On 20 April 2020, Kadlec transferred him from BARDA to the National Institutes of Health (NIH) to oversee a new project on COVID-19 diagnostics.

Bright recognized the importance of ramping up diagnostics, but he charged that the transfer amounted to retaliation. In his richly detailed, 63-page complaint—which included Navarro's memo as one of 61 exhibits—Bright made allegations of fraud, cronyism, waste, and abuse of power within the federal COVID-19 response. According to Bright's lawyers, the Office of Special Counsel—an independent federal agency that oversees the Whistleblower Protection Act—promptly concluded that Bright's complaint documented a "substantial likelihood of wrongdoing" and referred the matter to HHS for an investigation.

Rick Bright drew the ire of former President Donald Trump with a whistleblower complaint and congressional testimony about the federal response to COVID-19.

On the morning of 14 May 2020, shortly before Bright was set to testify at a congressional hearing, Trump attacked. "I don't know the so-called Whistleblower Rick Bright, never met him or even heard of him," Trump tweeted. "But to me he is a disgruntled employee, not liked or respected by people I spoke to and who, with his attitude, should no longer be working for our government!"

At the hearing, Bright warned that the U.S. response to the pandemic had gone awry. "Without better planning, 2020 could be the darkest winter in modern history," he said. Asked about Trump's downplaying the pandemic's threat in the preceding months, Bright minced his words—but the criticism was plain: "I believe Americans need to be told the truth," he said. "And I believe that the best scientific guidance and advice was not being conveyed to the American public during that time."

Kadlec, who was not allowed to respond to Bright's allegations while he was ASPR, was outraged. "You want to talk about some hurt feelings? You got it here, buddy," says Kadlec, who now works on the minority Republican staff of the U.S. Senate health committee. He acknowledges that they had different views about BARDA's role from the start, but says his mandate required addressing both bioterror and infectious disease equally. Kadlec says shifting Bright to NIH wasn't retaliation, but rather part of the war on COVID-19. "This is my military background. The mission was to save lives, and the immediate mission is, 'We need diagnostics—Rick, go over there.'"

"None of [Bright's] allegations have been substantiated," asserts a former HHS lawyer who helped evaluate the whistleblower complaint and asked not to be identified. Many come down to what that lawyer, who was appointed by the Trump administration, sees as professional judgment calls. "Rick did not play well with others," the lawyer says. "He wanted to be the guy that called the shots, and he didn't want any criticism or oversight or accountability or checks on that authority."

After Bright's congressional testimony, Navarro called his former ally "a deserter

Pandemic preventers multiply

COVID-19 has illustrated the many problems the world has in spotting and responding to a global outbreak of a killer pathogen. A variety of new initiatives are forming to reduce the odds that history will repeat itself.

NEW EFFORT	LOCATION	FUNDER	AMOUNT (MILLIONS $)	FOCUS
Pandemic Prevention Institute	Rockefeller Foundation, Washington, D.C.	Rockefeller Foundation	150	Fill in gaps in global data collection and sharing, speed sequencing
Center for Forecasting and Outbreak Analytics	U.S. Centers for Disease Control and Prevention, Washington, D.C.	U.S. government's American Rescue Plan	Up to 500	Help U.S. mayors and governors respond to outbreaks with better data, analytics
World Health Organization Hub for Pandemic and Epidemic Intelligence	Robert Koch Institute and Charité University Hospital in Berlin	Germany	100	Cross-disciplinary effort to link surveillance data, improve analytics
Global Pandemic Radar	TBD	U.K. government, Wellcome Trust, G-20	TBD	Identify new COVID-19 variants, track new diseases, create surveillance hubs
Global Pandemic Prevention & Biodefense Center	Washington, D.C., area	U.S. government, philanthropy, global coalitions	Up to 2500	Initially make monoclonal antibodies against top 100 pathogens

in the war on the China virus." When Bright appeared on *60 Minutes* a few days later, Trump lashed out at him again, tweeting that he "fabricates stories," "spews lies," and is "a creep." Bright says unknown people subsequently began to call his relatives about his personal life, trying to dig up dirt about boyfriends, even though he is openly gay. "It was disgusting," he says.

Bright's fall from grace didn't last long. President-elect Joe Biden made him an adviser on a COVID-19 transition team. In August 2021, HHS settled the whistleblower suit with Bright, agreeing to back pay and damages for "emotional stress and reputational damage," according to his lawyers. They add that HHS has a separate, ongoing investigation into his allegations about contract improprieties and inappropriate responses to the pandemic. (HHS would not confirm or deny this.) And Bright is trying again to head off pandemics, this time from outside government.

PPI'S VAST OFFICE SPACE in Washington, D.C., isn't just pandemic empty—it's startup empty. As Bright begins to fill its cubicles with disease modelers, global health specialists, political scientists, epidemiologists, and health economists, he recognizes that his vision for PPI also still has many blanks to fill in—and knows he is entering an increasingly crowded and well-funded field. With backing from Germany, WHO will supplement its long-standing outbreak alert network with a hub in Berlin to analyze the incoming data and better plan responses. CDC is similarly launching a new group to aid local U.S. officials facing a spreading pathogen. "No one can do it all," Bright says. "We have to now come together to decide how we divide and conquer this ecosystem."

His recipe—building trust, collecting and sharing data, creating a reliable early warning system, and applying modern analytics—

is far from unique. But PPI should be able to react more quickly to gathering threats than other entities, says Manisha Bhinge, a health economist and computer scientist on the team. Organizations like WHO, she says, "don't have the space to be wrong or fail, so they often take a lot of time," Bhinge says. "Our ability as a nonstate actor, and having some degree of independence, is to complement them and be wrong occasionally."

Dylan George, an infectious disease modeler and a leader of CDC's new center, suggests PPI has another advantage: more freedom than government or quasi-governmental agencies like WHO to quickly fund projects that, say, test wastewater or scale up new diagnostics. "A little bit of money early in an outbreak can have an outsized impact," George says.

PPI now receives data from 30 partners that track diseases, including universities in several countries, the African and U.S. CDCs, large hospital systems, a livestock research institute in Kenya, a genomics center in Nigeria, and a molecular biology group in India. In addition to giving several of those partners a total of $20 million in grants, PPI has invested $4.5 million in a new South African center that aims to strengthen genomic surveillance of pathogens throughout that continent. "Today, health care providers, labs, and health departments are the primary source of information on new disease threats," says disease modeler Sam Scarpino, head of PPI's pathogen surveillance. "This traditional approach misses large swaths of the population who do not have access to quality health care. It also means that the first signs of an outbreak are detected weeks, if not months, after the emergence."

The next step is to make sense of the data, Bright says. "What no one has ever done is put together the new, brilliant architectural system to connect all this disparate data together and analyze it the way a hedge fund

manager analyzes all sorts of different data to understand where to invest," Bright says. "Yes, a lot of data are now being generated, but I liken it to severed arteries: There is just blood spurting everywhere."

The ultimate challenge remains translating the improved surveillance into earlier detection and better tracking. "After all the billions of dollars that I have been instrumental in spending to make a better vaccine, a better drug, or a better diagnostic, I've learned that none of that is effective if we don't have a global early warning system that is agnostic of politics, without financial pressure, without someone waiting for another government to share information, without us waiting until it gets to our borders," Bright says.

"It's not just for governments and policymakers," he says. "How do we reach my mother with a tool that can say, 'It's not safe to go outside today,' or 'I need to take extra precautions,' or 'I need to change my behavior?'"

In August 2021, Bright's mother was living with a relative who tested positive for SARS-CoV-2. She had received two doses of a vaccine, but she was at high risk of severe COVID-19, so she isolated herself at a different home. When she fell ill, Bright urged her to get a coronavirus test, but her doctor said it wasn't necessary because she was vaccinated. But, he recalls, "She would call and she could barely talk."

On 17 August, his mother died of what he is certain was COVID-19. "I'm heartbroken beyond words that this pandemic has now taken my mom," Bright tweeted. "Please get vaccinated, tested & please wear a high quality mask (over mouth & nose). I'm in so much pain from losing my mother, trust me, I don't want anyone else to feel this pain." ∎

INSIGHTS

LETTERS

NEXTGEN VOICES: ASK A PEER MENTOR

Back in person, back to the races

After months of remote work, many scientists have recently returned to their labs and classrooms. We asked young scientists to act as peer mentors by providing advice to this researcher, who feels left behind after a difficult pandemic experience. In response, young scientists ask questions to help this scientist reflect, share their own experiences, and offer advice about how to move forward. Read a selection of their thoughts below. Follow NextGen Voices on Twitter with hashtag #NextGenSci. **—Jennifer Sills**

Dear NextGen Voices peer mentors,

I am so excited that my university is finally allowing us to come back to work in a semi-normal way. However, I know that some of my labmates have been exceptionally productive during our time working remotely, publishing papers and finding new grants. Meanwhile, I have been overwhelmed by the stress of being away from my family, who live in a poorer country that has been hit especially hard, while also trying to care for my young children. I haven't published anything, and applying for grants feels impossible. As we move forward, how can I get my already sparse early-career CV back on track and compete with all the people who have been so productive during this difficult time?

Sincerely, Playing Catchup

Look back with pride

What happened over these past 20 months to make you feel proud? When the pandemic began, I worried constantly about my students' health; whether we could keep their projects going as they took on more responsibility in tighter quarters at home; and whether my career could survive the slowdown that would come from setting humane expectations. I still don't know the answer, but I'm alive. My family is healthy. And though my productivity did take a serious hit, I kept my group going and mentored a master's student to graduation! These were real successes, and I'm sure you've had some, too. Cherish them, and try not to dwell on the comparisons.

Shervin Fatehi
Department of Chemistry, The University of Texas Rio Grande Valley, Edinburg, TX 78539, USA. Email: shervin.fatehi@utrgv.edu

Have you considered that your productivity shouldn't be judged against your labmates with different circumstances? I've also felt like my research progress has been stagnant at times during the pandemic. However, I bet you made more progress

than you are giving yourself credit for! I suggest that you make a list of all the things you've accomplished during the pandemic—both academic and non-academic—and take a moment to be proud that you persisted. Then, compile a list of short- and long-term goals that you can use to improve your CV, and start going after these goals with a fresh mindset. I've found that taking one step at a time makes the transition back to semi-normal work more manageable.

Ashley Barbara Heim
Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 18453, USA. Email: abh229@cornell.edu

Share stories

Have you considered reaching out to the labmates that have been, from your perspective, exceptionally productive? The academic environment can sometimes be more competitive than cooperative, but talking to your colleagues might help. They might be struggling as well in ways that you can't see. In my experience, comparing myself to others' achievements without actively trying to understand how they got there was not constructive. I also recommend that you acknowledge and respect your own limits. Doing your best should not mean sacrificing your physical or mental health.

Camila Fonseca Amorim da Silva
School of Arts, Science, and Humanities, University of São Paulo, Araçazal, São Paulo, SP 3821000, Brazil. Email: camilafonsecaamorim@usp.br

What do you think about looking for a mentoring program or a mentor whose career was not smooth and easy? I recently joined a mentoring program as a mentee, and meeting others who have struggles similar to mine has been therapeutic. Their stories have made me realize that there is no single, correct path to career advancement. I am sure it feels like you fell far behind everyone else, but most of us struggled like you. Every path to career advancement is different, and talking to someone who has overcome challenges may be just what you need to get inspired.

Monika Lewinska
Biotech Research & Innovation Centre, University of Copenhagen, 2100 København, Denmark. Twitter: @LewinskaMonika

Reconsider priorities

Have you considered finding a science job outside of academia? In my experience, a career in academia is highly competitive and demanding: You need to be good at writing, presenting, and teaching; you have to study abroad and travel frequently; and you face long working hours and the whims of luck. A position as a scientific adviser at a company, a researcher at a start-up, or a writer or editor at a scientific journal or science communication organization would give you the satisfaction of scientific work without all the stress and pressure to publish and compete with your peers.

Anna Uzonyi
Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel. Email: anna.uzonyi@weizmann.ac.il

As you reflect on the past months, what life priorities have emerged as being most important to you? The demands in my work life increased exponentially during the pandemic. Simultaneously, my wife and I moved out of state to provide in-home care to my 100-year-old mother-in-law. This experience reminded me that we all go through dormant periods in our careers and lives. Rather than making social comparisons, try working with your most trusted advisers to establish achievable and realistic benchmarks and timelines that will allow you to make meaningful advancements toward accomplishing your most important personal and professional goals.

Bradley J. Cardinal
Kinesiology Program, School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR 97331, USA. Email: brad.cardinal@oregonstate.edu

Adjust expectations

Please do not compare yourself to others. Keeping those kids alive during the pandemic is a tough job; do not worry about work too much at a time like this. Just know you will feel your life is back on track soon. Focus on your priorities and find your rhythm again slowly. I am a mother of twin toddlers. My co-workers have also fallen behind, and we all show compassion toward each other. Don't focus on what else you should be doing. Just focus on what is important now.

Dieddra Atondo
STEM Program, College of the Sequoias, Visalia, CA 93292, USA. Email: dieddraa@cos.edu

When we worry about other people's progress, we often forget that all of them started from scratch at some point. My daughter was born in 2021. It has been tough to look after her while lecturing in an online environment. She is the priority, of course, so I'm not focused on making progress as fast as my colleagues; I'm focused on the progress I'm able to make. We endure and we will succeed because we have a family, not in spite of it. With that in mind, get a good cup of coffee, select your best idea, and focus on it. A good, well-written idea will always be welcomed by funding boards, no matter how much or how little you have accomplished in the past year.

Wagner Eduardo Richter
Federal University of Technology–Paraná, Paraná, Brazil. Email: richter@utfpr.edu.br

Be pragmatic

Have you thought about listing and ranking the stresses you faced? In my experience, solving problems in order of importance is a good way to cope with complex challenges. As a graduate student, I faced a series of stressors, including experiments, writing, job searching, and family. My mentor suggested that I write down these challenges in order of priority and then solve them one by one. In your case, you might rate your lack of papers and grants most important. To address this, you could organize your existing experimental results and ideas, prepare and submit manuscripts based on existing results, and apply for some grants. Next, you could do some experiments, which will cover the shortage of papers and provide preliminary results for grant applications.

Bo Cao
Core Research Laboratory, The Second Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xincheng District, Xi'an, Shaanxi 710004, China. Email: bocao@vip.qq.com

The experience you have described has been pervasive. Have you considered gaining some feedback by networking virtually? As the time in research decreased, I tried to increase my online presence, especially by reviewing papers and grants and doing outreach activities like Skype-A-Scientist. These options are still available. You may also want to ask your university administration if they can provide any extensions and ask your departmental mentors to help you extend your network to gain speaking engagements and collaborations. Remember to schedule time to think, read, and plan. This will allow you to take care of yourself as an individual, not as a scientist or a parent or a family member.

Naga Rama Kothapalli
Trudeau Institute, Saranac Lake, NY 12983, USA. Twitter: @rama_k_n

PERSPECTIVES

CORONAVIRUS

Using correlates to accelerate vaccinology

Correlates and surrogates of desired outcomes are valuable but have limitations

By **Peter J. M. Openshaw**

There have been many reports of large-scale vaccine studies showing that various severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines give almost complete protection against severe COVID-19 and incomplete but dwindling protection against infection of the nose and lungs. Given the costs and difficulty of field studies involving thousands of people, which are necessary to show vaccine efficacy, the hunt for immune correlates of protection (COPs; laboratory measurements that predict the outcomes of large-scale studies) has become intense. On page 43 of this issue, Gilbert *et al.* (*1*) report the use of a technique called mediation analysis to examine data from a trial of the mRNA-1273 vaccine from Moderna to infer that virus neutralizing antibody (VN-Ab) accounts for ~60% of protection. They propose that VN-Abs might provide a reliable COP, which could be used to support the approval of future COVID-19 vaccines, bypassing the need for large trials.

Vaccines create resistance to infection by creating a set of immunological barriers that are deployed at different phases of pathogen entry and replication. These immune responses control engagement of the virus with target host cells in the nose and lungs, suppress the propagation of infection within the mucosa, modulate the inflammatory response to viral invasion, and inhibit the dissemination of the virus with consequent extrapulmonary involvement in those with severe disease (see the figure). Understanding the immunology that underlies protection is of great value to enable monitoring of vaccine duration and protective efficacy, as well as in aiding the development of new or improved vaccines that induce key elements of the protective immune response. Any single measure of immunity, however convenient, is never going to reflect this multilayered response. Therefore, can one measure sufficiently reflect the full palette of immune responses and thus reliably predict protection?

The discovery of reliable COPs can be game-changing in the production and licensure of vaccines. For example, influenza vaccines are updated twice a year according to the prevalent and projected circulating strains of influenza virus, based on the ability of an updated vaccine to induce an antibody response in a small group of volunteers, bypassing the need for large, slow, and expensive field trials. The use of such a correlate depends on a clear definition of the type of protection that is required and a solid relationship between the measure and the desired outcome.

In assessing the effects of vaccines designed to prevent COVID-19, VN-Abs, as described by Gilbert *et al.*, are a logical and measurable target. VN-Abs coat virus particles, thereby preventing virus entry into cells. However, protection induced by two doses of messenger RNA (mRNA) vaccine may not last, declining to nonprotective levels in some vaccinees after 5 or 6 months (*2*). By focusing on an early serological COP, the opportunity may be missed to develop vaccines that mediate broad and durable protection. T cell responses may be important in supporting VN-Ab–mediated protection and helping to maintain long-term protection, but each component of immune memory

National Heart and Lung Institute, Imperial College London, London, UK. Email: p.openshaw@imperial.ac.uk

Sequential events during SARS-CoV-2 infection

The immunopathogenesis of COVID-19 can be depicted as an early viral phase followed by an inflammatory phase, which may be restricted or spread outside the respiratory tract. Immune control occurs at various stages (teal arrows), whereas inflammation is important in disease pathogenesis (red arrows). Local and systemic immunoglobulins operate at multiple levels, direct and indirect; VN-Abs are detectable from approximately day 10 for at least 3 months.
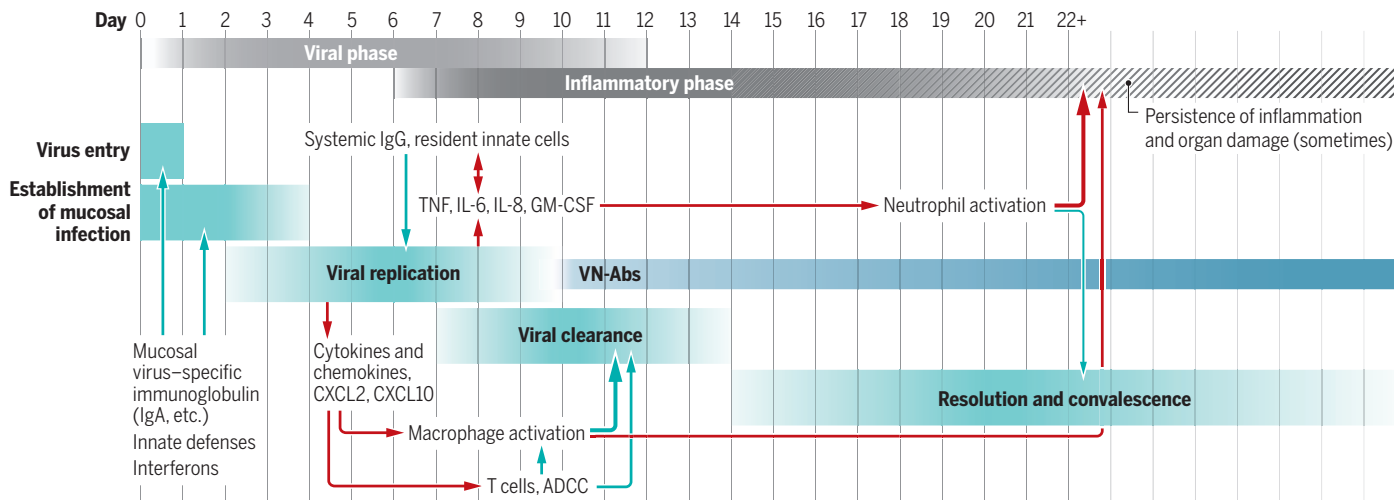
ADCC, antibody-dependent cellular cytotoxicity; CXCL, C-X-C chemokine ligand; GM-CSF, granulocyte-macrophage colony-stimulating factor; Ig, immunoglobulin; IL, interleukin; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; TNF, tumor necrosis factor; VN-Abs, virus-neutralizing antibodies.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

exhibits distinct kinetics, which makes measures of such protection challenging (3). Cell-mediated immunity has the added benefit of conferring protection against a wide variety of viral strains and may provide additional protection against emerging variants (4). The protective role of mucosal T cells, including resident memory T cells (5), emphasizes the importance of studying immune responses in relevant tissue sites, not just in the blood (6).

There have been other attempts to define COPs against SARS-CoV-2 infection and disease across a range of vaccines (7, 8). These studies generally support the premise that circulating VN-Ab is a good indicator of protection, not only against severe disease but also (if amounts are very high) against viral replication in the mucosa (with or without symptoms), thereby reducing transmission. This does not mean that it is circulating antibody that is causing protection: Antibody in the serum does not normally diffuse (or get transported) into mucosal fluids. Protection against superficial infection depends on mucosal antibodies [including immunoglobulin A (IgA)] and possibly on antiviral T cells that reside within the linings of the respiratory tract. T cells floating free in the blood cannot have antiviral action; they only act through direct contact with infected cells. It could be that serum VN-Ab initially correlates with protection but fades and is replaced by other forms of immunity (such as T cells) that become important later. So, is the COP stable over time?

With the potential for a disconnection to develop between a COP and the desired effect of future vaccines, such correlates require constant scrutiny and reevaluation. For example, the live attenuated influenza virus vaccine that has been widely adopted for use in children does not induce acceptable concentrations of serum N-Abs. Instead, protective immune responses that have been demonstrated to result from the use of this vaccine require techniques of site-specific mucosal monitoring (9). COPs should therefore be validated with vaccines that work in a variety of ways, not just one type of vaccine. COPs should also be robust in diverse settings, including after natural infection and at delayed time points.

There are clear advantages and pitfalls to the use of correlates and surrogates in diverse medical fields (10). For example, measuring viral load to determine the effect of antiviral drugs in those living with hepatitis C or HIV is a practical shortcut in tailoring treatment to prevent disease and early death. The relationship between viral suppression and clinical benefit is so clear that there is no need to doubt viral load as a surrogate for long-term benefit. However, the relationship between a marker and the desired outcome can be insecure or even misleading. For example,

the use of forced airflow measurements in evaluating the reversibility of airflow obstruction induced by β2 agonists in patients with asthma might distract clinicians and patients from also treating airway inflammation that is fundamental to pathogenesis. Indeed, long-acting β2 agonists may enhance inflammation (11) and so must be paired with anti-inflammatory treatments, such as inhaled steroids. Another example is in the treatment of osteoporosis: The measurement of bone density is widely accepted as a predictive marker of future bone fractures, but measures that increase bone density may not necessarily reduce fractures and might even increase them (12).

An accurate and stable COP can save vaccine developers from performing large and expensive trials to demonstrate the efficacy of new or updated vaccines, can predict the effects of established vaccines against new variants of the pathogen, and may expedite regulatory approval of updated or improved vaccines. For example, the emergence of the Omicron variant of SARS-CoV-2 (13) required urgent investigation into the likely protective efficacy of existing vaccines and possibly the redesign of vaccines to match Omicron's mutated spike protein. A reliable COP could allow this to be done without large field trials, accelerating vaccine rollout and saving many lives. However, uncritical adoption of a COP may have the perverse effect of focusing future vaccine development on meeting the correlate rather than preventing infection and reducing disease and death. All is well if the correlate is reliable in the face of new viral variants and vaccines, but direct clinical outcomes need to be kept firmly in focus. ∎

REFERENCES AND NOTES

1. P. B. Gilbert et al., Science **375**, 43 (2022).
2. E. G. Levin et al., N. Engl. J. Med. **385**, e84 (2021).
3. J. M. Dan et al., Science **371**, eabf4063 (2021).
4. Y. Peng et al., Nat. Immunol. **21**, 1336 (2020).
5. A. Guvenel et al., J. Clin. Invest. **130**, 523 (2020).
6. D. L. Faber, Nature **593**, 506 (2021).
7. D. S. Khoury et al., Nat. Med. **27**, 1205 (2021).
8. D. Cromer et al., Lancet Microbe 10.1016/S2666-5247(21)00267-6 (2021).
9. J. Dunning et al., Mucosal Immunol. **13**, 566 (2020).
10. US Food and Drug Administration, "Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure" (2021); https://bit.ly/3oW1EEu.
11. A. I. Ritchie et al., Am. J. Respir. Cell Mol. Biol. **58**, 128 (2018).
12. S. Ma et al., Sci. Rep. **7**, 43399 (2017).
13. World Health Organization, "Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern" (2021); https://bit.ly/3dFgLNh.

T cells to fix a broken heart

In vivo engineered T cells provide a promising approach to treat cardiac diseases

By **Torahito A. Gao**[1] and **Yvonne Y. Chen**[1,2,3]

Fibrosis that results from excessive extracellular matrix protein deposition by activated cardiac fibroblasts is a hallmark of heart disease and plays a critical role in disease progression to heart failure (1). However, therapies targeting this cardiac fibrosis remain limited. On page 91 of this issue, Rurik et al. (2) describe a new approach to eliminate activated fibroblasts by harnessing the power of engineered T cells. Lipid nanoparticles (LNPs) carrying messenger RNA (mRNA) that encodes a chimeric antigen receptor (CAR) are used to generate CAR T cells in mice, yielding a therapeutic T cell population that is capable of ablating pathogenically activated fibroblasts and attenuating cardiac fibrosis.

CARs are synthetic receptors that allow immune cells—usually T cells—to recognize targeted antigens and initiate antigen-specific immune responses (3). In conventional CAR T cell therapy, T cells are isolated from the blood of patients or healthy donors, genetically modified to express a CAR targeting a disease-associated antigen (e.g., a surface protein found on tumor cells), and expanded ex vivo before infusion into the patient. CAR T cell therapy has shown substantial clinical efficacy against various hematological malignancies, becoming the first genetically modified cell therapy to receive US Food and Drug Administration approval (4). Following these successes, efforts are underway to expand CAR T cell therapy to other indications, such as infectious and autoimmune diseases (5).

Previously, CAR T cells targeting fibroblast activation protein (FAP) were shown to specifically target activated fibroblasts and significantly reduce cardiac fibrosis in a

[1]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. [2]Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA, USA. [3]Parker Institute for Cancer Immunotherapy Center at UCLA, Los Angeles, CA, USA. Email: yvchen@ucla.edu

In vivo generation of transient engineered T cells

Messenger RNA (mRNA) that encodes fibroblast activation protein (FAP)–targeted chimeric antigen receptor (CAR) is carried by lipid nanoparticles (LNPs). These are injected into mice and delivered to T cells, resulting in transient anti-FAP CAR expression. The anti-FAP CAR directs T cells to eliminate activated fibroblasts and alleviate cardiac fibrosis in a mouse model of hypertensive cardiac injury.
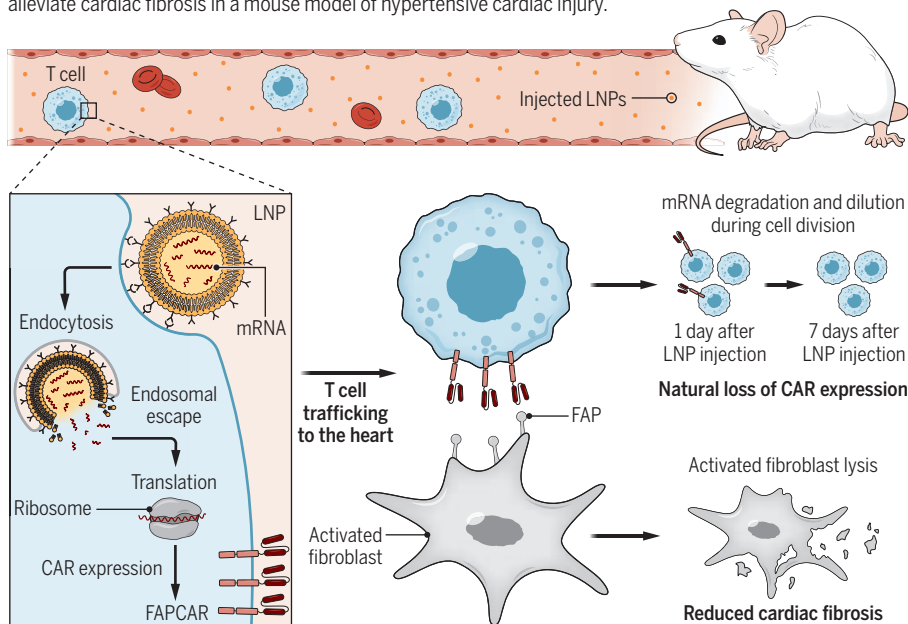
mouse model of hypertensive cardiac injury (6). In this prior work, T cells were stably integrated ex vivo with the CAR-encoding transgene through retroviral transduction, resulting in long-term CAR expression. Extensive clinical experience in the context of cancer therapy has shown that autologous CAR T cells can persist for months or years after adoptive transfer into the patient, and this long-term persistence plays an important role in therapeutic efficacy and durability of response (7). However, cardiac injury has a different temporal profile compared with cancer. Although FAP is generally expressed at low levels in healthy tissue, it is up-regulated during the normal wound-healing process, which involves fibroblast activation (8). As such, long-term persistence of FAP-targeting CAR T cells could present a safety risk if the patient sustains injury after treatment.

To balance the need for robust CAR T cell activity against pathogenic fibroblasts while minimizing long-term safety risks, Rurik et al. developed a method to generate transient CAR T cells in vivo, bypassing the need for ex vivo cell manufacturing (see the figure). In this approach, mRNA that is chemically modified for stability and encodes the FAPCAR is packaged into LNPs decorated with CD5-targeting antibodies to facilitate selective uptake by T cells. Upon LNP uptake, the mRNA is translated to generate CAR T cells. Given the inherent instability of mRNA, CAR expression is transient, and no T cell will remain per-

sistently CAR⁺ by design. Rurik et al. confirmed the emergence of anti-FAP CAR T cells in mice 24 hours after LNP injection, as well as the decline of these CAR T cells to undetectable levels within 7 days. Using a mouse model of hypertensive cardiac injury and fibrosis, the authors demonstrated that transient FAPCAR T cells substantially reduce ventricular fibrosis and improve various cardiac functions, as assessed by echocardiography. These findings provide a rationale for the use of transient CAR T cells and demonstrate the applicability of CAR T cell therapy beyond oncology applications.

Heart disease remains the leading cause of death globally, requiring therapies that can be manufactured at large scale and at reasonable cost. The success of the COVID-19 mRNA-LNP vaccines (9) suggests that LNP-based mRNA delivery could be a viable path to the development of scalable immunotherapies, providing a more economical alternative to the conventional method of ex vivo CAR T cell manufacturing that requires extensive infrastructure and high-cost reagents (10). In vivo CAR T cell generation had previously been reported with the use of adeno-associated viruses instead of mRNA-LNPs (11). A potential advantage of the mRNA-LNP platform is the ability to flexibly decorate the LNP with antibodies to increase target specificity, as well as greater ease of clinical-grade reagent production. Furthermore, compared with conventional ex vivo CAR T cell manufacturing methods, the virus-free, in vivo CAR T cell manufacturing method detailed

by Rurik et al. eliminates the risk of unintended host-genome alterations mediated by lentiviral or retroviral integration (12) as well as the need for lymphodepleting chemotherapy before adoptive transfer, thus removing a source of considerable toxicity that is typically associated with CAR T cell therapies.

FAP-targeting CAR T cells were previously evaluated, with some groups reporting minimal side effects (13) and others reporting lethal bone toxicity and cachexia (muscle wasting) in mice due to interactions with bone marrow stem cells (14). No overt toxicity was observed by Rurik et al., and LNP-mediated anti-FAP CAR expression was largely restricted to the spleen. Nevertheless, these discrepancies in toxicity must be carefully evaluated before translation into the clinic, and a more in-depth characterization of the interaction between FAPCAR T cells and activated fibroblasts within the fibrotic tissue will be necessary to demonstrate antigen-specific CAR T cell activity at the disease location. Additionally, although Rurik et al. demonstrated that interstitial fibrosis was largely reduced with mRNA-LNP treatment, it was noted that perivascular fibrosis persisted beyond the treatment duration owing to the presence of activated fibroblasts that do not express FAP. Perivascular fibrosis and associated inflammation can decrease the contact area of neighboring cardiac tissue with blood vessels and reduce oxygen and nutrient availability that may exacerbate pathological cardiac remodeling (1). Further investigation into the biology of perivascular fibrosis could uncover additional therapeutic targets.

The work of Rurik et al. provides a strong rationale for the broadening of immunotherapies into disease areas with unmet needs, and FAPCAR T cells have already been evaluated in the context of cancer (15). Other fibrotic diseases or associated disorders, such as chronic inflammatory diseases with FAP⁺ fibroblasts, may benefit from this approach as well. This work represents an exciting step toward translating personalized immunotherapies into accessible and affordable "off-the-shelf" immunotherapies. ∎

REFERENCES AND NOTES

1. J. G. Travers et al., Circ. Res. 118, 1021 (2016).
2. J. G. Rurik et al., Science 375, 91 (2022).
3. M. Hong et al., Cancer Cell 38, 473 (2020).
4. M. C. O'Leary et al., Clin. Cancer Res. 25, 1142 (2019).
5. M. Seif et al., Front. Immunol. 10, 2711 (2019).
6. H. Aghajanian et al., Nature 573, 430 (2019).
7. L. Jafarzadeh et al., Front. Immunol. 11, 702 (2020).
8. P. Bainbridge, J. Wound Care 22, 407, 410 (2013).
9. W. C. Koff et al., Sci. Transl. Med. 13, eabd1525 (2021).
10. S. Rafiq et al., Nat. Rev. Clin. Oncol. 17, 147 (2020).
11. W. Nawaz et al., Blood Cancer J. 11, 119 (2021).
12. R. M. David, A. T. Doherty, Toxicol. Sci. 155, 315 (2017).
13. L. C. S. Wang et al., Cancer Immunol. Res. 2, 154 (2014).
14. E. Tran et al., J. Exp. Med. 210, 1125 (2013).
15. S. Kakarla et al., Mol. Ther. 21, 1611 (2013).

GRAPHIC: A. MASTIN/SCIENCE

Paleobiology provides glimpses of future ocean

Fossil records from tropical oceans predict biodiversity loss in a warmer world

By **Moriaki Yasuhara**[1] and **Curtis A. Deutsch**[2]

In the ocean, the most powerful forces of climate change may not be as visually striking as the storms and wildfires on land. Yet small changes in seawater temperature, acidity, or oxygen content can have substantial impacts on the marine biota (1). Some species will relocate, whereas those with limited mobility will either adapt to the changing conditions or perish (2). The difference between these responses across thousands of species will determine the structure of ecological communities for centuries to come. However, the outcomes are notoriously difficult to detect in sparse and patchy historical data, and even harder to attribute, without a time machine. On page 101 of this issue, Salvatteci et al. (3) document changes in the abundance of fish vertebrae, alongside indicators of the physical and chemical environments, by using sediments off the west coast of South America deposited over the past 130,000 years.

The new data provide a window into a crucial time: the last interglacial period, when Earth's climate was warmer than it is today. The data reveal a propitious place:

a warm tropical region with low oxygen—properties that synergistically shape species physiology and biogeography (4). The region's Humboldt Current also lies at the heart of the El Niño phenomenon (the periodic warming of the ocean surface in the central and eastern tropical Pacific Ocean) whose effect on fish populations was known to indigenous fishing communities long before modern scientists formally identified the phenomenon (5).

The fossil records obtained by Salvatteci et al. reveal that the Humboldt Current during the last interglacial period was not dominated by the anchovy species that drive the extraordinary fishery productivity today. Instead, the most abundant fossils were of smaller, goby-like fishes. This shift to smaller species was not accompanied by changes in the productivity of the area, so the change was unlikely to be driven by food supply. Instead, the shift coincides with indicators of a more extreme oxygen minimum zone, suggesting that regional O_2 content was lower. Warmer water and lower O_2 concentration are both factors that can produce smaller fish sizes, either within a population or among different taxa as water properties change (6). The shift to smaller

fish in a warmer Humboldt Current was not achieved by an existing species' adaptation to reach smaller body sizes, but by an ecological replacement of anchovy with a dominant species of smaller fishes.

If organismal growth is limited by O_2 concentration in seawater, one may expect species to adapt by having a smaller size when fully grown (7). However, physiologists have argued that tolerance to lower environmental O_2 concentration is independent of body size both within and among species and is an evolutionary necessity for growth in water with limited O_2 availability (8). Salvatecci et al. also found that the size of fossil vertebrae within species showed no signs of shrinking, which goes against the prediction for smaller organism sizes as an intraspecific adaptation to maintain aerobic balance in a warmer and lower O_2 environment. However, the evidence may also be interpreted as a reflection of the migratory strategy of more-mobile species, such as the anchovy (9). Among less-mobile species, fossil records show that body-size reductions within the taxa often accompany climate warming (10). Whether the small size of the goby-like fishes was essential or coincidental, their interglacial ascendance is crucial to the interpretation and generalization of the results. Answering that question will require more explorations in paleobiology, loaded with more data on the modern distribution and physiological traits of fossilized species.

Whether body sizes played a causal or incidental role in species turnover in the Humboldt Current, the fossil records add to the evidence that the community structure of tropical marine ecosystems responds strongly to climate across a wide range of magnitudes and time scales. Species displacements may have a domino effect, with migrations out of the tropics, because of increasing ocean temperature and/or declining oxygen concentration, affecting the competitive fitness of extratropical species, which then propagate toward higher latitudes near the poles. These changes may increase diver-

The future of ecosystems in tropical, temperate, and polar oceans

A warming and deoxygenating ocean will make species smaller and push them from the tropical zone to the temperate zone, from the temperate zone to the polar zone, and from the polar zone to extinction, resulting in a loss of biodiversity in the tropics and higher biodiversity in higher latitudes. Overexploitation and other direct human impacts are not considered in this conceptual model.

GRAPHIC: KELLIE HOLOSKI/SCIENCE
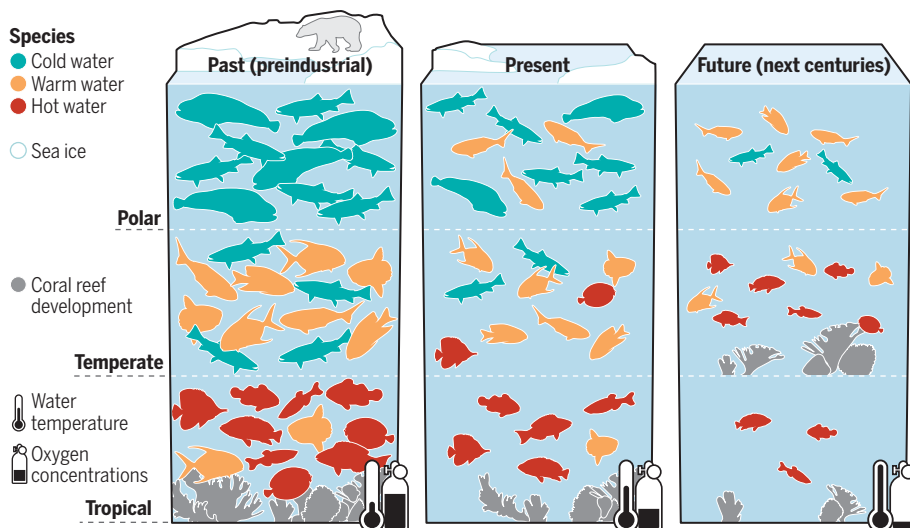
sity in the polar region with migrations of diverse warmer-water species, while wiping out less-diverse but distinct polar endemics.

This domino effect of species displacements leads to the prediction that warming may reduce tropical diversity while causing extinction for polar endemic species (see the figure). Although the reality is certainly far more complex, growing evidence at multiple time scales suggests that the broad brushstrokes of this simple picture are correct. Indeed, warming-induced tropical biodiversity losses have been reported in modern marine biological records (*11*), and they are largely consistent with reports from paleobiology (*12, 13*).

Different areas of the tropics may respond differently. Although the tropical oceans contain the warmest and lowest O_2-containing waters, the two extremes are not ubiquitous in all tropical oceans. The eastern tropical Pacific may contain very low O_2 below the surface but is relatively cool for the tropics, allowing replacement with warmer-water fishes. Such replacement may not be possible on the other side of the tropical Pacific, where the waters are warmer. In addition, this climate change–triggered loss of tropical diversity may not be confined to the oceans, as the exceedance of thermal tolerance on land is also projected to be the most severe in the tropics (*14*). The climate-driven reshuffling of the ecosystem structure will be profound. The findings of Salvatteci *et al.* are the latest addition to the emerging evidence that a warmer future will alter ecological communities in tropical oceans, which disproportionally affect developing countries, where the reliance on small-scale fishing is especially high. ∎

REFERENCES AND NOTES

1. L. Alexander *et al.*, *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
2. M. L. Pinsky, R. L. Selden, Z. J. Kitchel, *Annu. Rev. Mar. Sci.* **12**, 153 (2020).
3. R. Salvatteci *et al.*, *Science* **375**, 101 (2022).
4. H.-O. Pörtner, *J. Exp. Biol.* **213**, 881 (2010).
5. F. P. Chavez, A. Bertrand, R. Guevara-Carrasco, P. Soler, J. Csirke, *Prog. Oceanogr.* **79**, 95 (2008).
6. W. C. E. P. Verberk *et al.*, *Biol. Rev. Camb. Philos. Soc.* **96**, 247 (2021).
7. W. W. L. Cheung, D. Pauly, in *Explaining Ocean Warming: Causes, Scale, Effects and Consequences* (IUCN, Gland, 2016), pp. 239–253.
8. S. Lefevre, D. J. McKenzie, G. E. Nilsson, *Glob. Change Biol.* **23**, 3449 (2017).
9. E. M. Howard *et al.*, *Sci. Adv.* **6**, eaay3188 (2020).
10. G. Hunt, K. Roy, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1347 (2006).
11. C. Chaudhary, A. J. Richardson, D. S. Schoeman, M. J. Costello, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015094118 (2021).
12. W. Kiessling, C. Simpson, B. Beck, H. Mewis, J. M. Pandolfi, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21378 (2012).
13. M. Yasuhara *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 201916923 (2020).
14. C. A. Deutsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6668 (2008).

Setting boundaries for tissue patterning

Engineers point the way toward more complex and homogeneous intestinal organoids
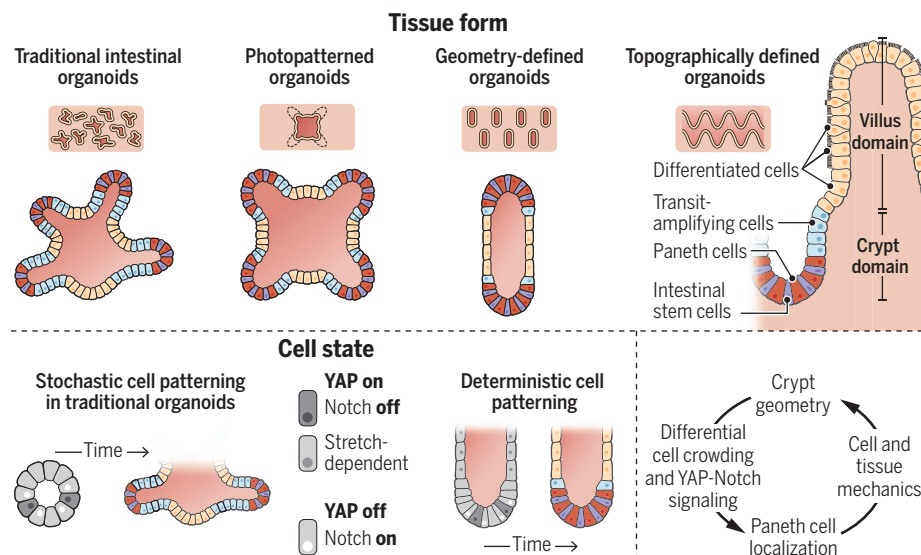
By **Tyler R. Huycke** and **Zev J. Gartner**

T he functional units of structure in the small intestine are crypts and villi, where compartmentalized zones of stem, progenitor, and differentiated cell types carry out essential roles in tissue homeostasis, nutrient absorption, and barrier maintenance. These structures can be modeled in vitro by using organoids that self-organize from isolated crypts or intestinal stem cells (ISCs). However, organoids suffer from heterogeneity in size, geometric patterning, cell type composition, and overall morphology, which impairs reproducibility and limits their applications. On page 40 of this issue, Gjorevski *et al.* (*1*) report the use of multiple bioengineering strategies to generate more structurally complex and reproducible intestinal organoids. They find that tissue geometry instructs organoid self-organization into patterned crypt and villus domains by influencing regional differences in cell crowding and downstream Yes-associated protein (YAP)–Notch signaling. These findings add to evidence that tissue structure is sufficient to specify cell state and behavior.

To spatially control organoid crypt formation, the authors use photochemistry to pattern regions of soft and stiff extracellular matrix (ECM) under the basal surface of the epithelium. Organoids bud precisely in the softened regions into crypt-like domains that harbor ISCs and Paneth cells (secretory cells located between ISCs), whereas the nonbudding regions resemble villus-like domains with differentiated absorptive cells. Gjorevski *et al.* reasoned that altering substrate geometry might act analogously to altering substrate stiffness to pattern cell states within the intestinal epithelium. To test this idea, they use a microscale engineering approach originally designed to study branching in the mammary epithelium (*2*), which involves seeding organoids in pill-shaped cavities within a collagen hydrogel. They find that ISCs localize to the curved ends of the cavities, whereas differentiated cells localize to the flat sides. This demonstrates that simple

Engineering intestinal organoids

Methods to engineer more complex and homogeneously patterned organoids reveal how cell patterning arises. Key parameters in defining cell state are differential cell crowding and Yes-associated protein (YAP)–Notch signaling. A dynamic reciprocity exists between tissue geometry and cell state that patterns intestinal organoids.

Tissue form

Traditional intestinal organoids

Photopatterned organoids

Geometry-defined organoids

Topographically defined organoids

Villus domain

Differentiated cells

Transit-amplifying cells

Paneth cells

Intestinal stem cells

Crypt domain

Cell state

Stochastic cell patterning in traditional organoids

— Time →

YAP on Notch off

Stretch-dependent

YAP off Notch **on**

Deterministic cell patterning

— Time →

Crypt geometry

Cell and tissue mechanics

Paneth cell localization

Differential cell crowding and YAP-Notch signaling

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

boundary conditions—in this case, provided by tissue geometry—can instruct cells to self-organize deterministically into domains that resemble the crypt-villus axis (see the figure).

An instructive role for tissue boundaries in cell patterning has long-standing roots in developmental biology, and recent efforts have begun applying these concepts to engineering organoids and other developmental processes in vitro (3). For example, the initial geometry of mammary epithelial tubules determines branching patterns (2), and the geometry of human pluripotent stem cell colonies influences their self-organization into structures that resemble the neural tube (4). How is the geometry of the tissue boundary mechanistically linked to patterning? Broadly, these effects can be attributed to the dynamic reciprocity that exists between a cell and the mechanical and molecular components of its microenvironment—for example, the ECM (5).

A cell's ability to sense the microenvironment can occur through myriad mechanisms, including through its physical deformation. In intestinal organoids, this appears to be mediated by the mechanosensitive transcription factor YAP. Preceding the up-regulation of activity of ISC markers in geometrically constrained organoids, regions of high curvature develop increased cell crowding and cytoplasmic (inactive) YAP, relative to neighboring flat regions with nuclear (active) YAP. Suppression of crypt identity occurs in flat regions of active YAP. Subsequently, patterning of the crypt domain begins in the curved ends where some cells then activate YAP and the Notch ligand delta-like 1 (DLL1), which triggers Notch signaling in neighboring cells that retain inactive YAP. This symmetry-breaking lateral inhibition event culminates in the differentiation of Paneth cells that maintain ISCs locally. A similar mechanism acts stochastically in traditional organoid culture to determine the site of crypt formation (6), where curvature is subsequently generated through differential actomyosin contraction and luminal pressure (7).

Although regions with villus-like cell composition exist in organoids, the morphology of these domains does not resemble villi in vivo, and thus, their function cannot be accurately modeled. Having identified engineering principles for deterministic patterning of crypt-like domains, Gjorevski et al. turn to the patterning of villus domains. They seed cells onto engineered hydrogels that mimic the surface topography of the small intestine, with villus-shaped pillars located above well-shaped crypts. Cells on these topographically rich surfaces self-organized into both crypts

and villi, further supporting the idea that intestinal epithelia can innately pattern in a manner dependent on the tissue geometry and independent of additional cues, such as signals derived from stromal cells that are absent in these epithelial cell-only organoids.

The findings of Gjorevski et al. raise the question of how well intestinal organoid morphogenesis recapitulates intestinal morphogenesis during development—a point of recent debate (8, 9). For example, during normal development, villi emerge before crypts, suggesting that the location of villi may determine the location of crypts. Indeed, the morphogen gradients that emerge during folding of the epithelium shape signaling gradients that localize ISC progenitors to the intervillus spaces that later give rise to crypts (10, 11). These same signaling gradients are also important for homeostasis in the adult (12). How are these two determinants of cell state, one geometry-based and the other signaling-based, normally integrated during development, homeostasis, and regeneration? Whether these are redundant mechanisms or they work in concert to ensure that cell identity and tissue structure mutually reinforce one another is unclear.

What is now clear, however, is that both geometry and epithelial-autonomous signaling pathways are sufficient to pattern cell identity in the small intestine. For example, previous work established that intestinal organoid monolayers grown in the absence of any surface contours seem to display zones with crypt and villus identity that are likely determined by the same bone morphogenetic protein (BMP) and Wnt signaling axes that set up these domains in vivo (13, 14). Therefore, future studies should investigate whether and how these mechanisms work in concert to ensure that tissue form, and thus tissue function, are robust outputs of development and homeostasis. ∎

REFERENCES AND NOTES

1. N. Gjorevski et al., Science **375**, eaaw9021 (2021).
2. C. M. Nelson et al., Science **314**, 298 (2006).
3. I. Martyn, Z. J. Gartner, Dev. Biol. **474**, 62 (2021).
4. E. Karzbrun et al., Nature **599**, 268 (2021).
5. M. J. Bissell et al., J. Theor. Biol. **99**, 31 (1982).
6. D. Serra et al., Nature **569**, 66 (2019).
7. Q. Yang et al., Nat. Cell Biol. **23**, 733 (2021).
8. J. Guiu, K. B. Jensen, Cell. Mol. Gastroenterol. Hepatol. **13**, 1 (2021).
9. S. Sugimoto, T. Sato, Cell. Mol. Gastroenterol. Hepatol. **13**, 195 (2021).
10. A. E. Shyer et al., Cell **161**, 569 (2015).
11. J. Guiu et al., Nature **570**, 107 (2019).
12. N. McCarthy et al., Cell Stem Cell **26**, 391 (2020).
13. C. A. Thorne et al., Dev. Cell **44**, 624 (2018).
14. C. Pérez-González et al., Nat. Cell Biol. **23**, 745 (2021).

Department of Pharmaceutical Chemistry, University of California San Francisco, CA, USA. Email: zev.gartner@ucsf.edu

Archaeal nitrification without oxygen

The single-cell organism can self-produce oxygen for ammonia oxidation

By **Willm Martens-Habbena**[1] *and* **Wei Qin**[2]

Ammonia-oxidizing archaea (AOA) constitute up to 30% of the microbial plankton in the oceans and play a key role in the marine nitrogen and carbon cycle. Together with nitrite-oxidizing bacteria (NOB), they oxidize ammonia to nitrate—the predominant inorganic nitrogen source in the sea. AOA and NOB are thought to rely on molecular oxygen (O_2), but recent studies revealed their presence in strictly anoxic marine oxygen minimum zones, challenging this paradigm (1–3). On page 97 of this issue, Kraft et al. (4) demonstrate that the marine AOA *Nitrosopumilus maritimus*, under anoxic conditions, produces O_2 for ammonia oxidation by itself while simultaneously reducing nitrite to nitrous oxide (N_2O) and dinitrogen (N_2). These results provide a possible explanation for the presence of AOA in marine oxygen minimum zones, where they may have an important role in nitrogen loss. The findings may further have implications for the evolution of the nitrogen cycle on Earth.

The discovery of ammonia oxidation within the domain Archaea has substantially changed the understanding of the global nitrogen cycle (5). Phylogenomic and molecular dating analyses using different clock models have suggested that ammonia oxidation in the Archaea first arose at least a billion years ago (6, 7). Modern AOA are among the most abundant and ecologically successful microbial groups, showing extraordinary adaptations to low nutrient and energy fluxes in the deep oceans (8, 9). They also possess the most energy-efficient aerobic pathway for carbon fixation (10). As predominant ammonia oxidizers in the oceans, AOA are considered a primary source of ocean-

[1]Department of Microbiology and Cell Science, University of Florida, Institute for Food and Agricultural Sciences, Fort Lauderdale Research and Education Center, Davie, FL 33314, USA. [2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA. Email: w.martenshabbena@ufl.edu

derived fluxes of the potent greenhouse gas N_2O, which is a side product of their metabolism (11, 12). However, their presence in anoxic environments has remained perplexing.

Despite the global biogeochemical importance of AOA, their physiology and metabolism are poorly understood. This is partly because of their limited homology to genetically tractable model microorganisms and because of the practical challenges in cultivation and biomass production for studies.

Kraft et al. add a dimension to the metabolic repertoire and adaptation of marine AOA, invoking a pathway for O_2 and N_2 production. By using sensors that can detect O_2 at concentrations as low as 1 nanomole per liter, and ^{15}N stable isotope label-

trite, a comparison with AOA cell numbers and rate measurements shows that AOA in oxygen minimum zones could contribute to N_2 production in a similar range as the previously known N_2-producing processes, denitrification and anaerobic ammonia oxidation (see the figure).

Current biogeochemical models assume that canonical nitrification cannot proceed in anoxic environments. The results of Kraft et al. suggest that the nitrogen cycle in marine oxygen minimum zones could be more complex than previously thought. Notably, a mechanism of oxygen production by nitrite disproportionation was recently suggested to account for NOB activity in these anoxic marine zones (1). However, these

the dismutation of chlorite to O_2 and chloride ions during microbial (per)chlorate respiration; and the dismutation of NO to O_2 and N_2 during nitrite-dependent anaerobic methane oxidation (13). However, besides two putative nitrite reductase paralogs, the N. maritimus genome encodes no enzyme implicating any of these processes, and ROS detoxification under anoxic conditions was ruled out experimentally.

Lacking biochemical precedent, Kraft et al. could only speculate on the potential mechanism. Based on the isotopic evidence, they postulate the most parsimonious, thermodynamically favorable pathway involving nitrite reduction and dismutation of NO to O_2 and N_2O, followed by the reduction of N_2O to N_2. Although this model provides a canonical function for the putative nitrite reductase found in all mesophilic AOA, it adds questions to the unresolved ammonia oxidation pathway by invoking further unprecedented biochemistry. Future studies should reveal whether this metabolism is confined to anoxic conditions or if it is integral to ammonia oxidation in general, and whether this metabolism can be found in other AOA, such as those that live in soils, sediments, or marine and terrestrial subsurface ecosystems. The detection of internal oxygen-producing pathways in distantly related nitrite-dependent methane-oxidizing bacteria and ammonia-oxidizing archaea suggests that internal oxygen production might be more widespread. Internal oxygen production could have had a key role in enabling nitrification and accelerating the evolution of the modern nitrogen cycle before the rise of molecular O_2 after the great oxidation event 2.4 billion years ago (14, 15). ∎

Nitrogen cycle in the oxygen minimum zones

Conceptual diagram of the role of ammonia-oxidizing archaea (AOA) and nitrite-oxidizing bacteria (NOB) in the marine nitrogen cycle inside oxygen minimum zones. Nitrification occurs below the photic zone in the oceanic water column. But thus far, it was believed not to be possible in anoxic zones. The paper by Kraft et al. changes that possibility.

REFERENCES AND NOTES

1. X. Sun, C. Frey, E. Garcia-Robledo, A. Jayakumar, B. B. Ward, ISME J. 15, 1317 (2021).
2. X. Peng et al., Global Biogeochem. Cycles 29, 2034 (2015).
3. L. A. Bristow et al., Proc. Natl. Acad. Sci. U.S.A. 113, 10601 (2016).
4. B. Kraft et al., Science 375, 97 (2022).
5. M. Könneke et al., Nature 437, 543 (2005).
6. M. Ren et al., ISME J. 13, 2150 (2019).
7. Y. Yang et al., Mol. Biol. Evol. 38, 3637 (2021).
8. W. Martens-Habbena, P. M. Berube, H. Urakawa, J. R. de la Torre, D. A. Stahl, Nature 461, 976 (2009).
9. Y. Zhang et al., Proc. Natl. Acad. Sci. U.S.A. 117, 4823 (2020).
10. M. Könneke et al., Proc. Natl. Acad. Sci. U.S.A. 111, 8239 (2014).
11. A. E. Santoro, C. Buchwald, M. R. McIlvin, K. L. Casciotti, Science 333, 1282 (2011).
12. L. Y. Stein et al., Environ. Sci. Technol. 55, 797 (2021).
13. C. U. Welte et al., Environ. Microbiol. Rep. 8, 941 (2016).
14. D. E. Canfield, A. N. Glazer, P. G. Falkowski, Science 330, 192 (2010).
15. J. Jabłońska, D. S. Tawfik, Nat. Ecol. Evol. 5, 442 (2021).

ing, the authors show that N. maritimus continues to oxidize ammonia to nitrite under anoxic conditions while simultaneously reducing nitrite to N_2O and N_2. After external O_2 was exhausted, these microbes would begin producing O_2 and bring the O_2 concentration in the surrounding media back up to 50 to 200 nanomoles per liter, indicating the concurrent production and consumption of O_2. The authors performed an extensive set of control experiments to allow for correction of interference from nitric oxide (NO) and exclude interferences from other conceivable reactive nitrogen and oxygen species. Although the ammonia oxidation activity under anoxic conditions was low relative to activity with external O_2 and relied on an external supply of ni-

theoretical mechanisms remain to be experimentally verified. Detection of free O_2 in the N. maritimus culture raises the question of whether AOA-produced O_2 could serve nitrite oxidation in oxygen minimum zones. Further disentangling of the complex microbial nitrogen cycling network in low-oxygen marine environments is needed for a more robust understanding of major nitrogen loss processes in the global oceans.

N. maritimus appears to produce O_2 and N_2 using a biochemical process yet to be understood. Known biological pathways for O_2 production include the oxidation of H_2O in the photosystem II protein complexes found in plants, algae, and cyanobacteria; the detoxification of reactive oxygen species (ROS) by catalase or superoxide dismutase;

ZEOLITES

Regulation of zeolite particle morphology

Specific particle morphology endows zeolites with improved molecular recognizability

By **Weibin Fan** *and* **Mei Dong**

Zeolites as catalysts or adsorbents have been extensively applied in petro-, coal-, and fine-chemical industries. To decrease the pressure drop and heat effects in reaction and separation processes, zeolites are typically shaped into bars or spheres, with binders to facilitate diffusion of reactants, adsorbates, and products. Inspired by this, the intrinsic adsorptive and catalytic properties of zeolites are enhanced by regulating their morphology and size (*1*). On page 62 of this issue, Korde *et al.* (*2*) report the synthesis of single-walled zeolitic nanotubes with an inner building layer that is topologically identical to the ZSM-5 *ac* plane and an outer layer identical to that of zeolite beta. This material has the potential to fabricate molecular recognition-based devices applicable to catalysis, separations, sensing, and drug delivery by regulating molecules' diffusion pathway and length.

A simple method to improve zeolite diffusivity is to decrease crystal size. This can substantially increase the catalytic stability and product selectivity by reducing secondary reactions and shortening diffusion length (*3*). The shape-selectivity of zeolite can be maximized through appropriate exposure of certain crystal facets. The large, highly intergrown, and twinned ZSM-5 crystals exhibit *p*-xylene selectivity of >99%, in contrast to the 24.1% obtained on a commercial sample in toluene methylation, as a result of exposing many more (100) planes that possess sinusoidal channel pore openings (*4*).

Another approach is the creation of meso- or macropores in zeolite particles. With three-dimensionally ordered mesoporous carbon as a template, highly ordered mesoporous aggregates of zeolite nanocrystallites are synthesized by means of confined crystallization (*5*). Recently, an ordered macro-mesoporous hierarchical carbon template with a face-centered cubic structure led to the formation of macro-meso-microporous single zeolite crystals (*6*). The ZSM-5 that was obtained showed

a catalytic lifetime as long as 13 times that of commercial microsized mesoporous ZSM-5 in a methanol-to-olefins reaction.

A more interesting strategy is the direct synthesis of hierarchically porous zeolites by using the soft template method because this avoids the preparation and removal of the carbon mold (see the figure). Two-dimensional (2D) MFI-type nanosheets are synthesized and assembled into 3D mesoporous particles with house-of-cards or 90°-rotational intergrowth shape by using di- or multi-quaternary ammonium surfactants with the hydrophilic head directing

Crystal assembly into designed architecture

One strategy to assemble zeolite crystals into membrane-like particles, zeolite nanotubes, or hierarchical three-dimensional (3D) architecture is to apply surfactant templates or organic additives through coulomb force, hydrogen bonding, π−π interactions, and/or van der Waal forces. Further control of crystal structure, habit, and size will produce highly ordered, pore-tunable hierarchical zeolite materials.

Zeolite crystals → Organic additives → **Nanotubes** **3D architecture**
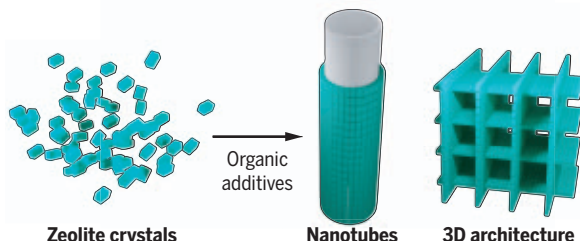
zeolite crystallization and the hydrophobic alkyl chain suppressing crystal growth along the *b* axis (*7, 8*).

Assembly of the 2D nanosheets depends on the quaternary ammonium structure. Bolaform or triply branched cationic surfactants with aromatic groups in hydrophobic regions, such as $C_6H_5–2N–C_6H_4–O–C_{10}H_{20}–N^+(CH_3)_2–C_6H_{12}–N^+(CH_3)_2–C_6H_{13}(2Br^-)$ ($C_{azo-10-6-6}$), enhance formation of lamellar micelle and 2D zeolite nanosheets through π–π interaction and lead to the generation of elongated octahedral crystals with ordered 2D square mesochannels fabricated with MFI nanosheets expanding along the *ac* plane through vertical conjunctions (*9*). Further change of $C_{azo-10-6-6}$ to $C_{azo-12-6-6}$ and $C_{azo-8-6-6}$ produces deformed rhombohedral and tetragonal-shaped aggregates, respectively (*10*). Additionally, alteration of the hydrophilic head of bolaform surfactant from the trialkyl group to quinuclidinium creates zeolite nanotubes, through π–π stacking of

surfactant molecules into a hydrophobic core along the nanotube axis and a hydrophilic periphery with quinuclidinium groups. This eliminates the diffusion limitation, resulting in high activity and selectivity in reactions that involve bulky molecules (*1, 7*). But these nanotubes exhibit low hydrothermal stability because of the large number of surface Si–OH groups.

Using well-crystallized zeolite crystal building units is an effective way to increase their hydrothermal stability. However, the rigid framework and the presence of few numbers of Si–OH groups on external surface makes assembly difficult. Silicalite-1 crystals have been assembled into 1D, 2D, and 3D architectures by selectively adsorbing ethanol and benzaldehyde (*11*). Unfortunately, the framework chemical compositions, crystal size, and morphology cannot be adjusted. In addition, oriented assembly of zeolite crystals into self-supported membrane along a desired axis and further regulation of size and thickness are required to simultaneously achieve high separation factor and flux upon being supported on a porous matrix. Further enhancement of its adsorptive separation and/or catalytic performance requires synthesis of zeolite nanotubes with walls assembled by perfectly arrayed crystals.

Another challenge is to tailor the macro- or mesopore structure and size of 3D architecture by modulating the construction scheme of (hierarchical) crystals. With the emerging of new techniques and the understanding of crystallization mechanism, solving these issues should be possible. Further rational design and assembly of hierarchical hollow particles and incorporation of metallic atoms, clusters, and nanoparticles will offer zeolites more interesting functions. ∎

REFERENCES AND NOTES
1. L.-H. Chen *et al.*, *Chem. Rev.* **120**, 11194 (2020).
2. A. Korde *et al.*, *Science* **375**, 62 (2022).
3. Q. Zhang *et al.*, *J. Am. Chem. Soc.* **141**, 3772 (2019).
4. C. Wang *et al.*, *Nat. Commun.* **10**, 4348 (2019).
5. W. Fan *et al.*, *Nat. Mater.* **7**, 984 (2008).
6. M.-H. Sun *et al.*, *Matter* **3**, 1226 (2020).
7. M. Choi *et al.*, *Nature* **461**, 246 (2009).
8. K. Möller, T. Bein, *Science* **333**, 297 (2011).
9. X. Shen *et al.*, *Angew. Chem. Int. Ed.* **57**, 724 (2018).
10. X. Shen *et al.*, *Chem. Eur. J.* **24**, 8615 (2018).
11. S. Li *et al.*, *ACS Appl. Mater. Interfaces* **13**, 58085 (2021).

State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, 27 South Taoyuan Road, Taiyuan 030001, China. Email: mdong@sxicc.ac.cn

RETROSPECTIVE

Richard Lerner (1938–2021)

Fearless scientist and leader of Scripps Research

By **Peter Schultz**

Richard Lerner, physician-scientist and president of Scripps Research from 1987 to 2012, died on 2 December 2021. He was 83. Richard had an enormous impact on science, both through his powerful intellect and creativity in the laboratory and as the bold, inspirational leader who grew Scripps Research into one of the world's preeminent research institutes.

Richard was born on 26 August 1938 and raised on the south side of Chicago, where he was a lifeguard on the beaches of Lake Michigan. After attending Northwestern University as an undergraduate, he earned his MD at Stanford University in 1964 and then interned for a year at Stanford before accepting a postdoctoral position in the Department of Experimental Pathology at Scripps Research (then known as Scripps Clinic and Research Foundation) in La Jolla, California. After his fellowship, during which he discovered the role of anti–glomerular basement membrane antibodies in Goodpasture syndrome (a rare autoimmune disease), he joined the Wistar Institute in Philadelphia but returned to Scripps Research a year later, in 1970, as a faculty member. He was named president of the institute in 1987 and served in that capacity for 25 years before retiring to his laboratory, where he continued doing research until his death.

Every conversation with Richard began with some fresh scientific insight or observation—he loved science above all. His own science showed remarkable creativity and breadth, spanning organic chemistry, immunology, and medicine. Despite being trained as a physician, Richard always thought on the molecular level, starting with his early work on using synthetic peptides as vaccine antigens. My first interaction with Richard was in 1986 when we independently generated the first catalytic antibodies—Richard as an immunologist learning chemistry and I as a young assistant professor learning immunology. When I serendipitously discovered that we were competitors, I nervously called Richard and introduced myself. Richard easily could have overwhelmed me with the resources at his disposal, but instead he suggested that we copublish and offered to help my lab at the University of California, Berkeley, make monoclonal antibodies. That phone call started a lifelong collaboration and close friendship—and it typified the scientific spirit and generosity that made him a great leader of Scripps.

Richard's work in catalytic antibodies, which reflected his early insights into the power of bridging chemistry and biology, was a tour de force. He used mechanistic principles to generate antibodies that catalyzed reactions ranging from aldol condensations

to disfavored cyclization reactions, and in doing so he proved fundamental notions of catalysis. He was one of the first to understand and exploit molecular diversity (in this case, immunological diversity) as a powerful new approach to advance biology, chemistry, and medicine. He then went on to create the earliest antibody libraries and later was the first to publish the concept of DNA-encoded libraries. These technologies have had an enormous impact on science and on the discovery of novel therapeutics in the pharmaceutical and biotechnology industries. Richard himself used these tools to create new medicines and played a key role in the development of two Food and Drug Administration–approved drugs. He continued to push the boundaries of science even after being incapacitated by a major stroke; just recently, he published a highly cited paper on the molecular mechanism of anesthetic drugs.

While doing all of this, Richard was president of Scripps. During his tenure, he grew the institute into a scientific power-house—creating a world-class chemistry department from scratch, founding a top-10 graduate school, and recruiting some of the world's leading scientists. He created an environment that enabled innovation and groundbreaking science. Richard funded people, not projects. He was remarkably adept at identifying both established and young scientists who would thrive in the "push the envelope" environment at Scripps. I believe most of us came to Scripps because of Richard's energy, enthusiasm, and focus on enabling great science.

Richard was also a leader in creating industry–academia partnerships and international collaborations. He used these relationships to build the institute while advancing science and medicine. He also played a leading role in building San Diego's now-thriving life sciences community. Foremost, Richard was fearless in pursuing exciting new ideas and directions despite the risks, and he magically seemed to make things work—always with the goal of bringing together great ideas, people, and resources with a minimum of distractions from the science. Very few institutions can boast of cell biologists, neurobiologists, immunologists, and chemists working so closely together to develop an understanding of, and treatment for, disease at the molecular level.

Known for his sharp intellect and quick wit, Richard sometimes pushed the boundaries of political correctness. He had a complete disregard for convention. His address at the Wolf Prize ceremony in the Knesset included a somewhat irreverent and wonderful joke about scientific discovery involving a physicist, a chemist, a rabbi, and a thermos bottle.

Richard also loved to push the boundaries outside the lab. I remember the day he laughingly parked his new turbo Porsche next to my old Volvo. Our friendly competition had clearly taken on a new direction, "forcing" me to buy an even faster turbo Porsche. The next thing I knew, there was a Ferrari Maranello parked in his space (with both our wives shaking their heads). Richard lived his life as a scientist and the leader of Scripps in sixth gear, never looking backward in the rearview mirror and never taking his foot off the gas pedal. He was a scientific giant, and his contributions were enormous. I only hope we can continue his legacy at Scripps, pushing the frontiers of science and the scientific enterprise with his example in mind. ∎

Scripps Research, La Jolla, CA, USA.
Email: schultz@scripps.edu

PHOTO: SCRIPPS RESEARCH

BIORISK MANAGEMENT

Eroding norms over release of self-spreading viruses

Risky research on lab-modified self-spreading viruses has yet to present credible paths to upsides

By **Filippa Lentzos[1], Edward P. Rybicki[2], Margret Engelhard[3], Pauline Paterson[4], Wayne Arthur Sandholtz[5], R. Guy Reeves[6]**

An evidence-based norm collectively established and reinforced through the work of generations of virologists is that laboratory modifications of self-spreading viruses are genetically too unstable to be used safely and predictably outside contained facilities. That norm now seems to be challenged. A range of transformational self-spreading applications have been put forward in recent years. In agriculture, for example, self-spreading viruses have been proposed as insecticides, or as vectors to modify planted crops. In health care, self-spreading viruses have been promoted as vaccines (*1*, *2*). Yet, glossed over by these proposals is that the self-spreading dynamics of a virus repeatedly passing from host-to-host (passaging) give it substantial potential to alter its biological properties once released into the environment (see the box). We explore the consequences of this apparent norm erosion in the context of recent proposals to develop self-spreading genetically modified viruses, in wildlife management and in self-spreading vaccines.

Wildlife control using self-spreading viruses is not a new idea. In the late 1980s, Australian researchers started to develop multiple approaches to sterilize or kill pest wildlife (foxes, mice, and rabbits) using self-spreading viruses (*3*). A decade later, Spanish researchers began limited field-testing of self-spreading viruses for the opposite purpose: to protect native wild rabbits (*4*). Concerns about self-spreading viruses were apparent from the inception of these programs and were exacerbated by the unrelated escape of a rabbit hemorrhagic disease virus from a limited field trial at an Australian high-security island laboratory. The accidental release was followed by widespread irreversible transmission within Australia and subsequent illegal international transportation to New Zealand (*5*). By 2007, funding for the Australian research had ceased and, despite approximately 15 years of work, no applications for field trials were ever made to Australian regulators. The Spanish efforts to license their self-spreading rabbit vaccine with the European Medical Agency also ceased. As part of a special issue of *Wildlife Research* that represented a self-written requiem to the Australian efforts, an article concluded:

"It is clear that a single unwanted introduction of a GM [genetically modified viral] biocontrol agent could have serious conse-

> " ...current developers and funders of this research should articulate comprehensive and credible regulatory paths..."

quences. Once a persisting transmissible GMO is released (whether intentionally, legally, or otherwise), it is unlikely that it could be completely removed from the environment. The scientific community involved in developing GM biocontrols therefore needs to demonstrate a highly precautionary attitude. Scientists also have an ethical responsibility to consider the full implications of the solutions they are researching: they must be seen to be acting openly, collaboratively and responsibly" (*6*, p. 583).

Indeed, as far back as 1993 both the World Organization for Animal Health (OIE) and the World Health Organization (WHO) expressed explicit concern at using self-spreading agents for wildlife management. Many regulatory issues arising from self-spreading approaches are widely acknowledged to have remained unresolved—such as who is responsible, or liable, if self-spreading viruses don't behave as expected or cross national borders (*3*)?

SUPPRESSED VIRAL EVOLUTION AND PREDETERMINED LIFETIMES

In 2016, interest in self-spreading vaccines reignited. Proposals seem to have been largely motivated by wildlife immunization (*1*, *2*, *7–9*), but a whole range of applications have been proposed (*8*). Agencies funding projects that incorporate or focus on such approaches include the European Union (EU) through its Horizon 2020 program, the US National Institutes of Health, and the US Defense Advanced Research Projects Agency (*10*). This time, proposals were accompanied by repeated assertions from funders and scientists that approaches exist that enable suppression of viral evolution and that "researchers can fine-tune vaccines to have predetermined lifetimes, which could eliminate concerns over unwanted mutations or ongoing evolution of the vaccine organism" (*7*). It is hypothesized that this could be achieved by long-established laboratory manipulations of viral genomes, namely, synonymous codon replacement, genome rearrangement, and deletions (*11*). However, it remains to be experimentally tested if any combinations of these manipulations could simultaneously limit viral transmissibility to the extent that they could be perceived as controllable while maintaining sufficient transmissibility to be considered useful as vaccines in continually dynamic environments.

One of the purported uses of lab-modified self-spreading viruses is as wildlife vaccines to limit the risk of spillover events generating previously undescribed human pathogens like severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Though an outwardly attractive application, there are notable hurdles that have been glossed over. First, the vast majority of virus species that currently exist are undescribed by science (*12*). This makes it very difficult to imagine how the considerable effort necessary to develop and test self-spreading vaccines could identify and then prioritize single viral species circulating in wildlife.

Second, the dynamic nature of mutation and recombination events in wild global

viromes, which are speculated to play a defining role in many spillover events, makes it extraordinarily difficult to mitigate spillover risk using wildlife vaccines. Although a massive increase in viral monitoring in wildlife might in some theoretical circumstances provide a time-limited degree of insight upon which to base preemptive vaccine design, it could only have a very indirect impact on prioritizing what genetic event, in which wildlife species, and at what location might present a substantive risk for the emergence of a new virus.

In addition, the extraordinary practical complexity of wildlife vaccination, particularly in terms of sustaining and monitoring the immune response in wildlife populations, has not been explicitly addressed by funders or scientists promoting self-spreading vaccines (13). The combination of these concerns in the context of emerging new viral pathogens has led most virologists to consistently advocate for surveillance at the human–animal interface particularly in regions of ecological disturbance, rather than the riskier mass prospective development of self-spreading vaccines (8).

A more extreme application touted is the use of self-spreading viruses in human vaccination (8). This is often paired with an acknowledgment that use in humans would likely come with insurmountable ethical and safety concerns, as well as public outcry (1, 2). From a security perspective, it is important to note that, with the exception of heightened safety standards, it is broadly true that whatever self-spreading vaccines could be developed for wildlife could be more easily generated, from a technical perspective, for humans. This is because monitoring populations for levels of immunity and viral evolution is much easier in humans than in wildlife, and critically, the hurdles outlined above would not apply in quite the same way, as by definition, the target virus will be well known to science.

Theoretical claims of suppressed viral evolution and their assumed predetermined lifetimes in the environment currently remain peripheral within the scientific community. However, this seems paradoxical if the accumulating collection of more than 15 publications over the past 5 years does indeed outline a genuinely innovative path to safely achieving and maintaining high levels of immunity, using a fraction of the resources and time that conventional vaccination programs require. Why would recognition of this (re)emerging field be so low, if there is even a remote possibility of the promoted thoroughly transformational goals being realized?

One reason might be that experts from across relevant disciplines will view claims of suppressing viral evolution, or sustainably

What are self-spreading viruses?

To date, proposed modified self-spreading viral approaches can usefully be placed in one of three types:

1. Experimental approaches to kill or sterilize mammalian wildlife or pests as a means to reduce their population sizes, also called wildlife management (3).

2. Experimental approaches to vaccinate mammalian wildlife to protect them from disease (4, 15) or to limit their capacity to act as reservoirs for vectored diseases (1, 2).

3. Speculations about applications in humans as vaccines (1, 2, 8).

The terms "self-spreading," "transmissible," "self-disseminating," "contagious," and "horizontally transferable" have all been used interchangeably to describe artificially modified viruses developed for applied uses that intentionally retain the capacity to transmit between individual hosts upon their release into the environment. Here we adopt the term "self-spreading virus," defined as satisfying both of the following criteria:

1. Intentionally developed to be transmissible between individual hosts in the environment, where safety testing, efficacy testing, and regulatory approval incorporate the numerous consequences arising from this property.

2. Possessing deployment strategies that fundamentally rely upon transmission between individual hosts for their successful application (see the figure).

Viral transmissibility between individual hosts is in almost all circumstances dynamic, particularly in complex environments. For example, coinfection of wild-type and genetically modified released viruses has the capacity to enhance transmission rates of the latter through viral complementation. Furthermore, spontaneous recombination has the potential to alter the transmissibility of parts or all of the released viral genomes. In some recent publications the vague concept of "transferable vaccines" has been introduced (9), which are proposed to be in some respects intermediate between conventional and self-spreading vaccines. However, it is unclear to us if their hypothecated existence has any basis in fact, and as such their consideration is potentially unhelpful.

Currently, none of the licensed genetically modified viruses for use in the environment are transmissible, including the various widely applied oral-bait rabies vaccines for wildlife.

fine-tuning transmissibility in complex, dynamic environments, with a high degree of informed skepticism. Alternatively, perhaps they may see nothing genuinely new in the claims because it has been technically possible to generate such vaccines for decades (4). Indeed, the pervasive challenge had always been to minimize or eliminate transmission of human vaccines based on live viruses between individuals. This is for reasons of safety and ethics (e.g., the infection of immunocompromised individuals or nursing infants), but also for practical considerations (e.g., retaining the capacity to suspend or geographically restrict trials). It has been argued that existing live poliovirus vaccines are "self-spreaders" (2, 11). Although they are not self-spreading vaccines as the term is used here (see the box), the vaccine-associated paralytic poliomyelitis (caused by uncontrolled community transmission of Sabin type 2 polio vaccine) does present an object lesson in the risk of transmissibility. An ongoing example of efforts to further minimize the impact of unwanted transmission is the early-stage development of a Sabin type 2 polio vaccine strain that aims to limit back-mutation to virulence (14).

Self-spreading vaccine research continues to proceed despite a lack of new information that would compellingly refute long-standing evidence-based norms in virology, evolutionary biology, vaccine development, international law, public health, risk assessment, and other disciplines. Providing such evidence, along with anticipated benefits, possible harms and risks, and appropriate precautionary measures, should have been considered a critical first step in undertaking self-spreading vaccine research. Furthermore, there are currently no fully, or even partially, articulated proposals for regulatory pathways that could establish self-spreading vaccines as not only safe, effective, and useful but also, crucially, as the patchy uptake of COVID-19 vaccines has shown, publicly trusted.

And if, as claimed, self-spreading vaccines do indeed represent a flexible and transformational technology in areas as diverse as conservation, human health, and agriculture, then additional justification is needed for why efforts would not be exclusively focused on addressing pressing needs in the countries funding or developing these approaches. This was the case for

the earlier Australian and Spanish efforts and remains so for ongoing EU-funded efforts to address African swine fever (*7, 15*). But one self-spreading vaccine development program—to prevent Lassa virus transmission to humans in West Africa—is focused on *Mastomys* rats, which are only distributed in sub-Saharan Africa (*10*), and motivations for other studies also imply target areas outside North America, where most of this research is currently taking place.

Potential for repurposing these technologies is also a concern, where adversaries could, for instance, deliberately use self-spreading viruses, or vaccines, to cause harm. This,

peer-reviewed description of the Spanish self-spreading rabbit vaccine and the submission for publication of the results of a successful field trial (*4*) was just 12 months.

URGENT NEXT STEPS

Although earlier work on self-spreading vaccines coincided with, or was preceded by, regulatory engagement in multiple international forums, like the OIE, WHO, the International Plant Protection Convention (IPPC), and the Convention on Biodiversity (CBD), current developments appear to be taking place without similar international efforts. The previous regulatory discussions ultimately concluded

Echoing international efforts, national governments should clarify and, if necessary, update any relevant legislation and guidance. In parallel, current developers and funders of this research should articulate comprehensive and credible regulatory paths through which they believe the safety and efficacy of self-spreading approaches could be established and through which publics may accept the inherently coercive and mandatory nature of self-spreading vaccines.

Only a concerted, global governance effort with coherent regional, national, and local implementation can tackle the challenges of self-spreading viruses that have the potential to radically transform both wildlife and human communities. This is because, as the case for the rabbit hemorrhagic disease virus in Australia showed, for self-spreading techniques, there is a real possibility that the first regulatory approval for a limited field trial could turn into an unapproved international release (*5*). ∎

Deployment strategies for self-spreading vaccines

At time T1, four members of a hypothetical population are directly vaccinated by injection (●). Only with a self-spreading vaccine will immunity potentially expand to those not directly vaccinated (● ●). Transmission of the self-spreading vaccine occurs spatially (top) and temporally (bottom, to subsequent generations not shown). The self-spreading vaccine provides some outwardly attractive opportunities if there is a need for rapid vaccination of whole populations or difficulties in accessing individuals (this relies on the rather unrealistic assumption that all individuals in the population remain naïve to infection by the self-spreading vaccine).

Conventional vaccines

T1 T2 T3

T1 T2 T3 T4 T5

Only vaccinated individuals are affected and maintained.

Self-spreading vaccines

T1 T2 T3

T1 T2 T3 T4 T5

Vaccine spreads throughout the population over time.
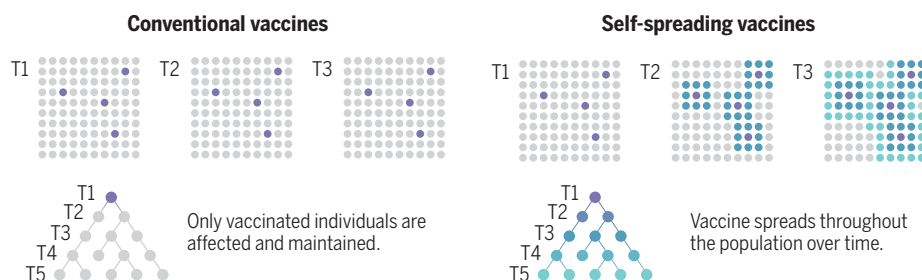
too, needs to be addressed by those pushing self-spreading vaccine development.

The technologies available to the Australian and Spanish researchers' decades ago were sufficient to develop multiple candidates, one of which progressed to field trials and a licensing application (*4*). Molecular biology tools have advanced since then, and arguably, little to no technological development is necessary to produce self-spreading viral vaccines today. Without open and inclusive engagement about potential benefits, risks, and appropriate precautionary measures from the scientific and international communities, self-spreading viruses for environmental release could arguably be developed very quickly, with limited funding or expertise and with potentially irreversible consequences for the planet's biodiversity, ecosystems, and environments. With only modest technological innovation required, typical risk mitigation measures, such as increased education of scientists or the creation of new international forums to address governance, are likely to prove too slow to have a constructive impact. It is notable that the time between the very first

without resolving key safety and regulatory questions (*3, 6*), but there was a clear consensus that, given the near inevitability of transboundary movement, the appropriate forums to consult were international ones.

A clear priority for the international community must be to update existing phytosanitary, medical, and veterinary regulations to reflect contemporary societal values for responsible stewardship of science—and specifically with respect to environmental releases of self-spreading viruses. Key principles that should be endorsed and actively promoted include safety, intergenerational justice, accountability, and public engagement. Immediate opportunities are the CBD Conference of the Parties (COP) and the meeting of its Subsidiary Body on Scientific, Technical, and Technological Advice (SBSTTA) in spring 2022. Each provides a chance to build on earlier CBD work (*3*).

Additional steps would be to establish and implement a robust horizon-scanning process and to develop a global consensus on the criteria for safe, secure, and responsible research and the evidence needed to meet those criteria.

REFERENCES AND NOTES

1. M. W. Smithson, A. J. Basinki, S. L. Nuismer, J. J. Bull, *Vaccine* **37**, 1153 (2019).
2. S. L. Nuismer *et al.*, *Proc. Biol. Sci.* **283**, 20161903 (2016).
3. Conference of the Parties to the Convention on Biological Diversity, report of the Canada-Norway expert workshop on risk assessment for emerging applications of living modified organisms UNEP/CBD/BS/COP-MOP/4/INF/13, 39 (2007).
4. J. M. Torres *et al.*, *Vaccine* **19**, 4536 (2001).
5. P. O'Hara, *Rev. Sci. Tech.* **25**, 119 (2006).
6. W. R. Henderson, E. C. Murphy, *Wildlife Res.* **34**, 578 (2007).
7. R. P. Ortega, "Can vaccines for wildlife prevent human pandemics?" *Quanta Mag.* (2020); www.quantamagazine.org/can-vaccines-for-wildlife-prevent-human-pandemics-20200824/.
8. M. Cogley, "Could self-spreading vaccines stop a coronavirus pandemic?" *The Telegraph (UK)* (2020); www.telegraph.co.uk/technology/2020/01/28/could-self-spreading-vaccines-stop-global-coronavirus-pandemic/.
9. K. M. Bakker *et al.*, *Nat. Ecol. Evol.* **3**, 1697 (2019).
10. PREEMPT, Prediction of Spillover and Interventional En Masse Animal Vaccination to Prevent Emerging Pathogen Threats in Current and Future Zones of US Military Operation (2021); www.preemptproject.org/about.
11. J. J. Bull, M. W. Smithson, S. L. Nuismer, *Trends Microbiol.* **26**, 6 (2018).
12. M. Wille, J. L. Geoghegan, E. C. Holmes, *PLOS Biol.* **19**, e3001135 (2021).
13. K. M. Barnett, D. J. Civitello, *Trends Parasitol.* **36**, 970 (2020).
14. P. Van Damme *et al.*, *Lancet* **394**, 148 (2019).
15. C. Gallardo *et al.*, *Transbound. Emerg. Dis.* **66**, 1399 (2019).

ARCHAEOLOGY

Deciphering Egypt's Great Pyramid

An ancient archive offers new insights into the construction of a man-made wonder

By **Andrew Robinson**

Two centuries after the decipherment of Egyptian hieroglyphs by Jean-François Champollion, there is still much to be learned about ancient Egypt. Most prominently, perhaps: Who built the pyramids at Giza and when, and what techniques were used? The astonishing Great Pyramid—the only survivor of the Seven Wonders of the Ancient World—is enigmatic about its origins and construction. Until very recently, no contemporaneous documentary references to its creation had been found.

Without a doubt, the Great Pyramid was commissioned by the Old Kingdom pharaoh Khufu (Cheops). The British Museum and Cairo's Egyptian Museum give his regnal dates as 2589 to 2566 BCE. Egyptologists Mark Lehner, who has conducted fieldwork at Giza for four decades, and Zahi Hawass, a former Egyptian government official in charge of Giza, argued for the later range of 2509 to 2483 BCE in their massive 2017 book, *Giza and the Pyramids*. But another high-profile Egyptologist, Pierre Tallet, whose pioneering fieldwork on the Red Sea coast of Egypt began in 2011, favors the earlier range of 2633 to 2605 BCE, derived from a recent astronomically based chronological model for the Old Kingdom.

In *The Red Sea Scrolls*, written by Tallet in close collaboration with Lehner, the authors use this latter date range. Egyptian dates before the Late Period (circa 660 BCE), they

The Red Sea Scrolls
*Pierre Tallet
and Mark Lehner*
Thames and Hudson,
2022. 320 pp.

note, "are much debated, especially so for the earliest periods."

This landmark, elegantly illustrated book is the first to reveal how the raw materials used in the Great Pyramid's construction—copper, for instance, derived from pharaonic mining expeditions to the remote deserts of Sinai—were transported to Giza via Egyptian ports during the reigns of Sneferu and his son Khufu. Boats to transport miners and materials were built in the Nile Valley, dismantled, and then arduously conveyed to the Red Sea via tracks across the Eastern Desert. After use, the boats were stored at Red Sea ports in artificial galleries. "Egyptians, rather than being inexperienced and reluctant sailors, seem to have acquired a high level of experience in maritime navigation," note the authors.

Timber, meanwhile, was likely ferried from the Nile Valley for use in furnaces to smelt copper ore on a vast scale. At one site in southern Sinai, discovered by Tallet in 2009, at least 3000 smelting units are estimated to have existed, one of them up to 80 m long.

The papyrus archive found by Tallet in 2013 at Wadi el-Jarf on the western Red Sea coast greatly excited Egyptologists. Old Kingdom papyri are extremely rare because papyrus does not last long in humid conditions. Yet more than a thousand fragments were discovered by Tallet's team, probably deriving from at least 30 rolls—"the oldest known explicitly dated Egyptian documents," which frequently mention Khufu and the pyramid-building project. They survived because they were abandoned in the galleries instead of being officially archived in the Nile

A papyrus fragment sheds light on the construction of Egypt's Great Pyramid.

Valley—presumably because they were no longer regarded as being of any use.

Having pieced together evidence written in both hieroglyphic and hieratic script, Tallet and Lehner expertly reveal the archives of a single work gang, 160 men in total, which cover slightly more than a calendar year, probably during the final year of Khufu's reign. The crew worked under the naval designation "The Escort Team of 'The Uraeus of Khufu is its Prow'"—the Uraeus being the stylized image of an upright, rearing cobra, a symbol of Egyptian sovereignty. Divided into four sections, they undertook at least five different tasks in different locations, all of which were recorded on separate papyri. These tasks included transporting by boat heavy blocks from Nile Valley limestone quarries to Giza's harbor for use in the Great Pyramid, voyaging across the Red Sea, maintaining a network of artificial canals, and building harbor facilities. One section leader, Inspector Merer, drafted "meticulous reports," naming Khufu's half brother as the recipient of the stone, which provide a detailed picture of the functioning of the state at this iconic time in Egyptian history.

The archive reveals that Khufu's pyramid was made not by large teams of ruthlessly exploited, unpaid slaves, as seminally proposed by the ancient Greek historian Herodotus, but by a relatively small, "highly skilled, well-rewarded workforce." The papyri reveal nothing definitive, however, about building techniques, such as how materials were lifted onto ramps as the pyramid rose from its base. This mystery remains to be deciphered. ∎

The reviewer is the author of *Cracking the Egyptian Code: The Revolutionary Life of Jean-François Champollion* (Oxford Univ. Press, 2012). Email: andrew@andrew-robinson.org

Confronting age-old questions, big and small

A wry romp through science's greatest hits and challenges entertains and illuminates

By **Megan Engel**

Hidden in the number π (pi)—that mysterious and elusive ratio between a circle's circumference and its diameter—is every word of the book review you're about to read. As geneticist Adam Rutherford and mathematician Hannah Fry, authors of *The Complete Guide to Absolutely Everything (Abridged): Adventures in Math and Science,* point out, "every possible string of numbers *must* appear [in pi] eventually." By encoding English letters as the numbers 0 to 25, we arrive at the fact that every possible book review appears in there, too, including one in which said book is panned. This, however, is not that review. Rutherford and Fry's eclectic tour through "life, the universe, and everything" is a lighthearted, thoroughly entertaining read that delivers many laugh-out-loud moments.

With wry, irreverent British humor, the authors tackle a host of big, albeit seemingly random, questions, devoting a chapter to each. These range from the age-old and deep (Do we have free will? What would aliens look like?) to the droll (Does my dog love me? Are we all suckers?). Most are not definitively answered, but the quest for solutions is engaging and enjoyable, peppered with pop culture, literary, and classical references—from Hollywood to Jane Eyre to Cicero—and easy-to-digest analogies. There is not a boring page in the bunch, but the book does feel at times like an assortment of trivia with some non sequiturs thrown in rather than a coherent whole.

The authors' stated theme is that "much of what you might think you know is not quite what it seems," and science comprises the "best tools ever invented [to] see things as they really are." Rutherford and Fry focus less on objective scientific truths and more on us: who we are, how we experience reality, the questions we ponder, and our historical pursuit of answers. For example, readers journey through human timekeeping, from sundials to "leap seconds," and learn how the modern financial system hinges precariously on subterranean fiber-optic cables that deliver atomic time to banks.

Instead of Einstein's time dilation, expect to learn about time dilation inside the human body (time seems to slow when we are in peril or cut off from sunlight). And Darwin's theory of evolution takes a backseat to his debunked theories of human emotion. If you are looking for a hard science explainer book, this is not it.

The Gravity Probe B space vehicle, designed to verify Einstein's theory of general relativity, is assembled in Sunnyvale, California.

Despite mostly glossing over technical details, *The Complete Guide* does highlight some fascinating recent scientific findings—the fact that rats appear to feel regret, for example, and a case in which a woman appeared to be able to smell Parkinson's disease on a patient's clothing months before they began exhibiting symptoms—as well as a number of state-of-the-art experiments. Among the most awe-inspiring of the latter is the authors' account of the Gravity Probe

The Complete Guide to Absolutely Everything (Abridged): Adventures in Math and Science
Adam Rutherford and Hannah Fry
Norton, 2022. 304 pp.

B experiment, designed to verify Einstein's theory of general relativity in space and featuring liquid helium and man-made gyroscopes that constitute the closest things to perfect spheres ever found in the universe.

Rutherford and Fry also shine a light on important issues that plague modern science, such as the "File Drawer Problem," wherein "shiny novelties" are prioritized for publication over essential but less flashy experiments that verify existing results, and the related "replication crisis." The authors' discussion of the cognitive biases that underlie these problems, such as "belief perseverance"—continuing to believe in something (e.g., a vaccine hoax) in the face of "overwhelming evidence to the contrary"—and confirmation bias, which undergirds the spread of misinformation online, is timely and topical. These biases are a product of our neurobiology, and no one is immune to them, so Rutherford and Fry implore us all to "be vigilant… because your own brain is trying to trick you."

The authors close with a bold, slightly hubristic assertion that science is "the only way to compose the ultimate guide to everything," which is ironic, given their own acknowledgment of the limitations of science ("science cannot help you to reliably infer how someone feels," for example). Still, readers are likely to finish *The Complete Guide* with a pocketful of intriguing anecdotes with which to entertain at cocktail parties (when they resume) and a feeling of endearment toward humankind, who dare to ask questions they are "singularly ill-equipped to answer" and who, despite the "peculiar failings of human intuition," have proved in many respects fairly "clever sausages." ∎

The reviewer is at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. Email: mcengel@seas.harvard.edu

RESEARCH

NEUROSCIENCE
Dendritic spines' electrical function?

Dendritic spines are small protrusions that cover the dendrites of most neurons in the brain. Their electrical properties are still controversially discussed. Cornejo et al. used an array of techniques to investigate the degree of voltage attenuation by dendritic spine necks in pyramidal neurons of the mouse neocortex. Spines not only synchronously depolarized in response to backpropagating action potentials, but local and transient depolarization also occurred. Isolated depolarization in individual spines reflected localized synaptic activation. A significant voltage gradient between dendritic spine and dendrite indicated that spines may constitute elementary electric compartments. The spine neck resistance is thus not negligible and may substantially contribute to the regulation of synaptic efficacy in the central nervous system. —PRS   *Science*, abg0501, this issue p. 82

Dendritic spines, seen here on a Purkinje neuron, experience compartmentalized voltage dynamics.

PLANETARY EVOLUTION
Abrupt oxygenation of Earth

The amount of oxygen in Earth's early atmosphere increased abruptly 2.3 billion years ago during the "Great Oxygenation Event," but questions remain as to whether there were earlier, transient "whiffs" of oxygen before then, perhaps related to early photosynthesis. A new study by Slotznick et al. examined samples from an area in western Australia where others found what they interpreted as evidence of transient appearances of oxygen as early as 2.5 billion years ago. The new, microanalytical studies reveal no evidence of early oxygenation, leading the authors to conclude that oxygen levels on Earth were negligible until the Great Oxygenation Event. —KVH
Sci. Adv. 10.1126/sciadv.abj7190 (2022).

DNA DAMAGE
Double duty for DNA-PK

DNA-dependent protein kinase (DNA-PK) coordinates the DNA damage response and also promotes cell survival and resistance to chemotherapy through activation of the kinase AKT. Liu *et al.* found that DNA-PK mediates a phosphorylation event that activates AKT in a manner dependent on mechanistic target of rapamycin complex 2 and the guanine nucleotide exchange factor ECT2 (see the Focus by Shin *et al.*). Disrupting this mechanism increased the toxicity of DNA-damaging agents in cultured cells. These findings may lead to new ways to overcome tumor resistance to DNA-damaging chemotherapeutics. —LKF
Sci. Signal. **15**, eabh2290, eabm6211 (2022).

HEAVY FERMIONS
A gentle reconstruction

Changes in the volume of a material's Fermi surface are typically associated with the breaking of symmetry. Maksimovic *et al.* found evidence for an unusual type of this Fermi surface reconstruction, one without symmetry breaking, in the heavy fermion compound cerium cobalt indium ($CeCoIn_5$). Doping the material with tin led to abrupt changes in Hall effect, quantum oscillation, and photoemission signals. The researchers interpret the findings in terms of the delocalization of f-electrons associated with cerium atoms. —JS
Science, aaz4566, this issue p. 76

BATTERIES
Preservation of cycling behavior

Understanding the changes in interfaces between electrode and electrolyte during battery cycling, including the formation of the solid-electrolyte interphase (SEI), is key to the development of longer-lasting batteries. Z. Zhang *et al.* adapted a thin-film vitrification method to ensure the preservation of liquid electrolyte so that the samples taken for analysis using microscopy and spectroscopy better reflect the state of the battery during operation. A key finding is that the SEI is in a swollen state, in contrast to current belief that it only contains solid inorganic species and polymers. The extent of swelling can affect transport through the SEI, which thickens with time, and thus might also decrease the amount of free electrolyte available for battery cycling. —MSL
Science, abi8703, this issue p. 66

SPLICEOSOME
Intron recognition in the spotlight

Excision of noncoding introns from pre–messenger RNAs is catalyzed by the spliceosome, a large RNA-protein complex that recognizes specific sequences at the exon-intron boundaries (splice sites). These sequences are highly degenerate in humans, and it has remained elusive how they are recognized by the spliceosome. Tholen *et al*. report a series of high-resolution structures of the human U2 small nucleolar ribonucleoprotein, the component of the spliceosome that recognizes branch sites. The structures explain how SF3B6 helps to stabilize the branch helix in the absence of extensive sequence complementarity. A newly identified spliceosome assembly intermediate suggests a mechanism for fidelity control of branch site recognition. —DJ

Science, abm4245, this issue p. 50

GEOPHYSICS
Hotspot cooldown

Deep-seated mantle plumes are responsible for volcanic island chains such as Hawai'i. Upwelling from the deep interior requires that the plumes are hotter than the surrounding mantle to make it all the way up to the surface. However, Bao *et al*. found that some of these "hotspots" are surprisingly cool. The temperature is actually low enough to challenge a deep mantle origin for some hotspots. In these specific cases, deep plumes may be entrained and cooled or possibly originate in the upper mantle instead. —BG

Science, abj8944, this issue p. 57

STRUCTURAL BIOLOGY
Not your typical GPCR

Among the large family of G protein–coupled receptors (GPCRs) are many orphans, so called because their signaling reactions remain poorly understood. Among these is GPR158, which is highly expressed in the nervous system and implicated in processes from cognition to memory to mood. Patil *et al*. determined a high-resolution structure of GPR158 alone and bound to a regulator of the G protein–signaling (RGS) complex. GPR158 has an unusual dimerization mode with an extensive interaction interface that locks it in a conformation that likely prevents G protein activation. RGS binds to the homodimer at a site that substantially overlaps the surface that binds G proteins, again preventing canonical G protein signaling. Binding of a ligand to the extracellular domain may regulate signaling through the RGS complex. —VV

Science, abl4732, this issue p. 86

IN OTHER JOURNALS

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

DEVELOPMENT
A molecular map of the eye

The iris is the colored part of the eye that regulates the level of light entering the eye by making the pupil either smaller (constricted) or larger (dilated). Wang *et al*. performed single-nucleus RNA sequencing to compile a molecular map of the mouse iris. They characterized all major cell types within the iris and examined cell-type-specific transcriptome signatures that differed among the resting, constricted, and dilated states. The most striking differences were evident in the dilated state, particularly in a gene called early growth response 1 (EGR1), which is expressed in dilator muscles. Four new cell types were also identified, along with insights into how the iris develops in early life. Next steps include comparisons with the human eye with the goal of understanding iris disorders and potentially developing treatments for eye disease. —PNK *eLife* **10**, e73477 (2021).

A single-cell transcriptome atlas of the mouse iris reveals molecular and cellular underpinnings of the mechanical functioning of the eye.

NEUROSCIENCE
The brain behind novelty and reward

Novelty seeking is a key feature of intelligent behavior and adaptive cognition. However, we know little about the circuits that regulate our attraction to novel objects for novelty's sake. Ogasawara *et al*. discovered that a brain nucleus called the zona incerta was causally related to novelty seeking. A region in the anterior medial temporal lobe projected to the zona incerta and sent motivational signals required to control novelty seeking through the zona incerta circuit. A novelty-seeking task, in which monkeys were motivated by the receipt

Satellite image of Ascension Island, which formed from volcanism above a particularly cool hotspot.

Fluorescence microscopy image showing ocular microglial precursors (green) colonizing the zebrafish retina
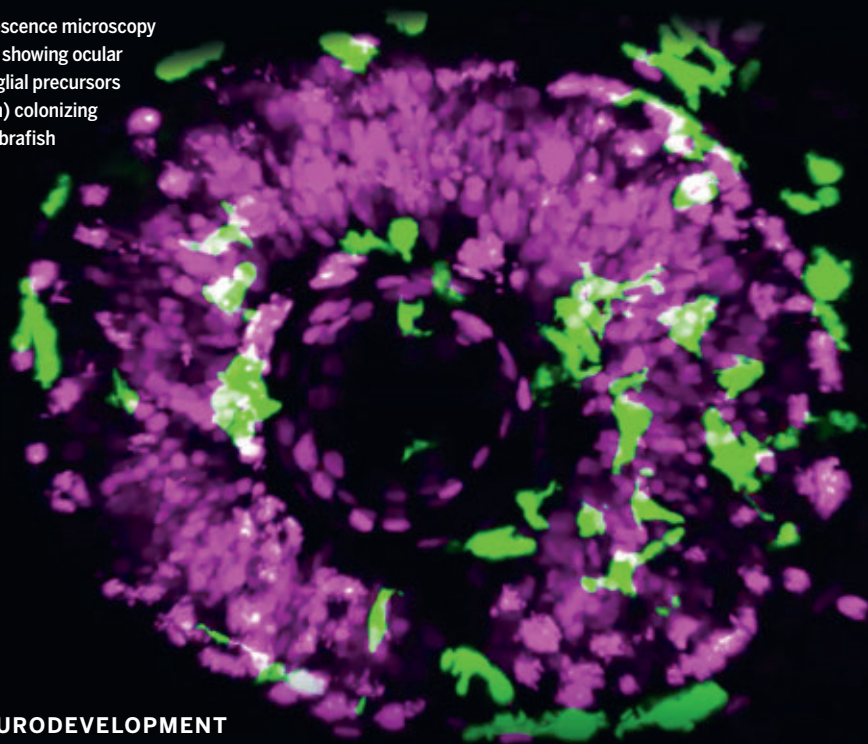
NEURODEVELOPMENT
Highways into the retina

Microglia, the macrophages in the brain, consume cellular debris and remove unnecessary synapses. These handy cells do not originate from the neuroectoderm but rather are generated through hematopoiesis. How, then, do they get into the brain? Ranawat and Masai tracked fluorescently labeled cells through transparent zebrafish larva to follow ingress of microglia during development. In zebrafish, the retina is among the first brain tissues to be populated by microglia. The microglia are encouraged to migrate by cytokine signals, use the surfaces of blood vessels as a pathway through the choroid fissure, and invade the neurogenic region of the developing retina. —PJH   *eLife* **10**, e70550 (2021).

of novel objects, showed that this behavior was not regulated by the dopamine reward-seeking circuitry. This work provides evidence for a clear dissociation in the brain circuitry between reward seeking and novelty seeking. —PRS

Nat. Neurosci. 10.1038/s41593-021-00950-1 (2021).

DECISION MAKING
Diagnosing heart attacks

A machine-learning algorithm for heart attack risk that draws upon patients' health data suggests that physicians substantially misallocate tests for heart attacks. Mullainathan and Obermeyer studied 246,265 visits to a large urban hospital emergency department. Compared with doctors' aggregated decisions,

a preferred allocation informed by the risk model would have cut 62% of total tests, added 61% of new tests, and reduced cost per life year by 34%. Three limitations in physicians' decision making were identified: Their model of risk is too simple, they over-weight salient risks (e.g., demographics and symptoms), and they over-weight symptoms that are stereotypical of heart attack (e.g., chest pain). —BW

Q. J. Econ. 10.1093/qje/qjab046 (2021).

MICROBIOLOGY
Keeping host cell death at bay

The agent that causes Legionnaire's disease, *Legionella pneumophila*, takes up residence

in human macrophages and hijacks their machinery to set up a replicative niche. They are helped in this task by injecting more than 300 effector proteins into the host cell using a so-called type IV secretion system. One intracellular target of some of these effectors is the host cell mitochondrion. Escoll *et al.* found that infected macrophages, while exhibiting reduced mitochondrial oxidative phosphorylation, still maintained a healthy mitochondrial membrane potential. Within infected cells *L. pneumophila* effectively reversed the activity of the mitochondrial F0F1-ATPase using cytosolic ATP to preserve the mitochondrial membrane potential. Blocking this activity collapsed the mitochondrial membrane potential and the infected cells then died. Thus, *L.*

pneumophila puts its host cells on a form of life support to preserve its own replicative niche. —SMH

eLife **10**, e71978 (2021).

QUANTUM OPTICS
Coupling quantum emitters

Single photons from quantum emitters can encode quantum information across long distances using optic fibers and free space. For integrated on-chip quantum optical processing, arrays of quantum emitters would form a local network that could process the single photons propagating through the network. Yu *et al.* present a photonic crystal-based platform in which two quantum emitters placed in nanocavities spaced several micrometers apart can be coupled. Their photonic structure was designed and fabricated so that the nanocavities are in resonance, which enhances the coupling between the two remote emitters. By demonstrating mutual excitation and the possibility of synchronization between the two emitters, the scalable approach provides a route to realize larger on-chip quantum networks. —ISO

Optica **8**, 1605 (2021).

GEOPHYSICS
A powerful cloth

Ferropericlase, one of the primary mantle minerals, is known to have an electronic spin transition. However, evidence of this transition in the Earth's mantle has been challenging to find. Shephard *et al.* found changes in seismic wave speeds at two depth ranges that correspond to the iron spin transition in ferropericlase. The authors compared compressional and sheer-wave velocities in tomographic models, finding relative changes between the two types of waves that they could attribute to the transition. Using tomographic models is important because nonuniform thermochemical variations wash the signal out in global, one-dimensional models. —BG

Nat. Commun. **12**, 5905 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CANCER
Tertiary lymphoid structures in cancer

Tertiary lymphoid structures (TLSs) are lymphoid formations that are found in nonlymphoid tissues. TLSs can develop in inflamed tissues and are associated with chronic inflammatory disorders, autoimmunity, and cancer. In the setting of tumors, TLSs facilitate the influx of immune cells into the tumor site and have therefore attracted interest as a means of improving anticancer immunity and favorable treatment response in patients. Schumacher and Thommen review the biology of TLSs and outline recent advances in TLS research. They discuss how TLSs are detected and defined, the mechanism(s) of formation in cancer, and the potential of targeting TLSs for therapeutic benefit. —PNK

Science, abf9419, this issue p. 39

ORGANOIDS
Spatial and temporal organoid control

Stem cell–derived organoids form through self-organization and serve as models for organ development, function, and disease, with potential applications in drug development and personalized medicine. However, in the absence of external guidance, developmental processes are stochastic, resulting in variable end products that differ significantly from the native organ. Gjorevski et al. developed approaches for specifying the initial organoid geometry to build intestinal organoids of defined shape, size, and cell distributions, forming structures that are predictable, more similar to normal organs, and reproducible (see the Perspective by Huycke and Gartner). These methods identify symmetry-breaking mechanisms in intestinal morphogenesis and have potential for standardizing organoid-based therapies and facilitating the refinement of mechanistic studies. —BAP

Science, aaw9021, this issue p. 40; see also abn3054, p. 26

NEUROSCIENCE
Cell type diversity in the neocortex

The ability to link molecularly identified neurons with their function during behavior requires monitoring the activity of these cell types in vivo. Condylis et al. developed a platform for population functional imaging with spatial transcriptomics. Using newly acquired transcriptomic cell census data from the Allen Institute for Brain Science, the authors studied the function of cell types in primary somatosensory cortex in mice performing a tactile working memory task. Task-related properties of both excitatory and inhibitory neurons continued to differentiate as they were segregated into increasingly discrete molecular types. A new excitatory cell type, Baz1a, formed a sensory-driven circuit hub that orchestrated local sensory processing in superficial layers of the neocortex. This approach opens new venues for exploring information processing in the brain. —PRS

Science, abl5981, this issue p. 41

NEUROSCIENCE
The amygdala and goal-directed actions

Almost everything we do in our daily lives is goal directed. The brain can maintain a motivational state to direct actions to achieve desired outcomes. Using deep-brain calcium imaging, electrophysiology, and optogenetics in mice, Courtin et al. observed that at the time of goal-directed action, basolateral amygdala principal neurons integrate and encode pursued outcome identity, pursued outcome value, and action-outcome contingency information. At the time of consumption, basolateral amygdala neuronal firing represents current outcome identity and value. Together, action- and consumption-associated activity integrate behaviorally relevant information at distinct time points along goal-directed action-consumption sequences. —PRS

Science, abg7277, this issue p. 42

CORONAVIRUS
Antibody levels predict vaccine efficacy

Symptomatic COVID-19 infection can be prevented by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines. A "correlate of protection" is a molecular biomarker to measure how much immunity is needed to fight infection and is key for successful global immunization programs. Gilbert et al. determined that antibodies are the correlate of protection in vaccinated individuals enrolled in the Moderna COVE phase 3 clinical trial (see the Perspective by Openshaw). By measuring binding and neutralizing antibodies against the viral spike protein, the authors found that the levels of both antibodies correlated with the degree of vaccine efficacy. The higher the antibody level, the greater the protection afforded by the messenger RNA (mRNA) vaccine. Antibody levels that predict mRNA vaccine efficacy can therefore be used to guide vaccine regimen modifications and support regulatory approvals for a broader spectrum of the population. —PNK

Science, abm3425, this issue p. 43; see also abn0007, p. 22

NANOMATERIALS
Zeolitic nanotubes

Nanotubes generally have solid walls, but a low-dimensional version of zeolites now introduces porosity into such structures. Korde et al. used a structure-directing agent with a hydrophobic biphenyl group center connecting two long alkyl chains bearing hydrophilic bulky quaternary ammonium head groups to direct hydrothermal synthesis with silicon-rich precursors (see the Perspective by Fan and Dong). The nanotubes have a mesoporous central channel of approximately 3 nanometers and zeolitic walls with micropores less than 0.6 nanometers. Electron microscopy and modeling showed that the outer surface is a projection of a large-pore zeolite and the inner surface is a projection of a medium-pore zeolite. —PDS

Science, abg8, this issue p. 62; see also abn2048, p. 29

SOLAR CELLS
Directing efficient hole transport

Surface defects in three-dimensional perovskites can decrease performance but can be healed with coatings based on two-dimensional (2D) perovskite such as Ruddlesden-Popper phases. However, the bulky organic groups of these 2D phases can lead to low and anisotropic charge transport. F. Zhang et al. show that a metastable polymorph of a Dion-Jacobson 2D structure based on asymmetric organic molecules reduced the energy barrier for hole transport and their transport through the layer. When used as a top layer for a triple-cation mixed-halide perovskite, a solar cell retained 90% of its initial power conversion efficiency of 24.7% after 1000 hours of operation at approximately 40°C in nitrogen. —PDS

Science, abj2637, this issue p. 71

STRUCTURAL VIROLOGY
A block to viral cell entry

Crimean-Congo hemorrhagic fever virus is a tickborne virus that can cause severe disease

and even death in humans. Disease occurrence is linked to the geographic range of the tick vector, and climate change may increase this range. Infection of host cells requires the fusion glycoprotein Gc, which is the main target of neutralizing antibodies. Mishra *et al*. build on previous work that identified a combination of two Gc-targeting antibodies that gave postexposure protection in an animal model. The authors determined the structure of the antigen-binding fragments of the two antibodies bound to a prefusion form of Gc and also the structure of Gc after the conformational change into the trimeric postfusion form. The structures show how the antibodies work together to block membrane fusion. —VV

Science, abl6502, this issue p. 104

NITROGEN CYCLE
Consuming oxygen, but making it too

For marine microbes, there are myriad biological reactions involved in the cycling of nutrients and the generation of energy. Availability of oxygen is crucial for many species' metabolism. Kraft *et al*. were surprised to find that pure cultures of an ammonia-oxidizing archaean (AOA) (see the Perspective by Martens-Habbena and Qin), *Nitosopumilus maritimus*, were able to regenerate small amounts of oxygen when placed under anoxic conditions. Isotope labeling of nitrogen species revealed a series of reactions transforming nitrite, the expected metabolic end product, into nitric oxide, nitrous oxide, and, eventually, dinitrogen. Oxygen was also formed, likely from nitric oxide disproportionation, but was mostly consumed, which is consistent with the overall aerobic metabolism of AOA. These organisms can be found in oxygen-depleted waters and may benefit from producing oxygen from nitrite under these conditions. —MAF

Science, abe6733, this issue p. 97; see also abn0373, p. 27

INFLAMMATION
Complement coordination in macrophages

The serum complement system performs well-established functions in innate immune system recognition and destruction of invading pathogens, but complement proteins produced by immune cells can also execute distinct functions intracellularly. In human macrophages, Niyonzima *et al.* demonstrated that the complement protein C5a signals on mitochondrial membranes through its receptor, C5aR1, to control interleukin-1B (IL-1B) production in response to cholesterol crystal uptake. C5a–C5aR1 signaling altered mitochondrial activity and increased reactive oxygen species production, promoting IL-1B gene expression and processing. In a mouse model of atherosclerosis in which cholesterol crystal accumulates in arterial walls, deletion of C5aR1 in myeloid cells reduced the severity of cardiovascular disease. These results identify a function for intracellular complement proteins in myeloid cell mitochondrial metabolism and response to sterile inflammation. —CO

Sci. Immunol. **6**, eabf2489 (2021).

CELL AND GENE THERAPY
Making CAR T cells in vivo

Cardiac fibrosis, the stiffening and scarring of heart tissue, can be fatal. Rurik *et al*. designed an immunotherapy strategy to generate transient chimeric antigen receptor (CAR) T cells that can recognize the fibrotic cells in the heart (see the Perspective by Gao and Chen). By injecting CD5-targeted lipid nanoparticles containing the messenger RNA (mRNA) instructions needed to reprogram T lymphocytes, the researchers were able to generate therapeutic CAR T cells entirely inside the body. Analysis of a mouse model of heart disease revealed that the approach was successful in reducing fibrosis and restoring cardiac function. The ability to produce CAR T cells in vivo using modified mRNA may

have a number of therapeutic applications. —PNK

Science, abm0594, this issue p. 91; see also abn0851, p. 23

PALEOECOLOGY
Species shifts

Our anthropogenically warmed climate will lead to a suite of organismal changes. To predict how some of these may occur, we can look to past warm (interglacial) periods. Salvatteci *et al*. used this approach and looked at a marine sediment record of the Humboldt Current system off the coast of Peru (see the Perspective by Yasuhara and Deutsch). They found that previous warm periods were dominated by small, goby-like fishes, whereas this ecosystem currently is dominated by anchovy-like fishes. Such a shift is not only relevant to ecosystem shifts but also to fisheries because anchovies are heavily fished as a food source and gobies are much less palatable than anchovies. —SNV

Science, abj0270, this issue p. 101; see also abn2384, p. 25

NEUROINFLAMMATION
Denying entry

Blocking infiltration of pathogenic immune cells into the central nervous system (CNS) can prevent autoimmune neuroinflammation. However, selectively targeting pathogenic cells while still allowing the infiltration of protective cells has proved challenging. Charabati *et al*. identified dual immunoglobulin domain–containing cell adhesion molecule (DICAM) as a mediator of pathogenic, T helper 17 (T_H17) cell trafficking into the CNS. DICAM was highly expressed on T_H17 cells isolated from patients with multiple sclerosis. Blockade of DICAM with an antibody in mice prevented T_H17 cell infiltration, reducing disease severity in models of experimental autoimmune encephalomyelitis, suggesting that DICAM merits further exploration as a therapeutic target for neuroinflammation. —CSM

Sci. Transl. Med. **13**, eabj0473 (2022).

CANCER

Tertiary lymphoid structures in cancer

Ton N. Schumacher* and Daniela S. Thommen*

BACKGROUND: Tertiary lymphoid structures (TLSs) are organized aggregates of immune cells that form postnatally in nonlymphoid tissues. TLSs are not found under physiological conditions but arise in the context of chronic inflammation, such as in autoimmune disease, chronic infection, and cancer. With few exceptions, the presence of TLSs in tumors correlates with better prognosis and clinical outcome upon immunotherapy, but, in spite of their presumed importance, the drivers of TLS formation in cancer and the contribution of these structures to intratumoral immune responses remain incompletely understood.

ADVANCES: TLSs resemble secondary lymphoid organs (SLOs) anatomically, and it was originally assumed that their formation would largely be induced by the same stimuli. However, the cell pools and signals that provide inductive stimuli for TLS formation are at least partially different. For instance, several observations suggest that tumor-specific T and B cell immunity may induce some of the molecular factors required for TLS formation and maintenance, and heterogeneity in these drivers may result in distinct TLS states.

It has been speculated that TLSs recapitulate SLO functions at the inflamed tissue site, and available evidence suggests that a contribution of TLSs to the strength of tumor-specific immune responses is plausible. However, whether such a contribution primarily involves the boosting of T cell responses generated in SLOs or the development of new T and B cell reactivities remains a key unanswered question. In addition, the presence of TLSs at the tumor site may offer the possibility for the generation of qualitatively distinct immune responses. Specifically, because TLSs are not encapsulated, exposure of TLS-resident immune cells to macromolecules from the inflamed microenvironment appears to be a realistic possibility, and this could potentially sculpt the nature of intratumoral immune responses. Finally, recent studies suggest a role for TLSs in the clinical response to immune checkpoint blockade, which may make these structures attractive therapeutic targets. However, the development of such strategies should take into account the possible consequences of ectopic formation of lymphoid tissue at other body sites.

OUTLOOK: The prognostic and predictive value of TLSs in cancer has strengthened the interest in these structures as potential mediators of antitumor immunity. Although TLSs have been identified in many cancer types, the markers used to define and characterize TLSs have often varied across studies, complicating efforts to compare predictive value and to assess TLS heterogeneity between cancer types. Thus, the development of standardized approaches to measure TLS number and composition is likely to further reveal their predictive and prognostic value in different disease settings. Related to this, a more comprehensive characterization of TLSs may potentially lead to the identification of a spectrum of TLS states, based on aspects such as cellular composition, location, maturation, and function. Similar to the definition of T cell states in cancer, which has substantially improved our understanding of the role of specific T cell populations in tumor-specific immunity, the molecular definition of TLS states may help to improve their value as prognostic and predictive markers. Finally, a better appreciation of TLS function and the potential contribution of TLSs to autoimmune toxicity will be important to maximize their value as therapeutic targets. ∎

READ THE FULL ARTICLE AT

Defining TLS states. The molecular definition of TLS states may advance their use as prognostic and predictive markers. Characteristics that will provide insight into the diversity of TLS states include their cellular composition, location, and maturation; properties of their cytokine and chemokine environment; and their B cell receptor (BCR) and T cell receptor (TCR) repertoires. AID, activation-induced cytidine deaminase; CTL, cytotoxic T lymphocyte; FRCs, follicular reticular cells; FDCs, follicular dendritic cells; HEVs, high endothelial venules; Ig, immunoglobulin; T_{FH}, T follicular helper cell; T_H, T helper cell; T_{reg}, regulatory T cell.
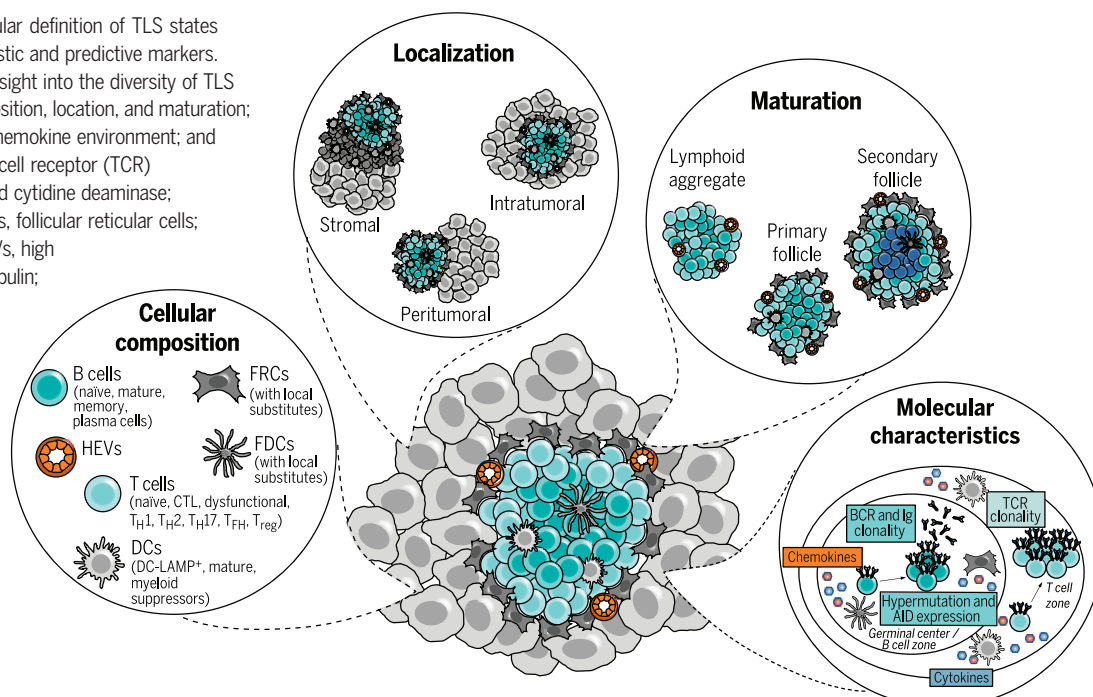
CANCER

Tertiary lymphoid structures in cancer

Ton N. Schumacher[1]* and Daniela S. Thommen[2]*

Ectopic lymphoid aggregates, termed tertiary lymphoid structures (TLSs), are formed in numerous cancer types, and, with few exceptions, their presence is associated with superior prognosis and response to immunotherapy. In spite of their presumed importance, the triggers that lead to TLS formation in cancer tissue and the contribution of these structures to intratumoral immune responses remain incompletely understood. Here, we discuss the present knowledge on TLSs in cancer, focusing on (i) the drivers of TLS formation, (ii) the function and contribution of TLSs to the antitumor immune response, and (iii) the potential of TLSs as therapeutic targets in human cancers.

Work over many years has documented that the presence of certain immune infiltrates in tumor lesions is associated with better prognosis in a number of cancer entities (1–3). In more recent years, efforts to increase tumor-specific T cell reactivity, either through infusion of ex vivo expanded intratumoral T lymphocytes (TILs) (4) or through blockade of immune checkpoint molecules on T cells (5–7), have provided causal evidence for a role of T cell immunity as a modifier of cancer growth. Furthermore, the observation that the presence of brisk immune infiltrates correlates with response to immune checkpoint blockade (ICB) (8–11) unites these two lines of research. Although the above data argue for the routine assessment of immune infiltrates in cancer lesions, there is increasing evidence that additional information may be gleaned from analysis of not just the presence but also the localization and interaction of immune cells at cancer sites.

A first, relatively straightforward refinement is the subdivision of T cells based on their location at the tumor border or in the tumor parenchyma (9). As may be expected, the presence of T cells in the tumor parenchyma is associated with improved clinical outcome, but whether this reflects increased attraction of T cells in those tumors that harbor an ongoing tumor-specific T cell response, or the active repulsion of T cells in other cancers, remains an important open question. Next to the location of intratumoral immune cells, the clustering of intratumoral immune infiltrates also appears of relevance. As a first example, an analysis of immune infiltrates in breast cancer has revealed that tumors with comparable immune infiltrates displayed distinct spatial distributions, referred to as mixed and compartmentalized organization (12). Importantly, compartmentalized organization, defined by the physical separation of clusters of immune cells and clusters of cancer cells, was associated with increased survival, independent from TIL density. Although in this study the prognostic potential was not formally coupled to the presence of tertiary lymphoid structures (TLSs), other recent studies have reported the association of TLSs in cancer lesions with improved prognosis (13, 14), and with response to ICB (15–17), in a number of human malignancies. Collectively, these observations suggest that not only the presence of an immune infiltrate in a tumor but also the organization of tumor-infiltrating immune cells in TLSs may be crucial. Main questions that should be further addressed in the coming years concern the molecular processes that lead to TLS formation in cancer, the types of cancer-associated TLSs that exist, and the consequences of their presence for the generation or maintenance of tumor-specific immunity.

Composition and organization of immune infiltrates in cancer

TLSs, sometimes also referred to as tertiary lymphoid organs or ectopic lymphoid structures, are organized aggregates of immune cells that arise postnatally in nonlymphoid tissues. TLSs are not present under physiological conditions but form in chronically inflamed environments, for instance, in autoimmune diseases (18), allograft rejection (19), chronic inflammation (20), and cancer (14, 21). TLSs have been reported in a number of cancer types such as non–small cell lung cancer (NSCLC), colorectal cancer (CRC), ovarian cancer, and melanoma (22–26). The occurrence of TLSs is likely to differ between cancer types, but with the presently available datasets, in which a number of different markers have been used to identify TLSs, a direct comparison has not been possible.

TLSs are characterized by an inner zone of CD20+ B cells that is surrounded by CD3+ T cells, similar to the lymph follicles in secondary lymphoid organs (SLOs) (14, 27). Although the specific composition of TLSs may vary, within the T cell compartment, CD4+ T follicular helper (T_{FH}) cells often represent the dominant subset (28), but CD8+ cytotoxic T cells, CD4+ T helper 1 (T_H1) cells, and regulatory T cells (T_{regs}) can also be present (24, 29, 30). Whereas B and T cell populations make up the bulk of TLS-associated immune cells, TLSs are also populated by distinct dendritic cell (DC) populations, for instance, CD21+ follicular dendritic cells (FDCs), which are of mesenchymal origin and play a critical role in the selection of memory B cells during germinal center (GC) reactions in SLOs (25, 31), or CD83+ mature DCs [in some studies also described as dendritic cell–lysosomal associated membrane protein (DC-LAMP)+ (24)], which predominantly localize in the T cell zone (32). The follicles can further contain scattered CD68+ macrophages for clearance of apoptotic cells, similar to their role in SLOs (33). A dense stromal network, similar to the one formed by follicular reticular cells (FRCs) in SLOs, anchors the TLSs at the chronically inflamed tissue site (34). Finally, peripheral node addressin (PNAd)–positive high endothelial venules (HEVs) provide the specialized vasculature associated with TLSs that is thought to mediate lymphocyte recruitment (31).

Recently, an additional type of structured immune infiltrate in cancers has been described (35). Specifically, intratumoral immune or antigen-presenting cell (APC) niches in renal cell carcinoma have been defined as small, APC-dense regions with more than five MHC II+ cells per 10,000 μm² that harbor tumor-reactive stem-like CD8+ T cells, crucial mediators of durable immunotherapy responses in mouse models (36–38). Of note, the absence of APC niches was associated with tumor progression, consistent with the possibility that these structures may play a critical role in maintaining tumor control. Although APC niches are distinct from TLSs, with the latter consisting of larger organized aggregates densely packed with both B and T lymphocytes, it is presently unclear whether APC niches could reflect a very early stage of TLS formation.

The drivers of TLS formation

SLOs (including lymph nodes, spleen, tonsils, Peyer's patches, and mucosa-associated lymphoid tissue) are situated throughout the body to allow antigen sampling from different tissues and thereby promote the induction of adaptive immune responses. In settings of ongoing chronic inflammation, extranodal seeding of lymphoid tissue occurs, resulting in the formation of TLSs at organ sites. To understand the development of such TLSs, it may be useful

[1]Division of Molecular Oncology and Immunology, Oncode Institute, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [2]Division of Molecular Oncology and Immunology, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands.

*Corresponding author. Email: t.schumacher@nki.nl (T.N.S.); d.thommen@nki.nl (D.S.T.)

to contrast it to the formation of SLOs during embryogenesis. The seeding and organization of SLOs, specifically of lymph nodes and Peyer's patches, results from a highly ordered series of events that involves an interplay between hematopoietic cells and nonlymphoid stromal cells, with critical roles for cytokines, chemokines, adhesion molecules, and survival factors as molecular components (39, 40). SLO formation is initiated early during embryogenesis by the colonization of the lymph node anlagen by hematopoietic lymphoid tissue inducer (LTi) cells, $CD4^+$ $CD3^-$ $CD45^+$ innate lymphoid cells that differentiate from fetal liver precursors and are characterized by the expression of the RORγt and Id2 transcription factors (41). Clustering of LTi cells drives the initial steps of SLO formation in a tumor necrosis factor (TNF) family member–dependent fashion, with central roles for lymphotoxin α1β2 (LTα1β2) and, to some extent, TNF (42), and in the absence of either LTi cells or LTα1β2, formation of both lymph nodes and Peyer's patches is precluded (41). LTα1β2 and TNF bind to their respective receptors, LTβR and TNFR1, on mesenchymal lymphoid tissue organizer (LTo) cells, thereby promoting the expression of adhesion molecules such as vascular cell-adhesion molecule 1 (VCAM1), intercellular adhesion molecule 1 (ICAM1), mucosal addressin cell-adhesion molecule 1 (MAdCAM1), and PNAd, as well as the production of a set of chemokines known as lymphoid or homeostatic chemokines, including CC-chemokine ligand 19 (CCL19), CCL21, and CXC-chemokine ligand 13 (CXCL13) (39, 42). Together, these molecules regulate the subsequent recruitment of immune cells to the lymphoid niche (42–44) and the vascularization by HEVs (45–48). Finally, compartmentalization of the nascent lymph follicle is achieved by the segregated expression of homeostatic chemokines with, for instance, $CCL19^+$ and/or $CCL21^+$ FRCs and $CXCL13^+$ FDCs guiding the distribution of lymphocytes that express the corresponding CCR7 and CXCR5 receptors, thereby allowing the formation of T cell and B cell zones (42, 49–51). Of note, lymphoid chemokine secretion also induces a positive feedback loop that is crucial for the maintenance of the lymphoid niche, as signaling through CXCR5, which is expressed on B cells and on LTi cells, has been found to induce LTα1β2 expression (45).

TLSs display a pronounced anatomical resemblance to SLOs but, in most tissues, lack the surrounding capsule (52). This absence of encapsulation may permit direct access of their cellular components to the surrounding tissue but also creates the possibility of exposure of TLS-resident immune cells to macromolecules from the inflamed microenvironment. Although the formation of TLSs and SLOs was initially thought to be induced by the same molecular factors, with roles for LTα1β2-LTβR

signaling and local expression of adhesion molecules and lymphoid chemokines, the cellular components involved are, at least partially, different, and the precipitating events that drive TLS generation are still only partly understood. In addition, a number of molecular inducers of TLSs that are independent of lymphotoxin signaling have been described [(14, 27, 53); see below]. Importantly, much of our understanding of the cellular and molecular processes that drive TLS formation has been obtained in models of autoimmune disease and chronic infection, and findings made in these disease models should thus only be seen as hypothesis-generating with regard to TLS formation in cancer tissue.

With respect to the upstream initiation of TLSs, it is, as of now, unclear whether bona fide LTi cells are required for the priming of the local mesenchyme or whether locally accumulated immune cells can substitute for LTi cells. In favor of the latter hypothesis, several immune cell populations—including T_H17 cells (54, 55) and innate lymphoid cell-3 [ILC3 (56)], which both share the RORγt transcription factor with classical LTi cells, effector $CD8^+$ T cells and natural killer cells (57, 58), B cells (59), and M1-polarized macrophages (60)—have all been reported to act as potential surrogate LTi cells in murine and human settings of either allograft rejection (54), autoimmunity (55), chronic inflammation (59, 60), or cancer (56, 57). Of note, unlike SLO formation, TLS induction may not always depend on lymphotoxin, as, for instance, interleukin-17 (IL-17) produced by T cells could induce CXCL13 and CCL19 expression in murine stromal cells in response to microbial stimulation, thereby promoting the formation of induced bronchus-associated lymphoid tissue (iBALT), a type of TLS that is formed in lung tissue (61). By the same token, lymphoid aggregates do develop in $LT\alpha-/-$ mice, although these structures do not show a segregation of T and B cell zones and lack HEVs (62), and thus may not be considered proper TLSs.

Similar to the presumed role of surrogate LTi cells in TLS generation, it is likely that certain local stromal and immune populations can act as surrogate LTo cells. Specifically, as has been shown for synovial fibroblasts from patients with rheumatoid arthritis, lymphotoxin and TNF signaling can induce tissue-resident fibroblasts to produce lymphoid chemokines such as CXCL13, CCL19, and CCL21, as well as survival factors including BAFF, IL-7, and April (34). In a B16-OVA melanoma model, a population of intratumoral cancer-associated fibroblasts could likewise act as LTo cells to induce TLS formation (58). Similarly, chemokine secretion by adipocytes and by vascular smooth muscle cells have been shown to explain TLS formation in the mesenteric adipose tissue of patients with Crohn's

disease (63) and in atherothrombotic arteries (64), respectively. With respect to the role of different SLO-associated chemokines in TLS formation, local TLS formation in the pancreas could be induced by tissue-specific expression of chemokines such as CXCL13 (51), CCL21, CCL19, and CXCL12 in murine models (50), suggesting that the importance of these downstream chemokines is shared between SLOs and TLSs. Of note, although each chemokine was able to independently induce TLSs, their individual presence yielded structures with slightly distinct characteristics (see below). If cells located within inflamed cancer tissue can produce any of these chemokines in a lymphotoxin- and TNF-independent fashion, this may be predicted to allow TLS formation without a requirement for LTi cells. Of note, a number of intratumoral T cell subsets, including T_{FH} cells in breast cancer (28) and programmed cell death-1 (PD-1) bright $CD8^+$ T cells in NSCLC (65), as well as macrophages and B cells in ovarian cancer (66) and fibroblasts in triple-negative breast cancer (67) express CXCL13, suggesting that immune and stromal cells may be able to function as LTo cells and contribute to TLS formation and/or maintenance. Next to LTi and LTo cells, HEVs play a role in TLS formation because they can regulate lymphocyte entry and control the type of lymphocytes that are recruited to the lymphoid tissue through the expression of vascular addressins (68).

In support of a role for the local tissue context in determining TLS composition, transgenic expression of different cytokines and chemokines in murine models has been shown to induce TLSs with distinct characteristics. For instance, tissue-specific expression of CXCL13 induced B cell aggregates that lacked FDC networks (51), whereas TNF and CXCL12 expression induced small lymphocytic infiltrates consisting predominantly of B cells, few T cells, and, in the latter case, DCs (50, 69). In addition, whereas the CCR7 ligands CCL19 and CCL21 were shown to induce similarly composed aggregates, the structures induced by CCL21 expression were both larger and more organized (50).

Differences in the detected TLS components have also been reported in distinct human cancer types (21, 27, 70), as, for instance, DC-LAMP$^+$ DCs have been described in TLSs in NSCLC but less in other cancer types (22, 24). In addition, T_{FH} cells have mostly been documented in TLSs in breast cancer (28, 71) and, more recently, in sarcoma (17). It is, however, important to note that much of the available data is derived from studies that used inconsistent markers to describe TLS components, and large-scale analyses using the same set of parameters, thereby allowing a rigorous assessment of TLS heterogeneity across cancers, are lacking as of now. A specific type of

heterogeneity for which a reasonable amount of evidence does exist relates to the extent of TLS maturation. Three maturation stages of lymphoid structures based on their structural similarity to SLOs have been defined in NSCLC, hepatocellular carcinoma (HCC), and CRC (72–74). The least-organized stage consists of dense lymphocytic aggregates without the presence of FDCs and with a lack of segregated T and B cell zones characteristic of bona fide TLSs. Primary follicle-like TLSs do contain FDCs but lack GC reactions. On the contrary, fully mature, secondary follicle–like TLSs also display active GCs, likely reflecting their full functional capacity.

Next to the evidence supporting the existence of different types of TLSs, arguably the most important difference between tumors is that whereas some are permissive for TLS formation, others are not, and it is important to understand under which conditions tumors do or do not support TLS formation. TLSs arise at sites of chronic inflammation, and several observations indicate that their formation is linked to antigen recognition by B and T cells at those sites. For instance, GC formation in SLOs, which is crucial for the generation of high-affinity, long-lived plasma cells and memory B cells (75), is regulated by the antigen-driven interaction between B cells and T_{FH} cells, and the fact that GC can form in TLSs suggests that a similar process of antigen recognition takes place in these structures. In human melanoma and ovarian cancer metastases, and in primary breast and gastric-esophageal cancers, clonal amplification of B cells and somatic hypermutation and isotype switching of immunoglobulins have been observed in microdissected TLSs, further reinforcing the concept of a local antigen-driven B cell response (76–81). Formation of TLSs has likewise been associated with the presence of antigen-specific T cell responses. Specifically, in NSCLC, the number of TLSs was shown to correlate with clonal dominance in both CD8+ and CD4+ T cells (82). In addition, tumor reactivity in human lung cancer was enriched in a subset of oligoclonal dysfunctional PD-1high CD8+ T cells and these PD-1high T cells were predominantly observed in TLSs (65), consistent with the hypothesis that tumor-reactive T cell responses are present in TLSs. Of note, unlike other CD8+ subsets, this PD-1high CD8+ subset has acquired the capacity to constitutively secrete CXCL13, one of the major chemoattractants involved in TLS formation (65). Furthermore, in patients with ovarian and uterine cancer, the presence of CXCL13+ CD103+ CD8+ T cells correlated with TLS abundance and with predicted neoantigen burden (83). Collectively, these human data provide strong evidence for continued antigen recognition in TLSs and suggest that antigen-specific cells present in TLSs can produce the molecular factors required for TLS induction and main-

tenance. Definitive evidence for a role of antigen recognition in TLS formation has been obtained in murine tumor models. Specifically, work from Engelhard and colleagues has shown that the CD8+ T cell pool is required for the efficient induction of lymph node–like vasculature, characterized by expression of PNAd and CCL21, in transplantable tumor models. In addition, PNAd expression on the intratumoral vasculature was higher in tumors that expressed a well-presented CD8+ T cell antigen (57). Recent work by the same group furthermore demonstrated that intratumoral CD8+ T cells and B cells jointly drive local fibroblast organization and TLS formation (58). Similarly, in a carcinogen-induced murine tumor model, HEV formation after T_{reg} depletion was dependent on CD8+ T cells (84). Next to the role for antigen recognition in the formation of TLSs, such antigen recognition may also be required for TLS maintenance, because TLS numbers rapidly go down after pathogen clearance in lungs of mice infected with influenza virus (20, 85–87) or in patients after clearance of gastric *Helicobacter pylori* infection (88). As discussed further below, the link between antigen recognition and TLS formation complicates the interpretation of the association between TLS presence and clinical response to ICB.

The immune infiltrate in tumors in which TLSs are present is often skewed toward a T_H1 or cytotoxic effector state (89), with expression of genes relating to chemoattraction (*CXCL9*, *CXCL10*, *CXCL11*) and cytotoxicity (*GZMB*, *GZMH*, *GNLY*), and characterized by expression of a series of immune checkpoint molecules (*PDCD1*, *CTLA4*, *LAG3*, *HAVCR*, *TIGIT*) (26, 90, 91). The fact that expression of many of these molecules is induced by (chronic) T cell activation is consistent with the notion that antigen encounter forms a driver of TLS formation. However, it also remains possible that the presence of TLSs promotes such a T_H1 or cytotoxic effector cell–like response or that both processes are stimulated by a common upstream event. Of note, the presence of T_{regs} has been suggested to impede TLS formation by preventing HEV induction and immune infiltration in murine fibrosarcoma models (84, 92). In addition to the emerging evidence suggesting that the nature of intratumoral T cell responses may influence TLS formation, it will be useful to determine how tumor cell–intrinsic characteristics can influence TLS formation. Work by Cabrita et al. did not observe an association between TLSs and either tumor mutational burden or specific driver mutations (16), but it will be important to further explore this relationship in additional cohorts, as well as the relationship between the presence of these tumor cell–intrinsic properties and the maturation state of TLSs. In breast cancer, evidence for FOXP1

expression by cancer cells as a determinant of their capacity to express lymphoid chemokines has been obtained (93). Although the above work identifies a number of factors that influence TLS formation in the tumor microenvironment, our understanding of the specific molecular determinants that create a local milieu that is or is not conducive to TLS formation is likely to be far from complete. As a framework to classify the different tumor microenvironments in which TLS formation does not occur, we propose to distinguish "restrictive tissue environments," in which TLS formation is actively suppressed, from "inadequate tissue environments," in which essential drivers, such as perhaps antigen, are lacking.

Role of TLSs in the regulation of tumor-specific immune responses

Because of their anatomical resemblance to SLOs, it has been suggested that TLSs recapitulate SLO functions at the inflamed tissue site. SLOs, specifically lymph nodes, foster the encounter of antigen-laden APCs from tissues and naïve lymphocytes from blood by providing a specialized niche that maximizes cell-cell contacts and thereby enables the generation of adaptive immune responses (45, 94). Accumulating evidence suggests that adaptive immune responses can also be generated or boosted in TLSs. One of the main effector functions associated with B cells in TLSs is the production of disease-relevant antibodies that can mark antigen-expressing cells for opsonization, complement-mediated lysis, or antibody-dependent cellular cytotoxicity (95). GC formation in TLSs has been found to correlate with serum autoantibody concentrations, disease severity, and decreased organ function in several autoimmune diseases, including Sjögren's syndrome (96), myasthenia gravis (97), and Hashimoto's thyroiditis (98), suggesting a potential contribution of TLSs to disease progression. Furthermore, evidence that clonal proliferation, isotype switching, and B cell effector differentiation actively take place in TLSs is provided by the detection of activation-induced cytidine deaminase, the enzyme driving somatic hypermutation and class switching, and BCL6, the transcription factor contributing to GC entry and late-stage B cell maturation, in TLSs (77, 99). Similarly, increased expression of activation markers has been observed on T cells within TLSs, as compared with other tumor-resident T cells in melanoma (15).

An important unresolved matter is whether TLSs mainly serve to reactivate or reeducate effector T cells, or whether they mostly support the priming of naïve T cells. Recruitment of effector T cells has been reported, particularly in the earliest phases of TLS formation (34), although this could also reflect the role of such effector T cells in TLS generation. Furthermore, TLS-associated T_H17 cells can acquire

phenotypic characteristics of T_{FH} cells in experimental autoimmune encephalomyelitis models (55), suggesting that a reeducation of effector populations may take place in TLSs. On the other hand, recruitment of naïve T cells to TLSs in pancreatic islets of nonobese diabetic mice has also been described. Of note, such naïve T cells do undergo proliferation in situ, consistent with local priming (100). Similarly, in the inflamed central nervous system, TLSs have been found to be involved in local priming of autoreactive T cell responses to endogenous myelin peptides (101). Additional evidence that TLSs can induce B and T cell responses in the absence of SLOs has been obtained using LTα-deficient mice that lack lymph nodes. Influenza A infection of such mice leads to the induction of iBALT at the time of viral clearance (102, 103). Furthermore, T and B cell responses to viral antigens in these mice were qualitatively similar to responses initiated in lymph nodes but caused less immunopathology (102, 104). In a modified vaccinia virus Ankara model, priming of antigen-specific T cell responses after blockade of lymphocyte egress from SLOs has been observed (87). Additionally, in a murine melanoma model, tumor-specific T cell responses driven by TLSs have been identified in the absence of SLOs and resulted in immune cell infiltration and tumor regression (105, 106). Collectively, these data provide compelling evidence that TLSs can replicate SLO functions locally.

With canonical SLOs having evolved as sites to efficiently generate antigen-specific adaptive immune responses, one may wonder what the value is of replicating the process of lymphoid neogenesis at the inflamed tissue site. With detailed information on the function of TLSs presently lacking, a number of models may be proposed to explain why it may be advantageous to create a lymphoid niche at the site of infection or cancer during conditions of chronic inflammation (Fig. 1): (i) Speed: Local priming of T and B cell pools may lead to faster immune responses because it circumvents the trafficking of DCs and lymphocytes to and from SLOs. In line with this, entry of naïve T cells into tumors has been described and was dependent on the development of PNAd- and CCL21-expressing vasculature in mouse models (57). As a counterargument, considering that TLSs are particularly prominent in the context of a chronic inflammation, the argument of "speed" seems less than compelling. (ii) Efficiency: Generation of a local lymphoid niche may increase the likelihood of encounters between disease-associated antigens and rare matching lymphocytes, thereby, perhaps, enabling the induction of more vigorous or more broad immune responses. In this model, the functional and phenotypic properties of T and B cells induced in TLSs and SLOs may be identical, but the antigens that they target

could partially differ. (iii) Control: Having a lymphoid niche that is in direct contact with the inflamed tissue site may provide an additional opportunity to steer the immune response. For instance, cytokines and/or metabolic factors produced in the surrounding tumor tissue may potentially percolate through the TLSs and thereby influence the nature of the immune response that is created. (iv) Survival: The presence of large numbers of dysfunctional PD-1^{high} CD8$^+$ T cells, a phenotype that has been linked to preferential tumor reactivity (65, 107, 108), in some TLSs is consistent with a model in which TLSs are important not only for the induction but also for the maintenance of immune responses. The secretion of survival factors by TLS-associated fibroblasts and other cell subsets supports lymphocyte homeostasis in TLSs (34, 109) and, by analogy, may contribute to the long-term persistence of tumor-reactive T cells at tumor sites. To better understand the benefits, and possibly also detriments, of generating ectopic lymphoid niches, it will be important to determine whether immune responses induced in TLSs and SLOs are similar or distinct, with respect to either the antigens that they target or the properties of the antigen-specific lymphocyte pool that is created. Experimental approaches to induce the formation or disassembly of TLSs on command may be of value to dissect their effects on the tumor-resident immune cell pool.

Prognostic and predictive potential of TLSs

TLSs are associated with favorable prognosis in many cancer types (27), and the prognostic value of TLSs is often independent of TNM staging, as, for instance, documented in lung (110), colorectal (74), and pancreatic (111) cancer. TLS density as well as the presence of their components, such as T_{FH} cells, follicular B cells, DC-LAMP$^+$ mature DCs, and HEVs, have been shown to correlate with better survival in many different tumor types (22, 111–113). In addition, multiple gene expression signatures associated with TLSs have shown positive prognostic value, including a plasma cell signature in ovarian cancer (29), a T_{FH} signature in head and neck squamous cell carcinoma (114), and various gene signatures associated with lymphoid chemokines (including CCL5, CXCL9, CXCL10, and CXCL13) in CRC (26), melanoma (115), and breast cancer (28, 116). Furthermore, the presence of TLSs in tumors is frequently accompanied by a general increase in immune infiltration, as, for instance, shown in human NSCLC and in triple-negative breast cancer (24, 117, 118). Finally, the combination of TLSs and brisk intratumoral CD8$^+$ T cell infiltrates correlates with superior prognosis compared with CD8$^+$ T cell infiltration alone (29, 119), an observation that has been used as an argument for a superior quality of the immune response generated in tumors that harbor TLSs.

Although it is tempting to interpret the observed correlations as evidence for a central role of TLSs in the induction or maintenance of tumor-specific immunity, it is important to realize that the formation of TLSs appears to depend on antigen recognition. As such, part of the observed prognostic value is likely to be explained by the fact that TLS formation indicates the presence of an ongoing immune response. To obtain further insight into this matter, it may be valuable to compare the cell states of tumor-reactive T cells that reside in lesions that harbor or lack TLSs. In addition, it may be of interest to identify (tumor) cell parameters that influence TLS formation in mouse models independent of antigen load, to subsequently test their effect on tumor control.

As discussed above, with the caveat that available studies in some cases vary with respect to the markers used for TLS identification, there are indications for TLS heterogeneity between cancer types as well as between patients (21, 27), and work in chemokine-transgenic mouse models provides reasonable mechanistic support for this notion (50, 51, 69). It has been hypothesized that heterogeneity in TLS maturation state or location influences their prognostic value. In support of this, risk of recurrence was lower in patients with HCC or CRC harboring TLSs with primary or secondary follicle–like differentiation, compared with those with lymphoid aggregates (73, 74). With regard to TLS location, most studies have reported the presence of peritumoral TLSs, but intratumoral TLSs have been described in HCC (73), germ cell tumors (120), and in lung metastases of renal cell carcinoma (121). In HCC, the presence of peritumoral TLSs was associated with a higher risk of cancer recurrence and unfavorable outcome as compared with intratumoral TLSs (73). However, in most cancers, no clear association between peri- or intratumoral location of TLSs and prognosis has been established. It should be noted though that the definition of peritumoral TLSs often does not differentiate between TLSs located in the stroma with clear separation from the tumor parenchyma and TLSs at the invasive margin (122), and it may be postulated that the association of these two subtypes of peritumoral TLSs with disease prognosis could differ.

Although TLSs are generally associated with good prognosis in most cancer types, their presence has been linked to tumor development or progression in some cancer types (123–125). A number of potential immunosuppressive mechanisms have been invoked to explain this observation. First, depletion of T_{regs}, which were predominantly present within TLSs, improved tumor control in a murine lung adenocarcinoma model, suggesting that TLS-associated T_{regs} may suppress endogenous antitumor T cell responses (124). Second, next
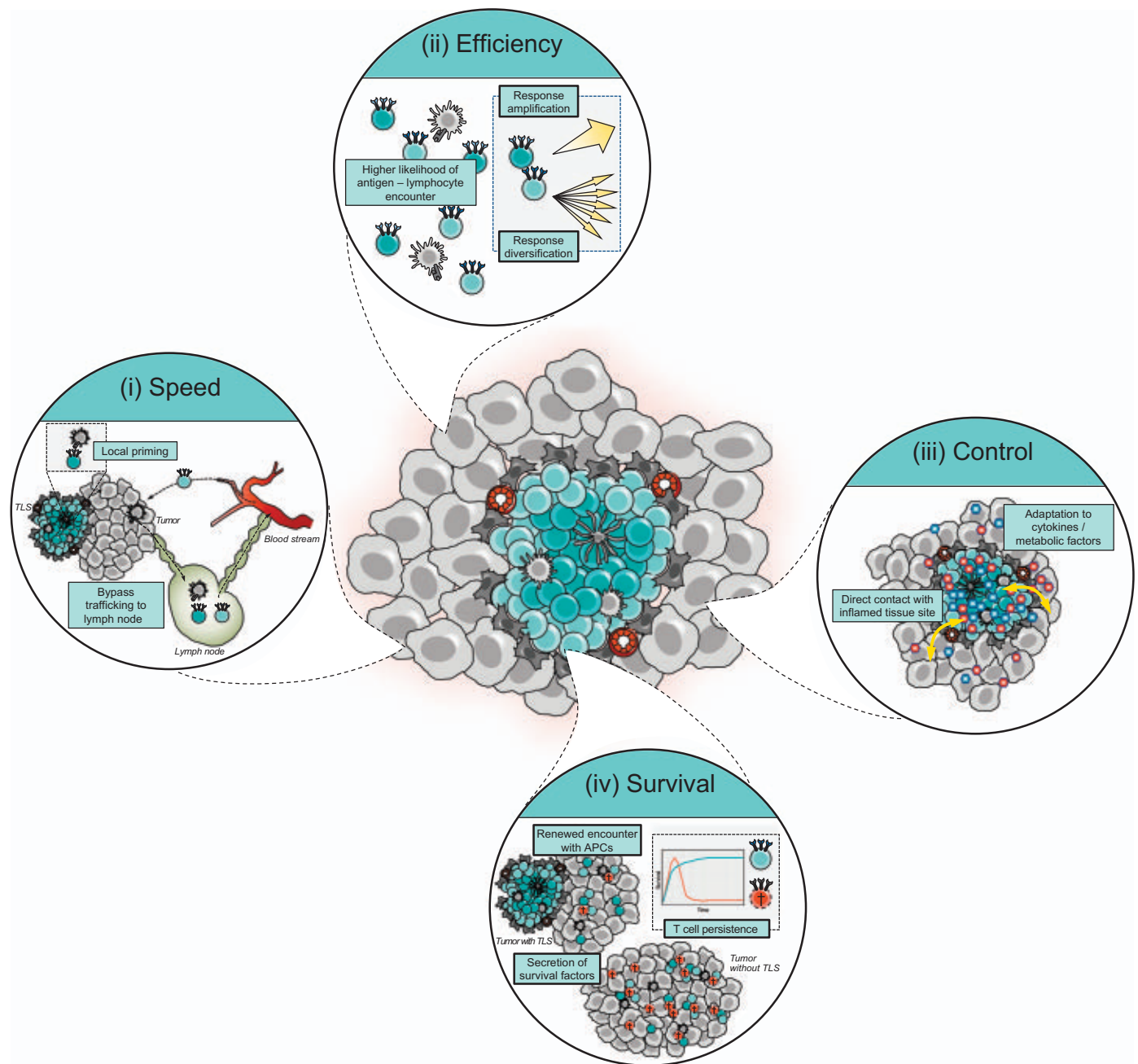
Fig. 1. Potential contributions of TLSs to antitumor immunity. The presence of TLSs in cancer tissue could support antitumor immune responses in different ways: (i) Speed: The priming of T and B cells at the tissue site may shorten the time to generate immune responses because it bypasses the trafficking of DCs and lymphocytes to and from SLOs. (ii) Efficiency: The formation of a local lymphoid niche may foster the induction of stronger or broader immune responses because lymphocytes may be more likely to encounter cognate antigen. (iii) Control: The direct exposure of TLS-associated immune cells to the inflamed tissue milieu may enable the fine-tuning of immune responses toward specific output signals. (iv) Survival: Lymphocyte homeostasis and survival could be promoted by survival factors that are secreted by TLS-associated cell populations or by repeated APC encounter by effector T cells.

to their possible role as producers of opsonizing tumor-specific antibodies, TLS-resident B cells may also suppress tumor-specific immunity, for instance, through IL-10 secretion (126, 127). In addition, depending on the antibody isotypes produced and immune cell types present, tumor-specific antibodies may conceivably also dampen tumor-specific immune responses by signaling through inhibitory Fc receptors (126). Finally, TLSs have been proposed as microniches that may foster the transformation and outgrowth of malignant cells, based on the observation that HCC progenitor cells first appear in TLSs before egressing and forming liver tumors (125). Likewise, clusters of cancer cells have been detected within TLSs in human breast cancers, and their presence was associated with lymphatic invasion and higher nodal stage (128). Although the above observations provide evidence for heterogeneity in TLS composition that influences their prognostic value, direct evidence for the existence of "suppressive" TLSs that promote tumor progression is still limited. Methods to

specifically alter TLS properties in situ could offer a powerful approach to address such questions of causality.

A number of recent studies have also provided evidence for a predictive value of TLSs in response to ICB. Specifically, the presence of TLSs and brisk B cell infiltrates in pretreatment biopsies of melanoma, renal cell carcinoma, soft tissue sarcoma, and urothelial carcinoma has been shown to correlate with response to PD-1 or combined PD-1 plus cytotoxic T lymphocyte–associated protein 4 (CTLA-4) blockade (15–17, 129). Similarly, a number of TLS components, including memory-like B cells and plasmablast-like cells, were enriched in pretreatment biopsies of ICB responders in melanoma (130). Furthermore, the presence of the PD-1^{high} dysfunctional CD8$^+$ T cells that predominantly localize within TLSs was shown to predict response to PD-1 blockade in late-stage NSCLC (65). Interestingly, TLS abundance correlated with programmed cell death ligand-1 (PD-L1) expression on immune cells (131, 132) but not on tumor cells (16). In mice, combination therapy that led to the induction of TLSs also sensitized tumors to ICB in a checkpoint blockade–resistant tumor model and resulted in the generation of effector and memory T cells (133), suggesting that TLSs either directly contribute to the ICB response or report on a tumor microenvironment that is permissive to ICB.

Intriguingly, analysis of on-treatment tumor biopsies has shown that ICB treatment can also promote the formation of TLSs. After neoadjuvant ICB in high-risk melanoma and urothelial carcinoma, tumors of responding patients showed a higher number of TLS-associated B cells relative to matched pretherapy samples (15). In several studies exploring neoadjuvant PD-1 blockade in NSCLC and PD-1 plus CTLA-4 blockade in urothelial cancer, an increase in TLSs has been observed in regressing lesions (134, 135). Similarly, ICB treatment increased the number and size of TLSs in a murine melanoma model, which correlated with superior tumor control (58). Considering the role of TLSs in promoting antigen-specific T and B cell responses, it may be speculated that ICB enhances not only TLS formation but also TLS functionality. Although a number of observations are consistent with a model in which ICB affects TLS functionality (15, 136, 137), the evidence is still circumstantial. Spatial analysis of tumors at very early time points (hours to days) after start of treatment, or analysis of ICB-treated ex vivo human tumor cultures (138), should be helpful to gain further insight into this matter.

Therapeutic induction of TLSs

In view of the reported association between TLSs and disease outcome in a number of settings, the induction of TLSs could form an attractive therapeutic strategy. The feasibility of local TLS induction by tissue-specific expression of TLS-associated cytokines and chemokines, including lymphotoxin (139), TNFα (140, 141), LIGHT (100), CXCL13 (51), CCL21, CCL19, and CXCL12 (50), has been demonstrated in murine models. In addition, evidence that TLSs can be therapeutically induced and associate with tumor control was obtained in mouse models of breast and neuroendocrine pancreatic cancers, in which the combination of PD-L1 blockade with anti-angiogenic therapies resulted in the transformation of tumor blood vessels into HEVs followed by TLS formation, increased CD8$^+$ T cell stimulation, and tumor destruction (133, 142). TLS induction independent of ICB has been observed in human cancers (Fig. 2), for instance, in patients with high-grade cervical intraepithelial neoplasia (CIN2/3), where TLS formation and clonal expansion of TLSs could be observed in regressing lesions after vaccination against the human papillomavirus oncoproteins (143). Similarly, therapeutic vaccination with an irradiated, allogeneic granulocyte-macrophage colony-stimulating factor–secreting pancreatic tumor vaccine (GVAX) in combination with cyclophosphamide led to TLS formation in pancreatic cancers in a large majority of patients (144). As a side note, the observed induction of TLSs upon these different types of vaccinations provides strong evidence that the strength of the antigen-specific immune responses forms a determinant of TLS formation in human disease. With respect to the effect of conventional therapies on TLS formation, induction of TLSs has been observed after neoadjuvant chemotherapy in NSCLC (145) and hepatoblastoma related to APC mutations (146). Notably, opposing findings were obtained in squamous cell lung cancer, in which neoadjuvant chemotherapy treatment resulted in impaired TLS maturation and loss of GCs (72). In addition, a similar observation has been made after steroid treatment, which is often coadministered with chemotherapy, in lung cancer (72) and urothelial cancer (135). At present, it is not entirely clear whether the negative effect on TLS organization depends

Chemotherapy
Vaccination
(ICB)

ICB

Induction of TLS

Increase in TLS numbers
Shift towards more mature TLS
Enhanced TLS function
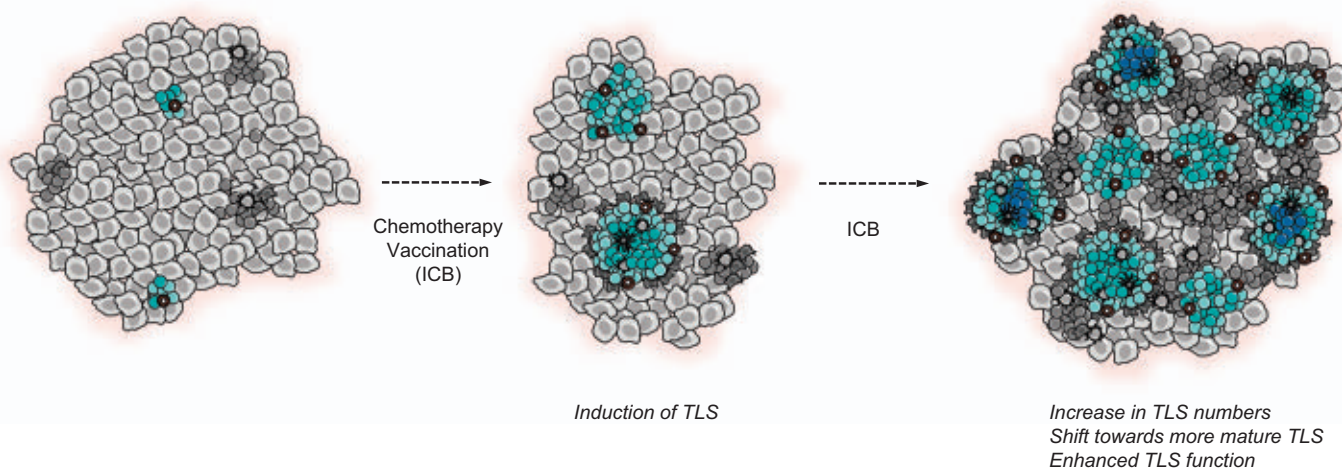
Fig. 2. Potential impact of cancer treatment on TLSs. Several therapeutic strategies have been found to induce or boost TLS formation in cancer. For instance, neoadjuvant chemotherapy was shown to promote de novo TLS development in NSCLC (145) and hepatoblastoma (146). Similarly, cancer vaccines were shown to promote TLS formation in pancreatic cancer (144) or CIN2/3 lesions (143). Although ICB has been shown to increase TLS numbers in several cancer types (15, 134, 135), it is unclear at present whether it can also induce de novo TLS formation. Additionally, ICB has been suggested to enhance TLS function in mice and humans by promoting the generation of effector and memory T cells (133), the activation of T$_{FH}$ cells (136, 137), and the induction of B cell class switching (15). Based on these observations, it is conceivable that ICB can also induce TLS maturation, although direct evidence is lacking as of now.
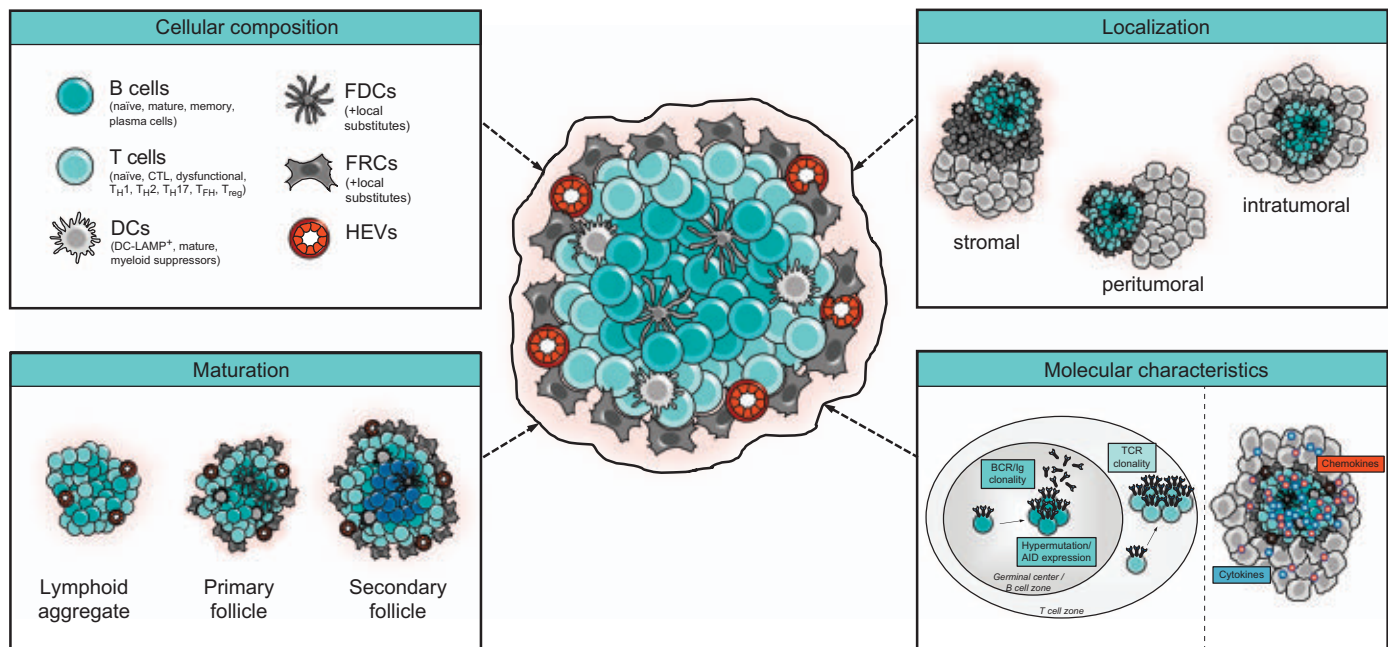
Fig. 3. Defining TLS states. A comprehensive molecular definition of TLS states using a consistent set of parameters should improve their value as prognostic and predictive biomarkers. Aspects that will provide more insight into the existence of TLS states include their cellular composition, location, and maturation; the molecular characteristics of their cytokine and chemokine milieu; and their B cell receptor (BCR) and T cell receptor (TCR) repertoires. AID, activation-induced cytidine deaminase; CTL, cytotoxic T lymphocyte; Ig, immunoglobulin.

on the type of chemotherapy used or on the concomitant treatment with steroids.

One important factor to consider is that although inducing or augmenting TLS function may improve tumor control, such interventions may at the same time boost autoreactive T and B cell responses at other tissue sites. Autoimmune reactions are observed as the main type of toxicity after ICB, and these so-called immune-related adverse events resemble the inflammatory processes that are often found in autoimmune diseases, including arthritis, myositis, thyroiditis, vasculitis, and colitis (*147*). Considering that TLSs have been found to support local inflammatory processes in many autoimmune diseases, it is conceivable that approaches that boost TLS numbers or TLS functionality could also increase ICB-induced autoimmune toxicity. Although, at present, few data exist on the role of TLSs in immune-related adverse events, an association between TLS formation and autoimmune myopathy upon PD-1 blockade has been reported. Specifically, biopsies from patients presenting with myalgia and muscle weakness after anti–PD-1 treatment revealed CD8$^+$ T cell–driven muscle tissue destruction that was associated with the formation of TLS-like structures expressing PNAd and CCL21 (*148*). Hence, induction or boosting of TLS function may promote not only antitumor responses but also the expansion of autoreactive T and B cells, and the risk-benefit ratio of such approaches therefore needs to be carefully evaluated.

Concluding remarks

Recent studies describing the prognostic and predictive value of TLSs in cancer have fueled interest in these structures as potential mediators of antitumor immunity. Based on available evidence, it is plausible that TLSs contribute to the strength of tumor-specific immune responses. However, whether this primarily involves the boosting of T cell responses generated in SLOs or the development of new T and B cell reactivities remains a key unanswered question. Similarly, the presence of TLSs at the tumor site offers a clear possibility for the generation of qualitatively distinct immune responses through the effects of local tissue factors. However, direct evidence for a distinct nature of immune responses that are formed or boosted in TLSs is presently lacking. Although TLSs have been described in numerous cancer types and their prognostic value is beyond doubt, the usage of consistent markers to define and characterize TLSs should form an area of future attention to maximize the value of these structures as potential biomarkers. Related to this, a more comprehensive characterization of TLSs would likely help provide a definition of a spectrum of "TLS states," based on aspects such as cellular composition, location, maturation, and function (Fig. 3). Much like the definition of T cell states has helped the field to better understand their role in cancer control, the molecular definition of TLS states could improve their value as prognostic and predictive mark-

ers. Finally, a more detailed understanding of TLS function and their potential role in autoimmune toxicity will be helpful to appreciate their value as therapeutic targets.

REFERENCES AND NOTES

1. J. Galon *et al.*, Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006). doi: 10.1126/science.1129139; pmid: 17008531
2. F. Pagès *et al.*, International validation of the consensus Immunoscore for the classification of colon cancer: A prognostic and accuracy study. *Lancet* **391**, 2128–2139 (2018). doi: 10.1016/S0140-6736(18)30789-X; pmid: 29754777
3. W. H. Fridman, F. Pagès, C. Sautès-Fridman, J. Galon, The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012). doi: 10.1038/nrc3245; pmid: 22419253
4. S. A. Rosenberg, N. P. Restifo, Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348**, 62–68 (2015). doi: 10.1126/science.aaa4967; pmid: 25838374
5. D. M. Pardoll, The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12**, 252–264 (2012). doi: 10.1038/nrc3239; pmid: 22437870
6. F. S. Hodi *et al.*, Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010). doi: 10.1056/NEJMoa1003466; pmid: 20525992
7. S. L. Topalian, C. G. Drake, D. M. Pardoll, Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Curr. Opin. Immunol.* **24**, 207–212 (2012). doi: 10.1016/j.coi.2011.12.009; pmid: 22236695
8. R. S. Herbst *et al.*, Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014). doi: 10.1038/nature14011; pmid: 25428504
9. P. C. Tumeh *et al.*, PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014). doi: 10.1038/nature13954; pmid: 25428505
10. F. Azimi *et al.*, Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J. Clin. Oncol.*

30, 2678–2683 (2012). doi: 10.1200/JCO.2011.37.8539; pmid: 22711850

11. P. Savas et al., Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat. Med. 24, 986–993 (2018). doi: 10.1038/s41591-018-0078-7; pmid: 29942092

12. L. Keren et al., A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. Cell 174, 1373–1387.e19 (2018). doi: 10.1016/j.cell.2018.08.039; pmid: 30193111

13. M. C. Dieu-Nosjean et al., Tertiary lymphoid structures, drivers of the anti-tumor responses in human cancers. Immunol. Rev. 271, 260–275 (2016). doi: 10.1111/imr.12405; pmid: 27088920

14. C. Sautès-Fridman et al., Tertiary lymphoid structures in cancers: Prognostic value, regulation, and manipulation for therapeutic intervention. Front. Immunol. 7, 407 (2016). doi: 10.3389/fimmu.2016.00407; pmid: 27752258

15. B. A. Helmink et al., B cells and tertiary lymphoid structures promote immunotherapy response. Nature 577, 549–555 (2020). doi: 10.1038/s41586-019-1922-8; pmid: 31942075

16. R. Cabrita et al., Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature 577, 561–565 (2020). doi: 10.1038/s41586-019-1914-8; pmid: 31942071

17. F. Petitprez et al., B cells are associated with survival and immunotherapy response in sarcoma. Nature 577, 556–560 (2020). doi: 10.1038/s41586-019-1906-8; pmid: 31942077

18. A. Manzo, M. Bombardieri, F. Humby, C. Pitzalis, Secondary and ectopic lymphoid tissue responses in rheumatoid arthritis: From inflammation to autoimmunity and tissue damage/remodeling. Immunol. Rev. 233, 267–285 (2010). doi: 10.1111/j.0105-2896.2009.00861.x; pmid: 20193005

19. O. Thaunat et al., Chronic rejection triggers the development of an aggressive intragraft immune response through recapitulation of lymphoid organogenesis. J. Immunol. 185, 717–728 (2010). doi: 10.4049/jimmunol.0903589; pmid: 20525884

20. K. Neyt, F. Perros, C. H. GeurtsvanKessel, H. Hammad, B. N. Lambrecht, Tertiary lymphoid organs in infection and autoimmunity. Trends Immunol. 33, 297–305 (2012). doi: 10.1016/j.it.2012.04.006; pmid: 22622061

21. E. J. Colbeck, A. Ager, A. Gallimore, G. W. Jones, Tertiary lymphoid structures in cancer: Drivers of antitumor immunity, immunosuppression, or bystander sentinels in disease? Front. Immunol. 8, 1830 (2017). doi: 10.3389/fimmu.2017.01830; pmid: 29312327

22. M. C. Dieu-Nosjean et al., Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. J. Clin. Oncol. 26, 4410–4417 (2008). doi: 10.1200/JCO.2007.15.0284; pmid: 18802153

23. L. de Chaisemartin et al., Characterization of chemokines and adhesion molecules associated with T cell presence in tertiary lymphoid structures in human lung cancer. Cancer Res. 71, 6391–6399 (2011). doi: 10.1158/0008-5472.CAN-11-0952; pmid: 21900403

24. J. Goc et al., Dendritic cells in tumor-associated tertiary lymphoid structures signal a Th1 cytotoxic immune contexture and license the positive prognostic value of infiltrating CD8+ T cells. Cancer Res. 74, 705–715 (2014). doi: 10.1158/0008-5472.CAN-13-1342; pmid: 24366885

25. F. Bergomas et al., Tertiary intratumor lymphoid tissue in colo-rectal cancer. Cancers 4, 1–10 (2011). doi: 10.3390/cancers4010001; pmid: 24213222

26. D. Coppola et al., Unique ectopic lymph node-like structures present in human primary colorectal carcinoma are identified by immune gene array profiling. Am. J. Pathol. 179, 37–45 (2011). doi: 10.1016/j.ajpath.2011.03.007; pmid: 21703392

27. C. Sautès-Fridman, F. Petitprez, J. Calderaro, W. H. Fridman, Tertiary lymphoid structures in the era of cancer immunotherapy. Nat. Rev. Cancer 19, 307–325 (2019). doi: 10.1038/s41568-019-0144-6; pmid: 31092904

28. C. Gu-Trantien et al., CD4+ follicular helper T cell infiltration predicts breast cancer survival. J. Clin. Invest. 123, 2873–2892 (2013). doi: 10.1172/JCI67428; pmid: 23778140

29. D. R. Kroeger, K. Milne, B. H. Nelson, Tumor-infiltrating plasma cells are associated with tertiary lymphoid structures, cytolytic T-cell responses, and superior prognosis in ovarian cancer. Clin. Cancer Res. 22, 3005–3015 (2016). doi: 10.1158/1078-0432.CCR-15-2762; pmid: 26763251

30. A. Hennequin et al., Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients. OncoImmunology 5, e1054598

(2015). doi: 10.1080/2162402X.2015.1054598; pmid: 27057426

31. G. Di Caro et al., Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. Clin. Cancer Res. 20, 2147–2158 (2014). doi: 10.1158/1078-0432.CCR-13-2590; pmid: 24523438

32. T. P. McMullen, R. Lai, L. Dabbagh, T. M. Wallace, C. J. de Gara, Survival in rectal cancer is predicted by T cell infiltration of tumour-associated lymphoid nodules. Clin. Exp. Immunol. 161, 81–88 (2010). doi: 10.1111/j.1365-2249.2010.04147.x; pmid: 20408858

33. M. Baratin et al., T cell zone resident macrophages silently dispose of apoptotic cells in the lymph node. Immunity 47, 349–362.e5 (2017). doi: 10.1016/j.immuni.2017.07.019; pmid: 28801233

34. F. Barone et al., Stromal fibroblasts in tertiary lymphoid structures: A novel target in chronic inflammation. Front. Immunol. 7, 477 (2016). doi: 10.3389/fimmu.2016.00477; pmid: 27877173

35. C. S. Jansen et al., An intra-tumoral niche maintains and differentiates stem-like CD8 T cells. Nature 576, 465–470 (2019). doi: 10.1038/s41586-019-1836-5; pmid: 31827286

36. I. Siddiqui et al., Intratumoral Tcf1+PD-1+CD8+ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. Immunity 50, 195–211.e10 (2019). doi: 10.1016/j.immuni.2018.12.021; pmid: 30635237

37. S. Kurtulus et al., Checkpoint blockade immunotherapy induces dynamic changes in PD-1−CD8+ tumor-infiltrating T cells. Immunity 50, 181–194.e6 (2019). doi: 10.1016/j.immuni.2018.11.014; pmid: 30635236

38. B. C. Miller et al., Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. Nat. Immunol. 20, 326–336 (2019). doi: 10.1038/s41590-019-0312-6; pmid: 30778252

39. T. D. Randall, D. M. Carragher, J. Rangel-Moreno, Development of secondary lymphoid organs. Annu. Rev. Immunol. 26, 627–650 (2008). doi: 10.1146/annurev.immunol.26.021607.090257; pmid: 18370924

40. S. A. van de Pavert, R. E. Mebius, New insights into the development of lymphoid tissues. Nat. Rev. Immunol. 10, 664–674 (2010). doi: 10.1038/nri2832; pmid: 20706277

41. T. Cupedo, G. Kraal, R. E. Mebius, The role of CD45+CD4+CD3− cells in lymphoid organ development. Immunol. Rev. 189, 41–50 (2002). doi: 10.1034/j.1600-065X.2002.18905.x; pmid: 12445264

42. R. E. Mebius, Organogenesis of lymphoid tissues. Nat. Rev. Immunol. 3, 292–303 (2003). doi: 10.1038/nri1054; pmid: 12669020

43. S. Nishikawa, K. Honda, P. Vieira, H. Yoshida, Organogenesis of peripheral lymphoid organs. Immunol. Rev. 195, 72–80 (2003). doi: 10.1034/j.1600-065X.2003.00063.x; pmid: 12969311

44. T. Cupedo, R. E. Mebius, Role of chemokines in the development of secondary and tertiary lymphoid tissues. Semin. Immunol. 15, 243–248 (2003). doi: 10.1016/j.smim.2003.08.002; pmid: 15001173

45. D. L. Drayton, S. Liao, R. H. Mounzer, N. H. Ruddle, Lymphoid organ development: From ontogeny to neogenesis. Nat. Immunol. 7, 344–353 (2006). doi: 10.1038/ni1330; pmid: 16550197

46. Y. Okabe, R. Medzhitov, Tissue-specific signals control reversible program of localization and functional polarization of macrophages. Cell 157, 832–844 (2014). doi: 10.1016/j.cell.2014.04.016; pmid: 24792964

47. S. A. Luther, K. M. Ansel, J. G. Cyster, Overlapping roles of CXCL13, interleukin 7 receptor α, and CCR7 ligands in lymph node development. J. Exp. Med. 197, 1191–1198 (2003). doi: 10.1084/jem.20021294; pmid: 12732660

48. H. Yoshida et al., Different cytokines induce surface lymphotoxin-αβ on IL-7 receptor-α cells that differentially engender lymph nodes and Peyer's patches. Immunity 17, 823–833 (2002). doi: 10.1016/S1074-7613(02)00479-X; pmid: 12479827

49. H. Fleige et al., IL-17-induced CXCL12 recruits B cells and induces follicle formation in BALT in the absence of differentiated FDCs. J. Exp. Med. 211, 643–651 (2014). doi: 10.1084/jem.20131737; pmid: 24663215

50. S. A. Luther et al., Differing activities of homeostatic chemokines CCL19, CCL21, and CXCL12 in lymphocyte and dendritic cell recruitment and lymphoid neogenesis.

J. Immunol. 169, 424–433 (2002). doi: 10.4049/jimmunol.169.1.424; pmid: 12077273

51. S. A. Luther, T. Lopez, W. Bai, D. Hanahan, J. G. Cyster, BLC expression in pancreatic islets causes B cell recruitment and lymphotoxin-dependent lymphoid neogenesis. Immunity 12, 471–481 (2000). doi: 10.1016/S1074-7613(00)80199-5; pmid: 10843380

52. N. H. Ruddle, Lymphatic vessels and tertiary lymphoid organs. J. Clin. Invest. 124, 953–959 (2014). doi: 10.1172/JCI71611; pmid: 24590281

53. C. Gago da Graça, L. G. M. van Baarsen, R. E. Mebius, Tertiary lymphoid structures: Diversity in their development, composition, and role. J. Immunol. 206, 273–281 (2021). doi: 10.4049/jimmunol.2000873; pmid: 33397741

54. C. Deteix et al., Intragraft Th17 infiltrate promotes lymphoid neogenesis and hastens clinical chronic rejection. J. Immunol. 184, 5344–5351 (2010). doi: 10.4049/jimmunol.0902999; pmid: 20357253

55. A. Peters et al., Th17 cells induce ectopic lymphoid follicles in central nervous system tissue inflammation. Immunity 35, 986–996 (2011). doi: 10.1016/j.immuni.2011.10.015; pmid: 22177922

56. P. Carrega et al., NCR+ILC3 concentrate in human lung cancer and associate with intratumoral lymphoid structures. Nat. Commun. 6, 8280 (2015). doi: 10.1038/ncomms9280; pmid: 26395069

57. J. D. Peske et al., Effector lymphocyte-induced lymph node-like vasculature enables naive T-cell entry into tumours and enhanced anti-tumour immunity. Nat. Commun. 6, 7114 (2015). doi: 10.1038/ncomms8114; pmid: 25968334

58. A. B. Rodriguez et al., Immune mechanisms orchestrate tertiary lymphoid structures in tumors via cancer-associated fibroblasts. Cell Rep. 36, 109422 (2021). doi: 10.1016/j.celrep.2021.109422; pmid: 34289373

59. M. Lochner et al., Microbiota-induced tertiary lymphoid tissues aggravate inflammatory disease in the absence of RORγt and LTi cells. J. Exp. Med. 208, 125–134 (2011). doi: 10.1084/jem.20100052; pmid: 21173107

60. K. Guedj et al., M1 macrophages act as LTβR-independent lymphoid tissue inducer cells during atherosclerosis-related lymphoid neogenesis. Cardiovasc. Res. 101, 434–443 (2014). doi: 10.1093/cvr/cvt263; pmid: 24272771

61. T. D. Randall, Bronchus-associated lymphoid tissue (BALT) structure and function. Adv. Immunol. 107, 187–241 (2010). doi: 10.1016/B978-0-12-381300-8.00007-1; pmid: 21034975

62. J. Rangel-Moreno et al., The development of inducible bronchus-associated lymphoid tissue depends on IL-17. Nat. Immunol. 12, 639–646 (2011). doi: 10.1038/ni.2053; pmid: 21666689

63. K. Guedj et al., Adipocytes orchestrate the formation of tertiary lymphoid organs in the creeping fat of Crohn's disease affected mesentery. J. Autoimmun. 103, 102281 (2019). doi: 10.1016/j.jaut.2019.05.009; pmid: 31171476

64. K. Guedj et al., Inflammatory micro-environmental cues of human atherothrombotic arteries confer to vascular smooth muscle cells the capacity to trigger lymphoid neogenesis. PLOS ONE 9, e116295 (2014). doi: 10.1371/journal.pone.0116295; pmid: 25548922

65. D. S. Thommen et al., A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. Nat. Med. 24, 994–1004 (2018). doi: 10.1038/s41591-018-0057-z; pmid: 29892065

66. M. Yang et al., CXCL13 shapes immunoactive tumor microenvironment and enhances the efficacy of PD-1 checkpoint blockade in high-grade serous ovarian cancer. J. Immunother. Cancer 9, e001136 (2021). doi: 10.1136/jitc-2020-001136; pmid: 33452206

67. S. Z. Wu et al., Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. EMBO J. 39, e104063 (2020). doi: 10.15252/embj.2019104063; pmid: 32790115

68. A. Ager, High endothelial venules and other blood vessels: Critical regulators of lymphoid organ development and function. Front. Immunol. 8, 45 (2017). doi: 10.3389/fimmu.2017.00045; pmid: 28217126

69. G. C. Furtado et al., TNFα-dependent development of lymphoid tissue in the absence of RORγt+ lymphoid tissue inducer cells. Mucosal Immunol. 7, 602–614 (2014). doi: 10.1038/mi.2013.79; pmid: 24129162

70. M. C. Dieu-Nosjean, J. Goc, N. A. Giraldo, C. Sautès-Fridman, W. H. Fridman, Tertiary lymphoid structures in cancer and beyond. Trends Immunol. 35, 571–580 (2014). doi: 10.1016/j.it.2014.09.006; pmid: 25443495

71. C. Gu-Trantien et al., CXCL13-producing TFH cells link immune suppression and adaptive memory in human breast cancer. JCI Insight 2, e91487 (2017). doi: 10.1172/jci.insight.91487; pmid: 28570278

72. K. Siliņa et al., Germinal centers determine the prognostic relevance of tertiary lymphoid structures and are impaired by corticosteroids in lung squamous cell carcinoma. Cancer Res. 78, 1308–1320 (2018). doi: 10.1158/0008-5472.CAN-17-1987; pmid: 29279354

73. J. Calderaro et al., Intra-tumoral tertiary lymphoid structures are associated with a low risk of early recurrence of hepatocellular carcinoma. J. Hepatol. 70, 58–65 (2019). doi: 10.1016/j.jhep.2018.09.003; pmid: 30213589

74. F. Posch et al., Maturation of tertiary lymphoid structures and recurrence of stage II and III colorectal cancer. Oncoimmunology 7, e1378844 (2017). doi: 10.1080/2162402X.2017.1378844; pmid: 29416939

75. J. Alsughayyir, G. J. Pettigrew, R. Motallebzadeh, Spoiling for a fight: B lymphocytes as initiator and effector populations within tertiary lymphoid organs in autoimmunity and transplantation. Front. Immunol. 8, 1639 (2017). doi: 10.3389/fimmu.2017.01639; pmid: 29218052

76. A. Cipponi et al., Neogenesis of lymphoid structures and antibody responses occur in human melanoma metastases. Cancer Res. 72, 3997–4007 (2012). doi: 10.1158/0008-5472.CAN-12-1377; pmid: 22850419

77. J. S. Nielsen, B. H. Nelson, Tumor-infiltrating B cells and T cells: Working together to promote patient survival. Oncoimmunology 1, 1623–1625 (2012). doi: 10.4161/onci.21650; pmid: 23264915

78. A. Montfort et al., A strong B-cell response is part of the immune landscape in human high-grade serous ovarian metastases. Clin. Cancer Res. 23, 250–262 (2017). doi: 10.1158/1078-0432.CCR-16-0081; pmid: 27354470

79. S. Nzula, J. J. Going, D. I. Stott, Antigen-driven clonal proliferation, somatic hypermutation, and selection of B lymphocytes infiltrating human ductal breast carcinomas. Cancer Res. 63, 3275–3280 (2003). pmid: 12810659

80. J. A. Coronella et al., Antigen-driven oligoclonal expansion of tumor-infiltrating B cells in infiltrating ductal carcinoma of the breast. J. Immunol. 169, 1829–1836 (2002). doi: 10.4049/jimmunol.169.4.1829; pmid: 12165506

81. H. A. Schlößer et al., B cells in esophago-gastric adenocarcinoma are highly differentiated, organize in tertiary lymphoid structures and produce tumor-specific antibodies. Oncoimmunology 8, e1512458 (2018). doi: 10.1080/2162402X.2018.1512458; pmid: 30546950

82. W. Zhu et al., A high density of tertiary lymphoid structure B cells in lung tumors is associated with increased CD4+ T cell receptor repertoire clonality. Oncoimmunology 4, e1051922 (2015). doi: 10.1080/2162402X.2015.1051922; pmid: 26587322

83. H. H. Workel et al., A transcriptionally distinct CXCL13+CD103+CD8+ T-cell population is associated with B-cell recruitment and neoantigen load in human cancer. Cancer Immunol. Res. 7, 784–796 (2019). doi: 10.1158/2326-6066.CIR-18-0517; pmid: 30872264

84. E. J. Colbeck et al., Treg depletion licenses T cell-driven HEV neogenesis and promotes tumor destruction. Cancer Immunol. Res. 5, 1005–1015 (2017). doi: 10.1158/2326-6066.CIR-17-0131; pmid: 28947544

85. S. Stranford, N. H. Ruddle, Follicular dendritic cells, conduits, lymphatic vessels, and high endothelial venules in tertiary lymphoid organs: Parallels with lymph node stroma. Front. Immunol. 3, 350 (2012). doi: 10.3389/fimmu.2012.00350; pmid: 23230435

86. C. H. GeurtsvanKessel et al., Dendritic cells are crucial for maintenance of tertiary lymphoid structures in the lung of influenza virus-infected mice. J. Exp. Med. 206, 2339–2349 (2009). doi: 10.1084/jem.20090410; pmid: 19808255

87. S. Halle et al., Induced bronchus-associated lymphoid tissue serves as a general priming site for T cells and is maintained by dendritic cells. J. Exp. Med. 206, 2593–2601 (2009). doi: 10.1084/jem.20091472; pmid: 19917776

88. R. M. Genta, H. W. Hamner, D. Y. Graham, Gastric lymphoid follicles in Helicobacter pylori infection: Frequency, distribution, and response to triple therapy. Hum. Pathol. 24, 577–583 (1993). doi: 10.1016/0046-8177(93)90235-9; pmid: 8505056

89. J. Goc, W. H. Fridman, C. Sautès-Fridman, M. C. Dieu-Nosjean, Characteristics of tertiary lymphoid structures in primary cancers. Oncoimmunology 2, e26836 (2013). doi: 10.4161/onci.26836; pmid: 24498556

90. P. Jézéquel et al., Identification of three subtypes of triple-negative breast cancer with potential therapeutic implications. Breast Cancer Res. 21, 65 (2019). doi: 10.1186/s13058-019-1148-6; pmid: 31101122

91. C. Pfannstiel et al., The tumor immune microenvironment drives a prognostic relevance that correlates with bladder cancer subtypes. Cancer Immunol. Res. 7, 923–938 (2019). doi: 10.1158/2326-6066.CIR-18-0758; pmid: 30988029

92. J. P. Hindley et al., T-cell trafficking facilitated by high endothelial venules is required for tumor control after regulatory T-cell depletion. Cancer Res. 72, 5473–5482 (2012). doi: 10.1158/0008-5472.CAN-12-1912; pmid: 22962270

93. P. De Silva et al., FOXP1 negatively regulates tumor infiltrating lymphocyte migration in human breast cancer. EBioMedicine 39, 226–238 (2019). doi: 10.1016/j.ebiom.2018.11.066; pmid: 30579865

94. T. Junt, E. Scandella, B. Ludewig, Form follows function: Lymphoid tissue microarchitecture in antimicrobial immune defence. Nat. Rev. Immunol. 8, 764–775 (2008). doi: 10.1038/nri2414; pmid: 18825130

95. T. C. Bruno, New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. Nature 577, 474–476 (2020). doi: 10.1038/d41586-019-03943-0; pmid: 31965091

96. S. Salomonsson et al., Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjögren's syndrome. Arthritis Rheum. 48, 3187–3201 (2003). doi: 10.1002/art.11311; pmid: 14613282

97. G. P. Sims, H. Shiono, N. Willcox, D. I. Stott, Somatic hypermutation and selection of B cells in thymic germinal centers responding to acetylcholine receptor in myasthenia gravis. J. Immunol. 167, 1935–1944 (2001). doi: 10.4049/jimmunol.167.4.1935; pmid: 11489973

98. M. P. Armengol et al., Thyroid autoimmune disease: Demonstration of thyroid antigen-specific B cells and recombination-activating gene expression in chemokine-containing active intrathyroidal germinal centers. Am. J. Pathol. 159, 861–873 (2001). doi: 10.1016/S0002-9440(10)61762-2; pmid: 11549579

99. C. Germain et al., Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. Am. J. Respir. Crit. Care Med. 189, 832–844 (2014). doi: 10.1164/rccm.201309-1611OC; pmid: 24484282

100. Y. Lee et al., Recruitment and activation of naive T cells in the islets by lymphotoxin beta receptor-dependent tertiary lymphoid structure. Immunity 25, 499–509 (2006). doi: 10.1016/j.immuni.2006.06.016; pmid: 16934497

101. E. J. McMahon, S. L. Bailey, C. V. Castenada, H. Waldner, S. D. Miller, Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. Nat. Med. 11, 335–339 (2005). doi: 10.1038/nm1202; pmid: 15735651

102. J. E. Moyron-Quiroz et al., Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity. Nat. Med. 10, 927–934 (2004). doi: 10.1038/nm1091; pmid: 15311275

103. J. Rangel-Moreno, J. E. Moyron-Quiroz, L. Hartson, K. Kusser, T. D. Randall, Pulmonary expression of CXC chemokine ligand 13, CC chemokine ligand 19, and CC chemokine ligand 21 is essential for local immunity to influenza. Proc. Natl. Acad. Sci. U.S.A. 104, 10577–10582 (2007). doi: 10.1073/pnas.0700591104; pmid: 17563386

104. J. E. Moyron-Quiroz et al., Persistence and responsiveness of immunologic memory in the absence of secondary lymphoid organs. Immunity 25, 643–654 (2006). doi: 10.1016/j.immuni.2006.08.022; pmid: 17045819

105. D. Schrama et al., Targeting of lymphotoxin-α to the tumor elicits an efficient immune response associated with induction of peripheral lymphoid-like tissue. Immunity 14, 111–121 (2001). doi: 10.1016/S1074-7613(01)00094-2; pmid: 11239444

106. D. Schrama et al., Immunological tumor destruction in a murine melanoma model by targeted LTα independent of secondary lymphoid tissue. Cancer Immunol. Immunother. 57, 85–95 (2008). doi: 10.1007/s00262-007-0352-x; pmid: 17605009

107. G. Oliveira et al., Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. Nature 596, 119–125 (2021). doi: 10.1038/s41586-021-03704-y; pmid: 34290406

108. J. X. Caushi et al., Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Nature 596, 126–132 (2021). doi: 10.1038/s41586-021-03752-4; pmid: 34290408

109. C. D. Buckley, F. Barone, S. Nayar, C. Bénézech, J. Caamaño, Stromal cells in chronic inflammation and tertiary lymphoid organ formation. Annu. Rev. Immunol. 33, 715–745 (2015). doi: 10.1146/annurev-immunol-032713-120252; pmid: 25861980

110. J. Goc, W. H. Fridman, S. A. Hammond, C. Sautès-Fridman, M. C. Dieu-Nosjean, Tertiary lymphoid structures in human lung cancers, a new driver of antitumor immune responses. Oncoimmunology 3, e28976 (2014). doi: 10.4161/onci.28976; pmid: 25083325

111. N. Hiraoka et al., Intratumoral tertiary lymphoid organ is a favourable prognosticator in patients with pancreatic cancer. Br. J. Cancer 112, 1782–1790 (2015). doi: 10.1038/bjc.2015.145; pmid: 25942397

112. K. Siliņa, U. Rulle, Z. Kalniņa, A. Linē, Manipulation of tumour-infiltrating B cells and tertiary lymphoid structures: A novel anti-cancer treatment avenue? Cancer Immunol. Immunother. 63, 643–662 (2014). doi: 10.1007/s00262-014-1544-9; pmid: 24695950

113. W. H. Fridman, L. Zitvogel, C. Sautès-Fridman, G. Kroemer, The immune contexture in cancer prognosis and treatment. Nat. Rev. Clin. Oncol. 14, 717–734 (2017). doi: 10.1038/nrclinonc.2017.101; pmid: 28741618

114. A. R. Cillo et al., Immune landscape of viral- and carcinogen-driven head and neck cancer. Immunity 52, 183–199.e9 (2020). doi: 10.1016/j.immuni.2019.11.014; pmid: 31924475

115. J. L. Messina et al., 12-Chemokine gene signature identifies lymph node-like structures in melanoma: Potential for patient selection for immunotherapy? Sci. Rep. 2, 765 (2012). doi: 10.1038/srep00765; pmid: 23097687

116. S. Prabhakaran et al., Evaluation of invasive breast cancer samples using a 12-chemokine gene expression score: Correlation with clinical outcomes. Breast Cancer Res. 19, 71 (2017). doi: 10.1186/s13058-017-0864-z; pmid: 28629479

117. A. Kim et al., The prognostic significance of tumor-infiltrating lymphocytes assessment with hematoxylin and eosin sections in resected primary lung adenocarcinoma. PLOS ONE 14, e0224430 (2019). doi: 10.1371/journal.pone.0224430; pmid: 31743333

118. C. Solinas et al., BRCA gene mutations do not shape the extent and organization of tumor infiltrating lymphocytes in triple negative breast cancer. Cancer Lett. 450, 88–97 (2019). doi: 10.1016/j.canlet.2019.02.027; pmid: 30797818

119. N. A. Giraldo et al., Orchestration and prognostic significance of immune checkpoints in the microenvironment of primary and metastatic renal cell cancer. Clin. Cancer Res. 21, 3031–3040 (2015). doi: 10.1158/1078-0432.CCR-14-2926; pmid: 25688160

120. S. N. Willis et al., The microenvironment of germ cell tumors harbors a prominent antigen-driven humoral response. J. Immunol. 182, 3310–3317 (2009). doi: 10.4049/jimmunol.0803424; pmid: 19234230

121. R. Remark et al., Characteristics and clinical impacts of the immune environments in colorectal and renal cell carcinoma lung metastases: Influence of tumor origin. Clin. Cancer Res. 19, 4079–4091 (2013). doi: 10.1158/1078-0432.CCR-12-3847; pmid: 23785047

122. L. Munoz-Erazo, J. L. Rhodes, V. C. Marion, R. A. Kemp, Tertiary lymphoid structures in cancer – considerations for patient prognosis. Cell. Mol. Immunol. 17, 570–575 (2020). doi: 10.1038/s41423-020-0457-0; pmid: 32415259

123. S. L. Figenschau, S. Fismen, K. A. Fenton, C. Fenton, E. S. Mortensen, Tertiary lymphoid structures are associated with higher tumor grade in primary operable breast cancer patients. BMC Cancer 15, 101 (2015). doi: 10.1186/s12885-015-1116-1; pmid: 25884667

124. N. S. Joshi et al., Regulatory T cells in tumor-associated tertiary lymphoid structures suppress anti-tumor T cell responses. Immunity 43, 579–590 (2015). doi: 10.1016/j.immuni.2015.08.006; pmid: 26341400

125. S. Finkin et al., Ectopic lymphoid structures function as microniches for tumor progenitor cells in hepatocellular carcinoma. Nat. Immunol. 16, 1235–1244 (2015). doi: 10.1038/ni.3290; pmid: 26502405

126. G. V. Sharonov, E. O. Serebrovskaya, D. V. Yuzhakova, O. V. Britanova, D. M. Chudakov, B cells, plasma cells and antibody repertoires in the tumour microenvironment. Nat. Rev. Immunol. 20, 294–307 (2020). doi: 10.1038/s41577-019-0257-x; pmid: 31988391

127. S. Shalapour et al., Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. Nature 521, 94–98 (2015). doi: 10.1038/nature14395; pmid: 25924065

128. X. Liu et al., Distinct tertiary lymphoid structure associations and their prognostic relevance in HER2 positive and negative breast cancers. Oncologist 22, 1316–1324 (2017). doi: 10.1634/theoncologist.2017-0029; pmid: 28701569

129. J. Gao et al., Neoadjuvant PD-L1 plus CTLA-4 blockade in patients with cisplatin-ineligible operable high-risk urothelial carcinoma. Nat. Med. 26, 1845–1851 (2020). doi: 10.1038/s41591-020-1086-y; pmid: 33046869

130. J. Griss et al., B cells sustain inflammation and predict response to immune checkpoint blockade in human melanoma. Nat. Commun. 10, 4186 (2019). doi: 10.1038/s41467-019-12160-2; pmid: 31519915

131. L. Buisseret et al., Tumor-infiltrating lymphocyte composition, organization and PD-1/PD-L1 expression are linked in breast cancer. OncoImmunology 6, e1257452 (2016). doi: 10.1080/2162402X.2016.1257452; pmid: 28197375

132. A. Cimino-Mathews et al., PD-L1 (B7-H1) expression and the immune tumor microenvironment in primary and metastatic breast carcinomas. Hum. Pathol. 47, 52–63 (2016). doi: 10.1016/j.humpath.2015.09.003; pmid: 26527522

133. A. Johansson-Percival et al., De novo induction of intratumoral lymphoid structures and vessel normalization enhances immunotherapy in resistant tumors. Nat. Immunol. 18, 1207–1217 (2017). doi: 10.1038/ni.3836; pmid: 28892469

134. T. R. Cottrell et al., Pathologic features of response to neoadjuvant anti-PD-1 in resected non-small-cell lung carcinoma: A proposal for quantitative immune-related pathologic response criteria (irPRC). Ann. Oncol. 29, 1853–1860 (2018). doi: 10.1093/annonc/mdy218; pmid: 29982279

135. N. van Dijk et al., Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: The NABUCCO trial. Nat. Med. 26, 1839–1844 (2020). doi: 10.1038/s41591-020-1085-z; pmid: 33046870

136. D. P. Hollern et al., B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. Cell 179, 1191–1206. e21 (2019). doi: 10.1016/j.cell.2019.10.028; pmid: 31730857

137. S. Sánchez-Alonso et al., A new role for circulating T follicular helper cells in humoral response to anti-PD-1 therapy. J. Immunother. Cancer 8, e001187 (2020). doi: 10.1136/jitc-2020-001187; pmid: 32900863

138. P. Voabil et al., An ex vivo tumor fragment platform to dissect response to PD-1 blockade in cancer. Nat. Med. 27, 1250–1261 (2021). doi: 10.1038/s41591-021-01398-3; pmid: 34239134

139. R. H. Mounzer et al., Lymphotoxin-alpha contributes to lymphangiogenesis. Blood 116, 2173–2182 (2010). doi: 10.1182/blood-2009-12-256065; pmid: 20566898

140. Q. Zhang et al., Increased lymphangiogenesis in joints of mice with inflammatory arthritis. Arthritis Res. Ther. 9, R118 (2007). doi: 10.1186/ar2326; pmid: 17997858

141. S. Rehal, P. Y. von der Weid, TNFΔARE mice display abnormal lymphatics and develop tertiary lymphoid organs in the mesentery. Am. J. Pathol. 187, 798–807 (2017). doi: 10.1016/j.ajpath.2016.12.007; pmid: 28183530

142. E. Allen et al., Combined antiangiogenic and anti-PD-L1 therapy stimulates tumor immunity through HEV formation. Sci. Transl. Med. 9, eaak9679 (2017). doi: 10.1126/scitranslmed.aak9679; pmid: 28404866

143. L. Maldonado et al., Intramuscular therapeutic vaccination targeting HPV16 induces T cell responses that localize in mucosal lesions. Sci. Transl. Med. 6, 221ra13 (2014). doi: 10.1126/scitranslmed.3007323; pmid: 24477000

144. E. R. Lutz et al., Immunotherapy converts nonimmunogenic pancreatic tumors into immunogenic foci of immune regulation. Cancer Immunol. Res. 2, 616–631 (2014). doi: 10.1158/2326-6066.CIR-14-0027; pmid: 24942756

145. R. Remark et al., Immune contexture and histological response after neoadjuvant chemotherapy predict clinical outcome of lung cancer patients. OncoImmunology 5, e1255394 (2016). doi: 10.1080/2162402X.2016.1255394; pmid: 28123901

146. G. Morcrette et al., APC germline hepatoblastomas demonstrate cisplatin-induced intratumor tertiary lymphoid structures. OncoImmunology 8, e1583547 (2019). doi: 10.1080/2162402X.2019.1583547; pmid: 31069152

147. L. H. Calabrese, C. Calabrese, L. C. Cappelli, Rheumatic immune-related adverse events from cancer immunotherapy. Nat. Rev. Rheumatol. 14, 569–579 (2018). doi: 10.1038/s41584-018-0074-9; pmid: 30171203

148. S. Matsubara, M. Seki, S. Suzuki, T. Komori, M. Takamori, Tertiary lymphoid organs in the inflammatory myopathy associated with PD-1 inhibitors. J. Immunother. Cancer 7, 256 (2019). doi: 10.1186/s40425-019-0736-4; pmid: 31533865

ORGANOIDS

Tissue geometry drives deterministic organoid patterning

N. Gjorevski†, M. Nikolaev†, T. E. Brown†, O. Mitrofanova, N. Brandenberg, F. W. DelRio, F. M. Yavitt, P. Liberali, K. S. Anseth, M. P. Lutolf*

INTRODUCTION: Stem cell–derived organoids are in vitro tissue and organ mimetics that have enormous potential as models for human organ development and disease, as platforms for drug discovery and diagnostics, and for the design of cell and gene therapies. However, the stem cell self-organization processes underlying organoid development are poorly controlled, leading to a general lack of reproducibility of most existing organoid cultures. For example, the location and number of crypt-like domains in mouse intestinal organoids—perhaps the best-described organoid system to date—cannot be controlled, nor can the shape, size, and multicellular composition of organoids. The high variability of current organoid models poses a major challenge for basic and translational organoid-based research.

RATIONALE: Applying control over organoid formation and the resulting structures would allow both understanding the underlying morphogenetic mechanisms and building models that bear higher likeness to the native counterparts. The final functional architectures of real organs are the product of interplay between epithelial self-organizing programs and extrinsic microenvironmental controllers. Taking inspiration from development in vivo, we complemented organoid self-organization with external regulation. In particular, we have attempted to control the patterning and morphogenesis of intestinal organoids via the physical properties and, in particular, the initial geometry of the tissue itself.

RESULTS: We developed bioengineering strategies to extrinsically control the self-organization process of intestinal stem cells through in situ photopatterning of hydrogel mechanics and hydrogel microfabrication. We found that localized patterning of microenvironmental mechanics and predefined hydrogel microtopography could be used to build organoids with a controlled initial size and shape, and we were able to predict and influence the course of their development, especially the number and location of crypt domains. We used the predictability of organoid development to identify the underlying mechanism of epithelial patterning. Our data suggest that in vivo–like tissue geometries can drive stereotypical epithelial patterning by establishing reproducible local differences in cell packing and morphology. These differences in cell shape lead to spatial heterogeneities in YAP mechanosensing/transduction and Notch signaling, which in turn specify "crypt"- and "villus"-like domains by localizing Paneth cell differentiation and suppressing stem cell fates, respectively. Spatial variations in cell morphology dictated by tissue geometry thus render a normally random process highly stereotypical. We exploited these insights to build macroscopic organoids resembling the periodic crypt-villus architecture of the native intestinal epithelium. These structures adopt physiologically accurate patterning and regionalization conferred by tissue geometry only, in the absence of extrinsically imposed biochemical gradients. The predictable and well-delineated villus regions and the accessibility to both the basal and luminal surface are key advantages that enable the study of pathophysiological processes such as intestinal cell shedding.

CONCLUSION: We present an approach to guiding stem cell–based organogenesis, a process otherwise driven entirely by "stochastic" self-organization. We also verify long-standing but underexplored paradigms of morphogenesis whereby the present shape of a tissue can help to pattern and specify the course of development, and therefore the future shape, of the tissue. In the case of intestinal crypt formation, we conclude that budding not only can follow Paneth cell appearance but can also precede it. Our organoid cultures can be used to answer questions not readily addressable by existing organoid and animal models, and they may enable the translation of organoid technology to real-world applications. ∎
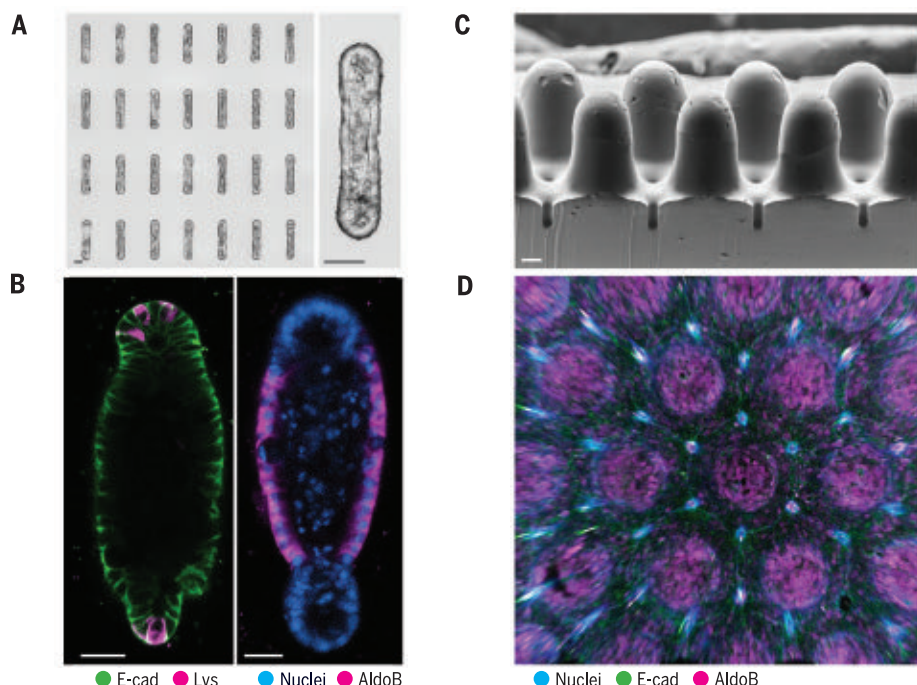
Extrinsic control of organoid development through engineered microenvironments. (A) Microengineering-based approaches for controlling the size and shape of intestinal organoids with micrometer-scale precision. **(B)** Organoids of controlled geometry get patterned in a predictable and reproducible manner. **(C** and **D)** Geometry-mediated organoid patterning can be used to produce macroscopic intestinal surfaces with crypt-villus architectures. Organoids are stained for E-cadherin (E-cad), lysozyme (Lys), and aldolase B (AldoB).

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aaw9021

ORGANOIDS

Tissue geometry drives deterministic organoid patterning

N. Gjorevski[1]†‡, M. Nikolaev[1]†‡, T. E. Brown[2,3]†, O. Mitrofanova[1], N. Brandenberg[1], F. W. DelRio[4], F. M. Yavitt[2,3], P. Liberali[5], K. S. Anseth[2,3], M. P. Lutolf[1,6]*‡

Epithelial organoids are stem cell–derived tissues that approximate aspects of real organs, and thus they have potential as powerful tools in basic and translational research. By definition, they self-organize, but the structures formed are often heterogeneous and irreproducible, which limits their use in the lab and clinic. We describe methodologies for spatially and temporally controlling organoid formation, thereby rendering a stochastic process more deterministic. Bioengineered stem cell microenvironments are used to specify the initial geometry of intestinal organoids, which in turn controls their patterning and crypt formation. We leveraged the reproducibility and predictability of the culture to identify the underlying mechanisms of epithelial patterning, which may contribute to reinforcing intestinal regionalization in vivo. By controlling organoid culture, we demonstrate how these structures can be used to answer questions not readily addressable with the standard, more variable, organoid models.

S tem cell–derived organoids are in vitro tissue and organ mimetics that hold promise as models of human development and disease, platforms for drug discovery, and material to repair diseased and damaged tissue, including personalized therapies (1–7). Although organoids display complex architecture and function, the structures generated are often variable, with methods lacking reproducibility. For instance, intestinal organoid formation is largely stochastic, resulting in structures that differ from the native organ in multiple aspects, such as the location and number of crypt-like domains and variation in the shape, size, and cellular composition of the overall organoid. This high variability poses a challenge in basic and translational organoid-based research (8, 9). Although the field of stem cell–based organoids was conceived at least a decade ago and the range of organoid types is continuously expanding, it is only very recently that researchers have introduced methods to control organoid formation (10–15). Our research group had previously used microfabrication and microfluidics to control the

macroscopic shape of intestinal organoids, which ultimately enabled the establishment of long-lived and perfused structures (13). This approach used fabrication to form the crypt-villus system, rather than relying on intrinsic morphogenetic programs, such as evagination and budding, which drive both organoid formation and intestinal development in vivo (16). However, the mechanisms whereby the macroscopic organoid shape can ultimately pattern the crypt-villus system were not elucidated. In this study, we set out to devise strategies for exerting extrinsic control over intestinal organoid symmetry breaking and crypt formation, and used the resulting models to shed light on the mechanisms by which tissue geometry can regulate intestinal morphogenesis.

Spatiotemporal control over organoid crypt formation by photopatterning

Previously, we showed that the transformation of a round intestinal stem cell (ISC) colony into a crypt-containing organoid within a synthetic hydrogel requires matrix softening (17). The global matrix-softening approach we used, however, resulted in stochastic and spatially uncontrolled budding, just as in the conventional organoid cultures that are based on native extracellular matrix (ECM)–derived three-dimensional (3D) matrices (18, 19). We postulated that by introducing localized matrix softening, thus restricting the regions permissive to budding, we might achieve spatially controlled crypt formation.

To this end, we embedded ISC colonies within arginine-glycine-aspartate (RGD)– and laminin-1–containing photosensitive poly(ethylene glycol) (PEG)–based hydrogels (20), which undergo degradation and softening when exposed to 405-nm light (fig.

S1). Localized light exposure allowed us to introduce softening at predefined regions within the hydrogel surrounding the colony (Fig. 1A). The initial stiffness of the hydrogel matched the value that we previously found to support ISC colony formation but not budding (17) (fig. S1), whereas the light dose supplied to specified regions was chosen to effect a drop of stiffness previously identified as crucial for organoid budding (17) (Fig. 1B). Shortly (<10 min) after photopatterning, the epithelium adjacent to the softened regions underwent an evagination-like event (movie S1), which we believe to be an attempt to establish mechanical balance. These pseudo-buds continued to extend over the next 72 hours, forming structures that morphologically resembled intestinal crypts (Fig. 1, C to E). Whereas crypt-like buds were frequent within softened regions, they were completely absent outside of these regions. Thus, we were able to control and predict the sites of bud formation with high ($84 \pm 6\%$) fidelity (Fig. 1F).

To ensure that the buds were bona fide crypts formed through epithelial symmetry breaking and patterning, rather than merely by differential growth, we considered the distribution of ISCs and differentiated intestinal cells throughout the structure. Lgr5-expressing ISCs were present exclusively at the end of the buds and were absent in the central epithelial cyst (Fig. 1, G and H). Cell division, indicated by incorporation of 5-ethynyl-2′-deoxyuridine (EdU), was similarly localized to crypt structures (Fig. 1I). Likewise, ISC-supporting Paneth cells were present within the buds, whereas enterocytes were confined to the central region of the cyst (Fig. 1, J and K). Enteroendocrine cells were also found within the structures (Fig. 1L). Thus, we used light-mediated matrix softening to control organoid symmetry breaking and direct crypt formation. Of note, the timing of matrix softening is important, because crypt formation was substantially reduced when softening was performed 2 days after the induction of differentiation (fig. S2).

In applying photo-mediated softening to control intestinal organoid formation, we noticed that the symmetry breaking and epithelial patterning were preceded by a change in epithelial shape (movie S1). Specifically, after the rapid evagination-like event that produced nascent buds, the colonies were uniformly composed of ISCs. The budded structure was patterned and transformed into an organoid in the subsequent 24 hours. Bearing in mind that the appearance of the bud preceded the molecular symmetry breaking, we postulated that the bud shape of the crypt epithelium itself represents an integral part of the ISC niche, helping to restrict the ISC zone and establish the crypt-villus axis. Indeed, our recent work demonstrated that ISCs grown in crypt-shaped cavities within microfluidic scaffolds can be

[1]Laboratory of Stem Cell Bioengineering, Institute of Bioengineering, School of Life Sciences (SV) and School of Engineering (STI), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [2]Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80309, USA. [3]BioFrontiers Institute, University of Colorado, Boulder, CO 80303, USA. [4]Material, Physical, and Chemical Sciences Center, Sandia National Laboratories, Albuquerque, NM 87185, USA. [5]Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland. [6]Institute of Chemical Sciences and Engineering, School of Basic Science (SB), EPFL, Lausanne, Switzerland.
*Corresponding author. Email: matthias.lutolf@epfl.ch
†These authors contributed equally to this work.
‡Present address: Institute for Translational Bioengineering (ITB), Roche Pharma Research and Early Development, Basel, Switzerland.
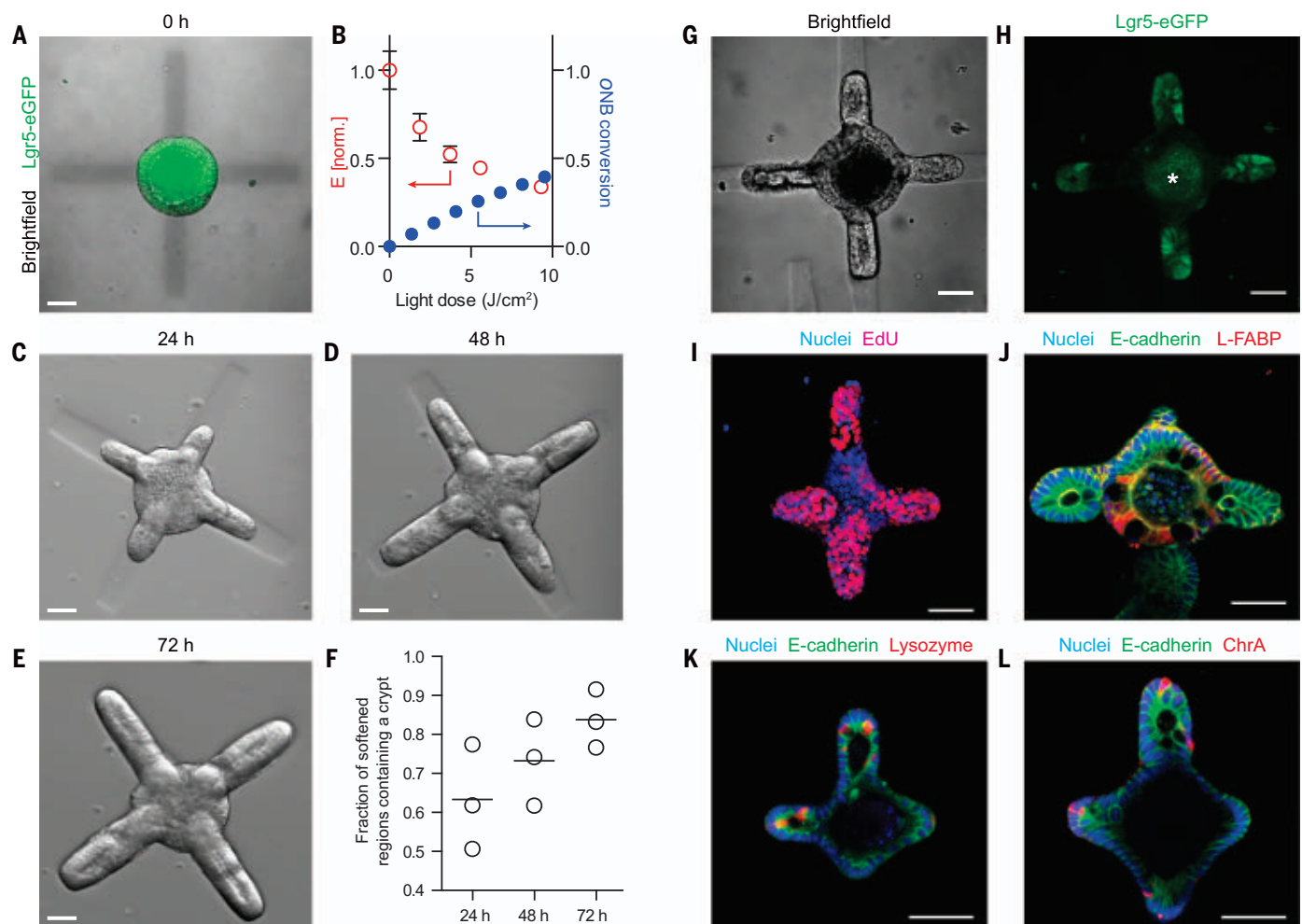
Fig. 1. Spatiotemporal control over organoid crypt formation through photopatterning. (**A**) Composite image showing Lgr5-GFP expression in a symmetric colony and photopattern visible with transmitted 405-nm light immediately after spatially restricted light exposure. (**B**) Mechanical characterization of hydrogels with atomic force microscopy reveals that the reduction in the Young's modulus (E) corresponds to conversion of photo-cleavable *ortho*-nitrobenzyl (*o*NB) moieties within the gel. Error bars denote SD. (**C** to **E**) Spatially defined crypt formation within photopatterned gels 24 hours (C), 48 hours (D), and 72 hours (E) after light-induced softening. (**F**) Quantification of fraction of photo-softened gel regions containing a crypt. Individual data points and mean are shown. (**G** to **I**) Lgr5-eGFP expression [(G), (H)] and proliferation (I) are localized within the buds, extending into the softened regions. (**J**) Enterocytes (L-FABP stain) are found in the central regions of the organoids. (**K**) Paneth cells (lysozyme stain) are restricted to the buds of the organoids. (**L**) Enteroendocrine cells (ChrA stain) are also present without clear morphological localization, consistent with their distribution in the real intestine. Scale bars, 30 μm.

maintained in long-term cultures (*13*), which suggests that tissue shape directly influences patterning. To test this hypothesis, we set out to build intestinal tissues of predefined size and geometry that mimic those of the crypt, and to monitor how the initial shape affects the spatial distributions of ISCs and the various differentiated cell types within the tissue (Fig. 2).

Microfabricated intestinal organoids of controlled geometry

We used soft lithography to microstructure a 3D hydrogel [composed of type I collagen (3 mg/ml) and 25% (v/v) Matrigel] with cavities of defined size and shape (*21, 22*), which were subsequently filled with purified Lgr5-eGFP⁺ ISCs (Fig. 2A). Initially randomly dispersed, the stem cells began to form contacts

with each other and the surrounding matrix, and within 48 hours they self-organized into a lumenized epithelial tissue conforming to the shape of the preexisting cavity. We used this approach to form 3D intestinal tissues of arbitrary sizes and shapes (fig. S3, A to C). Next, we sought to determine whether the differentiation of the engineered intestinal tissues followed a stereotypical pattern or occurred randomly. The method described above generates hundreds of regularly spaced tissues of identical size and shape, which permits rapid imaging of fluorescent markers or proteins visualized by immunofluorescence analysis, as well as quantification by automated image segmentation and analysis (fig. S3D). Stacking images of a high number (>80) of individual tissues in registration can provide information

about the average distribution of the molecular marker of interest, with high statistical confidence (fig. S3D). Although the tissues were formed from a cell suspension uniformly expressing Lgr5-eGFP, we found that within 4 days of culture, the signal became restricted to the curved ends of the tissues (Fig. 2, B and C), indicating that ISCs are confined to these regions in a pattern similar to that of the native crypt. The microfabricated tissues proceeded to extend crypt-like buds with a spatial bias reflecting that of Lgr5 expression: The curved ends of the tissues were significantly more likely to extend buds than the flat sides (Fig. 2, D and E). The spatial patterning also extended to differentiated intestinal cell types. Immunofluorescence analysis for Paneth cells and enterocytes revealed that the former
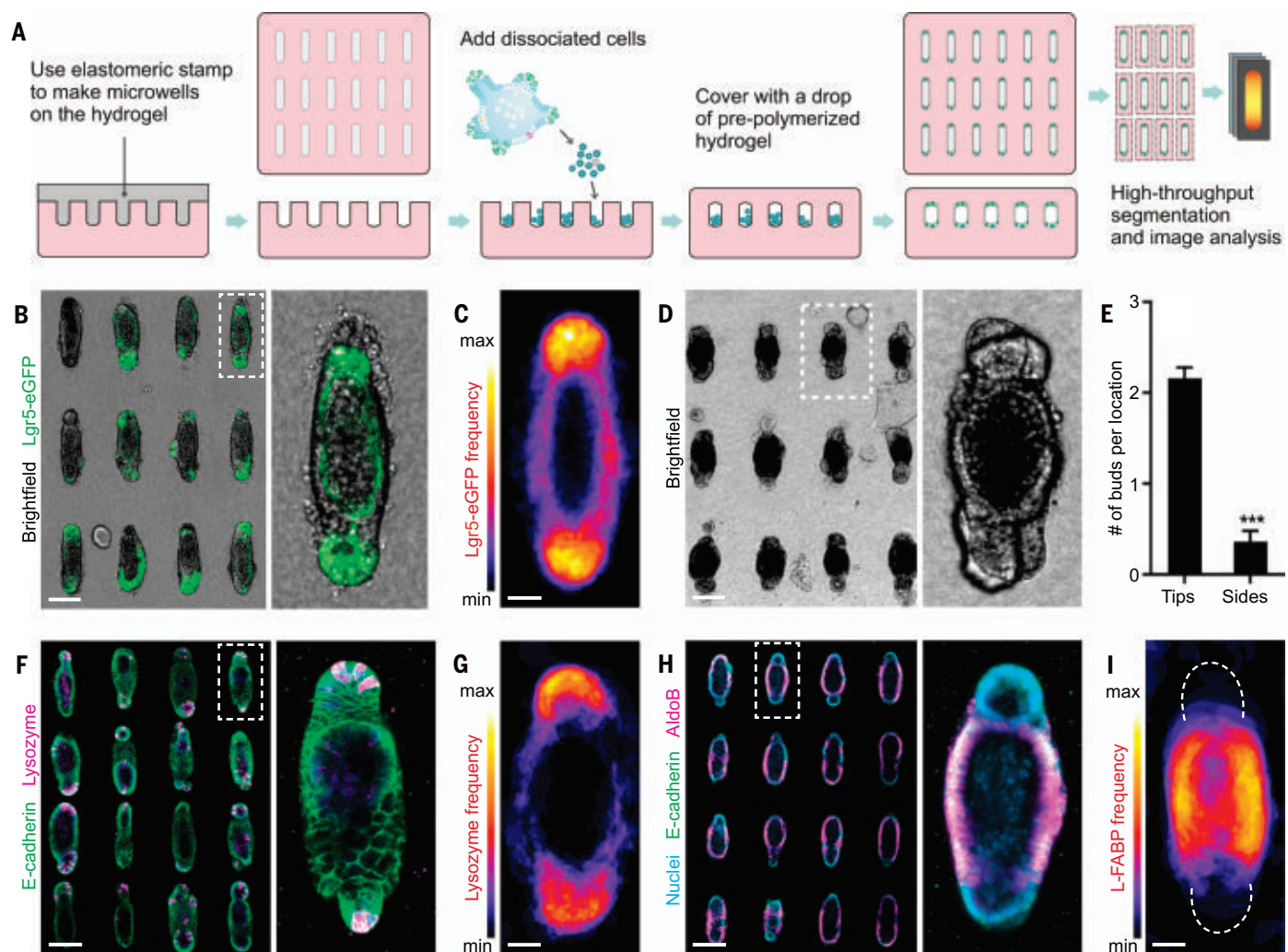
Fig. 2. Geometrically controlled symmetry breaking and epithelial patterning within intestinal organoids. (**A**) Schematic depicting the generation of microfabricated tissues of controlled size and shape. (**B**) An array of intestinal organoids formed from engineered intestinal tissues of rodlike geometry and magnification. (**C**) Frequency map showing average Lgr5 expression over ~80 tissues. (**D** and **E**) An array of intestinal organoids at day 5 (D) and quantification of the average number of buds per location (E) within tubular intestinal tissues. ***$P < 0.001$. (**F** and **G**) Paneth cell staining by lysozyme in the array of intestinal organoids (F) and average Paneth cell distribution (G). (**H** and **I**) AldoB-expressing enterocytes within rod-shaped organoids (H) and average enterocyte distribution (I). The dashed lines indicate the average contour of the tissues. Scale bars, 100 μm [(B), (D), (F), (H)], 25 μm [(C), (G), (I)].

were preferentially localized to the same end-locations as the ISCs (Fig. 2, F and G), whereas the latter were on average excluded from the ends and confined to the middle of the tissue (Fig. 2, H and I). These findings suggest that the spatial distribution of ISCs and differentiated cells within the engineered microtissues reflects the patterning of the crypt-villus axis in vivo (23). Thus, we could predict and control the site of budding and crypt formation within the organoids by merely manipulating their initial geometry.

Mechanism of geometry-controlled epithelial patterning

To elucidate how the initial epithelial geometry might dictate the patterning of intestinal tissue, we monitored the process using time-lapse microscopy (Fig. 3, A to D, and movie S2). Lgr5 expression at the time of seeding and shortly thereafter (<2 hours) appeared uniformly low. As the organoids formed in the crypt-like space, Lgr5 was reexpressed strongly at the ends of tissues, remaining low elsewhere (Fig. 3, C and D). Before the Lgr5 regionalization, we observed a morphological difference between the curved end regions and the flat sides. Within 24 hours of culture, and owing likely to crowding due to proliferation within a limited space at the ends, the cells in these regions became more packed, whereas cells in the lateral regions remained spread and flattened (Fig. 3E). Indeed, measurement of the internuclear distance at the different regions revealed a significant difference in cell packing (Fig. 3F).

Yes-associated protein 1 (YAP) is a regulator of ISC fate (24–27), which is strongly influenced by cell shape and mechanics (28, 29). To ascertain whether the differences in cell morphology between the different regions correlated with differences in YAP activity, we analyzed the subcellular distribution of YAP. Shortly after seeding, YAP was uniformly nuclear throughout the tissue, except in some cases where cell crowding was observed early in the curved regions as a result of (stochastic) variations in cell density (fig. S4A). Between 12 and 24 hours after seeding, corresponding to the time when spatial differences in cell shape appear (and preceding the patterning of Lgr5), nuclear YAP localization became restricted to the lateral regions of the tissues. At the ends of the tissues, cytoplasmic translocation, and

Fig. 3. Tissue geometry controls organoid patterning through cell shape–mediated regulation of YAP and Notch signaling.
(**A** and **B**) Bright-field and Lgr5-eGFP time-lapse imaging of the representative organoid development (A) and frequency maps showing average Lgr5 expression over ~80 tissues (B). (**C** and **D**) Relative changes in the Lgr5-eGFP expression in curved ends and flat sides of the organoids over time (C) and Lgr5-eGFP localization along length of the averaged tissue over time (D). (**E**) Immuno-fluorescence images showing the difference in internuclear distance, cell shape, and subcellular distribution of YAP between cells of the end and the side regions, 24 hours after cell loading. (**F**) Quantification of internuclear distance within the end and side regions of the tissues. Individual points, which represent the distance between neighboring nuclei, and means are shown. (**G**) Quantification of the nuclear localization of YAP within cells of the different organoid regions. Individual points and means are shown. (**H**) Immunofluorescence images showing the difference in the subcellular distribution of YAP cells between cells of the end and the side regions and appearance of the first DLL+, 36 hours after cell loading. Arrowheads denote adjacent pairs of YAP-ON/OFF cells; the YAP-ON cell expresses DLL1. (**I**) Quantification of DLL+ cell localization. (**J** and **K**) Immunofluorescence images showing YAP expression and localization of the enterocytes (AldoB), Paneth cells (Lys), and DLL+ cells in the representative organoids. (**L**) Schematic illustration summarizing the proposed mechanism of the geometry-driven organoid patterning. Scale bars, 25 μm. ****$P < 0.0001$ [(F), (G), (I)].
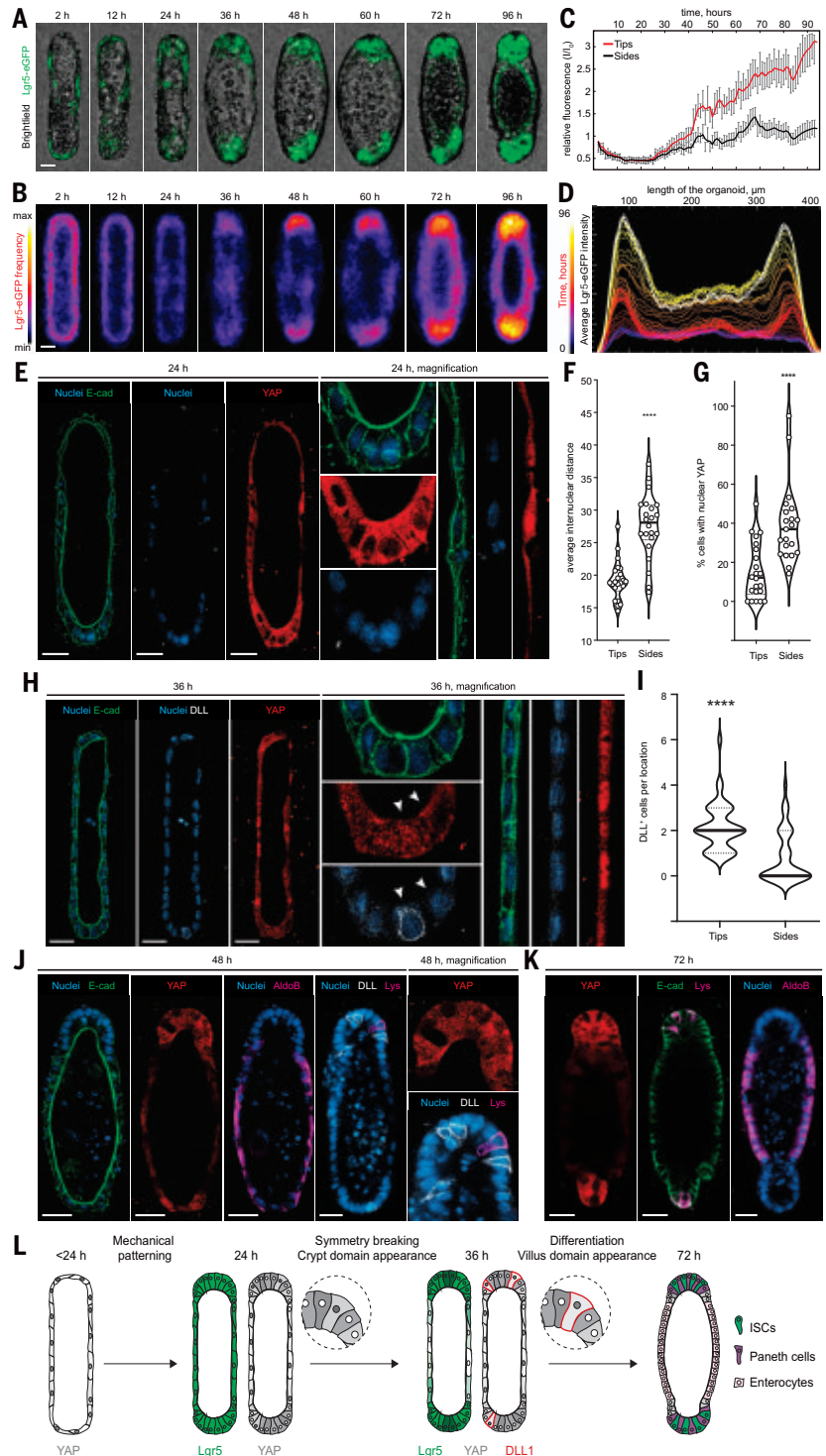
therefore inactivation, was observed (Fig. 3, E and G). This geometry-dependent region-alization of YAP activity was again lost at later time points (>36 hours) when cells throughout the tissue became uniformly packed (fig. S4A). To test whether differential cell spreading (and, conversely, cell crowding) is sufficient to drive differences in YAP activity, we cultured ISCs at equal cell-seeding density in microcavities of

small (50 μm) and large (100 μm) diameter, resulting in a packed or a spread system (fig. S4B). YAP activity strongly correlated with cell spreading: Nuclear translocation was sig-nificantly more frequent within 100-μm wells than in 50-μm wells, which suggests that tissue geometry per se controls the spatially patterned activation of YAP through differential cell spreading (fig. S4, C and D).

Several recent studies have implicated YAP activation in the repression of canonical ISC signatures, including *Lgr5*, *Olfm4*, and *EphB3*, during intestinal regeneration and cancer (*24*, *27*, *30*, *31*). Given the spatial and temporal correlation between YAP induction and Lgr5 suppression, we hypothesized that the pattern-ing of this system is driven by geometrically and mechanically established gradients in

YAP activity. In line with this model, abrogating spatial gradients in YAP activity by uniform induction [via treatment with the MST1/2 signaling inhibitor XMU-MP-1 (*32*)] or inhibition [via treatment with verteporfin (*33*, *34*)] both resulted in loss of intestinal tissue patterning (fig. S4, E to G). Furthermore, blocking cell spreading and the formation of mechanical gradients by treatment with the contractility inhibitor blebbistatin (*29*, *35*) abolished the patterns of both YAP activity and intestinal cell fate (fig. S5). A 36-hour delay of blebbistatin treatment did not interfere with Lgr5 patterning (fig. S5A); during this 36-hour period, treatment was most efficient only within a short time window in which the geometry-controlled cell spreading resulted in differential YAP activation (fig. S5, B and C). Thus, these gain- and loss-of-function experiments show that an early, geometrically induced YAP prepattern is necessary to promote later tissue regionalization. Previous studies by us and others have hinted that spatial heterogeneities in YAP are required for intestinal tissue morphogenesis (*17*, *30*). Here, we have shown that these heterogeneities can be governed by geometrically established gradients in cell mechanics to control downstream tissue patterning.

To explore whether the finding of geometrically and mechanically mediated YAP patterning translates to human intestinal epithelium, we generated arrays of human small intestinal organoids (fig. S6, A and B). The spatial gradients of YAP activity described above were also present in the human system and followed a similar temporal profile (fig. S6, C and D). Note, however, that the human organoids of controlled shape did not proceed to self-organize into a crypt-villus structure (fig. S6E). It is likely that the current culture conditions do not support robust Paneth cell differentiation and ISC niche establishment.

After determining that symmetry breaking in the system likely occurs through YAP-mediated spatial restriction of ISC maintenance, we wanted to further explore how the ISC niche at the ends is established. Multiple studies have shown that Paneth cells provide essential support to ISCs within intestinal organoids, and that their appearance is crucial for the establishment of the ISC niche (*18*, *19*, *30*, *36*). Recent work by Serra *et al.* revealed a YAP-Notch symmetry-breaking mechanism that is responsible for Paneth cell appearance and crypt initiation in classical intestinal organoids (*30*). In particular, it was found that Paneth cell differentiation and crypt initiation occur exactly at the sites where YAP ON cells are adjacent to YAP OFF cells. Bearing in mind the pattern of YAP activity described above, we reasoned that the ends of the tissue are the regions where YAP OFF and YAP ON cell pairs are more likely to coexist, triggering YAP-Notch symmetry breaking. In this model,

an early indication of Paneth cell differentiation is the expression of the Notch ligand DLL1 (delta-like ligand 1) in cells featuring inactive Notch and active YAP. We observed the appearance of this cell type at the end of the tissue around 36 hours (Fig. 3, H and I), along with localized Notch activation, as evidenced by the presence of the Notch intracellular domain (NICD) (fig. S7, A and B). Consistently with this model, Paneth cells were observed at these regions within 48 hours of culture (Fig. 3, J and K). Localized Notch activation and Paneth cell appearance were required for durable tissue patterning: Uniform inhibition of Notch by treatment with the γ-secretase inhibitor DAPT abolished patterns in Lgr5 and Paneth cell differentiation (fig. S7, C to E), as did treatment with valproic acid (fig. S7F), which, although a known pleiotropic histone deacetylase inhibitor, has been shown to activate Notch in this system (*37*). Taken together, these data show that the tissue shape leads to a mechanically controlled prepatterning of epithelial cells with different YAP activity, which in turn results in (i) suppression of crypt cell fates at the sides of the tissues and (ii) a stereotypical Notch-DLL1 lateral inhibition event at the ends, driving the formation of the first Paneth cell that constitutes the niche (movie S3; proposed mechanism summarized in Fig. 3L).

In this system, Wnt is likely a crucial factor in maintaining and reinforcing the cell fate pattern. Indeed, we believe that it is the Wnt (and Notch) ligands produced by the first Paneth cells appearing at the tips of the tissues that help locally establish and maintain the stem cell niche. In line with this model, inhibition of Wnt production and secretion by treatment with IWP2 depletes Lgr5$^+$ ISCs from the tips and abolishes niche appearance at 36 hours of culture and beyond (fig. S8A). Compensating for the blocked Wnt production by supplementing with exogenous Wnt3a led to the expected outcome: Lgr5$^+$ ISCs were no longer depleted from the system at 36 hours, but were distributed more uniformly throughout the tissue (fig. S8B). Exposure to exogenous Wnt3a in the absence of IWP2 did not significantly affect patterning (fig. S8C). These data support the role of localized Wnt production and signaling as an important factor in establishing and maintaining the patterning in the system. However, we believe that localized Wnt signaling, although necessary, is not sufficient to drive symmetry breaking in the tissues. Monitoring the Wnt activity within the tissues, we found that the restriction of Wnt signaling followed rather than preceded the patterning of Notch, Lgr5, and Paneth cells (Fig. 3L and fig. S8, G and H).

In addition to the mechanism introduced, we also considered others that had previously been described for the regionalization of in-

testinal and other epithelia. Investigating the potential role of geometrically established inhibitory gradients in sonic hedgehog (SHH) (*38*) or transforming growth factor–β (TGF-β) (*21*) (fig. S9), as well as the role of patterning through cell-cell repulsive interactions (*39–41*) (fig. S10), we found that, whereas these mechanisms may contribute to or reinforce the patterning, they are likely not responsible for the initial symmetry breaking.

Bioengineered tissues with an in vivo–like tissue architecture and pattern

Next, intrigued by the possibility that differences in the packing of cells inside and outside the intestinal crypt may help to locate ISCs at the bottom of the crypt and establish a crypt-villus–like axis, we used a simplified intestinal surface with indentations that mimic the size and shape of the crypt, surrounded by flat regions that approximate the much larger surface of the villi (fig. S11). Moving from the bottom of the engineered crypts to the exterior surfaces (at 24 to 36 hours after cell loading), we observed a gradual increase in cell spreading and YAP activation (movie S4). Conversely, Lgr5 expression was maximal at the bottom of the crypts and was virtually absent outside of the cavities, indicating that stem cells were preserved only at the bottom of the crypts, in the familiar in vivo–like pattern (fig. S11B), despite a homogeneous cocktail of soluble cell fate–determining factors.

Finally, we sought to exploit these mechanistic insights to engineer macroscopic organoids with an in vivo tissue architecture (Fig. 4). We microfabricated hydrogel substrates resembling the native intestinal mucosa, with crypts at the bottom and villi protruding outward (Fig. 4, A to D). Within 48 hours, ISCs seeded on these scaffolds expanded to establish a confluent monolayer of cells (Fig. 4E and movie S5). Induction of differentiation resulted in highly stereotyped organoid patterning, with stem cell–containing crypts (Fig. 4F) interspersed with villi composed of enterocytes and other differentiated cell types (Fig. 4G and fig. S12). These results demonstrate that the principles controlling cell fate patterning at the subtissue scale can be harnessed to engineer macroscopic intestinal surfaces that capture the cellular organization and periodicity of the crypt-villus system. Previous attempts to engineer intestinal surfaces (*10*, *11*) were dependent on complex chemical gradients to induce patterning (i.e., did not reveal or exploit tissue geometry as a patterning cue) and have not demonstrated recapitulation of the multicellular organization and periodicity of the in vivo crypt-villus system. A key advantage of the engineered crypt-villus surfaces over 3D organoids is the well-delineated and mature villus region and accessibility to the luminal intestinal surface. We leveraged
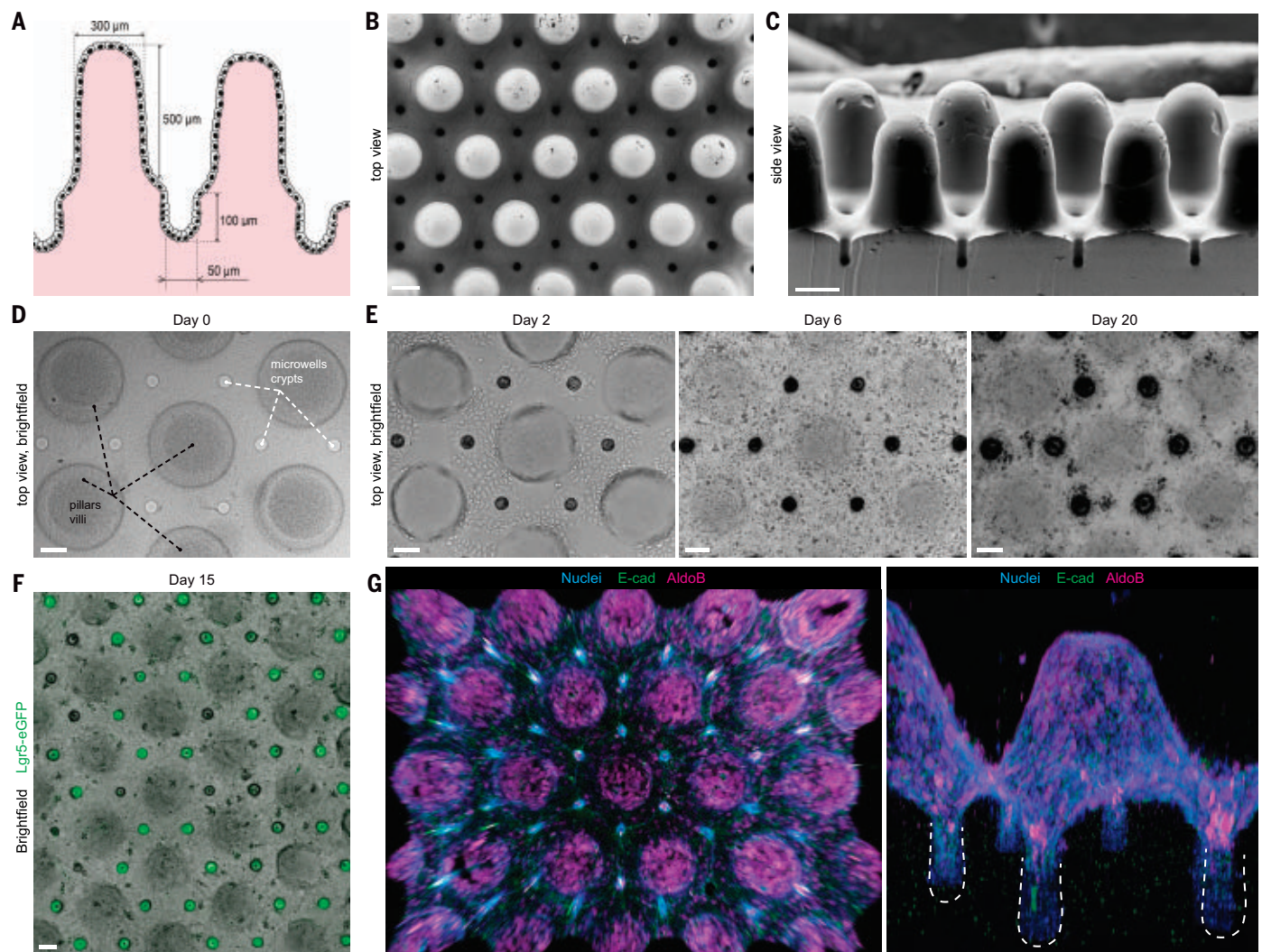
Fig. 4. Bioengineered organoids with an in vivo–like tissue architecture.
(**A**) Scheme of the designed topography resembling the native tissue with characteristic intestinal crypt-villus architecture. (**B** and **C**) SEM images of the poly(dimethylsiloxane) (PDMS) template used for fabricating bioengineered hydrogel substrates featuring a crypt-villus architecture (top and side views, respectively). Scale bars, 200 μm. (**D**) Top view of the hydrogel substrate shaped according to the topology of the native intestinal mucosa. (**E**) Bright-field time-course images of the intestinal epithelium development. (**F**) Localization of the Lgr5+ stem cells in the engineered crypts. In (D) to (F), extended depth of field for a z-stack ~ 600 μm; scale bars, 100 μm. (**G**) 3D reconstruction of the immunofluorescence images, showing confluent monolayer of E-cadherin–expressing epithelial cells covering hydrogel substrates, harboring villi composed of enterocytes (AldoB). See movie S5 for full time course of the epithelium growth and 3D immunofluorescence imaging.

these features, along with the possibility of live imaging, to explore the process of intestinal shedding.

Shedding is an important physiological process, whereby the pressure due to proliferative overcrowding is relieved while maintaining a tight and functional intestinal epithelial barrier. When dysregulated in pathological situations, such as inflammatory disease, shedding leads to loss of barrier function, exacerbating the disease (42, 43). Intestinal epithelial cell death and extrusion from the epithelium has been difficult to study in vitro, for the reasons mentioned above, whereas fixed ex vivo intestinal sections capture only snapshots of the process, precluding study of the temporal dynamics. We observed cell shedding within our engi-neered intestinal epithelia (Fig. 5 and movie S6) and showed that the process recapitulated pathophysiological hallmarks of the in vivo process. First, we found that cell shedding was always associated with the appearance of an actin ring at the interface of the shed cell and its neighbors (Fig. 5A and movie S7) (44, 45). Second, the system replicated the extensive shedding induced by the inflammatory cyto-kine tumor necrosis factor–α (TNF-α) (Fig. 5B), which has been shown to lead to loss of bar-rier function during intestinal inflammation (42, 46). Unlike 3D organoids and ex vivo approaches, our system allows the collection of shed cells for downstream analysis (Fig. 5C). We followed shed cells continuously before, during, and after extrusion and monitored apoptosis simultaneously. Using this approach, we demonstrated that apoptosis can occur before shedding, in situ, as well as after the cell has been extruded from the epithelium (Fig. 5D and movie S8).

We thus report a means to engineer external guidance in stem cell–based organogenesis, a process otherwise fully driven by stochastic self-organization. We show that localized pat-terning of ECM mechanics and topograph-ically structured hydrogel scaffolds can be used to build organoids of a controlled initial size and shape and to predict and influence the course of their development, in particular the breaking of symmetry and the number and location of crypt-like domains. This advance may help to overcome the lack of reproducibility
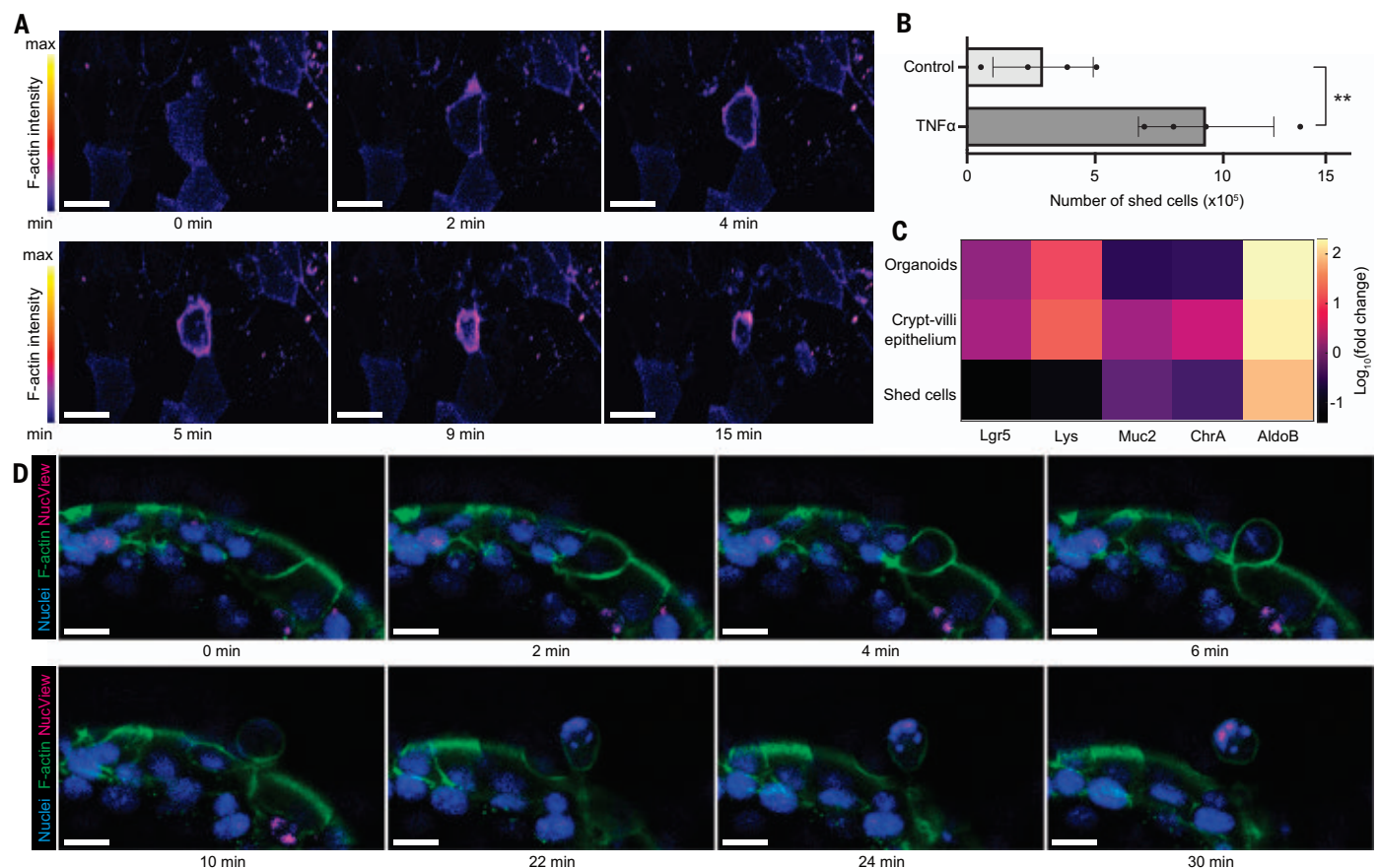
Fig. 5. Leveraging a bioengineered crypt-villus system to study the mechanisms of epithelial cell shedding in the small intestine. (**A**) Time-course imaging of epithelial cell undergoing shedding from the villus tip reveals gradual contraction of the of F-actin ring at the interface of the shed cell and its neighbors. (**B**) Quantification of the number of cells shed under homeostatic conditions and after treatment with TNF-α. Individual points and means with SD are shown. **P = 0.0025. (**C**) Heatmap of quantitative real-time PCR data showing the relative expression of intestine-specific genes (Lgr5, Lys, Muc2, ChrA, AldoB) in organoids, 15-day-old engineered crypt-villus epithelia, and shed cells over 8 hours. Data are average expression over n = 3 samples. (**D**) Time-course fluorescence imaging identifies caspase-3/7 activation after cell extrusion from the villus tip. See movies S6 to S8 for full time-lapse imaging of the shedding. Scale bars, 20 μm.

and control that are major limitations of organoids, undercutting their utility in basic and translational research. The controlled organoid development afforded by these approaches allowed us to identify a role for tissue geometry in intestinal tissue patterning, and to dissect the underlying mechanisms. Our data suggest that anisometric tissue geometries drive stereotypic epithelial patterning by establishing reproducible local differences in cell packing and morphology. It is the heterogeneities in YAP activity that ultimately prescribe "villus" and "crypt" domains by suppressing stem cell fates and localizing Notch-mediated Paneth cell differentiation, respectively. We believe that spatial variations in cell morphology alone are likely sufficient to induce symmetry breaking, and could be responsible for driving crypt initiation within classical organoids, where they would occur randomly. We have shown that tissue geometry can be used as a means to deterministically control the spatial distribution of cell morphology and, consequently, cell fate. The notion that morphogenesis is self-referential—that tissue form can serve as an independent input into its further development—is not new (*21*, *47*). In the case of the native intestine, the shape of the villi has been shown to influence the restriction of the stem cell zones by introducing diffusion-based spatial heterogeneities in signals communicated between the epithelium and the mesenchyme (*38*). Our experiments reveal that ISC restriction to the ends of crypt-like engineered tissue occurs in the absence of villi and mesenchyme; this suggests a complementary mechanism for intestinal regionalization, whereby the epithelial geometry per se allows for autonomous patterning of the tissue.

Methods summary

A full description of materials and methods is provided in the supplementary materials. Briefly, for organoid photopatterning experiments, ISC colonies were embedded within RGD- and laminin-1–containing photosensitive poly(ethylene glycol) (PEG)–based hydrogels (*20*), which undergo degradation and softening when exposed to 405-nm light. Photodegradation was performed on a laser scanning microscope (Zeiss LSM 710, 405-nm laser at a power of 1 mW) using four ROIs approximating the in vivo crypt dimensions (20 μm × 300 μm) arranged in a cross shape around the colony in a single plane. Approximately 30 colonies per gel were patterned and then the hydrogel was placed in differentiation medium (EGF, Noggin, and R-spondin) to induce differentiation and crypt-villus patterning.

Arrays of the microfabricated intestinal organoids were created using elastomeric stamps [PDMS, poly(dimethylsiloxane)] containing the desired geometries in bas relief. A hydrogel composed of type I collagen and 25% (v/v) Matrigel was cast on the bottom of the cell culture plate between two thin parallel PDMS spacers. A PDMS stamp with cavities of defined size and shape was placed on top of the PDMS spacers and hydrogel was polymerized in the incubator for 30 min. After removal of the stamp, hydrogels were covered with Advanced

DMEM:F12 medium and stored in the incubator. Intestinal organoids were cultured in Matrigel following previously published protocols (*37*). For cell loading in the microwells, concentrated suspension of dissociated intestinal stem cells was pipetted on top of the hydrogel surface. Cells were allowed to sediment within the microcavities of the gels for 5 to 7 min, and nonadherent cells were gently washed with medium. Partially gelled collagen "lid" was placed on top of the cell arrays and the two layers of collagen were allowed to seal for additional 15 min in the incubator. ISC expansion medium supplemented with 2.5 μM thiazovivin was added. The cells were allowed to self-organize into lumenized epithelial stem cell colonies adopting the geometry of the hydrogel microwells (typically 36 hours). To induce differentiation, budding, and organoid formation, the medium was replaced with fresh medium containing EGF, Noggin, and R-spondin.

Microfabricated crypt arrays and crypt-villi–shaped epithelium were produced in an identical manner, but the patterns were not sealed with a second layer of hydrogel and therefore cells were permitted to grow both within the cavities and atop the external surfaces.

REFERENCES AND NOTES

1. H. Clevers, Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597 (2016). doi: 10.1016/j.cell.2016.05.082; pmid: 27315476
2. J. Drost, H. Clevers, Organoids in cancer research. *Nat. Rev. Cancer* **18**, 407–418 (2018). doi: 10.1038/s41568-018-0007-6; pmid: 29692415
3. J. Drost *et al.*, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **358**, 234–238 (2017). doi: 10.1126/science.aao3130; pmid: 28912133
4. M. A. Lancaster, J. A. Knoblich, Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science* **345**, 1247125 (2014). doi: 10.1126/science.1247125; pmid: 25035496
5. G. Vlachogiannis *et al.*, Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018). doi: 10.1126/science.aao2774; pmid: 29472484
6. G. Rossi, A. Manfrin, M. P. Lutolf, Progress and potential in organoid research. *Nat. Rev. Genet.* **19**, 671–687 (2018). doi: 10.1038/s41576-018-0051-9; pmid: 30228295
7. Y. Sasai, Cytosystems dynamics in self-organization of tissue architecture. *Nature* **493**, 318–326 (2013). doi: 10.1038/nature11859; pmid: 23325214
8. M. Huch, J. A. Knoblich, M. P. Lutolf, A. Martinez-Arias, The hope and the hype of organoid research. *Development* **144**, 938–941 (2017). doi: 10.1242/dev.150201; pmid: 28292837
9. Advances in Organoid Technology: Hans Clevers, Madeline Lancaster, and Takanori Takebe [edited interview with Clevers, Lancaster, and Takebe]. *Cell Stem Cell* **20**, 759–762 (2017). doi: 10.1016/j.stem.2017.05.014
10. Y. Wang *et al.*, A microengineered collagen scaffold for generating a polarized crypt-villus architecture of human small intestinal epithelium. *Biomaterials* **128**, 44–55 (2017). doi: 10.1016/j.biomaterials.2017.03.005; pmid: 28288348
11. R. Kim *et al.*, Formation of arrays of planar, murine, intestinal crypts possessing a stem/proliferative cell compartment and differentiated cell zone. *Lab Chip* **18**, 2202–2213 (2018). doi: 10.1039/C8LC00332G; pmid: 29944153
12. Y. Wang *et al.*, Formation of Human Colonic Crypt Array by Application of Chemical Gradients Across a Shaped Epithelial Monolayer. *Cell. Mol. Gastroenterol. Hepatol.* **5**, 113–130 (2017). doi: 10.1016/j.jcmgh.2017.10.007; pmid: 29693040

13. M. Nikolaev *et al.*, Homeostatic mini-intestines through scaffold-guided organoid morphogenesis. *Nature* **585**, 574–578 (2020). doi: 10.1038/s41586-020-2724-8; pmid: 32939089
14. M. A. Lancaster *et al.*, Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017). doi: 10.1038/nbt.3906; pmid: 28562594
15. H. M. Poling *et al.*, Mechanically induced development and maturation of human intestinal organoids in vivo. *Nat. Biomed. Eng.* **2**, 429–442 (2018). doi: 10.1038/s41551-018-0243-9; pmid: 30151330
16. M. Zinner, I. Lukonin, P. Liberali, Design principles of tissue organisation: How single cells coordinate across scales. *Curr. Opin. Cell Biol.* **67**, 37–45 (2020). doi: 10.1016/j.ceb.2020.07.004; pmid: 32889170
17. N. Gjorevski *et al.*, Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016). doi: 10.1038/nature20168; pmid: 27851739
18. T. Sato *et al.*, Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009). doi: 10.1038/nature07935; pmid: 19329995
19. T. Sato, H. Clevers, Growing self-organizing mini-guts from a single intestinal stem cell: Mechanism and applications. *Science* **340**, 1190–1194 (2013). doi: 10.1126/science.1234852; pmid: 23744940
20. D. D. McKinnon, T. E. Brown, K. A. Kyburz, E. Kiyotake, K. S. Anseth, Design and characterization of a synthetically accessible, photodegradable hydrogel for user-directed formation of neural networks. *Biomacromolecules* **15**, 2808–2816 (2014). doi: 10.1021/bm500731b; pmid: 24932668
21. C. M. Nelson, M. M. Vanduijn, J. L. Inman, D. A. Fletcher, M. J. Bissell, Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. *Science* **314**, 298–300 (2006). doi: 10.1126/science.1131000; pmid: 17038622
22. C. M. Nelson, J. L. Inman, M. J. Bissell, Three-dimensional lithographically defined organotypic tissue arrays for quantitative analysis of morphogenesis and neoplastic progression. *Nat. Protoc.* **3**, 674–678 (2008). doi: 10.1038/nprot.2008.35; pmid: 18388950
23. H. Clevers, The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274–284 (2013). doi: 10.1016/j.cell.2013.07.004; pmid: 23870119
24. A. Gregorieff, Y. Liu, M. R. Inanlou, J. L. Wrana, Yap-dependent reprogramming of Lgr5+ stem cells drives intestinal regeneration and cancer. *Nature* **526**, 715–718 (2015). doi: 10.1038/nature15382; pmid: 26503053
25. M. Imajo, M. Ebisuya, E. Nishida, Dual role of YAP and TAZ in renewal of the intestinal epithelium. *Nat. Cell Biol.* **17**, 7–19 (2015). doi: 10.1038/ncb3084; pmid: 25531778
26. T. Panciera, L. Azzolin, M. Cordenonsi, S. Piccolo, Mechanobiology of YAP and TAZ in physiology and disease. *Nat. Rev. Mol. Cell Biol.* **18**, 758–770 (2017). doi: 10.1038/nrm.2017.87; pmid: 28951564
27. S. Yui *et al.*, YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22**, 35–49.e7 (2018). doi: 10.1016/j.stem.2017.11.001; pmid: 29249464
28. S. Dupont *et al.*, Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–183 (2011). doi: 10.1038/nature10137; pmid: 21654799
29. M. Aragona *et al.*, A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell* **154**, 1047–1059 (2013). doi: 10.1016/j.cell.2013.07.042; pmid: 23954413
30. D. Serra *et al.*, Self-organization and symmetry breaking in intestinal organoid development. *Nature* **569**, 66–72 (2019). doi: 10.1038/s41586-019-1146-y; pmid: 31019299
31. I. Lukonin *et al.*, Phenotypic landscape of intestinal organoid regeneration. *Nature* **586**, 275–280 (2020). doi: 10.1038/s41586-020-2776-9; pmid: 33029001
32. F. Fan *et al.*, Pharmacological targeting of kinases MST1 and MST2 augments tissue repair and regeneration. *Sci. Transl. Med.* **8**, 352ra108 (2016). doi: 10.1126/scitranslmed.aaf2304
33. E. A. Rosado-Olivieri, K. Anderson, J. H. Kenty, D. A. Melton, YAP inhibition enhances the differentiation of functional stem cell-derived insulin-producing β cells. *Nat. Commun.* **10**, 1464 (2019). doi: 10.1038/s41467-019-09404-6; pmid: 30931946
34. Y. Liu-Chittenden *et al.*, Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic

activity of YAP. *Genes Dev.* **26**, 1300–1305 (2012). doi: 10.1101/gad.192856.112; pmid: 22677547
35. W. R. Legant *et al.*, Measurement of mechanical tractions xerted by cells in three-dimensional matrices. *Nat. Methods* **7**, 969–971 (2010). doi: 10.1038/nmeth.1531; pmid: 21076420
36. H. F. Farin *et al.*, Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016). doi: 10.1038/nature16937; pmid: 26863187
37. X. Yin *et al.*, Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. *Nat. Methods* **11**, 106–112 (2014). doi: 10.1038/nmeth.2737; pmid: 24292484
38. A. E. Shyer, T. R. Huycke, C. Lee, L. Mahadevan, C. J. Tabin, Bending gradients: How the intestinal stem cell gets its home. *Cell* **161**, 569–580 (2015). doi: 10.1016/j.cell.2015.03.041; pmid: 25865482
39. J. Holmberg *et al.*, EphB receptors coordinate migration and proliferation in the intestinal stem cell niche. *Cell* **125**, 1151–1163 (2006). doi: 10.1016/j.cell.2006.04.030; pmid: 16777604
40. M. Genander *et al.*, Dissociation of EphB2 signaling pathways mediating progenitor cell proliferation and tumor suppression. *Cell* **139**, 679–692 (2009). doi: 10.1016/j.cell.2009.08.048; pmid: 19914164
41. E. Batlle *et al.*, β-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263 (2002). doi: 10.1016/S0092-8674(02)01015-2; pmid: 12408869
42. R. Kiesslich *et al.*, Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. *Gut* **61**, 1146–1153 (2012). doi: 10.1136/gutjnl-2011-300695; pmid: 22115910
43. J. J. Liu *et al.*, Mind the gaps: Confocal endomicroscopy showed increased density of small bowel epithelial gaps in inflammatory bowel disease. *J. Clin. Gastroenterol.* **45**, 240–245 (2011). doi: 10.1097/MCG.0b013e3181fbdb8a; pmid: 21030873
44. F. Wang *et al.*, Active deformation of apoptotic intestinal epithelial cells with adhesion-restricted polarity contributes to apoptotic clearance. *Lab. Invest.* **91**, 462–471 (2011). doi: 10.1038/labinvest.2010.182; pmid: 21042290
45. Y. Guan *et al.*, Redistribution of the tight junction protein ZO-1 during physiological shedding of mouse intestinal epithelial cells. *Am. J. Physiol. Cell Physiol.* **300**, C1404–C1414 (2011). doi: 10.1152/ajpcell.00270.2010; pmid: 21346149
46. A. M. Marchiando *et al.*, The epithelial barrier is maintained by in vivo tight junction expansion during pathologic intestinal epithelial shedding. *Gastroenterology* **140**, 1208–1218.e1 (2011). doi: 10.1053/j.gastro.2011.01.004; pmid: 21237166
47. C. M. Nelson, Geometric control of tissue morphogenesis. *Biochim. Biophys. Acta Mol. Cell Res.* **1793**, 903–910 (2009). doi: 10.1016/j.bbamcr.2008.12.014

Science

JOURNALS △△AAAS

Publish your research in the *Science* family of journals

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

Submit your research today!

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

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Dense functional and molecular readout of a circuit hub in sensory cortex

Cameron Condylis, Abed Ghanbari, Nikita Manjrekar, Karina Bistrong, Shenqin Yao, Zizhen Yao, Thuc Nghi Nguyen, Hongkui Zeng, Bosiljka Tasic, Jerry L. Chen*

INTRODUCTION: The diversity of cell types is a defining feature of the neuronal circuitry that makes up the areas and layers of the mammalian cortex. At a molecular level, the extent of this diversity is now better appreciated through recent efforts to census all potential cortical cell types through single-cell transcriptional profiling. Cortical populations can be hierarchically subdivided into multiple putative transcriptomic cell classes, subclasses, and types. This new catalog of neuronal subclasses and subtypes opens up new questions and avenues of investigation for how these cell types are collectively organized into circuits that function to process information and adapt to changes in experience.

RATIONALE: We investigated the function of newly identified cell types in layers 2 or 3 (L2/3) of the primary somatosensory cortex, a region that integrates bottom-up sensory information with top-down internal representations. Current in vivo methods primarily allow cell types to be investigated one at a time and have limited ability to label cell types defined by combinations of expressed genes. To densely survey these cell types and investigate how they interact during task behavior, we developed a platform, Comprehensive Readout of Activity and Cell Type Markers (CRACK), that combines population calcium imaging with subsequent multiplexed fluorescent in situ hybridization. Multiplexed labeling of mRNA transcripts is critical to deciphering the identity of cell types defined by combinatorial patterns of gene expression.

RESULTS: We profiled the functional responses of three excitatory cell types and eight inhibitory subclasses in L2/3 as mice performed a whisker-based tactile working memory task. Task-related properties of both excitatory and inhibitory neurons continue to differentiate as they are segregated into increasingly discrete molecular types. Our analysis revealed that the excitatory cell type, L2/3 intratelencephalic Baz1a (Baz1a), functions as a highly active detector of tactile features. Simultaneous imaging across identified cell types enabled measurements of functional connectivity between subpopulations. Functional connectivity analysis indicated that Baz1a neurons orchestrate local network activity patterns. We found that Baz1a neurons show strong functional connections with dendrite-targeting, somatostatin-expressing (Sst) inhibitory neurons. Trans-monosynaptic viral tracing confirmed that Baz1a neurons preferentially synapse onto Sst neurons. Baz1a neurons also show enrichment of select plasticity-related, immediate early genes, including *Fos*. To determine whether the expression pattern of immediate early genes is a stable property of Baz1a neurons and how this relates to neuronal plasticity, we tracked *Fos* expression and neuronal activity in mice subjected to whisker deprivation. We found that Baz1a neurons homeostatically adapt to sensory deprivation while stably maintaining *Fos* expression.

CONCLUSION: These results demonstrate that Baz1a neurons are a component of a molecularly defined circuit motif that is capable of recruiting local circuits for sensory processing when salient features are encountered during behavior. This cell type also functions to preserve sensory representations during ongoing and altered sensory experience. This builds on our knowledge for how local circuits in somatosensory cortex are implemented to negotiate bottom-up and top-down information. The ability to map functional and transcriptional relationships across neuronal populations provides insight into how the organizing principles of the cortex give rise to the computations it performs. ∎

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl5981
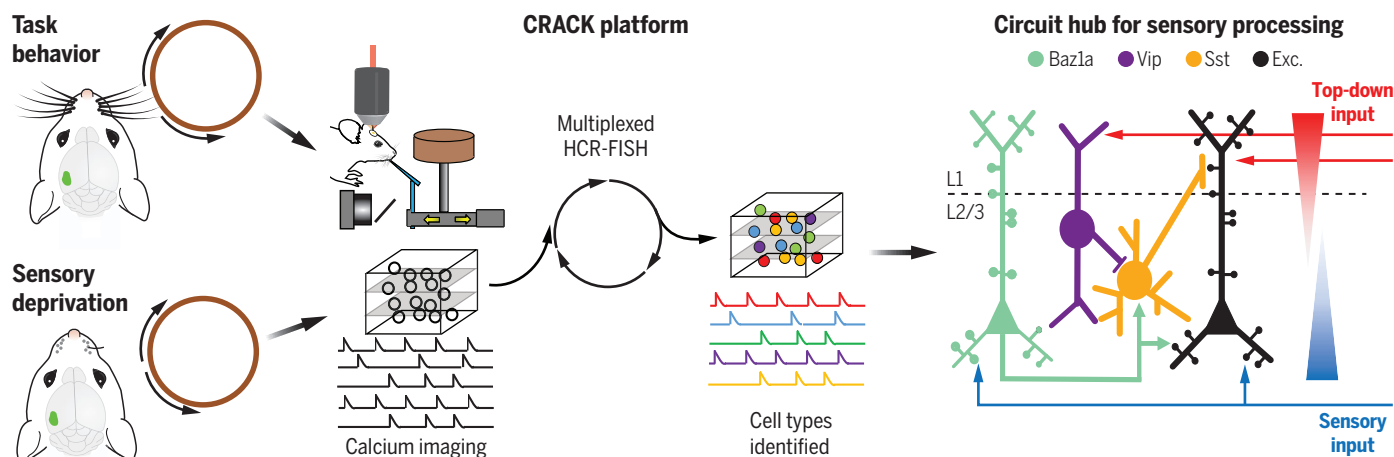
CRACK platform reveals a circuit hub for sensory processing. Functional profiling of molecularly defined cells was achieved with in vivo two-photon calcium imaging in L2/3 of the primary somatosensory cortex during task behavior or sensory deprivation followed by multiplexed fluorescent in situ hybridization. Excitatory Baz1a neurons form a connection motif capable of recruiting local circuits and preserving sensory representations during ongoing and altered sensory experience. HCR-FISH, hybridization chain reaction–fluorescence in situ hybridization; Vip, vasoactive intestinal peptide–expressing; Exc., excitatory.

NEUROSCIENCE

Dense functional and molecular readout of a circuit hub in sensory cortex

Cameron Condylis[1,2], Abed Ghanbari[3], Nikita Manjrekar[3], Karina Bistrong[3], Shenqin Yao[4], Zizhen Yao[4], Thuc Nghi Nguyen[4], Hongkui Zeng[4], Bosiljka Tasic[4], Jerry L. Chen[1,2,3,5]*

Although single-cell transcriptomics of the neocortex has uncovered more than 300 putative cell types, whether this molecular classification predicts distinct functional roles is unclear. We combined two-photon calcium imaging with spatial transcriptomics to functionally and molecularly investigate cortical circuits. We characterized behavior-related responses across major neuronal subclasses in layers 2 or 3 of the primary somatosensory cortex as mice performed a tactile working memory task. We identified an excitatory intratelencephalic cell type, Baz1a, that exhibits high tactile feature selectivity. Baz1a neurons homeostatically maintain stimulus responsiveness during altered experience and show persistent enrichment of subsets of immediately early genes. Functional and anatomical connectivity reveals that Baz1a neurons residing in upper portions of layers 2 or 3 preferentially innervate somatostatin-expressing inhibitory neurons. This motif defines a circuit hub that orchestrates local sensory processing in superficial layers of the neocortex.

Cells of the neocortex can be defined on the basis of their molecular composition, the diversity of which is reflected in their transcriptome. The transcriptional profiles observed across this brain region indicate that cortical populations can be hierarchically subdivided into multiple putative transcriptomic cell classes [such as γ-aminobutyric acid (GABA)–ergic or glutamatergic], subclasses (such as GABAergic Pvalb), and types (such as GABAergic Pvalb Vipr2) (1, 2). Even within a single layer of one cortical area, transcriptional diversity remains high (3). This organization may have developmental origins (4, 5) or reflect anatomical specificity (6, 7) or physiological properties (8, 9). The extent to which this diversity relates to information encoding during goal-directed behavior is unclear. In superficial layers of the neocortex, excitatory layer-2 or -3 (L2/3) pyramidal neurons can be disinhibited by subclasses of inhibitory vasoactive intestinal peptide–expressing (Vip) neurons through subclasses of inhibitory somatostatin-expressing (Sst) neurons. The degree to which this motif is part of a larger circuit composed of other transcriptomic cell types is unclear.

The ability to link molecularly identified neurons with their function during behavior requires monitoring the activity of cell types in vivo. Traditional approaches to label cell types by use of transgenic lines or post hoc immunohistochemistry are limited to one to three molecular markers (10, 11). This has restricted investigations to classes of excitatory and inhibitory neurons to the broadest hierarchical levels of cell type diversity. Techniques for spatial transcriptional profiling increase the number of genes that can be simultaneously identified in tissue (12–16). Combinatorial expression patterns of multiple genes can then be used to define finer divisions in the transcriptomic taxonomy that correspond to more specific neuronal subclasses and types. Further, spatial profiling of gene expression in intact tissue readily enables dense multimodal registration of anatomical and functional measurements across neurons within a single sample (17). We developed a platform, Comprehensive Readout of Activity and Cell Type Markers (CRACK), that combines in vivo two-photon calcium imaging with post hoc multiplexed fluorescence in situ hybridization. Using this platform, we sought to determine whether finer divisions in the transcriptomic taxonomy (subclasses and types) exhibit distinct functional characteristics and connection motifs. We focused on newly identified cell types in L2/3 of the primary somatosensory cortex (S1), a region involved in processing and integrating tactile information with motor and associative input.

CRACK platform

The CRACK platform uses a multi-area two-photon microscope (18) configured to perform simultaneous population calcium imaging across multiple tissue depths, providing three-dimensional (3D) spatial information of neuron location for later post hoc identification (Fig. 1A, fig. S1, and movie S1). After functional in vivo experiments, tissue encompassing the imaged volume was sectioned parallel to the imaging plane. The tissue was embedded in hydrogel and cleared (19) to facilitate labeling of mRNA transcripts by using hybridization chain reaction–fluorescence in situ hybridization (HCR-FISH) (13) and confocal imaging. Because HCR-FISH is a DNA-based labeling strategy, probes for different mRNA transcripts were labeled, imaged, and then stripped by using deoxyribonuclease (DNase) across multiple rounds. To reidentify and register in vivo neurons across multiple rounds of HCR-FISH, we dedicated one imaging channel (561) to repeated labeling and imaging of transcripts of the red genetically encoded calcium indicator, RCaMP1.07, which we used for functional imaging (fig. S2 and supplementary text S1) (20). Other imaging channels were used for labeling cell type–specific markers (table S1).

Although expression of a small number of genes can be detected through multiple rounds of sequential staining, a barcode readout scheme provides high read depth (100 to 1000 genes) in an error-robust manner. Using barcode readouts to decode arbitrary gene sets relies on single-molecule mRNA resolution, which is sensitive to image registration errors and has only been demonstrated in thin tissue sections (<40 μm) (12, 16). To obviate the need for single-molecule mRNA-resolution registration so that larger volumes of tissue (150 to 300 μm) could be imaged and analyzed, we programmed our barcode for cellular-resolution readout. This approach relies on prior knowledge of gene expression patterns so that binary decoding for each imaging channel and hybridization round could be programmed at cellular rather than mRNA resolution. This approach is highly compatible with identifying cell types defined by nonoverlapping gene expression patterns.

We analyzed single-cell RNA-sequencing (scRNA-seq) data from S1 that were acquired as part of a larger study of the molecular diversity of the isocortex (21). On the basis of combinatorial expression patterns, L2/3 intratelencephalic (IT) pyramidal neurons in S1 were observed to be segregated into three transcriptomic cell types: L2/3 IT Adamts2 (Adamts2), L2/3 IT Baz1a (Baz1a), and L2/3 IT Agmat (Agmat) (fig. S4). Excitatory neurons in L2/3 show both cell type–specific and area-specific gene expression patterns. When comparing S1 L2/3 cell types to those in the primary visual (V1) and anterior lateral motor (ALM) cortex, Baz1a and Agmat cells showed similarity to cell types identified in V1 and ALM, whereas Adamts2 cells were present in V1 but not ALM (1).

Inhibitory neuron cell types in S1 were shared with other cortical areas and found to be hierarchically organized. Although the major nonoverlapping inhibitory subclasses (Lamp5,

[1]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA. [2]Center for Neurophotonics, Boston University, Boston, MA 02215, USA. [3]Department of Biology, Boston University, Boston, MA 02215, USA. [4]Allen Institute for Brain Science, Seattle, WA 98109, USA. [5]Center for Systems Neuroscience, Boston University, Boston, MA 02215, USA.
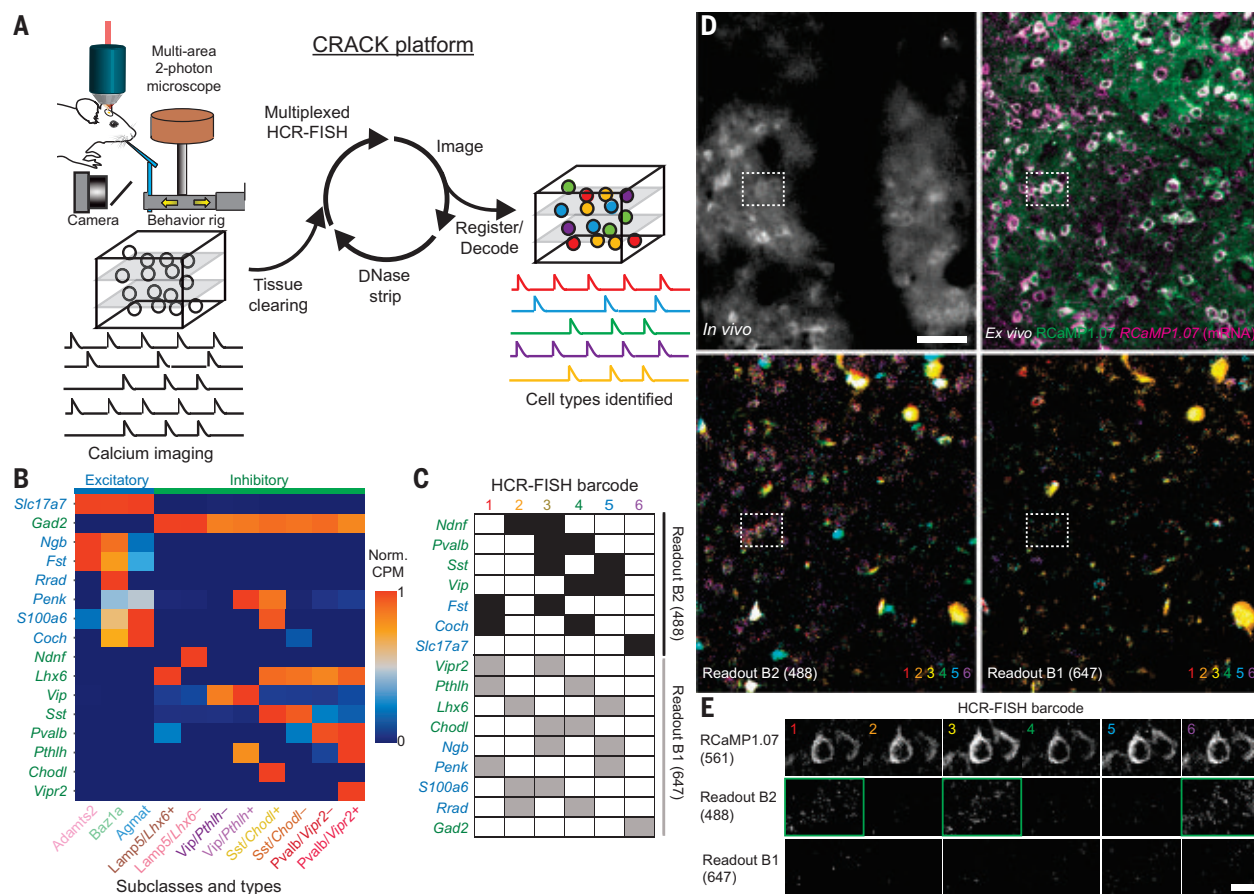*Corresponding author. Email: jerry@chen-lab.org

Fig. 1. Multiplexed identification of transcriptomic cell subclasses and types in functionally imaged neurons. (**A**) Schematic of the CRACK platform. (**B**) Expression patterns of genes selected to identify L2/3 S1 excitatory (blue) and inhibitory (green) cell subclasses and types. (**C**) Barcode scheme for multiplexed HCR-FISH of selected genes. (**D**) Registration of in vivo calcium-imaged neurons to ex vivo tissue section across multiple rounds of HCR-FISH. (Top left) In vivo two-photon images of RCaMP1.07+ neurons. (Top right) Ex vivo confocal images of reidentified RCaMP1.07+ neurons showing endogeneous protein (green) followed by HCR-FISH staining transcripts (magenta). (Bottom) Overlays of (left) B2-488 and (right) B1-647 readout channels across all HCR-FISH barcode rounds. (**E**) Decoding of in vivo imaged neuron [(D), dotted rectangle] identified as an Adamts2 cell type expressing *Fst* and *Slc17a7*. Positive readouts are identified with green rectangles. Scale bars, (D) 50 μm; (E) 20 μm.

Pvalb, Sst, and Vip) have each been investigated at the broadest level (*22, 23*), further subdivisions have not been investigated during task behavior. Thus, we selected gene markers that defined the next level of transcriptional subdivision (fig. S5). Lamp5 neurons were subdivided into two subclasses according to mutually exclusive expression of LIM homeobox 6 (*Lhx6*) or neuron-derived neurotrophic factor (*Ndnf*). Pvalb, Sst, or Vip neurons were subdivided according to expression of vasoactive intestinal peptide receptor 2 (*Vipr2*), chondrolectin (*Chodl*), or parathyroid hormone–like hormone (*Pthlh*), respectively. We devised a barcode scheme for detection of 16 mRNA species across six rounds of staining to resolve 11 transcriptomically defined cell populations (three excitatory types and eight inhibitory subclasses) (Fig. 1, B to E).

Task encoding across excitatory types

To identify functional differences between transcriptionally defined cell populations in

L2/3 of S1, two-photon calcium imaging was carried out on expert wild-type mice (*n* = 7) performing a head-fixed whisker-based delayed nonmatch to sample (DNMS) task (fig. S6) (*24*). In this context-dependent sensory processing task, a motorized rotor is used to deflect multiple whiskers in either an anterior or posterior direction during an initial "sample" and a later "test" period, separated by a 2-s delay (Fig. 2A). During the delay period and the intertrial interval, the rotor was withdrawn to prevent whisker-rotor contact. Behavior was reported as "go/no go," in which animals licked on "go" trials for a water reward ("hit") when the presented sample and test stimulus were nonmatching and withheld licking on "no go" trials ("correct rejection") when the presented sample and test stimulus were matching. High-speed videography was also performed to monitor whisking behavior.

We previously reported diverse task-related responses in L2/3 of S1 during the DNMS task (*24*). To characterize task-related responses for

each recorded cell in a more comprehensive manner, we fit a generalized linear model (GLM) to each neuron's estimated calcium event activity against a range of "task variables" (Fig. 2B, figs. S9 and S10, and supplementary text S2) (*25*). Task variables representing a related feature were grouped into "task factors" (such as stimulus direction and trial category). The ability for a neuron to encode a particular task factor was determined by calculating the difference in the Akaike information criterion (ΔAIC) between a full model and a partial model that excludes task variables representing that task factor. A positive ΔAIC value indicates reduced fit quality from the full to the partial model, revealing that the excluded task factor in the partial model is an important contributor to the modeled neuron's activity. Thus, we interpret significant, positive ΔAIC values to indicate neuronal encoding of the excluded task factor (Fig. 2C, fig. S11, and supplementary text S3). We analyzed 10 task-related factors. Six of the 10 task factors were defined by trial
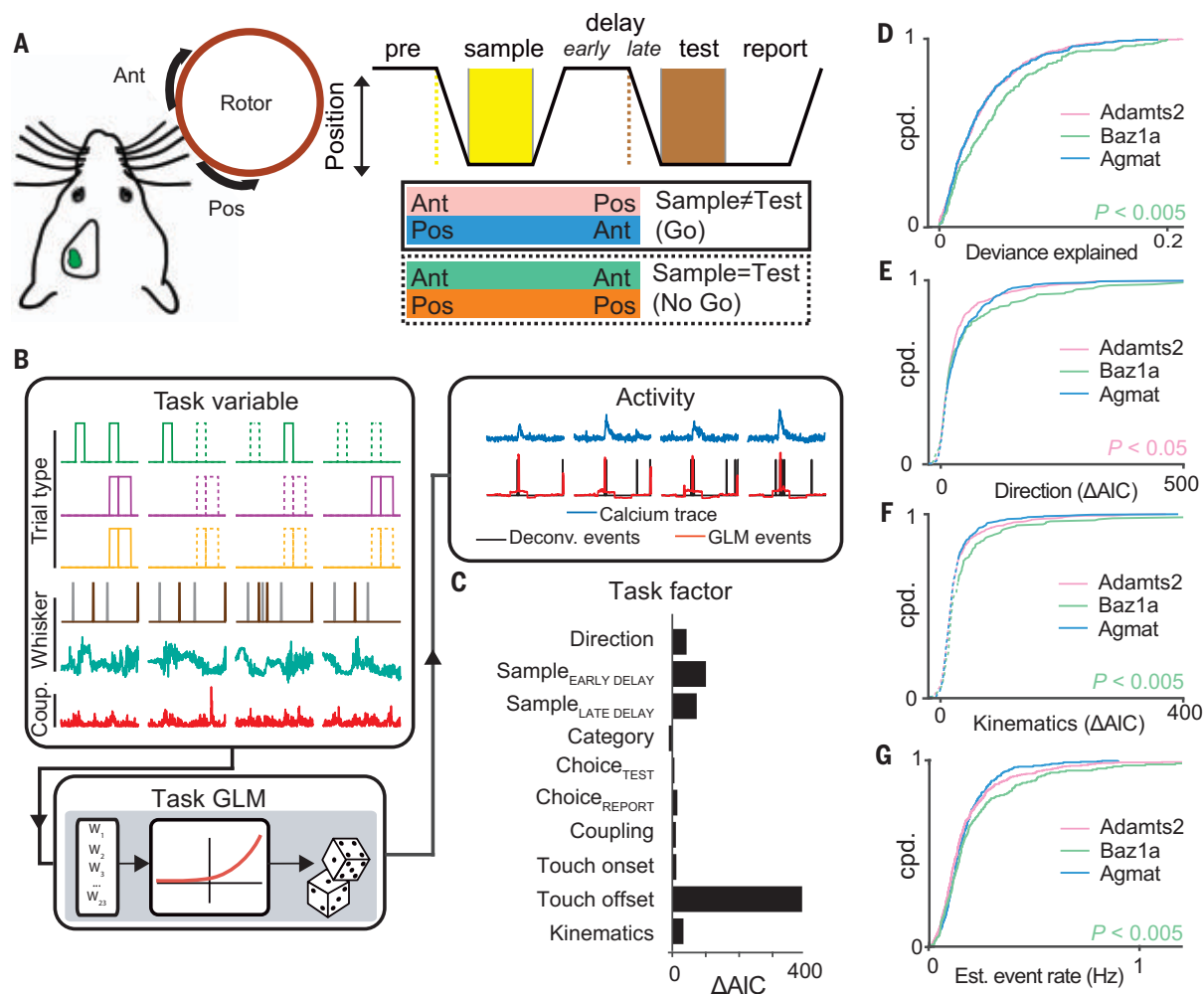
Fig. 2. Task encoding across L2/3 excitatory cell types. (**A**) Schematic of whisker-based delayed nonmatch to sample behavioral task. (**B**) Encoding of task-related activity in individual neurons using a GLM. (**C**) Encoding of task factors determined by comparing full and partial GLM fits (ΔAIC). (**D** to **G**) Cumulative probability distributions of (D) full model deviance explained, (E) encoding strength of stimulus direction, (F) encoding strength of whisker kinematics, and (G) estimated event rate across the three excitatory cell types. [(D) to (F)] Mann Whitney U test; (G) one-tailed Student's t test. In (E) and (F), solid and dotted lines indicate significant ($P < 0.01$) and nonsignificant encoding strengths, respectively, by means of χ^2 test. $n = 1107$ neurons from seven animals.

type information. This included information related to the direction of the task stimulus (direction), trial category defined by the combination of the sample and test stimulus (category), and the animal's choice during the test (choice$_{TEST}$) and report period (choice$_{REPORT}$). Although our previous study found no evidence of sustained activity in S1 during the delay period (*25*), we included task factors that represent the sample stimulus at later points in the trial (sample encoded early in the delay period, sample$_{EARLY\ DELAY}$; sample information late in the delay period, sample$_{LATE\ DELAY}$). Another set of task factors describing whisker movement and tactile-object interactions were derived from video analysis of whisker tracking and included whisker-object touch onset (touch onset), whisker-object touch offset (touch offset), and whisker kinematics (kinematics). A final task factor was derived from the activity of all other simultaneously recorded neurons

to assess the level of coupling the neuron had with overall network activity (coupling) (*26*). Overall, we identified neurons that were selective to a single or multiple task factors (fig. S12).

We first compared differences in task encoding across the three excitatory cell types. Baz1a neurons showed the best overall GLM fit (Fig. 2D) and more strongly encoded whisker kinematics compared with the other two excitatory cell types ($P < 0.005$, Mann Whitney U test) (Fig. 2F). By contrast, Adamts2 neurons more weakly encoded stimulus direction and touch offset (direction, $P < 0.05$; touch offset, $P < 0.02$; Mann Whitney U test) (Fig. 2E and fig. S14), whereas Agmat neurons more strongly encoded choice$_{REPORT}$ ($P < 0.05$, Mann Whitney U test). Baz1a neurons also showed overall higher event rates ($P < 0.005$, one-tailed Student's t test) and response reliability to sample and test stimuli (Fig. 2G and fig. S13). However, encoding of touch onset did not

differ between excitatory cell types, suggesting that Baz1a neurons are more tuned to specific kinematic features rather than more sensitive to nonspecific tactile input (fig. S14).

Persistent stimulus activity and *Fos* expression in Baz1a neurons

Highly active, sensory-driven L2/3 S1 neurons exhibit high expression of the immediate early gene, *Fos* (*27*, *28*). scRNA-seq analysis in naïve, untrained mice shows that whereas all cell types express some number of immediate early genes (IEGs), Baz1a neurons show consistent enrichment of *Fos* along with a subset of IEGs (Fig. 3A). Although *Fos* expression is dynamic and driven by experience-dependent plasticity (*29*), we speculated that *Fos* and other IEGs may be stably expressed in Baz1a neurons. To confirm this and address how it relates to neuronal function, we extended the CRACK platform to track *Fos* expression

Fig. 3. Persistent IEG expression and homeostatic plasticity in Baz1a neurons. (**A**) Examples of selectively (top) enriched and (bottom) not enriched immediate early genes in Baz1a cells. (**B**) Time course of bilateral whisker deprivation (BD) experiment. (**C**) (Top) Example of Baz1a neuron with stable high fosGFP expression across in vivo imaging sessions. (Middle) Average stimulus responses during calcium imaging. (Bottom) Post hoc identification of neuron and HCR-FISH for select genes. (**D**) HCR-FISH *Fos* spot density in high (1.2-fold above background) and low fluorescent fosGFP cells (two-tailed Student's *t* test). (**E**) Mean stimulus-evoked activity before and after BD across functionally imaged neurons (one-way ANOVA with post hoc multiple comparison test, *n* = 2569 cells from three animals). (**F**) Change in stimulus-evoked responses before BD versus at 1 day or 5 days BD across neurons with stable low, dynamic, and stable high fosGFP expression (two-tailed Student's *t* test, *n* = 790 cells from three animals). (**G**) Change in stimulus-evoked responses before BD versus at 1 day or 5 days BD across excitatory cell types (χ^2 test, *n* = 181 Adamts2, 136 Baz1a, and 153 Agmat cells from three animals). (**H**) Fraction of fosGFP neurons with stable high expression across all pre-BD sessions (days −2, −1, and 0) and across all BD sessions (days 1, 3, and 5) for excitatory cell types (two-tailed Student's *t* test, *n* = 3753 cells from three animals). * *P* < 0.05, ** *P* < 0.005 in (E) to (G). Error bars = SEM; (H) SD from bootstrap analysis.
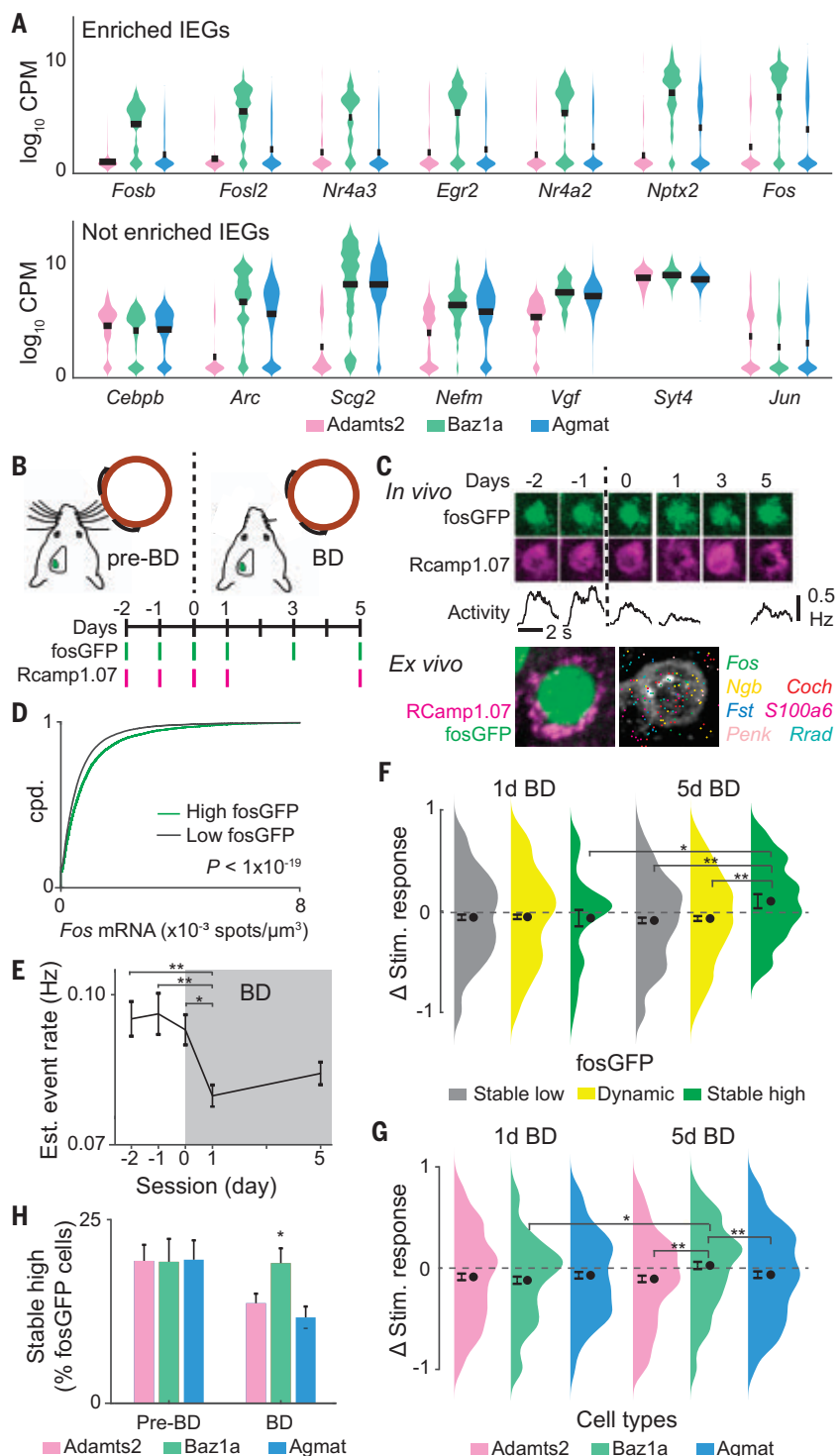
and stimulus-evoked activity during altered sensory experience using transgenic fosGFP mice (*30*) along with virally coexpressed Rcamp1.07 in S1 (*n* = 3) (Fig. 3, B and C). Ex vivo HCR-FISH confirmed that high fosGFP fluorescence corresponded with higher *Fos* mRNA and that green fluorescent protein (*GFP*) mRNA was also enriched in Baz1a neurons (Fig. 3D and fig. S18A). Although estimated event rates

correlated with FosGFP expression levels in Adamts2 and Agmat neurons, they did not in Ba1za neurons, confirming that *Fos* does not necessarily reflect ongoing activity in this cell type (fig. S18B).

FosGFP and sensory-evoked calcium responses were tracked before and for 5 days after bilateral whisker deprivation (BD). During BD, the principal whisker corresponding to

the imaged S1 barrel column was trimmed to a minimum length so that stimulus-evoked activity could still be tracked (fig. S17) (*31*). Overall, BD resulted in a decrease in stimulus-evoked activity after 1 day followed by a slow homeostatic compensation after 5 days [*P* < 0.0002, one-way analysis of variance (ANOVA), post hoc multiple comparison test] (Fig. 3E) (*32*). We first asked how fosGFP expression

related to functional response changes during BD. Cells were divided into three groups according to fosGFP expression: (i) stable low fosGFP expression across all imaging sessions; (ii) stable high fosGFP expression across all imaging sessions, and (iii) dynamic fosGFP expression between at least two imaging sessions (fig. S19). All groups showed decreased stimulus-evoked activity after 1 day of BD. However, after 5 days, responses in stable low and dynamic fosGFP neurons remained depressed, whereas stable high fosGFP neurons exhibited an enhancement in sensory response magnitude and reliability compared with pre-BD conditions ($P < 0.05$, Student's t test) (Fig. 3F and fig. S20).

Ex vivo cell type identification revealed that similar fractions of stable high fosGFP neurons were observed across all excitatory cell types before deprivation. However, during BD, there was increased fosGFP turnover in Adamts2 and Agmat neurons, whereas the fraction of stable high fosGFP cells remained unchanged in Baz1a neurons ($P < 0.05$, χ^2 test) (Fig. 3H). Functionally, all three cell types showed reduced stimulus activity after 1 day BD, whereas only Baz1a neurons showed recovery after 5 days of BD ($P < 0.005$, Student's t test) (Fig. 3G and fig. S20).

Task encoding in inhibitory subclasses and subdivisions

We next compared task encoding in three of the major subclasses of inhibitory neurons (Pvalb, Sst, and Vip). Lamp5 neurons were excluded from analysis because of their low numbers captured in the data set (table S2). Overall, Pvalb neurons exhibited the weakest coding of tactile-related features (Fig. 4, A to C, and fig. S15). However, the high firing rates of Pvalb neurons and associated difficulties in reliably inferring spiking-related calcium events in this subclass by calcium imaging may underestimate the strength of GLM-derived task responses (supplementary text S2) (11, 33). We therefore focused our analysis on Sst and Vip neurons.

We investigated whether more task-related differences emerge when inhibitory subclasses are further divided into finer transcriptomic subclasses or types. Among inhibitory subclasses, Sst showed the best overall GLM fit ($P < 0.005$, Mann-Whitney U test) (Fig. 4A) and strongly encoded stimulus direction ($P < 0.05$, Mann-Whitney U test) (Fig. 4B). We asked whether stimulus direction was encoded similarly in two subdivisions of Sst neurons. Sst/$Chodl^+$ neurons express nitric oxide synthase ($Nos1$) (34), display long-range axonal projection patterns (35, 36), and are active during slow-wave sleep (37). During the DNMS task, Sst/$Chodl^+$ encoded direction more weakly compared with Sst/$Chodl^-$ neurons ($P < 5 \times 10^{-5}$, Mann Whitney U test) (Fig. 4, D and F, and fig. S16A).

We next compared task differences between Vip/$Pthlh^+$ and Vip/$Pthlh^-$ neurons (Fig. 4E).

Vip neurons belonging to the $Pthlh^+$ subdivision coexpress choline acetyltransferase ($Chat$) and calretinin, typically have bipolar morphologies, and preferentially target Sst neurons (36, 38–40). Vip neurons belonging to the $Pthlh^-$ subdivision coexpress synuclein gamma ($Sncg$) and cholecystokinin (Cck), have multipolar and basket cell morphologies, and preferentially target Pvalb neurons (36, 41). Vip/$Pthlh^-$ neurons more strongly encoded direction, sample$_{EARLY\ DELAY}$, and touch onset than did Vip/$Pthlh^+$ neurons (direction, $P < 0.05$; sample$_{EARLY\ DELAY}$, $P < 0.01$; onset, $P < 0.05$, Mann-Whitney U test) (Fig. 4, G to I, and fig. S16B). Analysis of calcium events with respect to touch onset at the beginning of the sample and test period neurons showed elevated firing for Vip/$Pthlh^+$ neurons preceding touch onset, which correlated with an anticipatory increase in whisking amplitude (Fig. 4J). This pretouch activity suggested that Vip/$Pthlh^+$ neurons are driven by free whisking behavior. To disentangle movement-related from tactile-related whisker responses, we fit neuronal activity to a GLM with whisker kinematic variables using only time periods before touch onset during the prestimulus and delay period. Vip/$Pthlh^+$ neurons more strongly encoded whisker amplitude, angle, and phase task factors during free whisking periods compared with those of Vip/$Pthlh^-$ neurons (amplitude, $P < 0.02$; angle, $P < 0.001$; phase, $P < 0.05$; Mann-Whitney U test) (Fig. 4, K to M).

Network interactions between major subclasses and types

The ability to simultaneously record across all identified cell subclasses and types enables a comprehensive characterization of cell type–specific network structures that underlie coding of task information. Non-negative matrix factorization across varying ranks captures population dynamics across distinct functional subpopulations (supplementary text S4) (26). Neurons that exhibit strong population coupling with increasing ranks suggest functional relationships with multiple subpopulations. Compared with other excitatory neurons, Baz1a neurons consistently showed higher coupling across ranks, indicating that they are highly integrated into the local L2/3 network ($P < 0.02$, $F_{2,6}$, repeated measures ANOVA) (Fig. 5A). To investigate coupling between specific cell-type populations, we constructed a GLM that included all previously described task variables while subdividing the activity of other neurons into different "coupling factors" according to Adamts2, Baz1a, Agmat, Pvalb, Sst, and Vip transcriptomic populations (Fig. 5B). For a modeled neuron, the ∆AIC for each cell-type coupling factor constituted a measure of "functional connectivity" between that neuron and other simultaneously recorded cell types. Functional connections consist of positive and negative noise correlations that reflect either direct

interactions or common input from nonrecorded neurons (fig. S21). From these measures, a directional weighted network graph can be constructed composed of the six subpopulations as nodes and functional connectivity as edges. To assess interactions between cell populations encoding different task factors, task-specific networks were generated by selecting for neurons with significant ∆AIC ($P < 0.01$, χ^2 test) for a given task factor (Fig. 5C).

We observed different network patterns across task factors. All task factor networks exhibited population-specific functional connection weights that were greater than chance (with the exception of the network that contained noncoding neurons, which exhibited random connection weights) ($P < 0.05$, bootstrap test) (Fig. 5D). Functional connectivity was strongest among neurons encoding category and whisker kinematics (fig. S22A). We further investigated the structure of these networks. For each cell-type node, we used the input edge strengths to determine how other cell populations influence the activity of the measured node and the output edge strengths to determine how the measured node influences the activity of other cell populations (Fig. 5E and fig. S22B). Inhibitory neurons were more likely than excitatory neurons to be influenced by network activity patterns. Sst neurons exhibited the highest input node strength across all task conditions ($P < 0.05$, bootstrap test). This is in line with evidence that suggests that Sst neurons follow local network activity (23). By contrast, excitatory neurons had a greater influence on other cell types, with Baz1a cells showing high output node strength in seven out of the nine task factor networks ($P < 0.05$, bootstrap test).

Given the differences in node strengths across task factor networks, we asked whether functional connectivity between any two subpopulations varied across task factor networks. High variability suggests that functional connections between cell types are dynamic and depend on the information being processed, whereas low variability suggests a stable motif that is intrinsic to the underlying circuitry. We measured the overall strength of each connection by calculating the mean edge weight across task factor networks. The stability of this connection was reported as the coefficient of variation of the edge weight across task factor networks (Fig. 5F). The majority of connections exhibited variability between task factor networks that were equivalent to chance levels, suggesting that connection strengths were dynamic and depend on the encoded task factor. However, a subset of connections [Adamts2→Vip, Adamts2→Sst, Agmat→Baz1a, and Baz1a↔Sst (output node→input node)] were consistently strong and stable across task factor networks, suggesting that they represent intrinsic functional motifs between

Fig. 4. Task encoding across L2/3 inhibitory subclasses. (**A** to **C**) Cumulative probability distributions for (A) full model deviance explained, (B) encoding strength of stimulus direction, and (C) encoding strength of whisker kinematics for three major inhibitory subclasses (Mann Whitney *U* Test). (**D** and **E**) Estimated event rate responses to preferred stimulus direction for (D) Sst subclasses and (E) Vip subclasses. (**F** and **G**) Cumulative probability distribution of ΔAIC for task factor encoding direction for (F) Sst subclasses and (G) Vip subclasses (Mann Whitney *U* Test). (**H** and **I**) Cumulative probability distribution of ΔAIC for (H) task factors encoding sample$_{EARLY\ DELAY}$ and (I) touch onset for Vip subclasses (Mann Whitney *U* Test). (**J**) Estimated event rate for Vip subclasses along with mean whisking amplitude aligned to whisker-rotor touch onset preceding sample and test periods. (**K** to **M**) Cumulative probability distribution of ΔAIC for task factors encoding (K) free whisking amplitude, (L) angle, and (M) phase for Vip subclasses (Mann Whitney *U* Test). In (B), (C), (F) to (I), and (K) to (M), solid and dotted lines indicate significant (*P* < 0.01) and nonsignificant encoding strengths, respectively, by means of χ^2 test. Shaded regions in (D) and (E) indicate SEM. *n* = 48 Pvalb cells, 47 Sst/*Chodl*$^+$ cells, 88 Sst/*Chodl*$^-$ cells, 40 Vip/*Pthlh*$^+$ cells, and 49 Vip/*Pthlh*$^-$ cells from seven animals.
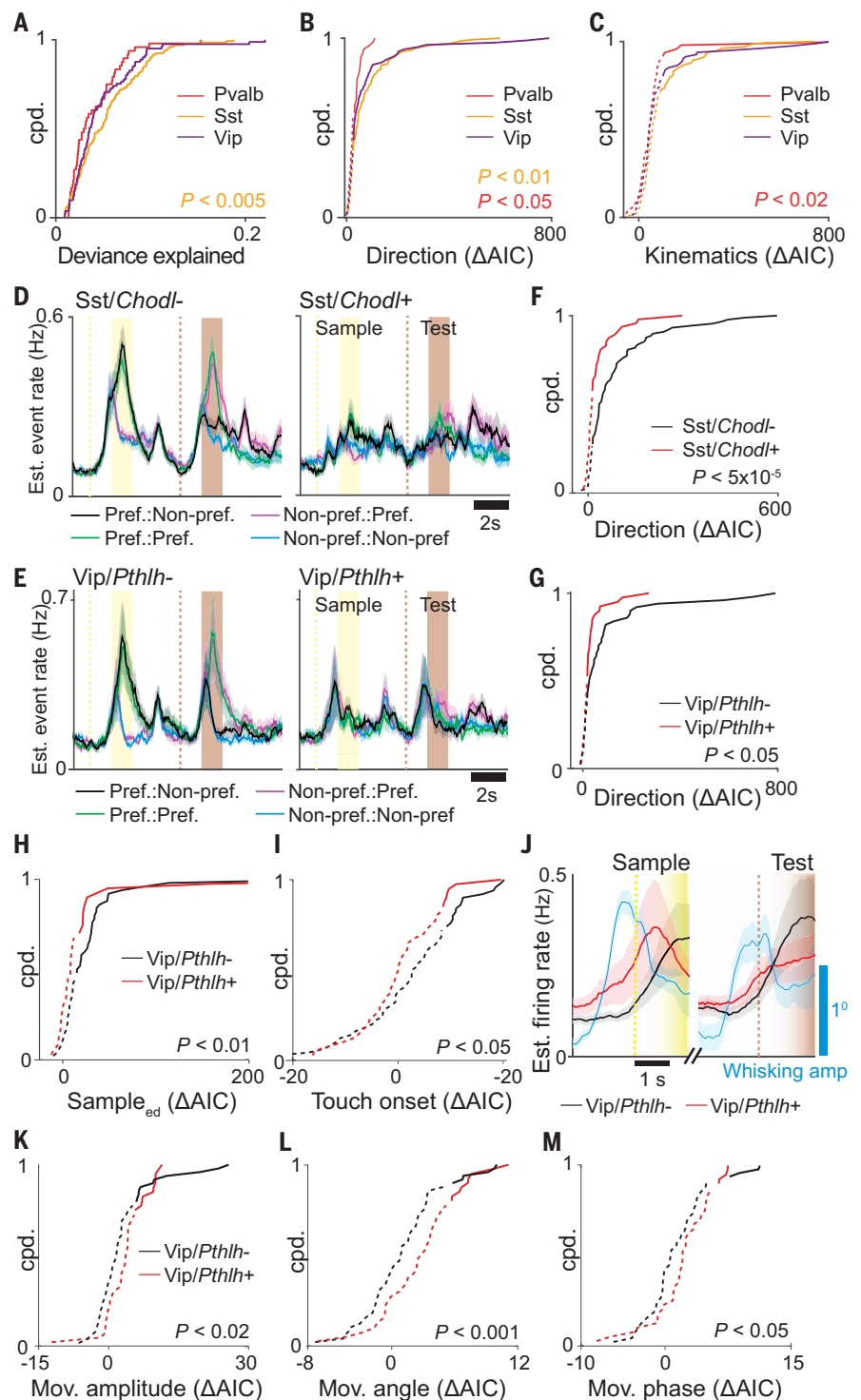
specific cell populations (*P* < 0.05, permutation test).

Cell type–specific tracing confirms intrinsic functional connectivity

The observed functional connections that persisted across task networks could be explained by cell type–specific synaptic connections. Trans-monosynaptic rabies tracing enables input mapping to specific cell types but requires genetic access for conditional infection. Because transgenic lines for the three excitatory cell types are not available, we focused on input patterns to Sst and Vip inhibitory classes. Because Baz1a neurons showed stable functional connectivity with Sst neurons but not Vip neurons, we compared Baz1a synaptic connectivity between these two inhibitory classes. Using Sst-IRES-Cre (*n* = 4) and Vip-IRES-Cre (*n* = 4) mice (*42*), L2/3 Sst and Vip starter cells were first labeled by using a Cre-dependent adeno-associated virus (AAV) expressing TVA, CVS-N2cG, and dTomato, followed by delivery of the EnvA-pseudotyped CVS-N2c(ΔG) rabies virus expressing histone-GFP (Fig. 6, A and B) (*43*). We examined the sublaminar distribution of histone-GFP–positive inputs (nGFP$^+$)
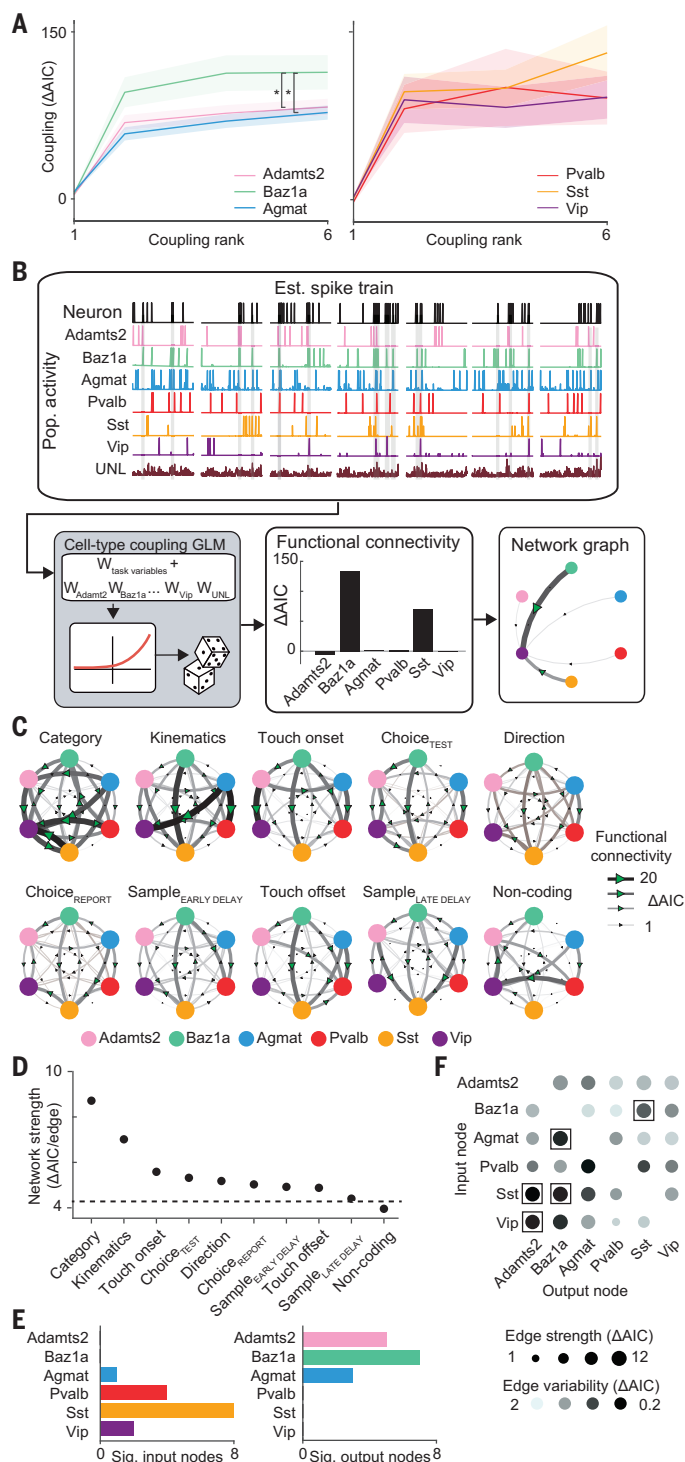
Fig. 5. Cell type functional connectivity across task networks. (**A**) Strength of coupling factor encoding across varying coupling ranks. *$P < 0.02$, repeated measures ANOVA test, $F_{2,6}$. (**B**) Schematic of network analysis for example neuron. (**C**) Task-specific networks generated by selecting for neurons with significant encoding for a given task factor in the task GLM. Networks are sorted according to average edge strength. (**D**) Network strength across task networks. The dotted line indicates strength of shuffled network. (**E**) Cell type and subclass differences in number of significant (left) input and (right) output nodes across task networks. (**F**) Strength and variability of functional connectivity in network edges across task networks. Network edges with significantly high strength and low variability are indicated with a box. $P < 0.05$, permutation test. Shaded region in (A) indicates SEM. n = 1996 neurons, direction; 1374 neurons, sample_EARLY DELAY; 1076 neurons sample_LATE DELAY; 360 neurons, category; 623 neurons, choice_TEST; 830 neurons, choice_REPORT; 898 neurons, touch onset; 1033 neurons, touch offset; 864 neurons, kinematics; and 273 neurons, noncoding from seven animals.

to Sst and Vip cells across L2/3. Overall, Sst and Vip neurons received a greater number of inputs from cells located in deeper L2/3 (>200 µm below the pia mater) as compared with superficial L2/3. However, Sst neurons received more of their inputs from superficial L2/3 neurons compared with Vip neurons (Sst, 29.1 ± 2.7%; Vip, 21.2 ± 2.6%; mean ± SEM; $P < 0.05$, one-tailed Student's t test) (Fig. 6C). We performed multiplexed HCR-FISH to identify cell type of nGFP+ input neurons. The overall density of Baz1a inputs was greater for Sst neurons as compared with Vip neurons (Sst, 12.8 ± 1.8%; Vip, 8.4 ± 1.2%; $P < 0.05$; one-tailed Student's t test) (Fig. 6D). This difference was greatest among cells in superficial L2/3 (Sst, 22.1 ± 0.9%; Vip, 17.3 ± 2.0%; $P < 0.05$, one-tailed Student's t test) (Fig. 6E).

Discussion

We developed a platform to densely survey the functional and molecular properties of neuronal populations in vivo and applied it to study cell types in L2/3 of S1. We found evidence for increasing functional specialization; excitatory and inhibitory neurons are divided into more discrete subclasses and types. We focused on the role of Baz1a neurons in neocortical function. Enriched *Fos* expression suggests that Baz1a neurons are members of a previously described, highly interconnected FosGFP population that operates as a network hub in S1 (*27*). S1 is important for both tactile feature discrimination as well as sensorimotor integration for object localization (*44*). Given their highly selective stimulus responsiveness, Baz1a neurons are well poised to act as feature detectors and recruit local circuits for sensory processing. Superficial L2/3 pyramidal neurons are laminarly situated to integrate both top-down motor and associative signals arriving in L1 onto apical dendrites with bottom-up sensory signal arriving from L4 onto basal dendrites (Fig. 6F) (*45*, *46*). Basal dendrites also contain highly recurrent, lateral connections between neighboring excitatory neurons (*47*). Integration of top-down signals and associative memory formation in L2/3 pyramidal neurons is mediated by Vip disinhibition (*40*, *48*) through apical dendrite-targeting Sst neurons (*49*). We propose that excitatory connections from superficial Baz1a neurons onto Sst neurons can counteract Vip-mediated disinhibition so as to inhibit top-down inputs and bias synaptic integration in local L2/3 pyramidal neurons toward bottom-up and recurrent inputs on basal dendrites. Therefore, these circuit motifs operate complementary to one another, allowing S1 to shift between gating long-range feedback inputs and engaging feed-forward computations.

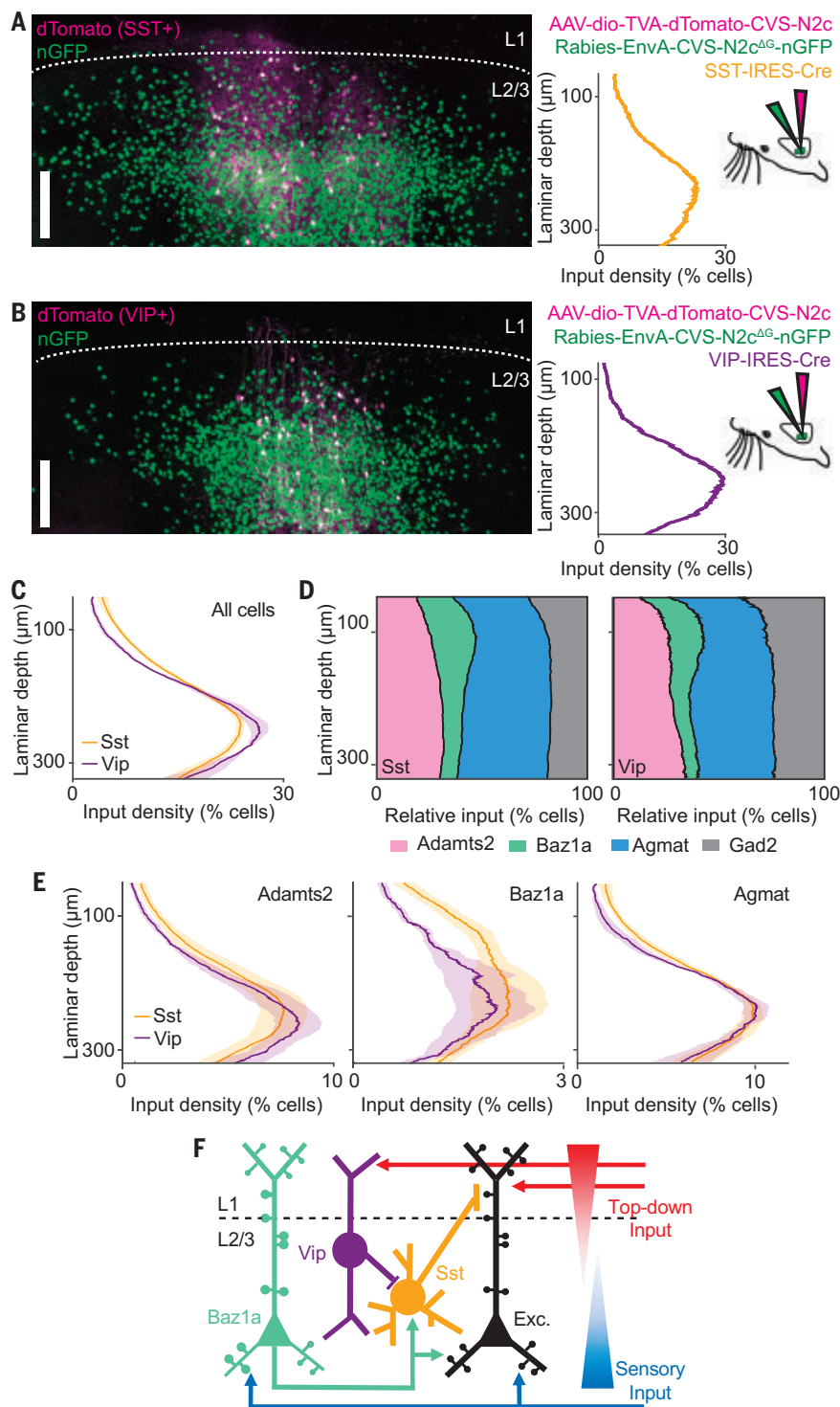Baz1a neurons are also functionally distinct in their ability to adapt to altered sensory

Fig. 6. Upper layer Ba1za neurons target Sst neurons. (**A** and **B**) Example of cell type–specific trans-monosynaptic tracing in (A) Sst-IRES-Cre and (B) Vip-IRES-Cre mice. (Left) Confocal images of starter cells (magenta) and nGFP⁺ input neurons (green). (Right) Sublaminar distribution of input density from left images, along with injection scheme. (**C**) Average sublaminar somatic density distribution of inputs across L2/3 for Sst and VIP neurons. (**D**) Relative proportion of excitatory cell types and *Gad2*⁺ inhibitory neurons as a function of laminar depth for Sst and Vip input neurons. (**E**) Density of excitatory cell types as a function of laminar depth for Sst and Vip input neurons. (**F**) Circuit model of L2/3 illustrating cell type–specific connectivity between Vip, Sst, Baz1a, and other local excitatory neurons. Shaded regions in (C) and (E) indicate SEM. *n* = 4 Sst-IRES-Cre animals, 16 slices, 33,957 neurons; and 4 Vip-IRES-Cre animals, 14 slices, 35,926 neurons. Scale bars, 100 μm.

experience by homeostatically maintaining their response to tactile stimuli. Sensory deprivation transiently induces changes in IEG expression, resulting in experience-dependent plasticity (*50*). We speculate that stable expression of *Fos* and other select IEGs in Baz1a cells primes this cell type to adapt to changes in experience through molecular mechanisms that could modulate excitatory-inhibitory balance, synaptic scaling, or intrinsic excitability. This plasticity suggests that Baz1a neurons serve additional roles in preserving existing sensory representations in the face of novel experiences. The presence of cell types in V1 and ALM with similar expression profiles as that of Baz1a neurons suggests that homologous circuits with common functions may exist across neocortical areas.

REFERENCES AND NOTES

1. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018). doi: 10.1038/s41586-018-0654-5; pmid: 30382198
2. A. Zeisel *et al.*, Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018). doi: 10.1016/j.cell.2018.06.021; pmid: 30096314
3. E. Klingler *et al.*, A translaminar genetic logic for the circuit identity of intracortically projecting neurons. *Curr. Biol.* **29**, 332–339.e5 (2019). doi: 10.1016/j.cub.2018.11.071; pmid: 30639110
4. L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, J. D. Macklis, Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013). doi: 10.1038/nrn3586; pmid: 24105342
5. L. Lim, D. Mi, A. Llorca, O. Marín, Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018). doi: 10.1016/j.neuron.2018.10.009; pmid: 30359598
6. S. A. Sorensen *et al.*, Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. *Cereb. Cortex* **25**, 433–449 (2015). doi: 10.1093/cercor/bht243; pmid: 24014670
7. A. Paul *et al.*, Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171**, 522–539.e20 (2017). doi: 10.1016/j.cell.2017.08.032; pmid: 28942923
8. A. Kepecs, G. Fishell, Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014). doi: 10.1038/nature12983; pmid: 24429630
9. K. D. Harris, G. M. Shepherd, The neocortical circuit: Themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015). doi: 10.1038/nn.3917; pmid: 25622573
10. T. L. Daigle *et al.*, A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480.e22 (2018). doi: 10.1016/j.cell.2018.06.035; pmid: 30007418
11. A. G. Khan *et al.*, Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat. Neurosci.* **21**, 851–859 (2018). doi: 10.1038/s41593-018-0143-z; pmid: 29786081
12. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi: 10.1126/science.aaa6090; pmid: 25858977
13. H. M. T. Choi *et al.*, Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018). doi: 10.1242/dev.165753; pmid: 29945988
14. S. Shah *et al.*, Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* **143**, 2862–2867 (2016). doi: 10.1242/dev.138560; pmid: 27342713
15. E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014). doi: 10.1038/nmeth.2892; pmid: 24681720
16. X. Wang *et al.*, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018). doi: 10.1126/science.aat5691; pmid: 29930089

17. P. R. Nicovich et al., Multimodal cell type correspondence by intersectional mFISH in intact tissues. bioRxiv [Preprint] 525451 (2019). doi: 10.1101/525451

18. J. L. Chen, F. F. Voigt, M. Javadzadeh, R. Krueppel, F. Helmchen, Long-range population dynamics of anatomically defined neocortical networks. eLife 5, e14679 (2016). doi: 10.7554/eLife.14679; pmid: 27218452

19. J. B. Treweek et al., Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. Nat. Protoc. 10, 1860–1896 (2015). doi: 10.1038/nprot.2015.122; pmid: 26492141

20. M. Ohkura, T. Sasaki, C. Kobayashi, Y. Ikegaya, J. Nakai, An improved genetically encoded red fluorescent Ca^{2+} indicator for detecting optically evoked action potentials. PLOS ONE 7, e39933 (2012). doi: 10.1371/journal.pone.0039933; pmid: 22808076

21. Z. Yao et al., A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell 184, 3222–3241.e26 (2021).

22. E. Abs et al., Learning-related plasticity in dendrite-targeting layer 1 interneurons. Neuron 100, 684–699.e6 (2018). doi: 10.1016/j.neuron.2018.09.001; pmid: 30269988

23. J. Yu, H. Hu, A. Agmon, K. Svoboda, Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior. Neuron 104, 412–427.e4 (2019). doi: 10.1016/j.neuron.2019.07.027; pmid: 31466734

24. C. Condylis et al., Context-dependent sensory processing across primary and secondary somatosensory cortex. Neuron 106, 515–525.e5 (2020). doi: 10.1016/j.neuron.2020.02.004; pmid: 32164873

25. J. W. Pillow et al., Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999 (2008). doi: 10.1038/nature07140; pmid: 18650810

26. C. A. Runyan, E. Piasini, S. Panzeri, C. D. Harvey, Distinct timescales of population coding across cortex. Nature 548, 92–96 (2017). doi: 10.1038/nature23020; pmid: 28723889

27. L. Yassin et al., An embedded subnetwork of highly active neurons in the neocortex. Neuron 68, 1043–1050 (2010). doi: 10.1016/j.neuron.2010.11.029; pmid: 21172607

28. J.-S. Jouhanneau et al., Cortical fosGFP expression reveals broad receptive field excitatory neurons targeted by POm. Neuron 84, 1065–1078 (2014). doi: 10.1016/j.neuron.2014.10.014; pmid: 25453844

29. E. L. Yap, M. E. Greenberg, Activity-regulated transcription: Bridging the gap between neural activity and behavior. Neuron 100, 330–348 (2018). doi: 10.1016/j.neuron.2018.10.013; pmid: 30359600

30. A. L. Barth, R. C. Gerkin, K. L. Dean, Alteration of neuronal firing properties after in vivo experience in a FosGFP transgenic mouse. J. Neurosci. 24, 6466–6475 (2004). doi: 10.1523/JNEUROSCI.4737-03.2004; pmid: 15269256

31. D. J. Margolis et al., Reorganization of cortical population activity imaged throughout long-term sensory deprivation. Nat. Neurosci. 15, 1539–1546 (2012). doi: 10.1038/nn.3240; pmid: 23086335

32. M. A. Gainey, D. E. Feldman, Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex. Philos. Trans. R. Soc. London B Biol. Sci. 372, 20160157 (2017). doi: 10.1098/rstb.2016.0157; pmid: 28093551

33. A. C. Kwan, Y. Dan, Dissection of cortical microcircuits by single-neuron stimulation in vivo. Curr. Biol. 22, 1459–1467 (2012). doi: 10.1016/j.cub.2012.06.007; pmid: 22748320

34. B. Tasic et al., Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. 19, 335–346 (2016). doi: 10.1038/nn.4216; pmid: 26727548

35. R. Tomioka et al., Demonstration of long-range GABAergic connections distributed throughout the mouse neocortex. Eur. J. Neurosci. 21, 1587–1600 (2005). doi: 10.1111/j.1460-9568.2005.03989.x; pmid: 15845086

36. M. He et al., Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. Neuron 91, 1228–1243 (2016). doi: 10.1016/j.neuron.2016.08.021; pmid: 27618674

37. D. Gerashchenko et al., Identification of a population of sleep-active cerebral cortex neurons. Proc. Natl. Acad. Sci. U.S.A. 105, 10227–10232 (2008). doi: 10.1073/pnas.0803125105; pmid: 18645184

38. A. Dudai et al., Barrel cortex VIP/ChAT interneurons suppress sensory responses in vivo. PLOS Biol. 18, e3000613 (2020). doi: 10.1371/journal.pbio.3000613; pmid: 32027647

39. A. Prönneke et al., Characterizing VIP neurons in the barrel cortex of VIPcre/tdTomato mice reveals layer-specific differences. Cereb. Cortex 25, 4854–4868 (2015). doi: 10.1093/cercor/bhv202; pmid: 26420784

40. S. Lee, I. Kruglikov, Z. J. Huang, G. Fishell, B. Rudy, A disinhibitory circuit mediates motor integration in the somatosensory cortex. Nat. Neurosci. 16, 1662–1670 (2013). doi: 10.1038/nn.3544; pmid: 24097044

41. H. Hioki et al., Preferential inputs from cholecystokinin-positive neurons to the somatic compartment of parvalbumin-expressing neurons in the mouse primary somatosensory cortex. Brain Res. 1695, 18–30 (2018). doi: 10.1016/j.brainres.2018.05.029; pmid: 29792869

42. H. Taniguchi et al., A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. Neuron 71, 995–1013 (2011). doi: 10.1016/j.neuron.2011.07.026; pmid: 21943598

43. T. R. Reardon et al., Rabies virus CVS-N2c(ΔG) strain enhances retrograde synaptic transfer and neuronal viability. Neuron 89, 711–724 (2016). doi: 10.1016/j.neuron.2016.01.004; pmid: 26804990

44. M. E. Diamond, M. von Heimendahl, P. M. Knutsen, D. Kleinfeld, E. Ahissar, 'Where' and 'what' in the whisker sensorimotor system. Nat. Rev. Neurosci. 9, 601–612 (2008). doi: 10.1038/nrn2411; pmid: 18641667

45. N. L. Xu et al., Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature 492, 247–251 (2012). doi: 10.1038/nature11601; pmid: 23143335

46. G. Doron et al., Perirhinal input to neocortical layer 1 controls learning. Science 370, eaaz3136 (2020). doi: 10.1126/science.aaz3136; pmid: 33335033

47. N. Spruston, Pyramidal neurons: Dendritic structure and synaptic integration. Nat. Rev. Neurosci. 9, 206–221 (2008). doi: 10.1038/nrn2286; pmid: 18270515

48. L. E. Williams, A. Holtmaat, Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition. Neuron 101, 91–102.e4 (2019). doi: 10.1016/j.neuron.2018.10.049; pmid: 30472077

49. Y. Wang et al., Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat. J. Physiol. 561, 65–90 (2004). doi: 10.1113/jphysiol.2004.073353; pmid: 15331670

50. S. Loebrich, E. Nedivi, The function of activity-regulated genes in the nervous system. Physiol. Rev. 89, 1079–1103 (2009). doi: 10.1152/physrev.00013.2009; pmid: 19789377

51. C. Condylis et al., Dense functional and molecular readout of a circuit hub in sensory cortex. G-Node (2021); doi: 10.12751/g-node.7q0lz0.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl5981
Materials and Methods
Supplementary Text
Figs. S1 to S22
References (52–70)
Tables S1 and S2
Movie S1
MDAR Reproducibility Checklist

Where Science Gets Social.

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY ◢AAAS

NEUROSCIENCE

A neuronal mechanism for motivational control of behavior

J. Courtin†*, Y. Bitterman†, S. Müller, J. Hinz, K. M. Hagihara, C. Müller, A. Lüthi*

INTRODUCTION: Acting to achieve goals underlies animals' behavioral flexibility and depends on the ability to motivate specific behaviors. Formal studies of goal-directed behavior are based on paradigms such as instrumental conditioning, in which subjects learn that specific actions result in a particular outcome. By definition, instrumental goal-directed actions are oriented toward specific outcomes and are sensitive to variations in outcome value and action-outcome contingency. However, the neuronal mechanisms that encode and maintain specific motivational control of self-paced behavior are poorly understood.

RATIONALE: Animal motivation to engage in goal-directed behavior is governed by a multitude of factors, ranging from the animal's general metabolic state to specific task parameters including outcome identity, outcome value,

and action-outcome contingency. Using Ca^{2+} imaging, electrophysiology, and optogenetic manipulations, we examined the relationship between noncued, self-paced goal-directed actions and neuronal responses in the amygdala. Although the amygdala is a brain structure well known for encoding associative learning in the context of cue-triggered outcome-seeking behaviors, here, we specifically focused on neuronal correlates of essential parameters supporting goal-directed actions, both during action execution and reward consumption.

RESULTS: Mice were trained in a self-paced, instrumental goal-directed task in which one action led to the delivery of a sucrose reward and another action to the delivery of a milk reward, with no explicit stimuli signaling trial start or reward availability. Based on an automatically derived description

of mouse behavior, we showed that after 5 days of training, mice exhibited stereotypical self-paced behavioral sequences, alternating between periods of action execution when mice pursued a goal and reward consumption when the goal was attained.

Using in vivo deep-brain Ca^{2+} imaging and electrophysiology, we found distinct outcome-specific ensembles of amygdala neurons activated during action-consumption sequences. These ensembles were spatially intermingled and broadcasted behaviorally relevant information to multiple striatum subregions. At the population level, distinct amygdala neuronal activity patterns tiled the entire action-consumption behavioral sequences. These internally generated action- and consumption-associated patterns maintained outcome-specific information along the session and across days.

Optogenetic inhibition of amygdala neurons during either action or consumption caused distinct behavioral phenotypes and indicated that action- and consumption-associated amygdala activity was necessary for the outcome-specific motivational control of goal-directed behavior.

Last, in agreement with experimental psychology models, we showed that action and consumption representations learned during training can be retrieved, truncated, or remapped upon manipulations of outcome value, such as after outcome devaluation or extinction or upon degradation of action-outcome contingency. More specifically, whereas outcome devaluation resulted in the immediate emergence of new action and outcome representations, action-outcome contingency degradation resulted in a gradual loss of the action representation with no effect on reward consumption representation.

CONCLUSION: We show here that the interplay between distinct outcome-specific and updatable neuronal activity patterns encodes and maintains specific motivational states necessary for the execution of adaptive, self-initiated action-consumption sequences. Amygdala neurons segregated behaviorally relevant information along different axes, such as outcome identity, value, and expectancy. Our study thus introduces the concept of specific motivation and extends the classic view that the amygdala controls outcome-seeking behavior by conferring value to outcome-predicting sensory stimuli. ∎

A Encoding of goal-directed behavior
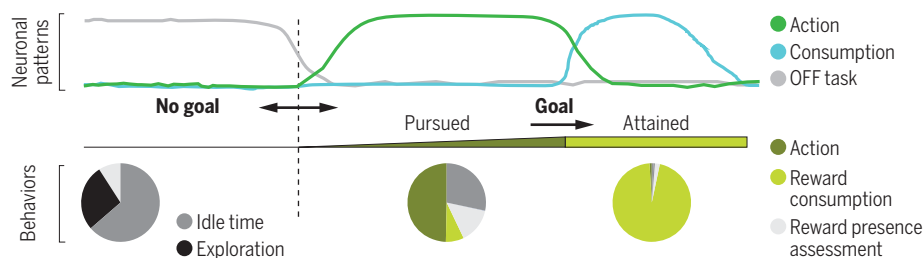
B Outcome-specific and updatable motivational state encoding
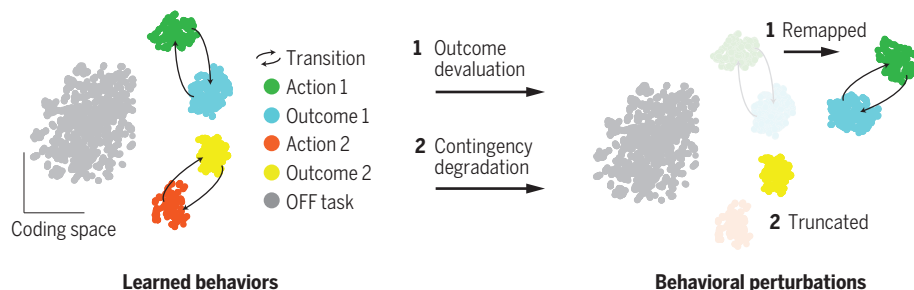
A neuronal mechanism for noncued goal-directed behavior. (**A**) The interplay between distinct neuronal activity patterns encodes and maintains behaviorally relevant information at distinct time points along goal-directed action-consumption behavioral sequences. (**B**) Outcome-specific action and consumption patterns reflect the maintenance of a specific motivational state. Upon behavioral perturbations, action and consumption patterns differentially adapt. This adaptive code might be key to direct behavior toward a pursued goal.

NEUROSCIENCE

A neuronal mechanism for motivational control of behavior

J. Courtin[1]*†, Y. Bitterman[1]†, S. Müller[1], J. Hinz[1,2], K. M. Hagihara[1,2], C. Müller[1], A. Lüthi[1,2]*

Acting to achieve goals depends on the ability to motivate specific behaviors based on their predicted consequences given an individual's internal state. However, the underlying neuronal mechanisms that encode and maintain such specific motivational control of behavior are poorly understood. Here, we used Ca²⁺ imaging and optogenetic manipulations in the basolateral amygdala of freely moving mice performing noncued, self-paced instrumental goal-directed actions to receive and consume rewards. We found that distinct neuronal activity patterns sequentially represent the entire action-consumption behavioral sequence. Whereas action-associated patterns integrated the identity, value, and expectancy of pursued goals, consumption-associated patterns reflected the identity and value of experienced outcomes. Thus, the interplay between these patterns allows the maintenance of specific motivational states necessary to adaptively direct behavior toward prospective rewards.

Maintaining motivational states allows animals to direct behavior to achieve desired goals (*1*). Formal studies of goal-directed behavior are based on paradigms such as instrumental conditioning, in which the subject learns through experience that specific actions result in a particular outcome. By definition, instrumental goal-directed actions are oriented toward specific outcomes and are sensitive to variations in outcome value and action-outcome contingency (*2*). However, current knowledge about the encoding of these parameters in the basolateral amygdala (BLA) is based on the use of cue-triggered outcome-seeking behaviors (*3–9*). Evidence of BLA involvement in noncued, self-paced instrumental goal-directed actions is mainly provided by system-level studies (*10–15*). Such studies suggest that the BLA confers specific motivational significance to goal-directed actions. However, the neuronal mechanisms encoding and maintaining such specific motivational states remain unknown.

Results

Self-paced action-consumption behavioral sequences

We trained food-restricted male mice in a self-paced instrumental goal-directed task in which one of two actions led to the delivery of a sucrose reward and the other to the delivery of a milk reward, with no explicit stimuli signaling trial start or reward availability. The number of actions required to obtain rewards increased over 5 training days on a variable ratio schedule

[1]Friedrich Miescher Institute for Biomedical Research, CH-4058 Basel, Switzerland. [2]University of Basel, CH-4000 Basel, Switzerland.
*Corresponding author. Email: julien.courtin@fmi.ch (J.C.); andreas.luthi@fmi.ch (A.L.)
†These authors contributed equally to this work.

of reinforcement [from constant reinforcement to variable ratio 5 (VR5); Fig. 1A]. During training, both instrumental actions were acquired at a similar rate (Fig. 1B). By day 5, performance of the actions was both hunger state dependent and goal directed, as demonstrated by its sensitivity to a change in hunger state, to specific satiety-induced outcome devaluation, and to action-outcome contingency degradation (Fig. 1, B and C). During each training session, mice were sequentially exposed to phases during which one or both instrumental actions was available or not (ON and OFF task phases, respectively). Using action and lick time stamps together with video tracking, we automatically derived a continuous description of mouse behavior during the entire session (Fig. 1D and fig. S1, A and B). We differentiated epochs of task-related behaviors such as instrumental actions, unrewarded lick and rewarded lick are epochs from epochs of non-task-related behaviors including idle times, when mice were in the task zone but did not perform any task-related behaviors, and context exploration idle times and context exploration are epochs, when mice ventured away from the task-related zone. Across training sessions, behaviors during ON task phases were increasingly composed of task-related behaviors. By contrast, behaviors during OFF task phases were stable across days (Fig. 1E and fig. S1, C and D).

By day 5, during ON task phases, mice repeatedly alternated between action and consumption periods. We thus defined a behavioral sequence as the time window between the first action of an action period and the last lick of the subsequent consumption period. Action periods, defined as the time from the first action to the last action of a behavioral sequence, included epochs of actions, unrewarded licks, and intermittent non-task-related behaviors,

whereas consumption periods were exclusively composed of rewarded lick epochs (Fig. 1F). These action-consumption sequences were self-paced, as made evident by the occurrence of intermittent unrewarded lick epochs and the occasional performance of actions even after outcome delivery (fig. S2A). As described classically for variable ratio reinforcement schedules, mice showed constant action response rates with brief pauses after consumption periods (*16*). Action rate, duration of the distinct behavioral epochs, action to lick latency, and interbehavioral sequence intervals were stable across the session (Fig. 1, F and G, and fig. S2, B and C).

Specific activation of distinct BLA ensembles during action and consumption

To evaluate the neuronal correlates of goal-directed behavior in the BLA, we performed deep-brain Ca²⁺ imaging of BLA Ca²⁺/calmodulin-dependent protein kinase 2 (CaMK2)–expressing principal neurons (PNs) using a miniaturized microscope and GCaMP6f as a Ca²⁺ indicator. The activity of single BLA PNs was extracted using constrained non-negative matrix factorization for microendoscopic data (CNMF-E) (see the materials and methods and fig. S3A; 113 ± 12 neurons per mouse, 905 total neurons from eight mice). On day 5, neuronal activity showed clear time-locked modulations to the different behavioral epochs during ON task phases (Fig. 1H). Across multiple action-consumption sequences, single neuron responses faithfully tracked the stereotypical behavioral switches between action and consumption periods (Fig. 2A). We therefore defined three functional types of BLA PNs. Action neurons (n = 137/905) displayed increased activity when mice were engaged in the performance of goal-directed actions (Fig. 2, A and B, and fig. S4, A to D). By contrast, consumption neurons (n = 170/905) exhibited increased activity time locked to rewarded lick epochs, when mice collected the reward. Finally, transition neurons (n = 39/905) showed increased activity during both action epochs and unrewarded lick or rewarded lick epochs (Fig. 2, A and B, and fig. S4, A to D). The three functional types of neurons were spatially intermingled, without clear anatomical organization either within each individual imaging field or along the anteroposterior (AP) or mediolateral (ML) BLA axes (fig. S3, B to F).

Of all BLA PNs recorded on day 5, 38% were significantly task modulated (n = 346/905 neurons from eight mice). In all three functional subsets, most neurons were outcome specific. That is, they discriminated between actions leading to, or the consumption of, a sucrose or milk reward. A small, but above chance level, proportion of neurons was not outcome specific (responding to milk and sucrose; $P < 0.001$, χ^2 test; Fig. 2C). Even though
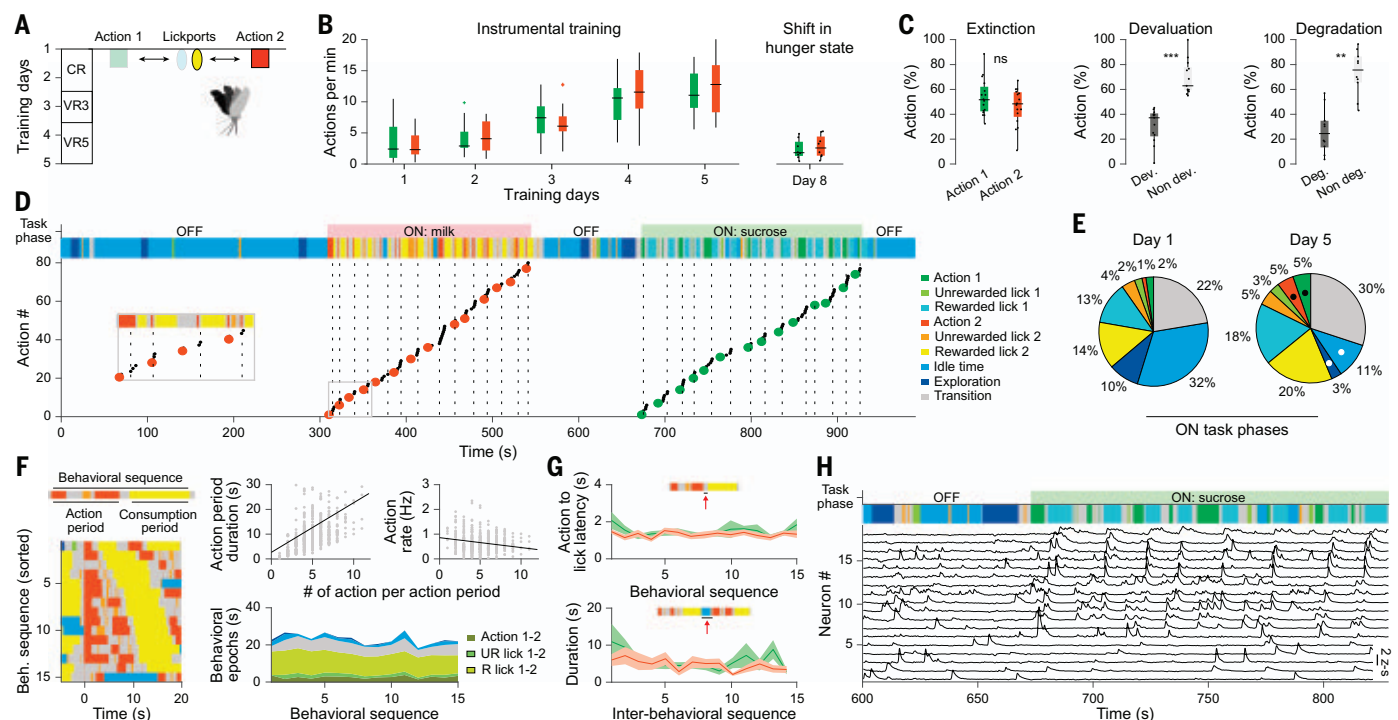
Fig. 1. Ca^{2+} imaging of BLA PNs during the learned, self-paced, instrumental goal-directed task. (**A**) Schematic of the behavioral context. Action availability is controlled and the required number of actions per outcome increases over training from CR to VR5. (**B**) Number of actions per minute over training (N = 16 mice in two cohorts) and during the free choice rewarded session in sated mice (N = 8 mice in two cohorts). (**C**) Action choice during the extinction test, satiety-induced outcome devaluation test, and action-outcome contingency degradation test (all tests were non-reinforced from two mice cohorts; N = 16, N = 13, and N = 10 mice, P = 0.72, P < 0.001, and P < 0.01, respectively; two-sided paired t tests comparing action choice). Box-and-whisker plots indicate median, interquartile, extreme data values, and outliers of the data distribution. (**D**) Task phases and behavioral epochs labeling during day 5 for one mouse. Black dots indicate individual actions; red and green dots, action periods onset (for milk and sucrose, respectively); vertical dashed lines, outcome delivery. Inset, initial behavioral sequences. Colors indicate the different behavioral epochs. (**E**) The proportion of time mice engaged in the different behavioral epochs during ON task phases on training days 1 and 5 (N = 16 mice; black or white dots denote significant increase or decrease in behavioral epochs duration, respectively; P < 0.01, two-sided Wilcoxon signed-rank tests comparing day 1 and 5 epoch durations). (**F**) Behavioral sequences aligned to action period onset (0 s) and sorted based on action period duration (left) while milk outcome was available. Duration of action periods (r = 0.21; P < 0.001) and action rate within action period (r = −0.07; P = 0.1) across action periods (top right). Mean duration of the distinct behavioral epochs (bottom right, N = 16 mice, day 5). (**G**) Action to lick latency and duration of the interbehavioral sequence intervals across behavioral sequences (N = 16 mice, day 5). Shading indicates SEM. (**H**) Ca^{2+} traces from 18 BLA CaMK2-expressing neurons simultaneously imaged during 4 min of day 5 [same mouse as in (D)]. Scale bar, 1 z-score. Colors at the top panels denote the different behaviors as in (D).

neuronal representations were outcome specific, behaviors of the two ON task phases were similar, suggesting comparable outcome values. Action and consumption neurons were oppositely modulated during ON task phases: Neurons activated during action were silent during consumption and vice versa (Fig. 2B). Accordingly, pairwise correlation and coactivation analysis of neuronal activity of action and/or consumption neurons indicated that coincident activity within each functional subset increased and coincident activity across functional subsets decreased during ON task phases (Fig. 2D). These patterns of coincidence were limited to the ON task phase, suggesting task-dependent functional interactions (fig. S4, E and F). We confirmed the results obtained with Ca^{2+} imaging using single-unit recordings (n = 78 neurons from four mice). Similar to the imaging results, 28% of all recorded neurons exhibited significant task-related modulation

(P = 0.10 and χ^2 = 2.7, Friedman test). Correlation and co-firing analysis of pairwise neuronal activity between action and/or consumption neurons confirmed task-dependent functional interactions (fig. S5).

Because the activity of action and consumption neurons was outcome specific, we tested the performance of a linear decoder trained with the imaging dataset on the five-way distinction among action periods, consumption periods (milk or sucrose), and periods of non-task-related behavior during the OFF task phase. Decoder performance was high when trained with activity traces of all PNs (F1 score = 87.8 ± 3.2%) and dropped to chance levels when trained on shuffled neuronal data (19.7 ± 0.1%, P < 0.01, paired t test; Fig. 2F and fig. S4G). When subsets of PNs were selected, performance was higher for training with the activity of all task-modulated neurons (69.7 ± 6.2%) than for training with random selec-

tions of non-task-modulated neurons (57.3 ± 5.1%; see the materials and methods).

BLA activity patterns tile the entire action-consumption sequences

To investigate population-level representational principles and contrast them with our observations at the level of single neurons, we obtained population activity vectors by binning single neuron activity in 200-ms bins. Population activity vectors time locked to the beginning of action and consumption epochs showed regularity and specificity (pairwise Pearson's correlations action-action, r = 0.24 ± 0.02; consumption-consumption, r = 0.52 ± 0.02; action-consumption, r = −0.004 ± 0.01 for this example mouse; Fig. 3A). Accordingly, pairwise correlations between population activity vectors along the entire behavioral session mirrored both the different task phases and the behavioral epochs in action-outcome
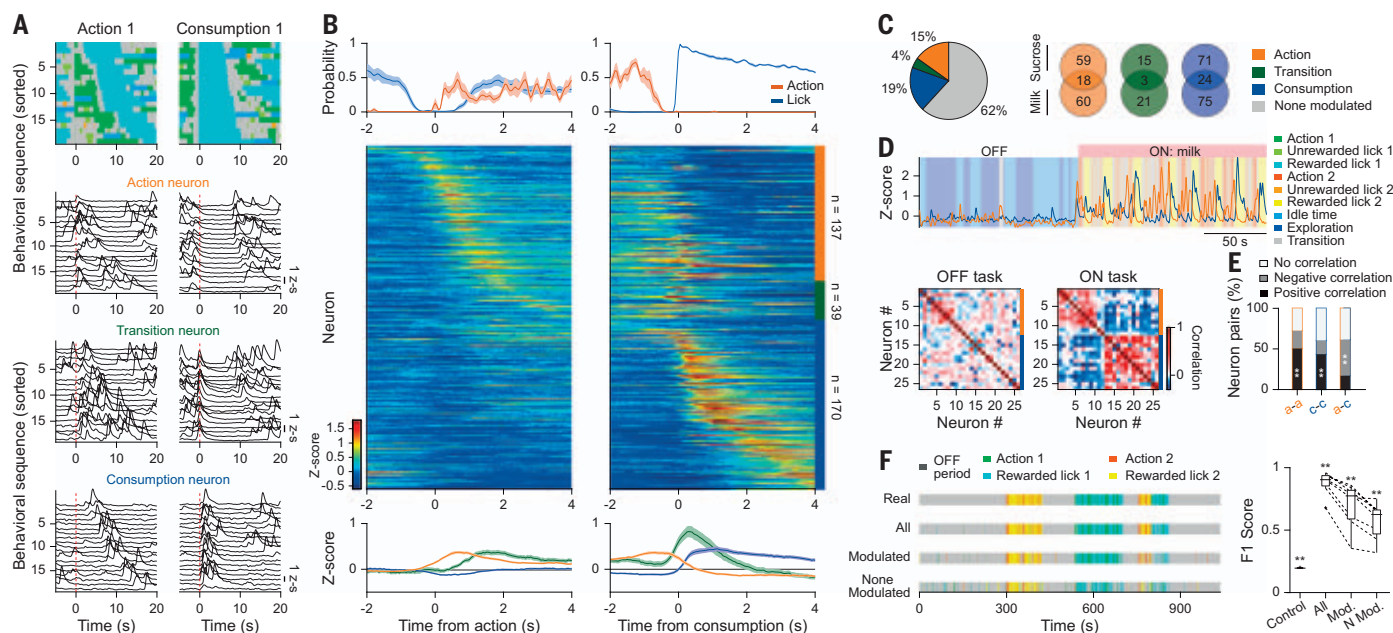
Fig. 2. Distinct sets of BLA PNs encode specific behavioral epochs.
(**A**) Mouse behavior (top) and corresponding activity of three neurons (bottom) during 20 behavioral sequences when sucrose outcome was available. The behavioral sequences were aligned to action period onset (left) or to consumption period onset (right) and sorted by action or consumption period duration [colors at the top panels denote behavioral epochs as in (D)]. (**B**) Action and lick probability (top, around action and consumption periods onsets), individual neuron response of all task-modulated neurons (middle), and average response by functional group (bottom; $n = 137$ action, $n = 39$ transition, and $n = 170$ consumption neurons) aligned to action periods (left) and consumption periods (right) onsets ($N = 8$ mice in two cohorts, $n = 905$ neurons recorded on day 5). Shading indicates SEM. (**C**) Proportion (left) and outcome selectivity (right) of the task-modulated neurons. (**D**) Average activity of 10 action neurons (orange) and 16 consumption neurons (blue) during OFF and ON phases (top)

aligned to the mouse behavior (background colors). Also shown is the pairwise correlation of activity for these neurons during the OFF and ON task phases (bottom). (**E**) Percentage of correlated neuron pairs during ON task phases on day 5 for action-action (a-a), consumption-consumption (c-c), and action-consumption (a-c) neuron pairs (**$P < 0.01$). (**F**) Performance of a linear decoder trained on the distinction between the five behavioral epochs depicted using responses of all PNs ($n = 80 \pm 10$, mean $\pm$ SEM neurons per mouse), only the task-modulated neurons (Mod., $n = 33 \pm 4$ neurons per mouse), or a random set of none-task-modulated neurons (N Mod., $n = 33 \pm 4$ neurons randomly chosen to match the number of modulated neurons). Left, Example mouse decoder results compared with the real behavioral description (top). Right, Mean F score for all mice ($N = 8$, two-sided paired t test, **$P < 0.01$). Box-and-whisker plots indicate median, interquartile, extreme data values, and outliers of the data distribution.

behavioral sequences (Fig. 3B). To quantify the mapping between behavior and the population activity vector, we averaged the pairwise correlations between all population activity vectors in action and consumption epochs. Intrabehavioral epoch correlations were positive for outcome-specific action and consumption epochs and higher than interbehavioral epoch correlations (action-consumption) and OFF task correlations (Fig. 3C and fig. S6B). Throughout the session, the population activity vectors during action and consumption epochs remained positively correlated and distinct from OFF task population activity vectors (Fig. 3D and fig. S6C), suggesting that specific neuronal activity patterns robustly define distinct behavioral epochs in an outcome-specific manner. We then used dimensionality reduction [t-distributed stochastic neighbor embedding (t-SNE); see the materials and methods] with correlation as the distance measure to visually demonstrate the distinction between the neuronal activity patterns associated with the OFF task phase and the two outcome-specific ON task phases (fig. S6A). Mapping the population

activity vectors associated with the different behavioral epochs to the embedded activity space revealed discrete clusters corresponding to distinct neuronal activity patterns occurring at specific behavioral epochs (Fig. 3E and fig. S6, B and C).

Defining action-associated neuronal activity patterns as all population activity vectors correlated with the population response occurring at the time of the first action of a behavioral sequence ($P < 0.01$; see the materials and methods), action-associated neuronal activity patterns were not restricted to action epochs but had already started before the first action of a behavioral sequence and lasted until reward detection (Fig. 3F). Indeed, the action-associated neuronal activity patterns that were maintained for several seconds throughout their respective action periods encompassed unrewarded lick and idle time epochs but not exploration epochs (Fig. 3G and fig. S6D). Therefore, BLA action-associated neuronal activity patterns persistently represent information about the pursued outcome from the initiation of the behavioral sequence

until reward detection, not just during action performance. By tracking neurons across days, we observed that BLA action-associated neuronal activity patterns were reactivated upon re-exposure to the task during both nonreinforced and reinforced tests, but their activity was maintained only in the latter case (fig. S7). By contrast, consumption-associated neuronal activity patterns were restricted to consumption epochs and were not detected during unrewarded lick epochs, suggesting that consumption-associated neuronal activity patterns reflect reward detection rather than licking behavior (Fig. 3, F and G, and fig. S6D). Using single-unit recordings and inferred spikes extracted from deconvoluted signal using CNMF-E, we confirmed that these results were not merely reflecting slow Ca^{2+} dynamics (figs. S8 and S9).

Specific motivational control of goal-directed behavior

To test whether action- and consumption-associated BLA activity is necessary for instrumental goal-directed behavior, we inhibited the
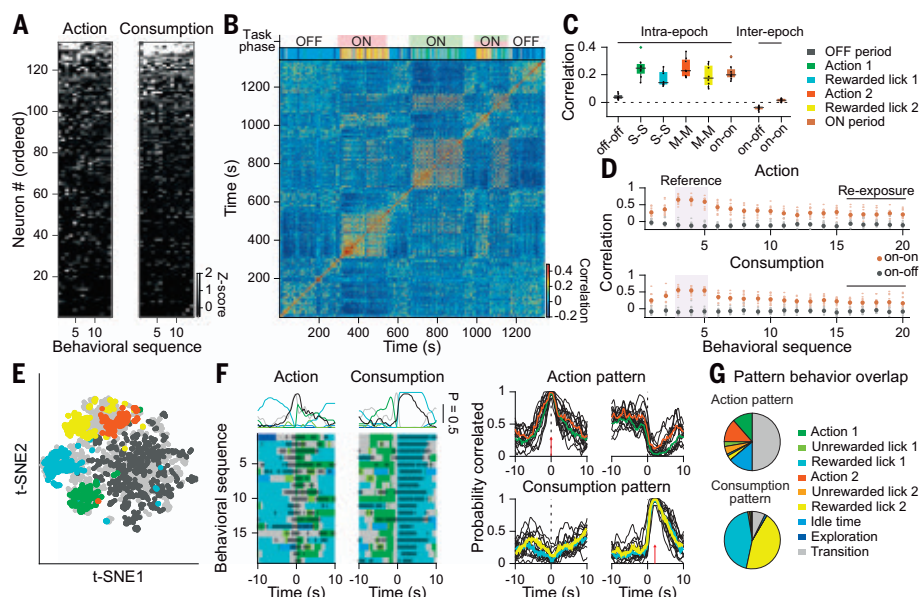
Fig. 3. Distinct BLA neuronal activity patterns maintain outcome-specific information along action-outcome behavioral sequences. (**A**) Population activity vectors from one mouse occurring at action period onset (left) and 2 s after the consumption period onset (right panel; bin size, 200 ms). Neurons ordered for each panel on the basis of their average activity over behavioral sequences. (**B**) Cross-correlation matrix of population activity vectors for the entire task. Same mouse as in (A) (bin size, 200 ms). Top, Task phases (red, milk; green, sucrose) and behavioral epochs labeling; colors are as in (G). (**C**) Intra- and interbehavioral epoch correlation between population activity vectors (N = 8 mice in two cohorts; colors denote the different behavioral epochs). Box-and-whisker plots indicate median, interquartile, extreme data values, and outliers of the data distribution. (**D**) Correlation between the reference activity vector and population activity vectors occurring at action (top) or rewarded lick (bottom) epochs along 20 behavioral sequences of day 5 session (N = 8 mice × two outcomes; the activity vectors selected as a mean reference for the correlation analysis are shaded). (**E**) 2D embedding (t-SNE) of all population activity vectors from (B) using correlation as a distance measure. (**F**) Left, Example of the maintenance of action (left)– or consumption (right)–locked activity patterns overlaid on behavior for 20 behavioral sequences. Black dots denote positive and significant correlation (P < 0.01) between population activity vectors and a reference activity pattern for each behavioral sequence (action: activity at action period onset; consumption: activity 2 s after consumption period onset). Probability of a correlated neuronal pattern and probability of behavioral epochs occurrences for each time bin is shown on the top (bin size, 200 ms; scale bars, probability of 0.5). Right, Probability of correlated action- or consumption-associated activity patterns locked to action and consumption period onsets (N = 8 mice × two outcomes). (**G**) Proportions of different behavioral epochs that occurred when the activity pattern is correlated with the action (top) or consumption (bottom) activity pattern.

activity of BLA PNs on day 5 using an optogenetic approach. We expressed archaerhodopsin (ArchT) or green fluorescent protein (GFP; control group) in PNs and delivered light (589 nm) in a closed-loop manner to independently inhibit neuronal activity during individual action or consumption periods (fig. S10). For each manipulation and each outcome, we compared behaviors between laser-OFF (behavioral sequences 1 to 8) and laser-ON periods (behavioral sequences 9 to 16) (Fig. 4A and fig. S11). Inhibition of PNs during action or consumption periods promoted non-task-related behaviors such as idle times and context exploration epochs, which consequently resulted in a general slowdown of instrumental performance (Fig. 4, B and C, and fig. S11, C and D). However, task-related behaviors such as instrumental actions, unrewarded lick epochs, and

rewarded lick epochs were generally not affected by the two types of manipulation (Fig. 4D and fig. S11, E to K).

Inhibition of BLA PNs during either action or consumption periods caused markedly distinct behavioral phenotypes. Inhibition of BLA PNs during consumption periods selectively prolonged the interbehavioral sequence duration, i.e., action initiation of subsequent behavioral sequences was delayed. Consumption behavior was not affected (fig. S11, E to K). Inhibition of BLA PNs during action periods prolonged action period duration, latency to reward consumption, and interbehavioral sequence duration, thereby both extending the time to obtaining the reward and delaying action initiation (Fig. 4B and fig. S11D). No significant differences were observed when comparing behaviors during laser-OFF periods

of both ON-task phases (all comparisons, P > 0.15), indicating that perturbing the neuronal activity when one of the rewards was available did not affect behavior in the subsequent phase, when the other reward was available. Moreover, all ArchT-expressing mice earned and consumed the maximum number of rewards.

During optogenetic inhibition of BLA PNs, mice did not show obvious aversive reactions, but rather engaged in non-task-related behaviors. Consistently, inhibition of BLA PNs did not result in real-time place avoidance (fig. S11L) or in mice switching preference to bigger rewards when they were allowed to choose between licking a 5% sucrose solution without laser stimulation and a 20% sucrose solution paired with laser (fig. S11M).

To confirm the causal role of action-associated activity in the motivational control of goal-directed behavior, we inhibited the activity of BLA PNs under non-reinforced conditions by delivering light (589 nm) in a closed-loop manner during individual action epochs (fig. S12). Inhibition of BLA PNs during a free choice non-reinforced test impaired instrumental performance (fig. S12, C to F). Moreover, inhibition of BLA PNs during specific satiety-induced outcome devaluation and action-outcome contingency degradation tests impaired both instrumental performance and action choice bias classically induced by both procedures (fig. S12, G to N).

Behavioral manipulations reveal an adaptive BLA code

To determine how BLA activity supporting goal-directed behaviors adapts to variations in outcome value and action-outcome contingency, we tracked how action- and consumption-associated activity was affected by manipulations of such behavioral parameters. First, we investigated the consequence of either specific satiety-induced outcome devaluation (six mice) or action-outcome contingency degradation (five mice) on BLA action-associated activity (fig. S13). During both non-reinforced tests, mice showed a clear bias in action choice toward the non-devaluated or non-degraded actions, and, concomitantly, only the activity associated with these actions was reactivated. This was evident at both the single-neuron level and the population level (fig. S13, C, D, G, and H).

We then simultaneously assessed how both action- and consumption-associated activity was affected by violating the expected outcome value and action-outcome contingency under reinforced conditions (Fig. 5, A and G, and fig. S14). After the completion of 5 training days, mice were first allowed to obtain eight rewards after the established VR5 schedule (referred to as the initial period), after which a perturbation was executed. During the violation of outcome value paradigm, outcomes
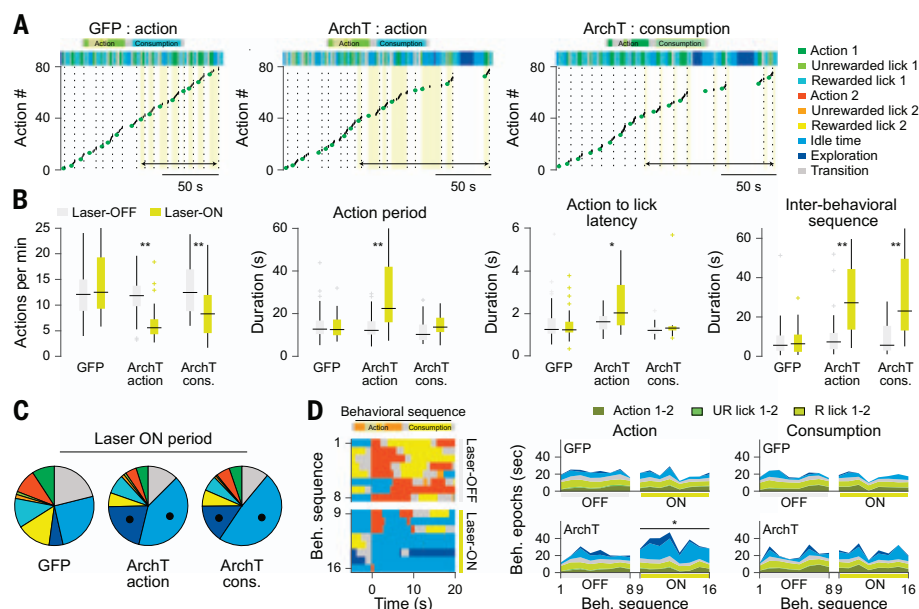
Fig. 4. BLA PNs activity is necessary for the maintenance of goal-directed actions. (**A**) Examples of the effects of optogenetic manipulations on goal-directed behaviors on day 5 for GFP- and ArchT-expressing mice. Black dots indicate individual actions; green dots, action period onset; vertical dashed lines, outcome delivery. Colors indicate behavioral epochs. Shaded yellow areas indicate laser deliveries; double-headed arrows, laser-ON periods (behavioral sequences 9 to 16). (**B**) Left to right, actions per minute, action period duration, action to lick latency, and interbehavioral sequence duration during laser-OFF (behavioral sequences 1 to 8, white bars) and laser-ON periods, respectively (yellow bars; two-sided paired *t* test for action per minute and two-sided Wilcoxon signed-rank tests for the other variables comparing laser-OFF and laser-ON periods, *$P < 0.05$; **$P < 0.01$. GFP, $N = 20$ mice in three cohorts × two outcomes ($N = 12$, $N = 8$ mice with laser stimulation during action periods and consumption periods, respectively, with data pooled). ArchT action, $N = 12$ mice in two cohorts × two outcomes (laser delivered during action periods); ArchT cons., $N = 8$ mice in two cohorts × two outcomes (laser delivered during consumption periods). Box-and-whisker plots indicate median, interquartile, extreme data values, and outliers of the data distribution. (**C**) Proportions of different behavioral epochs during laser-ON periods (color denote behavioral epochs as in (A); Black dots denote significant increase in behavioral epoch duration; $P < 0.01$, two-sided Wilcoxon signed-rank tests comparing laser-OFF and laser-ON periods; see raw values in fig. S11). (**D**) Mouse behavior during laser-OFF and laser-ON periods while milk outcome was available (left). Colors denote behavioral epochs as in (A). Also shown is the duration of the distinct behavioral epochs during the behavioral sequences for each sequence along laser-OFF and laser-ON periods (right, two-sided Wilcoxon signed-rank test).

9 to 16 were delivered on the usual VR5 schedule but were diluted by a factor of 4. This caused a disruption in the established goal-directed behavior without affecting consumption behavior. Mice exhibited significantly longer action period duration, interbehavioral sequence duration, and latency to reward consumption (Fig. 5, A and B) caused by increased non-task-related behaviors both inside and between behavioral sequences (fig. S14, A to C). Eventually, all mice obtained and consumed the maximum number of available outcomes.

We then tested the effect of the violation of the outcome value on the neuronal representations of action and consumption at both the single-neuron level and the population level (326 neurons from four mice). First, whereas action and consumption neurons detected during the initial period showed significantly reduced responses during the perturbed period (Fig. 5, C and D), additional consumption and

action neurons emerged during the perturbed period (fig. S15, C and D). Similarly, immediately after the start of outcome value violation (i.e., after consumption of the first diluted reward) new action- and consumption-associated neuronal activity patterns emerged that replaced the action- and consumption-associated neuronal activity patterns of the initial period (Fig. 5, E and F, and fig. S15F). This effect was also observed when considering only neuronal activity patterns associated with the first or the last action in an action period, indicating that it was not a consequence of the increased action period duration (fig. S15G, and see comparison with neuronal activity pattern stability on day 5 in fig. S16).

During the violation of action-outcome contingency paradigm, the initial eight rewards were delivered contingent on the actions on the usual VR5 schedule. Subsequently, outcomes 9 to 20 were delivered noncontingently

on the animal's actions. Whereas at the behavioral level, this test disrupted goal-directed behavior similarly to the violation of outcome value (Fig. 5, G and H, and fig. S14, D to F), the impact on BLA neuronal activity was different (434 neurons from five mice). Action neurons detected during the initial period showed a reduction of their responses during the entire perturbed period (Fig. 5I). However, unlike the immediate effect of violating outcome value, this change developed slowly after the start of the perturbation and was not accompanied by the emergence of additional action neurons (fig. S15H). In accordance with this finding, at the population level, the action-associated neuronal activity patterns gradually lost their correlation with the action-associated neuronal activity patterns of the initial period, but no new action-associated neuronal activity patterns emerged during the perturbed period (Fig. 5K and fig. S15J; also see the results for the first and last action periods in fig. S15L). In clear contrast to the impact of the violation of outcome value, consumption-associated activity during the initial period after violation of action-outcome contingency remained stable along the entire perturbed period (Fig. 5, J to L, and fig. S15K). As previously reported (*17*), we were able to detect a population of consumption neurons that gradually increased their activity over the course of unexpected reward deliveries (fig. S15I). However, the overall population activity pattern during consumption remained highly correlated throughout the entire session. Although changing outcome value resulted in the immediate emergence of new action and outcome representations, action-outcome contingency violation resulted in a gradual loss of the representation of the action with no emergence of a new stable action representation and no effect on reward consumption representation.

Discussion

Our findings support a crucial role for the BLA in the motivational control of goal-directed behavior. At the time of goal-directed action, BLA PNs integrated and encoded pursued outcome identity, pursued outcome value, and action-outcome contingency information. The maintenance of such prospective, outcome-specific, and updatable neuronal activity reflects a specific motivational state (*18*, *19*) that differs from a primary motivational state known to energize goal-directed actions in an unspecific manner (*1*, *20*). Furthermore, this state could be retrieved upon re-exposure to the task, but its maintenance depended on continued reinforcement. At the time of consumption, BLA PNs represented current outcome identity and value but not licking behavior. This is consistent with the observations that BLA neurons encode distinct reward features (*21*), including value (*22*), magnitude
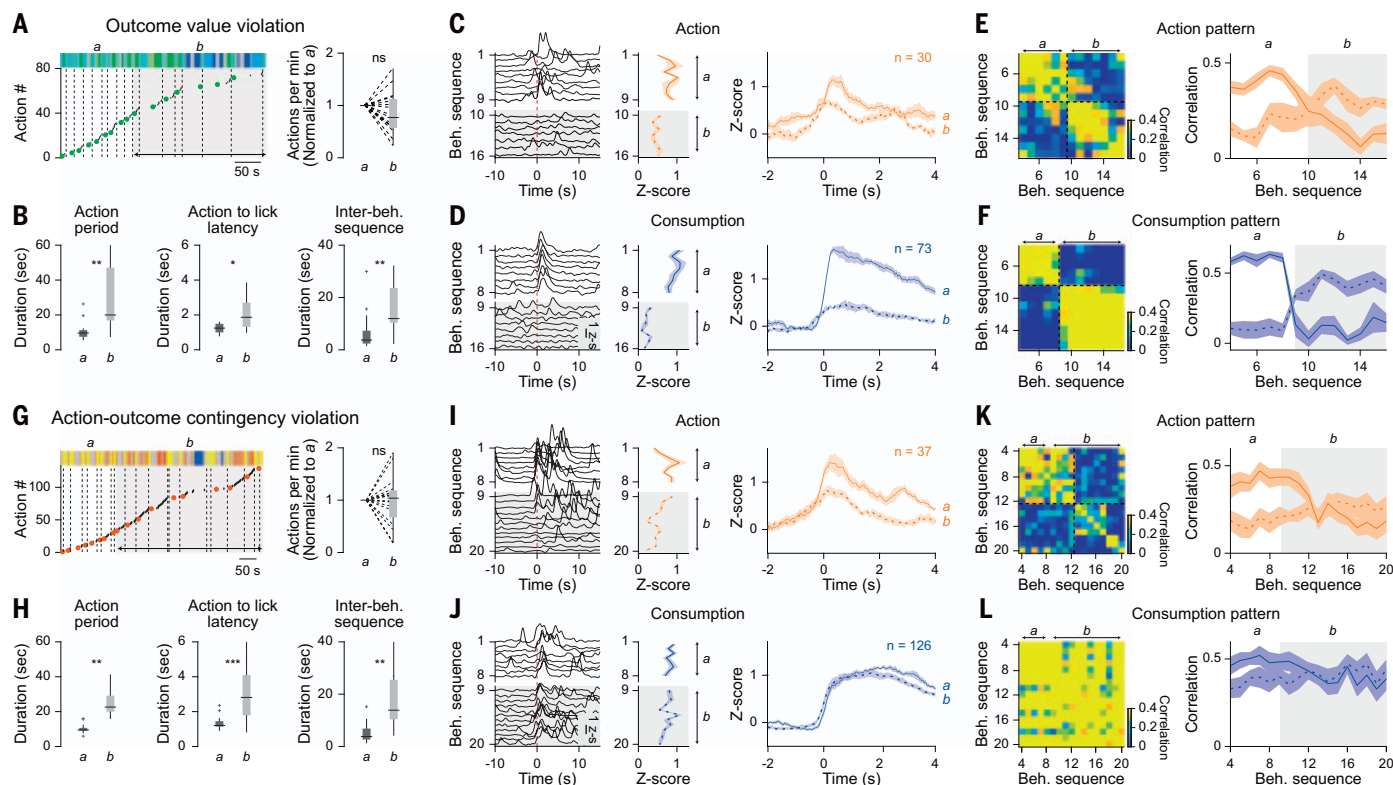
Fig. 5. BLA PNs adaptively control goal-directed actions. (**A**) Example of the effects of outcome value violation on goal-directed behaviors (left). Black dots indicate individual actions; green dots, action period onset; vertical dashed lines, outcome delivery (*a*, initial period, white; *b*, perturbed period, shaded area). Colors indicate behavioral epochs. Shown are actions per minute normalized to the initial period (right, *N* = 8 mice in two cohorts × two outcomes; *P* > 0.15, two-sided paired *t* test). (**B**) Action period duration (left), action to lick latency (middle), and interbehavioral sequence duration (right) during the initial and perturbed periods (**P* < 0.05, ***P* < 001, two-sided Wilcoxon signed-rank tests). (**C**) Left, Action neuron activity aligned to action periods onset during the initial and perturbed (shaded area) periods. Middle, Mean activation per sequence of initial action neurons (*n* = 30, with four mice in two cohorts × two outcomes averaged over 1 s after action initiation; *P* < 0.05, two-sided paired *t* test comparing initial and perturbed period *z*-score values). Right, Average response of action (orange) and

consumption (purple) neurons during the initial (full line) and perturbed (dashed line) periods. (**D**) Same as (C) but for initial consumption neurons aligned to consumption period onset (*n* = 73; averaged over 3 s after consumption initiation; *P* < 0.001, two-sided paired *t* test). Shading indicates SEM. (**E**) Left, Pairwise correlation between action activity vectors for one mouse. Dashed lines indicate the transition between neuronal patterns. Right, Dynamics of the correlation between action activity vectors (*N* = 4 mice × two outcomes), averaging correlations before (full line) and after (dashed line) the transition. (**F**) Same as (E) for consumption activity vectors. (**G** to **L**) Same as (A) to (F) but for action-outcome contingency violation (*N* = 8 mice in two cohorts for behavioral data; *N* = 5 mice in two cohorts for neuronal data). For (I), *n* = 37 action neurons (*P* < 0.01, two-sided paired *t* test). For (J), *n* = 126 consumption neurons (*P* = 0.28, two-sided paired *t* test). Box-and-whisker plots indicate median, interquartile, extreme data values, and outliers of the data distribution in all panels.

(*23*), and probability, (*17*) in a state-dependent manner. Such representations are thought to reflect the incentive value of outcomes (*24*). Optogenetic inhibition of BLA PNs during consumption delayed the initiation of action-consumption sequences without affecting consummatory behavior. Action and consumption PNs formed distinct functional populations showing opposite activity during action and consumption periods. This divergent feature was also observable at the population level, because action- and consumption-associated neuronal activity patterns were anticorrelated. Together, action- and consumption-associated activity integrated behaviorally relevant information at distinct time points along goal-directed action-consumption sequences. Upon behavioral perturbations, action- and consumption-associated activity differentially adapted.

Although changing outcome value resulted in the remapping of both activities, affecting action-outcome contingency resulted in the degradation of action-associated activity only. This adaptive code might be a central component to direct behavior toward a pursued goal. Overall, these results indicate that the interplay between action- and consumption-associated neuronal activity patterns maintains specific motivational states necessary for the execution of self-initiated, goal-directed action-consumption sequences, and demonstrate that the BLA is a key neuronal substrate for the control of voluntary behavior.

In contrast to a large body of work on the contribution of the BLA in outcome-seeking behaviors in Pavlovian tasks or in tasks combining instrumental actions with Pavlovian cues that signal trial start and/or reward availability, ex-

ecution of action consumption sequences in the present study was entirely self-paced. Our findings thus extend the classic model implicating BLA in controlling outcome-seeking behavior by conferring motivational significance to outcome-predicting sensory stimuli (*3–5*, *9*), showing encoding of the specific motivational state from before action initiation until reward attainment, consistent with findings from self-controlled behavior in monkeys (*18*). These results are consistent with BLA lesion, inactivation, or pharmacological studies that, despite notable discrepancies regarding the effect of pre- and post-training manipulations, support the view that noncued instrumental goal-directed action performance depends on BLA integrity (*12*).

We found that BLA PNs, at both the single-neuron level and the population level, differentiate

outcomes of similar value according to their identity (milk or sucrose). This distinction was encoded by distinct neuronal activity patterns during goal-directed action and consumption. Functional antagonism between BLA ensembles or circuits has been already observed in spatial exploration tasks (25) and in diverse Pavlovian tasks (26–30), where distinct BLA ensembles signal behaviors of opposite valence (avoidance and approach behaviors). Our results extend this view by showing that BLA population activity segregates behaviorally relevant information along different axes, such as outcome identity, value, and expectancy. Thereby, the BLA adaptively provides updated outcome-related information along the execution of goal-directed action-consumption sequences, with detailed representation of the motivational significance of action, suggesting a central role for the BLA in the circuit that supports the capacity for goal-directed action.

Although we found that most recorded BLA neurons were not significantly task modulated (31, 32), we observed that neurons that do not show significant task-modulated activity carry outcome-specific information regarding the different periods of a behavioral sequence. Thus, as a population, BLA neurons form neuronal activity patterns that maintain specific action- and consumption-related information along the entire temporal extent of action-consumption behavioral sequences. The underlying neuronal mechanisms may involve precise interactions between specific functional subpopulations through intra-BLA recurrent inhibitory and excitatory circuits (3, 33, 34) under the control of neuromodulatory inputs (35, 36). Moreover, BLA action and consumption neurons may be driven by differential external inputs from a variety of brain structures that process relevant information (37–40) and project to distinct targets, including the central nucleus of the amygdala (41), the nucleus accumbens (42), and the dorsomedial striatum (43). Action and consumption neurons could be found among BLA neurons projecting to the medial shell of the nucleus accumbens or to the posterior dorsomedial striatum (figs. S17 and S18). Although we cannot exclude that other BLA projections would be functionally more homogeneous, our data show that information encoded by BLA PNs along action-consumption behavioral sequences is broadcasted to distinct striatum subregions known to differentially control goal-directed behavior (42).

Materials and Methods
Animals
All animal procedures were performed in accordance with institutional guidelines at the Friedrich Miescher Institute for Biomedical Research and were approved by the Veterinary Department of the Basel-Stadt Canton. Male

mice (C57BL/6JRccHsd, Envigo) were used throughout the study. We chose to do experiments with male mice because they are on average heavier than female mice, which facilitates the carrying of the electrode or gradient-index (GRIN) lenses implants during locomotion. Moreover, this allowed us to directly compare our present results with previous studies. Mice were individually housed for at least 2 weeks before starting behavioral paradigms and were kept in a 12-hour light–dark cycle. Mouse well-being was monitored throughout the experimental period. All behavioral experiments were conducted during the light cycle.

Surgical procedures
Eight-week-old mice were anesthetized using isoflurane (3 to 5% for induction, 1 to 2% for maintenance; Attane, Provet) in oxygen-enriched air (Oxymat 3, Weinmann) and placed in a stereotactic frame (model 1900, Kopf Instruments). Injections of buprenorphine (Temgesic, Indivior UK; 0.1 mg/kg body weight subcutaneously 30 min before anesthesia) and ropivacain (Naropin, AstraZeneca; 0.1 ml locally under the scalp before incision) were provided for analgesia. Postoperative pain medication included buprenorphine (0.1 mg/kg body weight in the drinking water overnight) and injections of meloxicam (Metacam, Boehringer Ingelheim; 1 mg/kg subcutaneously) for 3 days after surgery. Ophthalmic gel was applied to avoid eye drying (Viscotears, Bausch + Lomb). The body temperature of the experimental animal was maintained at 36°C using a feedback-controlled heating pad (FHC).

Surgeries for deep-brain Ca²⁺ imaging
AAV2/5.CaMK2.GCaMP6f (University of Pennsylvania Vector Core) was unilaterally injected into the BLA (300 nl) using a precision micropositioner (model 2650, Kopf Instruments) and pulled glass pipettes (tip diameter ~20 μm) connected to a Picospritzer III microinjection system (Parker Hannifin) at the following coordinates from bregma: anterior–posterior (AP): –1.6 mm; medial–lateral (ML): –3.35 mm; dorsal–ventral (DV): 4.2 mm. retroAAV.Ef1a.GCaMP6f (FMI Vector Core) was unilaterally injected either into the caudal dorsomedial striatum (DMS; 100 nl) at the following coordinates from bregma: AP: +0.5 mm; ML: –1 mm; DV: 2 mm or along rostro-caudal axis of the medial shell of the nucleus accumbens (mShell; 100 nl) at the following coordinates from bregma: AP: +1.3 or +0.8 mm; ML: –0.5 mm; DV: 4.2 to 4.4 mm. After injection, a track above the imaging site was cut with a sterile needle to aid the insertion of the GRIN lens (0.6 × 7.3 mm, GLP-0673, Inscopix). The GRIN lens was subsequently lowered into the brain using a micropositioner through the track (4.3 mm from brain surface) with a custom-built lens holder and fixed to the skull using ultraviolet light-curable glue

(Henkel, Loctite 4305). Dental acrylic (Paladur, Heraeus) was used to seal the skull and attach a custom-made head bar for animal fixation during the miniature microscope mounting procedure. Mice were allowed to recover for 3 weeks after surgery before checking for GCaMP expression.

Surgeries for optogenetic sessions
AAV2/5.CaMK2.ArchT.GFP or AAV2/5.CaMK2.GFP (University of Pennsylvania Vector Core) was bilaterally injected into the BLA (200 nl per hemisphere, coordinates from bregma: AP: –1.6 mm; ML: –3.3 mm; DV: 4.2 mm). After injection, mice were bilaterally implanted with custom-made optic fiber connectors. Fiber tips were lowered to –3.9 mm below the brain surface using a micropositioner. Implants were fixed to the skull using cyanoacrylate glue (Ultra Gel, Henkel) and miniature screws (P.A. Precision Screws). Dental acrylic mixed with black paint (to avoid light spread) was used to seal the skull. Mice were allowed to recover for 3 weeks before behavioral training to ensure adequate virus expression.

Surgeries for in vivo electrophysiology
A set of mice was unilaterally implanted with a custom-built 32-wire electrode into the BLA using a micropositioner (coordinates from bregma: AP: –1.6 mm; ML: –3.3 mm; DV: 4.2 mm). Electrodes consisted of two bundles of 16 individually insulated, gold-plated nichrome wires (13 μm inner diameter, impedance 30 to 100 kΩ, Sandvik). Each wire in the bundles was attached to an 18-pin connector (Omnetics). Another set of mice was equipped with a custom-built optrode consisting of an optic fiber with an attached 16-wire electrode. Electrodes tips were cut at an angle to protrude ~300 to 500 μm beyond the fiber end. In this case, the surgical procedure was similar to that described above for the optogenetic section. Implants were fixed to the skull using cyanoacrylate glue (Ultra Gel, Henkel) and miniature screws (P.A. Precision Screws). Dental acrylic was used to seal the skull. Mice were allowed to recover for 2 weeks before recordings started.

Instrumental goal-directed behavioral context
The two behavioral contexts used in this study measured 26 cm L × 25 cm W × 40 cm H and each context was enclosed inside an acoustic foam isolated box. One context was equipped with two nose poke ports (left/right; H21-09M_CA, Coulbourn instruments) and the other one with two levers (left/right; ENV-312-2M, Med Associates). All *operanda* (two instrumental actions apparatus, left/right; and two lick ports, left/right) were located on the same wall. In each context, the two instrumental actions apparatus (left/right) were located at the extremes (5 cm from the wall limits). The lever edges and nose poke port centers were positioned 2 cm

from the floor. Lever extensions and retractions and nose poke port opening and closing (guillotine door) were remotely controlled. Custom-made lick ports were located at the middle of the wall, close to each other (4 cm between). The two lick ports were separated by a small barrier (to force mice to physically move to lick from one to another lick port). The lick ports were positioned 2 cm from the floor (fig. S1A). The lick port was composed of an empty cylinder (made of polyoxymethylene) positioned horizontally with an open window on the top where mice could access liquids (open window: ellipse of 6 × 3 mm). Liquids were delivered in a receptacle inserted in the cylinder (receptacle: half-ellipsoid of 6 × 3 × 2 mm). The receptacle was made of aluminum to measure tongue contacts through the analog input board of a neural recording data acquisition processor system (OmniPlex, Plexon). Lick onsets were inferred offline by detecting potential rise times. Each lick port allowed delivery of two outcomes through remotely controlled syringe pumps (PHM-107, Med Associates). A video camera recorded from above at 40 frames/s (fps) for video-tracking purposes using Cineplex Software (Plexon). All time stamps of camera frames, miniscope frames, nose pokes or lever presses, and analog signals from lick ports were recorded with the Omniplex neural recording data acquisition processor system at 40 kHz. Behavior, optogenetic, single-unit recordings, and miniscope were synchronized and controlled by a multi–input/output (I/O) processor (RZ6, Tucker Davis).

Instrumental goal-directed behavioral procedures
Food restriction

Mice were food restricted to 85% of their free-feeding body weight 4 days before and throughout the behavioral experiments and fed ~2 hours after their daily behavioral sessions with ~2.5 g of regular food.

Instrumental training

Instrumental training for the experimental set (N = 16 mice; N = 8 of 16 mice were equipped with a miniature microscope; Figs. 1 to 3 and figs. S1 to S9) was preceded with a session in which either sucrose (20%) or sweetened condensed milk (15%, Régilait) solutions were accessible, each at a fixed lick port (right and left lick port, respectively) upon licking (maximum duration 20 min or 20 of each outcome). The following day, instrumental training started with constant reinforcement (CR), in which outcomes were delivered after each action performed by the mouse. To speed up learning, only during this day was food dispensed in nose-poke ports or onto levers. Mice then started instrumental learning with 2 days of CR (without food in or onto actions, called day 1 and day 2). CR sessions were struc-

tured as follows: (i) 5 min without action or outcome availability (licking behavior was possible but not rewarded, OFF task); (ii) one of the two actions (left/right apparatus) was available for 30 min or until mice obtained 20 outcomes (milk/sucrose, ON task); (iii) both actions were not available for 2 min (OFF task); (iv) the second action (right/left apparatus) was available for 30 min or until mice obtained 20 outcomes (sucrose/milk, ON task); and (v) 2 min without action or outcome availability (OFF task). After CR sessions, mice went through variable action-outcome ratio training [variable response (VR)], first with average ratio 3 on day 3 (VR3, between 1 and 5 after normal distribution, μ = 3, σ = 1.5) and average ratio 5 on days 4 and 5 (VR5, between 1 and 11 after normal distribution, μ = 5, σ = 2.5). VR sessions were structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) one of the two actions (left/right apparatus) was available for 30 min or until mice obtained 15 outcomes (milk/sucrose, ON task); (iii) both actions were not available for 2 min (OFF task); (iv) the second action (right/left apparatus) was available for 30 min or until mice obtained 15 outcomes (sucrose/milk, ON task); (v) both actions were not available for 1 min (OFF task); (vi) both actions (left/right apparatus) were available for 30 min or until mice obtained 5 of each of the outcomes (milk/sucrose, ON task); and (vii) 2 min without action or outcome availability (OFF task). For CR and VR sessions, outcome order was counterbalanced.

Shift in hunger state test

To demonstrate the hunger-state dependency of instrumental actions, we exposed mice to a free choice rewarded test (N = 8 of 16 mice in the instrumental training group). A shift in hunger state was accomplished by interrupting food restriction from day 6 to day 8. The test was conducted on day 8 (Fig. 1B). The free choice reinforced test was structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 30 min or until mice obtained 20 of each of the outcomes (milk/sucrose, ON task); and (iii) 2 min without action or outcome availability (OFF task). Outcomes were delivered on a VR5 basis.

Free choice extinction test

After instrumental training, we exposed mice to a free choice extinction (non-reinforced) test on day 6 (N = 16 mice in Fig. 1C; N = 8 mice in fig. S7; N = 20 mice in fig. S12), structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 8 or 12 min (ON task; 8 min in fig. S12; 12 min in Fig. 1C and fig. S7); and (iii) 2 min without action or outcome availability (OFF task).

Free choice reinforced test

Two hours after the completion of the free choice extinction test, we exposed mice to a free choice reinforced test (N = 8 mice in fig. S7), structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 30 min or until mice obtained 20 of each of the outcomes (milk/sucrose, ON task); and (iii) 2 min without action or outcome availability (OFF task). Outcomes were delivered on a VR5 basis.

Satiety-induced outcome devaluation

Instrumental actions are characterized as goal-directed if the actions are sensitive to variations in outcome value. After training, outcome devaluation was accomplished by pre-feeding mice with one of the two outcomes (the devaluated one) for 1 hour in home cage (N = 13 mice in Fig. 1C; N = 6 mice in fig. S13; N = 20 mice in fig. S12). The identity of the devaluated outcome was counterbalanced between mice. The experimenter was blinded to the devaluated outcome. Animals were randomly allocated to experimental groups and were later identified by unique markers for group assignment. Mice were then exposed on day 6 or day 8 to a free choice non-reinforced test (day 6 in fig. S12; day 8 in fig. S12), structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 8 or 12 min (ON task; 8 min in fig. S12; 12 min in Fig. 1C and fig. S13); and (iii) 2 min without action or outcome availability (OFF task).

Action-outcome contingency degradation

Instrumental actions are characterized as goal directed if the actions are sensitive to variations in the action-outcome contingency. After instrumental training, action-outcome contingency degradation was accomplished by unpairing one of the two actions from its respective outcome (degraded action; N = 10 mice in Fig. 1C; N = 5 mice in fig. S13; N = 20 mice in fig. S12). The identity of the degraded action-outcome contingency was counterbalanced between mice. This training phase was conducted on day 6 and day 7, with the same structure as day 5 VR5 training (see above). Mice then were exposed on day 8 to a free choice non-reinforced test, structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 8 or 12 min (ON task; 8 min in fig. S12; 12 min in Fig. 1C and fig. S13); and (iii) 2 min without action or outcome availability (OFF task).

Optogenetic sessions

Mice were trained to obtained two outcomes as described earlier. For optogenetic manipulations performed on day 5 (N = 40 mice;

Fig. 4 and figs. S10 and S11), sessions were structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) one of the two actions (left/right apparatus) was available for 30 min or until mice obtained 16 outcomes (milk/sucrose, ON task); (iii) 2 min without action or outcome availability (OFF task); (iv) the second action was available for 30 min or until mice obtained 16 outcomes (ON task); and (v) 2 min without action or outcome availability (OFF task). For optogenetic manipulations performed during non-reinforced tests ($N = 40$ mice; fig. S12), sessions were structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) the two actions (left/right apparatus) were available for 8 min (ON task); and (iii) 2 min without action or outcome availability (OFF task). Laser delivery strategies are described below.

Outcome value violation

Outcome value violation session differs from the specific satiety-induced outcome devaluation classically used to characterize goal-directed behavior. Nevertheless, the decrease in outcome value was achieved during this session by diluting outcomes, as evident by the reduction in instrumental behavior, and by the clear preference for nondiluted outcomes mice showed during the two-sucrose choice challenge (see below). On the basis of these observations, we call this session outcome value violation. After mice were trained to obtain two outcomes as described earlier ($N = 8$ mice; $N = 4$ of 8 mice were equipped with a miniature microscope), they underwent an outcome value violation session on day 6 (Fig. 5 and figs. S14 and S15), structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) one of the two actions was available for 30 min or until mice obtained 16 outcomes (ON task); (iii) 2 min without action or outcome availability (OFF task); (iv) the second action was available for 30 min or until mice obtained 16 outcomes (ON task); and (v) 2 min without action or outcome availability (OFF task). During ON task phases, mice could obtain up to eight outcomes (called the initial period), and then outcomes 9 to 16 were diluted by a factor of 4 (called the perturbed period). During initial and perturbed periods, outcomes were delivered on a VR5 schedule.

Action-outcome contingency violation

After mice were trained to obtain two outcomes as described earlier ($N = 8$ mice; $N = 5$ of 8 mice were equipped with a miniature microscope), they underwent a retraining session on day 7 (in the same conditions as on day 5) followed by an action-outcome contingency violation session on day 8 ($N = 3$ mice; one mouse was excluded from analysis because it

bit the miniature microscope cable at the beginning of the session). An extra group of $N = 2$ mice that were submitted to an action-outcome contingency violation session on day 6 was then added to the analysis (Fig. 5 and figs. S14 and S15). Action-outcome contingency violation session was structured as follows: (i) 5 min without action or outcome availability (OFF task); (ii) one of the two actions was available for 30 min or until mice obtained 20 outcomes (ON task); (iii) 2 min without action or outcome availability (OFF task); (iv) the second action was available for 30 min or until mice obtained 20 outcomes (ON task); and (v) 2 min without action or outcome availability (OFF task). During ON task phases, mice were allowed to obtain eight outcomes delivered on a VR5 schedule (called the initial period) and then outcomes 9 to 20 were delivered noncontingently to action (called the perturbed period).

Instrumental goal-directed behavior analysis
Behavioral characterization

The instrumental training session was divided into OFF task phases (no action or outcome availability) and ON task phases (one or both actions available), as described in detail above. We collected action and lick time stamps together with mouse position tracking using Bonsai software (44). We then used these reading to automatically derive a continuous description of the mouse behavior during the entire session. We used nine descriptors differentiating epochs of task-related behaviors: action, unrewarded lick, and rewarded lick epochs from epochs of non-task-related behaviors: idle times and context exploration epochs, in an outcome-specific manner. Action epochs were defined as the time bin of a single action or series of actions (consecutive actions not interrupted by other behavioral epochs). Rewarded lick epochs were defined as the first lick bouts (series of two or more consecutive licks, with inter-lick interval <1 s) that occurred after an outcome delivery. Unrewarded lick epochs were defined as all lick bouts not defined as rewarded lick epochs (see above). Idle times epochs were defined as time windows when mice when mice were at the task zone (mouse center of mass located at ~8 cm (1/3 of the context) from the wall including instrumental actions and lick ports) but did not perform any task-related behaviors (actions/licks) for >4 s. Exploration epochs were defined as time windows when mice were away from the task zone. Transition epochs were defined as all time bins not allocated to any of the different behavioral epochs defined above. We further defined behavioral sequences consisting of action and outcome consumption periods. The action period was defined as the time from the first action until the last action between consumption epochs and included

epochs of instrumental actions or unrewarded lick and transition epochs. Consumption periods are identical to the rewarded lick epochs.

Behavioral quantification

We used different parameters to quantify goal-directed behavior. Action rate (actions per minute; total number of actions divided by the total ON task phase duration) was used as the quantifier of the general instrumental performance. Action to lick latency (in seconds; time between the last action of an action period and the first rewarded lick that followed) was used to quantify the latency to collect outcome. Interbehavioral sequence interval (in seconds; time between the last lick of a consumption period and the subsequent action) was used to quantify the latency to initiate a new behavioral sequence after a consumption period. We used these three parameters to estimate mice motivation to obtain outcomes. We used the continuous description of the mouse behavior to quantify the duration of behavioral sequences, action periods, consumption periods, and the different behavioral epochs.

Ca²⁺ imaging using the miniature microscope
Miniature microscope imaging

Four weeks after surgery, mice were head-fixed to check GCaMP expression using a miniature microscope (nVista HD, Inscopix Inc.). If the expression level was sufficient, mice were anesthetized with isoflurane (3 to 5% for induction, 1 to 2% for maintenance; Attane, Provet) in oxygen-enriched air (Oxymat 3, Weinmann) to fix the miniature microscope baseplate (BLP-2, Inscopix) on top of the cranium implant using blue-light-curable glue (Vertise Flow, Kerr). The miniature microscope was then detached, the base plate was capped with a base plate cover (Inscopix), and the mouse was returned to its home cage. During the 3 days preceding instrumental training, mice were habituated to head fixation (~5 min), followed by a free exploration session in home cage (~10 min). This daily procedure allowed mice to habituate to miniature microscope mounting and carrying. It allowed us to select mice with a high number of neurons and better signal quality (>50 neurons, $N = 19$ of 24 mice included in the analysis) and to set the miniature microscope light-emitting diode (LED) power and electronic focus (for the nVista3 system). The miniature microscope was mounted immediately before each behavioral session by head fixation. Imaging data were acquired using nVista HD software (Inscopix Inc.) at a frame rate of 20 Hz (exposure time, 50 ms) with an LED power of 0.6 to 0.8 mW/mm² (excitation irradiance at objective front surface), an analog gain of 1 to 4, and a field of view of 1080 × 1080 pixels (~648 × 648 μm

with the nVista2 system) and 1280 × 800 pixels (~768 × 480 μm, nVista3 system). Imaging frame time stamps were recorded with the Omniplex neural recording data acquisition processor system.

Motion correction

Imaging files produced by the nVista2 or nVista3 systems were imported into Matlab 2017b using custom code. Two regions of interest (ROIs) manually selected by the investigators were used to correct for translational motion. To address background noise during motion correction, we subtracted a Gaussian blurred image on a frame-by-frame basis from the raw imaging data. We used Fourier fast transform–based Image registration (45) to register the first 100 frames on one ROI, and then used the median image of these 100 frames as a template to register the rest of the frames in the imaging session until correction was lower than a user-defined threshold (<0.01 average pixel shift on average across all frames). The procedure was then repeated on the resulting motion-corrected movie using the second user-defined ROI to ensure minimal nonrigid motion. We applied calculated shifts to the raw movie and used the raw, motion-corrected movies for the extraction of Ca^{2+} traces and all subsequent analysis.

Ca^{2+} trace extraction

We processed each session independently and used CNMF-E, a recently developed algorithm based on nonnegative matrix factorization (46), for automatic extraction of Ca^{2+} traces. We then excluded from the analysis ROIs based on anatomy (ROI size, shape, or vicinity to the edge of lens), low signal-to-noise ratio or large overlap in signal and spatial location with other neurons (>60% spatial overlap and >0.6 Pearson's correlation between the traces across the entire session). Linear trends across an entire session were removed from the Ca^{2+} traces and further calculations were performed on the z-scores of the detrended traces.

Registration of neuron identities across imaging session

After the individual sessions were extracted and sorted, taking day 5 as a reference, we performed automatic neuron alignment between sessions using centroid and shape-matching algorithms (47). To ensure that the correct alignment visual inspection was performed to assess whether the imaged neuron was the same across days, the ROI shape and location in the field of view had to be consistent across sessions.

Histology

After the completion of behavioral experiments, mice were deeply anesthetized with urethane

(2 g/kg of body weight intraperitoneally) and transcardially perfused with PBS followed by 4% paraformaldehyde in PBS. Brains were removed and postfixed overnight at 4°C. Then, 80-μm coronal brain slices containing the BLA were cut with a vibratome (VT1000 S, Leica) and stored in PBS. Slices were washed for 10 min in PBS, given a 5-min exposure to 4′,6-diamidin-2-phenylindol (DAPI, 1:10,000, Sigma-Aldrich), and then washed 3× for 15 min each in PBS. Slices were mounted on glass slides, coverslipped, and imaged using an Axioscan Z1 slide scanner (Carl Zeiss AG), equipped with a 10× air objective (Plan-Apochromat 10×/0.45). Mice were excluded post hoc if the GRIN lens was not placed in BLA (N = 3 mice excluded).

Imaging plan location and anatomical reconstruction

We assessed the spatial location of each neuron within the imaging field and along AP or ML BLA axes (fig. S3). The position of the center of each GRIN lens was matched against a mouse brain atlas (48), and the resulting coordinates were used to infer the location of the center of mass of each ROI (or neuron) along the AP or ML BLA axes and to determine the relative concentration of the distinct functional neuronal populations, as previously described (49).

Freely moving electrophysiology
Single-unit recordings

Mice were habituated to the head-stage (Plexon) connection procedure by handling and short head restraining for at least 3 days before experiments. The head stages were connected to a 32-channel preamplifier (gain ×100, bandpass filter 150 Hz to 9 kHz for unit activity and 0.7 Hz to 170 Hz for field potentials, Plexon). Spiking activity was digitized at 40 kHz, bandpass filtered from 250 Hz to 8 kHz, and isolated by time-amplitude window discrimination and template matching using the Omniplex neural recording data acquisition processor system.

Single-unit analysis

Single-unit spike sorting was performed using Offline Spike Sorter software (Plexon). Principal component scores were calculated for unsorted waveforms and plotted in a three-dimensional principal component space; clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if the waveforms formed a discrete, isolated cluster in the principal component space and did not contain a refractory period of <1 ms, as assessed using autocorrelogram analyses. To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that

the reference neuron fired, only one of the two neurons was considered for further analysis. Putative inhibitory interneurons were separated from putative excitatory projection neurons using an unsupervised cluster algorithm based on Ward's method. In brief, the Euclidian distance was calculated between all neuron pairs based on the three-dimensional space defined by each neuron's average half-spike width (measured from trough to peak), the firing rate, and the area under the hyperpolarization phase of the spike (50). Further analyses were performed as for Ca^{2+} traces using z-score–transformed spike trains binned in 50-ms bins.

Histology

Before transcardial perfusion with 4% paraformaldehyde in PBS, an electrolytic lesion was made at the electrode tip by applying 0.1 μA for 10 s to one of the electrode wires in deeply anesthetized mice to mark the recording site. Brains were postfixed in 4% paraformaldehyde for at least 2 h at 4°C and cut in 80-μm coronal slices using a vibratome (VT1000S). Sections containing the BLA were washed three times for 10 min in PBS, exposed to DAPI (1:10,000, Sigma-Aldrich) for 10 min, and washed again three times for 10 min in PBS before mounting on glass slides. To verify the electrode placement, sections were scanned with an Axioscan Z1 slide scanner (Carl Zeiss AG), equipped with a 10× air objective (Plan-Apochromat 10×/0.45). Electrode placements were matched against a mouse brain atlas (48).

Optogenetic experiments

The experimenter was blinded to each animal's experimental cohort (virus condition, GFP, or ArchT). Animals were randomly allocated to experimental groups and were later identified by unique markers for group assignment. Before behavioral experiments, all mice were habituated to the optical fiber connection procedure by handling and brief head restraining for at least 3 days. On the experimental days, implanted fibers (fiber numerical aperture of 0.48, fiber inner core diameter of 200 μm; Thorlabs) were connected to a custom-built laser bench (Life Imaging Services) with custom fiber patch cables. An acousto-optic modulator (AA Opto-Electronic) controlled the laser intensity (MGL-F-589, 589-nm wavelength, CNI Lasers). Laser power at the fiber tip was measured before every subject with an optical power and energy meter (PM100D, ThorLabs) and adjusted to reach an irradiance value of ~4 mW at fiber tip.

Closed-loop optogenetic sessions

To inhibit BLA PNs during action periods, the laser was switched on from the first action until the last action of each action period. To inhibit BLA PNs during consumption periods,

the laser was switched on from the first lick after outcome delivery until the last rewarded lick of each consumption period. Licks were detected online, and rewarded licks were defined as the first lick bout (series of two or more licks, with inter-lick intervals <1 s) after outcome delivery. Maximal laser duration was fixed at 20 s. Laser was controlled by a multi-I/O processor (RZ6, Tucker Davis). The laser was switched on during individual action or consumption periods from the ninth to the 16th behavioral sequences, called the laser-ON period. The first to the eighth behavioral sequences were called the laser-OFF period (Fig. 4 and figs. S10 and S11). For extinction tests, to inhibit BLA PNs during action epochs the laser was switched on from the first action until the last action of each action epoch. Maximal laser duration was fixed at 4 s. Laser was delivered for each action epochs during the first 4 min of the test (laser-on period). No laser was delivered during the last 4 min of the test (laser-off period; fig. S12).

Two-sucrose choice challenge

Two-sucrose choice challenge was performed in a behavioral context measured 26 cm L × 25 cm W × 40 cm H. The context contained two lick ports located on the same wall and separated by 10 cm. For 12 min, food restriction mice freely chose between 5 and 20% sucrose solutions delivered from different lick ports. Licks triggered outcome delivery at slow rate (2 μl/s). Mice showed sustained licking bouts during the 12 min of the test (fig. S11M). The behavioral session was divided in two phases: In the first 6 min, the laser was off and in the last 6 min, each lick bouts of the 20% sucrose solution was paired with light. Laser was controlled by a multi-I/O processor (RZ6, Tucker Davis).

Real-time place avoidance paradigm

The real-time place avoidance experiment was performed in a behavioral context composed of two compartments (20 × 20 cm each) connected by an alleyway (5 × 5 cm). The two compartments differed in shape (circular and square) and visually (gray Plexiglas with black horizontal stripes or green Plexiglas). Mouse position was tracked outline using Cineplex software (Plexon). On day 1, mice were allowed to freely explore the entire behavioral context during a 12-min pre-exposure period. After pre-exposure, the compartment in which the mice spent the most time was designated as the most visited compartment. On day 2, mice were submitted to a 12-min test session during which light pulses (1-s pulse width, repeated at 1 Hz) were delivered while the mice occupied the most visited compartment but not when they occupied the less-visited compartment (fig. S11L). The laser was controlled by a multi-I/O processor (RZ6, Tucker Davis).

Histology

Mice were transcardially perfused (as described above) and optical fibers removed. Brains were postfixed in 4% paraformaldehyde for at least 2 h at 4°C and cut into 80-μm coronal slices using a vibratome (VT1000S). Sections containing the BLA were immediately mounted on glass slides and coverslipped. To verify the specificity of viral expression and fiber tip placement, sections were scanned with an Axioscan Z1 slide scanner (Carl Zeiss AG), equipped with a 10× air objective (Plan-Apochromat 10×/0.45). Fiber tip placements were matched against a mouse brain atlas (*48*). Mice were excluded from the analysis post hoc if they did not show bilateral expression of the virus, if virus expression (cell bodies expressing GFP) was detected outside of the BLA, or if they did not exhibit correct fiber placement (<300 μm away from the BLA; figs. S10 to S12). Twenty-four of 104 mice were excluded.

Data analysis

We used custom routines written in MATLAB (The MathWorks) to perform all data analyses.

Identification of task-modulated neurons

We identified, independently for each outcome, three functional types of neurons showing significant and consistent increase activity at action and/or consumption times: action neurons, consumption neurons, and transition neurons. We first collected the Ca²⁺ traces in a time window around action/consumption time stamps (see below). Neurons were classified as task modulated if the maximum of their average response in a more restricted reference time window fulfilled three criteria: (i) it exceeded a threshold on the maxima obtained for each neuron by a bootstrapping procedure using the same time window around random time stamps along the session (number of random time stamps matched the number of real time stamps, 1000 iterations, threshold set at the 99 percentile); (ii) it exceeded the average response outside of the restricted reference window; (iii) it exceeded the upper limit of the 95% confidence interval around zero (a z-score value of 0), computed using Student's t distribution with the empirical mean and standard error of the maxima of individual traces. To characterize BLA PNs as action neurons, we used a time window of 2 s before and 4 s after the onset of one of the following actions: (i) the first action in an action period (after a rewarded lick epoch), (ii) the last action before an unrewarded lick, or (iii) the last action in an action period. This allowed us to capture neurons locked to the distinct actions in action periods. In terms of behavior, these windows isolate the stereotypical behavioral switches between actions and licks (Figs. 1F and 2A and fig. S3A). The re-

stricted reference time window used for action neurons started 0.5 s before action onsets and lasted the duration of the shortest action epoch (determined per mouse; no shorter than 0.8 s). To characterize BLA PNs as consumption neurons, we used a time window of 2 s before and 8 s after the start of consumption periods (see the description of behavior in this time window in Fig. 2A and fig. S3B). The restricted reference time window used for consumption neurons started 0.1 s before consumption onset and lasted the duration of the shortest consumption epoch (determined per mouse). BLA PNs characterized as both action and consumption neurons were classified as transition neurons. During outcome value and action-outcome contingency violation sessions, task-modulated neurons were identified by taking into account time stamps of behavioral events occurring either during the initial or the perturbed periods.

Coincident activity between task-modulated neurons

For Ca²⁺ data, we used correlation and coactivation analysis between pairs of action and/or consumption neurons in high-activity frames (z-scored traces values <2 were filtered to 0) to assess the intrafunctional subset and interfunctional subset coincident activity independently for OFF and ON task phases. Correlation between neuron pairs was evaluated by using Pearson's correlation coefficient (r). We determined the proportion of noncorrelated, positively correlated, and negatively correlated neuron pairs ($P < 0.01$) for action-action, consumption-consumption, and action-consumption pairs. Then, we determined whether the proportions of correlated neurons were different from those obtained for random neuronal pairs (proportion bigger than the 99% percentile of proportions obtained with random pairs). The coactivation index for each pair of action and/or consumption neurons was calculated as the ratio of co-occurring high-activity frames to the total number of high-activity frames (*51*). For single-unit data, correlation and co-firing analyses between pairs of action and/or consumption neurons were performed after binning the spike train in 50-ms bins.

Decoder

We trained a linear discriminant analysis (LDA) classifier with fivefold cross-validation on the distinction between five behavioral epochs: action periods (left/right), consumption periods (milk/sucrose), and periods of non-task-related behavior during the OFF task phase (unrewarded licks epochs during the OFF task phase were excluded). Neuronal responses and behavior were binned in 200-ms bins, and bins with other behavioral descriptions were excluded from the analysis.

To address the unbalanced design (more non-task-related epochs) we used the F score calculated from the confusion matrices as a measure of the decoder performance. We repeated the analysis with the behavioral labels shuffled in relation to the real neuronal data (control, 100 repetitions) using only the responses of the task-modulated neurons (modulated) or random selections of non-task-modulated neurons with a sample size that matched the number of task-modulated neurons for each animal (non-modulated, 100 repetitions).

Pairwise correlation between population activity patterns

We binned the single neuron activity in 200-ms bins to obtain a single population activity vector per time point and quantified the similarity between two activity vectors by the pairwise Pearson's correlation coefficient (r). For the mean correlation analysis presented in Fig. 3C, we calculated intrabehavioral epoch mean correlations by first collecting the activity vectors in all the epochs of a specific behavior (two action epochs and two rewarded lick epochs) and then calculating the average pairwise Pearson's correlation between them (Fig. 3C). The mean on-on intra-epoch correlation was then calculated by averaging the four specific behavior intra-epoch correlations per mouse. Inter-epoch mean correlations were calculated per mouse by averaging the pairwise correlation between activity vectors during OFF task and ON task (off-on) or between on activity vectors during different ON task behaviors (on-on).

To track the consistency of the activity patterns along the session, we used as reference patterns the mean activity in two selections of time points: 24 OFF task time bins (every 10 s in the 4 min before the first ON task phase) and all time bins in three action/consumption epochs (behavioral sequences 3 to 5). We then calculated the pairwise correlations between the activity vectors during each action-consumption epoch and these two reference patterns (Fig. 3D).

To estimate the persistence of action- and consumption-associated patterns along the behavioral sequence and their relationship to behavior (Fig. 3, F and G), we selected as a reference for each sequence a representative action and consumption activity vector (time of the first action in an action epoch to follow action pattern consistency along the action period or 2 s after the first lick in a consumption period to differentiate from unrewarded lick epochs). We then marked all bins in the sequence with activity vectors that had a positive correlation with the reference (the lower bound of the 99% confidence interval of the Pearson's correlation >0). The same procedure was performed with single-unit recordings and inferred spikes after binning the spike trains in 50-ms bins (figs. S8 and S9).

Transitions in action- and consumption-associated patterns were detected by k-means clustering (n clusters = 2) of the mean pattern correlation matrices obtained by first averaging the pairwise correlations between the mean activity vectors during epochs in different sequences and then averaging the resulting correlation matrices ($N = 4/5$ mice, for outcome value/action-outcome contingency violation sessions, respectively, × 2 outcomes).

Dimensionality reduction

We used tSNE (tsne.m function from the MATLAB's Statistics Toolbox with perplexity = 60 and learning rate = 600) with correlation as the distance measure to visually demonstrate the distinct clusters of neuronal activity patterns and their relation to behavior.

Statistics

Statistical analysis was performed with MATLAB (The MathWorks). Normality of the data was assessed using Shapiro-Wilk test and either parametric (paired t) or nonparametric (Wilcoxon signed-rank or rank-sum) tests were used. In the figures, box-and-whisker plots indicate median (vertical line), interquartile (horizontal thick line), extreme data values (horizontal thin line), and outliers (plus sign) of the data distribution. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those generally used in the field. Statistical tests are mentioned in the figure legends (ns indicates $P > 0.05$; $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$).

REFERENCES AND NOTES

1. A. Dickinson, B. Balleine, Motivational control of goal-directed action. *Anim. Learn. Behav.* **22**, 1–18 (1994). doi: 10.3758/BF03199951
2. B. W. Balleine, The meaning of behavior: Discriminating reflex and volition in the brain. *Neuron* **104**, 47–62 (2019). doi: 10.1016/j.neuron.2019.09.024; pmid: 31600515
3. P. H. Janak, K. M. Tye, From circuits to behaviour in the amygdala. *Nature* **517**, 284–292 (2015). doi: 10.1038/nature14188; pmid: 25592533
4. K. M. Tye, P. H. Janak, Amygdala neurons differentially encode motivation and reinforcement. *J. Neurosci.* **27**, 3937–3945 (2007). doi: 10.1523/JNEUROSCI.5281-06.2007; pmid: 17428967
5. G. D. Stuber *et al.*, Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475**, 377–380 (2011). doi: 10.1038/nature10194; pmid: 21716290
6. J. J. Paton, M. A. Belova, S. E. Morrison, C. D. Salzman, The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006). doi: 10.1038/nature04490; pmid: 16482160
7. G. Schoenbaum, A. A. Chiba, M. Gallagher, Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* **1**, 155–159 (1998). doi: 10.1038/407; pmid: 10195132
8. F. Ambroggi, A. Ishikawa, H. L. Fields, S. M. Nicola, Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* **59**, 648–661 (2008). doi: 10.1016/j.neuron.2008.07.004; pmid: 18760700
9. A. Servonnet, G. Hernandez, C. El Hage, P. P. Rompré, A. N. Samaha, Optogenetic activation of the basolateral amygdala promotes both appetitive conditioning and the instrumental pursuit of reward cues. *J. Neurosci.* **40**, 1732–1743 (2020). doi: 10.1523/JNEUROSCI.2196-19.2020; pmid: 31953370
10. S. B. Ostlund, B. W. Balleine, Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *J. Neurosci.* **28**, 4398–4405 (2008). doi: 10.1523/JNEUROSCI.5472-07.2008; pmid: 18434518
11. D. A. Simmons, D. B. Neill, Functional interaction between the basolateral amygdala and the nucleus accumbens underlies incentive motivation for food reward on a fixed ratio schedule. *Neuroscience* **159**, 1264–1273 (2009). doi: 10.1016/j.neuroscience.2009.01.026; pmid: 19344638
12. R. Chesworth, L. Corbit, "The contribution of the amygdala to reward-related learning and extinction," in *The Amygdala: Where Emotions Shape Perception, Learning and Memories*, B. Ferry, Ed. (IntechOpen, 2017); https://www.intechopen.com/chapters/54509.
13. B. W. Balleine, A. S. Killcross, A. Dickinson, The effect of lesions of the basolateral amygdala on instrumental conditioning. *J. Neurosci.* **23**, 666–675 (2003). doi: 10.1523/JNEUROSCI.23-02-00666.2003; pmid: 12533626
14. S. E. Rhodes, E. A. Murray, Differential effects of amygdala, orbital prefrontal cortex, and prelimbic cortex lesions on goal-directed behavior in rhesus macaques. *J. Neurosci.* **33**, 3380–3389 (2013). doi: 10.1523/JNEUROSCI.4374-12.2013; pmid: 23426666
15. A. W. Johnson, M. Gallagher, P. C. Holland, The basolateral amygdala is critical to the expression of pavlovian and instrumental outcome-specific reinforcer devaluation effects. *J. Neurosci.* **29**, 696–704 (2009). doi: 10.1523/JNEUROSCI.3758-08.2009; pmid: 19158296
16. C. B. Ferster, B. F. Skinner, *Schedules of Reinforcement* (Appleton-Century-Crofts, 1957).
17. M. R. Roesch, D. J. Calu, G. R. Esber, G. Schoenbaum, Neural correlates of variations in event processing during learning in basolateral amygdala. *J. Neurosci.* **30**, 2464–2471 (2010). doi: 10.1523/JNEUROSCI.5781-09.2010; pmid: 20164330
18. I. Hernádi, F. Grabenhorst, W. Schultz, Planning activity for internally generated reward goals in monkey amygdala neurons. *Nat. Neurosci.* **18**, 461–469 (2015). doi: 10.1038/nn.3925; pmid: 25622146
19. W. Schultz, Multiple reward signals in the brain. *Nat. Rev. Neurosci.* **1**, 199–207 (2000). doi: 10.1038/35044563; pmid: 11257908
20. Y. Niv, D. Joel, P. Dayan, A normative perspective on motivation. *Trends Cogn. Sci.* **10**, 375–381 (2006). doi: 10.1016/j.tics.2006.06.010; pmid: 16843041
21. H. Nishijo, T. Ono, H. Nishino, Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J. Neurosci.* **8**, 3570–3583 (1988). doi: 10.1523/JNEUROSCI.08-10-03570.1988; pmid: 3193171
22. A. Fontanini, S. E. Grossman, J. A. Figueroa, D. B. Katz, Distinct subtypes of basolateral amygdala taste neurons reflect palatability and reward. *J. Neurosci.* **29**, 2486–2495 (2009). doi: 10.1523/JNEUROSCI.3898-08.2009; pmid: 19244523
23. M. A. Bermudez, W. Schultz, Reward magnitude coding in primate amygdala neurons. *J. Neurophysiol.* **104**, 3424–3432 (2010). doi: 10.1152/jn.00540.2010; pmid: 20861431
24. B. W. Balleine, S. Killcross, Parallel incentive processing: An integrated view of amygdala function. *Trends Neurosci.* **29**, 272–279 (2006). doi: 10.1016/j.tins.2006.03.002; pmid: 16545468
25. J. Gründemann *et al.*, Amygdala ensembles encode behavioral states. *Science* **364**, eaav8736 (2019). doi: 10.1126/science.aav8736; pmid: 31000636
26. P. Namburi *et al.*, A circuit mechanism for differentiating positive and negative associations. *Nature* **520**, 675–678 (2015). doi: 10.1038/nature14366; pmid: 25925480
27. J. Kim, M. Pignatelli, S. Xu, S. Itohara, S. Tonegawa, Antagonistic negative and positive neurons of the basolateral amygdala. *Nat. Neurosci.* **19**, 1636–1646 (2016). doi: 10.1038/nn.4414; pmid: 27749826
28. X. Zhang, B. Li, Population coding of valence in the basolateral amygdala. *Nat. Commun.* **9**, 5195 (2018). doi: 10.1038/s41467-018-07679-9; pmid: 30518754
29. P. Kyriazi, D. B. Headley, D. Pare, Multi-dimensional coding by basolateral amygdala neurons. *Neuron* **99**, 1315–1328.e5 (2018). doi: 10.1016/j.neuron.2018.07.036; pmid: 30146300
30. W. Zhang *et al.*, Functional circuits and anatomical distribution of response properties in the primate amygdala. *J. Neurosci.* **33**, 722–733 (2013). doi: 10.1523/JNEUROSCI.2970-12.2013; pmid: 23303950
31. X. Zhang, J. Kim, S. Tonegawa, Amygdala reward neurons form and store fear extinction memory. *Neuron* **105**, 1077–1093.e7 (2020). doi: 10.1016/j.neuron.2019.12.025; pmid: 31952856

32. C. Herry *et al.*, Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600–606 (2008). doi: 10.1038/nature07166; pmid: 18615015

33. S. Duvarci, D. Pare, Amygdala microcircuits controlling learned fear. *Neuron* **82**, 966–980 (2014). doi: 10.1016/j.neuron.2014.04.042; pmid: 24908482

34. S. Krabbe *et al.*, Adaptive disinhibitory gating by VIP interneurons permits associative learning. *Nat. Neurosci.* **22**, 1834–1843 (2019). doi: 10.1038/s41593-019-0508-y; pmid: 31636447

35. A. A. Grace, S. B. Floresco, Y. Goto, D. J. Lodge, Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci.* **30**, 220–227 (2007). doi: 10.1016/j.tins.2007.03.003; pmid: 17400299

36. J. A. da Silva, F. Tecuapetla, V. Paixão, R. M. Costa, Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018). doi: 10.1038/nature25457; pmid: 29420469

37. M. Malvaez, C. Shieh, M. D. Murphy, V. Y. Greenfield, K. M. Wassum, Distinct cortical-amygdala projections drive reward value encoding and retrieval. *Nat. Neurosci.* **22**, 762–769 (2019). doi: 10.1038/s41593-019-0374-7; pmid: 30962632

38. K. Lavi, G. A. Jacobson, K. Rosenblum, A. Lüthi, Encoding of conditioned taste aversion in cortico-amygdala circuits. *Cell Rep.* **24**, 278–283 (2018). doi: 10.1016/j.celrep.2018.06.053; pmid: 29996089

39. S. L. Parkes, B. W. Balleine, Incentive memory: Evidence the basolateral amygdala encodes and the insular cortex retrieves outcome values to guide choice between goal-directed actions. *J. Neurosci.* **33**, 8753–8763 (2013). doi: 10.1523/JNEUROSCI.5071-12.2013; pmid: 23678118

40. K. Yoshida, M. R. Drew, M. Mimura, K. F. Tanaka, Serotonin-mediated inhibition of ventral hippocampus is required for sustained goal-directed behavior. *Nat. Neurosci.* **22**, 770–777 (2019). doi: 10.1038/s41593-019-0376-5; pmid: 30988523

41. J. Kim, X. Zhang, S. Muralidhar, S. A. LeBlanc, S. Tonegawa, Basolateral to central amygdala neural circuits for appetitive behaviors. *Neuron* **93**, 1464–1479.e5 (2017). doi: 10.1016/j.neuron.2017.02.034; pmid: 28334609

42. G. Hart, B. K. Leung, B. W. Balleine, Dorsal and ventral streams: The distinct role of striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiol. Learn. Mem.* **108**, 104–118 (2014). doi: 10.1016/j.nlm.2013.11.003; pmid: 24231424

43. L. H. Corbit, B. K. Leung, B. W. Balleine, The role of the amygdala-striatal pathway in the acquisition and performance of goal-directed instrumental actions. *J. Neurosci.* **33**, 17682–17690 (2013). doi: 10.1523/JNEUROSCI.3271-13.2013; pmid: 24198361

44. G. Lopes *et al.*, Bonsai: An event-based framework for processing and controlling data streams. *Front. Neuroinform.* **9**, 7 (2015). doi: 10.3389/fninf.2015.00007; pmid: 25904861

45. M. Guizar-Sicairos, S. T. Thurman, J. R. Fienup, Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008). doi: 10.1364/OL.33.000156; pmid: 18197224

46. P. Zhou *et al.*, Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018). doi: 10.7554/eLife.28728; pmid: 29469809

47. G. Corder *et al.*, An amygdalar neural ensemble that encodes the unpleasantness of pain. *Science* **363**, 276–281 (2019). doi: 10.1126/science.aap8586; pmid: 30655440

48. G. Paxinos, K. Franklin, K. Franklin, *The Mouse Brain in Stereotaxic Coordinates* (Academic, ed. 2, 2001).

49. B. Engelhard *et al.*, Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019). doi: 10.1038/s41586-019-1261-9; pmid: 31142844

50. J. Courtin *et al.*, Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. *Nature* **505**, 92–96 (2014). doi: 10.1038/nature12755; pmid: 24256726

51. N. Karalis *et al.*, 4-Hz oscillations synchronize prefrontal-amygdala circuits during fear behavior. *Nat. Neurosci.* **19**, 605–612 (2016). doi: 10.1038/nn.4251; pmid: 26878674

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg7277
Figs. S1 to S18
MDAR Reproducibility Checklist

CORONAVIRUS

Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial

Peter B. Gilbert[1,2,3]*†, David C. Montefiori[4]†, Adrian B. McDermott[5]†, Youyi Fong[1,2], David Benkeser[6], Weiping Deng[7], Honghong Zhou[7], Christopher R. Houchens[8], Karen Martins[8], Lakshmi Jayashankar[8], Flora Castellino[8], Britta Flach[5], Bob C. Lin[5], Sarah O'Connell[5], Charlene McDanal[4], Amanda Eaton[4], Marcella Sarzotti-Kelsoe[4], Yiwen Lu[1], Chenchen Yu[1], Bhavesh Borate[1], Lars W. P. van der Laan[1], Nima S. Hejazi[1,9], Chuong Huynh[8], Jacqueline Miller[7], Hana M. El Sahly[10], Lindsey R. Baden[11], Mira Baron[12], Luis De La Cruz[13], Cynthia Gay[14], Spyros Kalams[15], Colleen F. Kelley[16], Michele P. Andrasik[1], James G. Kublin[1], Lawrence Corey[1,17], Kathleen M. Neuzil[18], Lindsay N. Carpp[1], Rolando Pajon[7], Dean Follmann[19], Ruben O. Donis[8]‡, Richard A. Koup[5]‡, on behalf of the Immune Assays Team§, Moderna, Inc. Team§, Coronavirus Vaccine Prevention Network (CoVPN)/ Coronavirus Efficacy (COVE) Team§, and United States Government (USG)/ CoVPN Biostatistics Team§

In the coronavirus efficacy (COVE) phase 3 clinical trial, vaccine recipients were assessed for neutralizing and binding antibodies as correlates of risk for COVID-19 disease and as correlates of protection. These immune markers were measured at the time of second vaccination and 4 weeks later, with values reported in standardized World Health Organization international units. All markers were inversely associated with COVID-19 risk and directly associated with vaccine efficacy. Vaccine recipients with postvaccination 50% neutralization titers 10, 100, and 1000 had estimated vaccine efficacies of 78% (95% confidence interval, 54 to 89%), 91% (87 to 94%), and 96% (94 to 98%), respectively. These results help define immune marker correlates of protection and may guide approval decisions for messenger RNA (mRNA) COVID-19 vaccines and other COVID-19 vaccines.

On the basis of their demonstrated efficacy to prevent COVID-19 in phase 3 clinical trials, to date, seven COVID-19 vaccines have been granted an emergency use listing by the World Health Organization (WHO) (1), three have been granted an emergency use authorization (EUA) by the US Food and Drug Administration (FDA) (2), and one has been formally approved by the FDA (3). However, the manufacturing challenges posed by the global demand for doses, the need for affordable and accessible options that are safe and effective in diverse populations, the current lack of efficacy data in certain populations (e.g., pediatrics, pregnant women, and autoimmune or immunocompromised individuals), and the emergence of more-transmissible viral variants all highlight the need for a large armamentarium of safe and effective COVID-19 vaccines (4, 5).

The coronavirus efficacy (COVE) phase 3 trial (NCT04470427) of the mRNA-1273 COVID-19 vaccine, which is being conducted in the US in adults aged 18 and over, showed estimated vaccine efficacy against COVID-19 of 94% in the primary analysis (6). These efficacy data supported the FDA's EUA of mRNA-1273 for the prevention of COVID-19 in adults (7). The mRNA-1273 vaccine has been shown to be highly effective in the elderly and in essential and frontline workers, including health care workers (8), and to have noninferior binding and neutralizing antibody responses in adolescents versus adults (9).

Correlates of protection, which are immunological markers that can be used to reliably predict the level of vaccine efficacy against a clinically relevant end point, such as COVID-19 (10–12), are highly sought in vaccine research. The identification and validation of a correlate of protection would expedite the clinical evaluation and regulatory approval process for existing vaccines for new populations, for vaccine regimen modifications, and for new

vaccines. Neutralizing antibodies (nAbs) or binding antibodies (bAbs) have been established as a correlate of protection for vaccines against many viral diseases (11). The hypothesis that antibodies, whether elicited by infection or by spike protein–based vaccines, are a correlate of protection against COVID-19 is supported by diverse lines of evidence (13–25). For the mRNA-1273 vaccine, multiple severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibody markers—including immunoglobulin G (IgG) bAbs to the spike protein, IgG bAbs to the spike receptor-binding domain (RBD), and 50% inhibitory dilution (ID_{50}) nAb titer—correlated with protection against SARS-CoV-2 replication after challenge in vaccinated rhesus macaques (24). Here, we assessed these same SARS-CoV-2 antibody markers as well as an 80% inhibitory dilution (ID_{80}) nAb titer as correlates of risk of COVID-19 and as correlates of mRNA-1273 vaccine protection against COVID-19 in the COVE trial.

Participant demographics

Table S1 describes demographics of the randomly sampled immunogenicity subcohort ($N = 1010$ vaccine, $N = 137$ placebo). Thirty-four percent of baseline SARS-CoV-2–negative per-protocol participants were age 65 or over, 40% were deemed to be at risk for severe COVID-19 illness (referred to as "at risk"), 47% were assigned female sex at birth, 32% were Hispanic or Latino, 46% were white and non-Hispanic, and 54% were from communities of color, with 18% Black or African American. Table S2 and figs. S1 and S2 describe the day 29 marker case-cohort set and the day 57 marker case-cohort set, which augment the immunogenicity subcohort with all vaccine breakthrough COVID-19 end point cases and make up the sets of participants included in the analyses of antibody markers measured at day 29 or day 57 as correlates, respectively.

COVID-19 end points

Analyses of day 29 and day 57 antibody markers as correlates included vaccine breakthrough COVID-19 end points starting 7 days after day 29 ($n = 46$) and after day 57 ($n = 36$), respectively (fig. S3). Average follow-up of vaccine recipients was 116 days after day 29 and 88 days after day 57. All immune correlates analyses were prespecified, as detailed in the supplementary file Statistical Analysis Plan (SAP).

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]Department of Biostatistics, University of Washington, Seattle, WA, USA. [4]Department of Surgery and Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC, USA. [5]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [6]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [7]Moderna, Inc., Cambridge, MA, USA. [8]Biomedical Advanced Research and Development Authority, Washington, DC, USA. [9]Division of Biostatistics, School of Public Health, University of California Berkeley, Berkeley, CA, USA. [10]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA. [11]Brigham and Women's Hospital, Boston, MA, USA. [12]Palm Beach Research Center, West Palm Beach, FL, USA. [13]Keystone Vitalink Research, Greenville, SC, USA. [14]Department of Medicine, Division of Infectious Diseases, UNC HIV Cure Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [15]Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [16]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine and the Grady Health System, Atlanta, GA, USA. [17]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [18]Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA. [19]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.
*Corresponding author. Email: pgilbert@fredhutch.org †These authors contributed equally to this work. ‡These authors contributed equally to this work. §The members of the Immune Assays Team; Moderna, Inc. Team; CoVPN/COVE Team; and USG/CoVPN Biostatistics Team and their affiliations are listed in the supplementary materials.
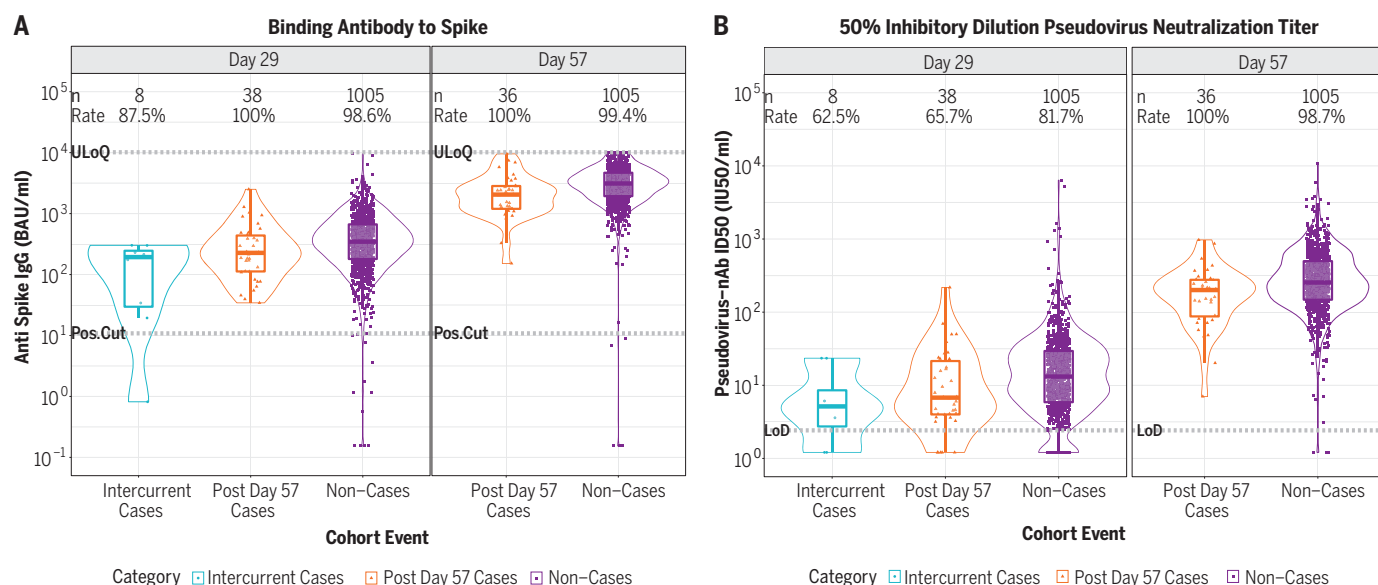
A — Binding Antibody to Spike

B — 50% Inhibitory Dilution Pseudovirus Neutralization Titer

Fig. 1. Anti-spike IgG concentration and pseudovirus neutralization ID$_{50}$ titer by COVID-19 outcome status. (**A**) Anti-spike IgG concentration. (**B**) Pseudovirus neutralization ID50 titer. Data points are from baseline-negative per-protocol vaccine recipients in the day 29 marker or day 57 marker case-cohort set. The violin plots contain interior box plots with upper and lower horizontal edges representing the 25th and 75th percentiles of antibody level and middle line representing the 50th percentile. The vertical bars represent the distance from the 25th (or 75th) percentile of antibody level and the minimum (or maximum) antibody level within the 25th (or 75th) percentile of antibody level minus (or plus) 1.5 times the interquartile range. Each side shows a rotated probability density (estimated by a kernel density estimator with a default Gaussian kernel) of the data. Positive response rates were computed with inverse probability of sampling weighting. Pos.Cut, positivity cut-off; LoD, limit of detection; ULoQ, upper limit of quantitation. ULOQ = 10,919 for ID$_{50}$ (above all data points). Positive response for spike IgG was defined by IgG > 10.8424 BAU/ml. Positive response for ID$_{50}$ was defined by value > LOD (2.42). Post–day 57 cases are COVID-19 end points starting 7 days after day 57 through the end of the blinded follow-up (last COVID-19 end point was 126 days after dose 2); intercurrent cases are COVID-19 end points starting 7 days after day 29 through 6 days after day 57.

COVE follows participants for 2 years, which will enable future analyses of how the current level of antibodies correlates with instantaneous risk of COVID-19. Such analyses may inform how vaccine efficacy wanes as antibody levels wane and as new variants emerge, which in turn may inform decisions about the timing of a potential third dose of vaccination and/or the need to update vaccine composition (*26*).

Antibody marker levels are lower in vaccine recipient cases versus noncases

At day 57, almost 100% of vaccine recipients had positive or detectable antibody response by all four markers (Table 1; table S3 shows assay limits for each marker). This was also true at day 29 for spike IgG and RBD IgG, whereas ID$_{50}$ and ID$_{80}$ titers were detectable in 82 and 64% of vaccine recipients, respectively. Each marker was moderately correlated between the day 29 and day 57 time points [Spearman rank correlation coefficient (*r*) = 0.53 to 0.62; fig. S4]. Together, the spike IgG and RBD IgG markers were tightly correlated (Spearman rank *r* = 0.94 and 0.97 at days 29 and 57, respectively; figs. S5 and S6) as were the ID$_{50}$ and ID$_{80}$ markers (*r* = 0.97 and 0.96 at days 29 and 57, respectively; figs. S5 and S6). Accordingly, some results focus on spike IgG and ID$_{50}$. Each bAb marker was correlated with each neutralization marker at each time point (*r* = 0.73 to 0.80).

Figure 1 and fig. S7 show the day 29 and day 57 marker distributions by case or noncase status in vaccine recipients (fig. S8 in placebo recipients), and figs. S9 and S10 show marker values by participant age. For all eight markers, the geometric mean was lower in vaccine breakthrough cases than in vaccine recipient noncases, with geometric mean ratios (cases/noncases) and their 95% confidence interval (CI) upper bounds all <1 (Table 1).
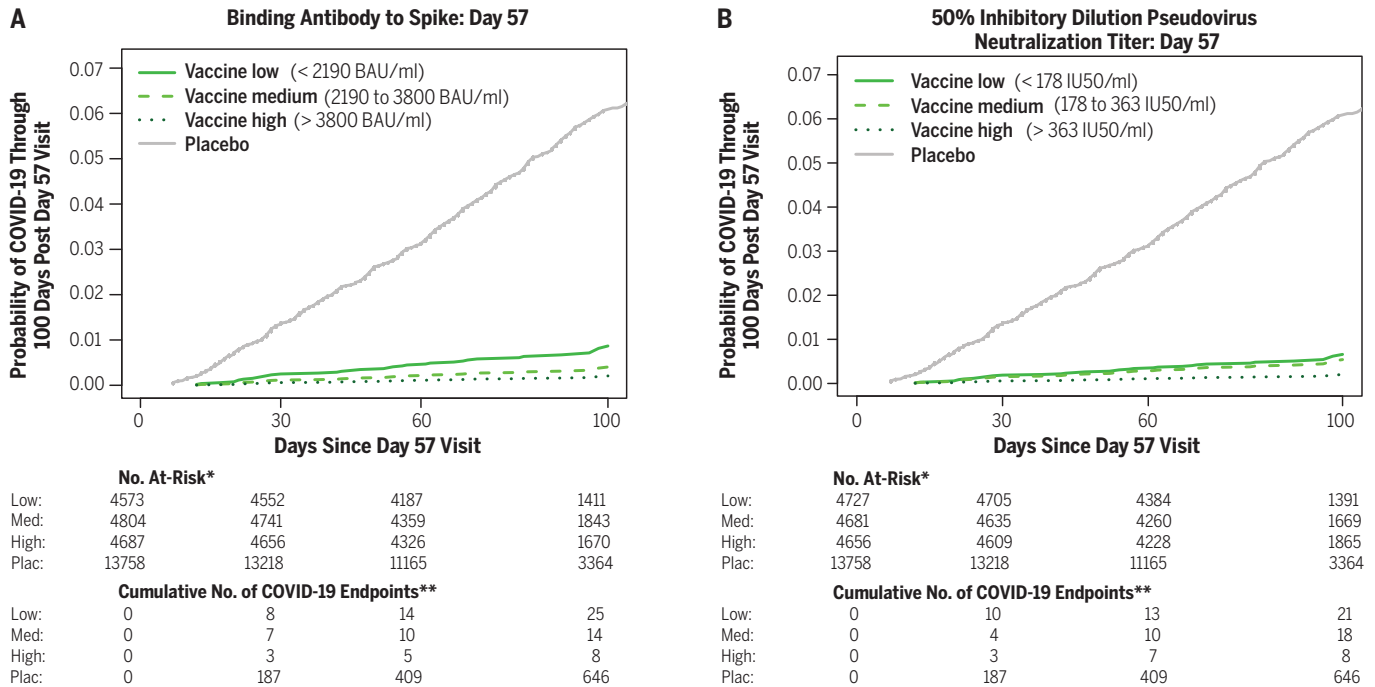
Figures S11 and S12 show reverse cumulative distribution function curves of the eight markers, in the context of the overall vaccine efficacy estimates (*27*). Figure S13 shows the day 29 and/or day 57 marker values of vaccine breakthrough cases by timing of COVID-19 end point diagnosis.

COVID-19 risk of vaccine recipients decreases as antibody marker levels increase

Figure 2 shows Cox model–based covariate-adjusted COVID-19 cumulative incidence curves for subgroups of vaccine recipients defined by tertile of day 57 IgG spike or ID$_{50}$ (Fig. 2, A and B). Corresponding results for IgG RBD and ID$_{80}$ are shown in fig. S14. (Details on covariate adjustment are given in the supplementary text, section S1; tables S4 to S7; and figs. S15 and S16.) Multiplicity-adjusted *P* values indicated significant inverse correlations with risk, with estimated hazard ratios for upper versus lower tertiles ranging between 0.20 and 0.31 (Fig. 2C). For quantitative day 57 markers, the estimated hazard ratio per 10-fold increase in marker value ranged between 0.35 and 0.66 (Fig. 3A), with multiplicity-adjusted *P* values indicating significant associations. Generally, similar results were obtained across prespecified vaccine recipient subgroups (Fig. 3, B and C, and fig. S17).

The four markers at day 29 were also significant inverse correlates of risk, with estimated hazard ratios for upper versus lower tertiles ranging between 0.19 and 0.32 (figs. S18 and S19) and estimated hazard ratios per 10-fold increase in marker value ranging between 0.19 and 0.54 (fig. S17). *P* values were smaller for day 29 markers than for day 57 markers, which indicates strengthened evidence for correlates of risk. If a day 29 immune marker in recipients of two mRNA-1273 doses becomes established as a correlate of protection, it could be a more practical surrogate marker than a day 57 marker. Notably, all participants in our correlates analysis received both dose 1 and dose 2, so the day 29 correlates results reflect the full effect of the two vaccine doses used in clinical practice.

A Binding Antibody to Spike: Day 57

B 50% Inhibitory Dilution Pseudovirus Neutralization Titer: Day 57

No. At-Risk*

Low:	4573	4552	4187	1411
Med:	4804	4741	4359	1843
High:	4687	4656	4326	1670
Plac:	13758	13218	11165	3364

Cumulative No. of COVID-19 Endpoints**

Low:	0	8	14	25
Med:	0	7	10	14
High:	0	3	5	8
Plac:	0	187	409	646

No. At-Risk*

Low:	4727	4705	4384	1391
Med:	4681	4635	4260	1669
High:	4656	4609	4228	1865
Plac:	13758	13218	11165	3364

Cumulative No. of COVID-19 Endpoints**

Low:	0	10	13	21
Med:	0	4	10	18
High:	0	3	7	8
Plac:	0	187	409	646

*No. At-Risk = estimated number in the population for analysis: baseline negative per-protocol vaccine recipients not experiencing the COVID-19 endpoint through 6 days post Day 57 visit.
**Cumulative No. of COVID-19 Endpoints = estimated cumulative number of this cohort with a COVID-19 endpoint.

C

COVE Immunologic Marker	Tertile[¶]	No. Cases/ No. At-Risk[§]	Attack rate	Hazard Ratio Pt. Est.	Hazard Ratio 95% CI	P-Value (2-Sided)	Overall P-value	FDR-adjusted p-value[†]	FWER-adjusted p-value[†]
Anti Spike IgG (BAU/ml)	Low	25/4,573	0.0055	1	N/A	N/A	0.006	0.014	0.010
	Medium	14/4,804	0.0029	0.45	(0.20, 1.01)	0.053			
	High	8 4,687	0.0017	0.23	(0.09, 0.60)	0.002			
Anti RBD IgG (BAU/ml)	Low	25/4,620	0.0054	1	N/A	N/A	0.009	0.014	0.014
	Medium	13/4,745	0.0027	0.45	(0.20, 1.01)	0.052			
	High	9/4,699	0.0019	0.28	(0.12, 0.67)	0.004			
Pseudovirus-nAb ID50 (IU50/ml)	Low	21/4,727	0.0044	1	N/A	N/A	0.052	0.042	0.054
	Medium	18/4,681	0.0038	0.82	(0.39, 1.72)	0.599			
	High	8/4,656	0.0017	0.31	(0.12, 0.80)	0.016			
Pseudovirus-nAb ID80 (IU80/ml)	Low	20/4,742	0.0042	1	N/A	N/A	0.012	0.014	0.015
	Medium	22/4,715	0.0047	1.00	(0.49, 2.03)	1.000			
	High	5/4,607	0.0011	0.20	(0.07, 0.61)	0.004			
Placebo		646/13,758	0.0470						

Baseline covariates adjusted for: baseline risk score, At Risk status, Community of color status.
Maximum failure event time 100 days post Day 57 visit.

[¶] Tertiles:
  Spike IgG: Low is < 2190 BAU/ml, Medium is 2190 to 3800 BAU/ml, High is > 3800 BAU/ml.
  RBD IgG: Low is < 3310 BAU/ml, Medium is 3310 to 5750 BAU/ml, High is > 5750 BAU/ml.
  ID50: Low is < 178 IU50/ml, Medium is 178 to 363 IU50/ml, High is > 363 IU50/ml.
  ID80: Low is < 407 IU80/ml, Medium is 407 to 661 IU80/ml, High is > 661 IU80/ml.
[§] No. cases = estimated number of this cohort with an observed COVID-19 endpoint. The total count (47) across all tertiles for each marker differs
  from 36 (Figure 1, Table 1), because the 47 includes all vaccine breakthrough cases including the 11 without Day 1, 29, 57 antibody marker data.
  No. at-risk = estimated number in the population for analysis: baseline negative per-protocol vaccine recipients not experiencing the COVID-19
  endpoint through 6 days post Day 57 visit. No. cases = estimated number of this cohort with an observed COVID-19 endpoint.
[†] FDR (false discovery rate)-adjusted p-values and FWER (family-wise error rate)-adjusted p-values are computed over the set of p-values both for
  quantitative markers and categorical markers (Low, Medium, High) using the Westfall and Young permutation method (10,000 replicates).

Fig. 2. COVID-19 risk by antibody marker level. The plots and table show covariate-adjusted cumulative incidence of COVID-19 by low, medium, and high tertiles of day 57 IgG concentration or pseudovirus neutralization titer in baseline SARS-CoV-2–negative per-protocol participants. (**A**) Anti-spike IgG concentration. (**B**) ID_{50} titer. (**C**) IgG (spike and RBD) and pseudovirus neutralization titer (ID_{50} and ID_{80}). The overall *P* value is from a generalized Wald test of whether the hazard rate of COVID-19 differed across the low, medium, and high subgroups. Baseline covariates are adjusted for baseline risk score, at risk status, and community of color status. Pt. Est., point estimate; FDR, false discovery rate; FWER, family-wise error rate.

The estimated cumulative incidence of COVID-19 by the end of blinded follow-up (100 days after day 57) for the entire vaccine group was 0.0033 (95% CI, 0.0022 to 0.0045). On the basis of nonparametric threshold regression, this cumulative incidence decreased across vaccinated subgroups with day 57 ID_{50} titer above a given threshold, with zero COVID-19 end points at ID_{50} titer above 1000 IU_{50}/ml (Fig. 4A). The shape of cumulative incidence across threshold subgroups tracked the reverse cumulative distribution function of ID_{50}

titer, which suggests a smooth incremental change in risk with titer (Fig. 4A). Based on the Cox model, Fig. 4B shows estimated cumulative incidence of COVID-19 by end of blinded follow-up across vaccinated subgroups with day 57 ID_{50} titer at specific titers, in contrast to Fig. 4A, which considers vaccinated subgroups with titers above specific values. For vaccine recipients with undetectable day 57 ID_{50} titer, estimated cumulative incidence was 0.030 (95% CI, 0.010 to 0.093), and it decreased to 0.014 (0.0067 to 0.028) at titer of 10, to 0.0056 (0.0039 to 0.0080) at titer of 100, and to 0.0023 (0.0013 to 0.0036) at titer of 1000 (Fig. 4B). The generalized additive model also supported inverse correlates of risk for all markers (figs. S20 and S21).

Vaccine efficacy increases as antibody marker levels increase

Figure 4C shows titer-specific vaccine efficacy across day 57 ID_{50} titer levels, which, for a given titer level, is the estimated covariate-adjusted percent reduction in cumulative incidence of COVID-19 by the end of blinded follow-up as a result of vaccination generating the given titer level compared with being un-vaccinated (28). Vaccine efficacy estimates increased with day 57 ID_{50} titer: At undetectable day 57 ID_{50}, vaccine efficacy was 51% (95% CI, −51 to 83%), and at day 57 ID_{50} value of 10, 100, and 1000 IU_{50}/ml, respectively, vaccine efficacy was 78% (54 to 89%), 91% (87 to 94%), and 96% (94 to 98%) (Fig. 4C). The increase in vaccine efficacy from 78 to 96% at ID_{50} values of 10 to 1000 IU_{50}/ml, respectively, represents a 5.5-fold increase in vaccine-risk reduction (1 – vaccine efficacy = 22 versus 4%). Vaccine efficacy estimates also increased with day 29 ID_{50} neutralization titers: 79% (−62 to 90%), 93% (90 to 95%), 97% (95 to 99%), and 99% (97 to 100%) at the same IU_{50} per milliliter values (fig. S22).

Figures S23 to S28 show these results for the other six antibody markers. Conclusions for bAbs were similar to those for nAbs, with vaccine efficacy increasing with IgG levels, for example at day 57 spike IgG of 33, 300, and 4000 binding antibody units (BAU)/ml, vaccine efficacy was 85% (31 to 92%), 90% (77 to 94%), and 94% (91 to 96%), respectively. Another conclusion of these analyses is that subgroups with neutralization titer 10 IU_{50}/ml (Fig. 4C) or with anti-spike IgG 33 BAU/ml (fig. S24C) have ~75 to 85% reduction in COVID-19 risk compared with being unvaccinated. Given the overall similarity of the bAb and nAb correlate of protection results, the potential value of the validated meso scale discovery (MSD) bAb assay for aiding vaccine approval decisions as a practical nonmechanistic correlate of protection (12) should be considered. This is because the MSD bAb assay is sensitive (table S3), robust, high-throughput, deployable, and

easily standardized across viral strains, even though validated sensitive bAb detection may lack the specific immune function, such as neutralization.

A sensitivity analysis further increases confidence that vaccine efficacy increases with antibody marker levels

A sensitivity analysis was conducted (supplementary text, section S2) assuming the existence of an unmeasured confounder associated with both the antibody marker and COVID-19 outcome that would make the estimated vaccine efficacy by marker curve flatter, with the specified amount of unmeasured confound-

ing detailed in the SAP (section 12.1.2). The analysis indicated that vaccine efficacy estimates still increased with day 57 ID_{50} titer [90% (95% CI, 69 to 96%) at undetectable day 57 ID_{50} titer, 95% (93 to 97%) at day 57 ID_{50} titer of 500, and 96% (93 to 97%) at day 57 ID_{50} titer of 1000] (fig. S29C). A similar pattern of results occurred for all other nAb markers (fig. S29D and fig. S30, C and D). By contrast, estimated vaccine efficacy appeared to vary only minimally with each bAb marker when unmeasured confounding was assumed (fig. S29, A and B, and fig. S30, A and B). The sensitivity analysis based on E-values (29) of the vaccine recipient antibody tertile subgroups

A

COVE Immunologic Marker	No. Cases/ No. At-Risk*	HR Per 10-fold Increase Point Est. (95% CI)		P-Value (2-Sided)	FDR-adjusted p-value**	FWER-adjusted p-value
Anti Spike IgG (BAU/ml)	47/14,064	0.66	(0.50, 0.88)	0.005	0.014	0.010
Anti RBD IgG (BAU/ml)	47/14,064	0.57	(0.40, 0.82)	0.002	0.008	0.008
Pseudovirus-nAb ID50 (IU50/ml)	47/14,064	0.42	(0.27, 0.65)	< 0.001	0.002	0.003
Pseudovirus-nAb ID80 (IU80/ml)	47/14,064	0.35	(0.20, 0.61)	< 0.001	0.003	0.003

Baseline covariates adjusted for: baseline risk score, At Risk status, Community of color status. Maximum failure event time 100 days post Day 57 visit.

*No. at-risk = estimated number in the population for analysis: baseline negative per-protocol vaccine recipients not experiencing the COVID-19 endpoint through 6 days Day 57 visit; no. cases = estimated number of this cohort with an observed COVID-19 endpoint starting 7 days post Day 57 visit. The count 47 differs from 36 (Figure 1, Table 1), because the 47 includes all vaccine breakthrough cases including the 11 without Day 1, 29, 57 antibody marker data.

**FDR (false discovery rate)-adjusted p-values and FWER (family-wise error rate)-adjusted p-values are computed over the set of p-values both for quantitative markers and categorical markers (Low, Medium, High) using the Westfall and Young permutation method (10,000 replicates).

B

Binding Antibody to Spike: Day 57

Group	No. Events		HR (95% CI)
All Vaccine	47		0.66 (0.50, 0.88)
Age ≥ 65	7		0.52 (0.19, 1.43)
Age < 65	40		0.68 (0.51, 0.91)
At risk	15		0.72 (0.43, 1.22)
Not at risk	32		0.63 (0.44, 0.90)
Comm. of color	12		0.15 (0.06, 0.35)
White Non–Hispanic	35		0.71 (0.52, 0.97)
Male	23		0.26 (0.07, 1.00)
Female	24		0.71 (0.53, 0.96)

Hazard Ratio

C

50% Inhibitory Dilution Pseudovirus Neutralization Titer: Day 57

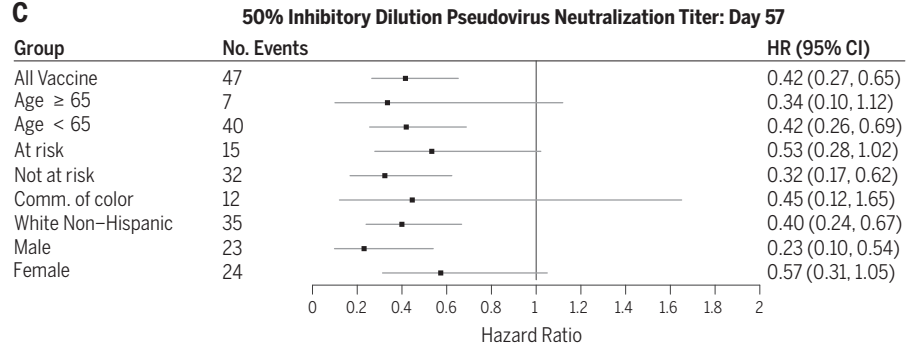| Group	No. Events		HR (95% CI)
All Vaccine	47		0.42 (0.27, 0.65)
Age ≥ 65	7		0.34 (0.10, 1.12)
Age < 65	40		0.42 (0.26, 0.69)
At risk	15		0.53 (0.28, 1.02)
Not at risk	32		0.32 (0.17, 0.62)
Comm. of color	12		0.45 (0.12, 1.65)
White Non–Hispanic	35		0.40 (0.24, 0.67)
Male	23		0.23 (0.10, 0.54)
Female	24		0.57 (0.31, 1.05)

Hazard Ratio

Fig. 3. Hazard ratio of COVID-19 as antibody marker level increases. The table and plots show covariate-adjusted hazard ratios of COVID-19 per 10-fold increase in each day 57 antibody marker in baseline-negative per-protocol vaccine recipients overall and in subgroups. (**A**) Inferences for IgG (spike and RBD) and pseudovirus neutralization titer (ID_{50} and ID_{80}). (**B**) Forest plots for spike IgG. (**C**) Forest plots for ID_{50}. Baseline covariates are adjusted for baseline risk score, at risk status, and community (Comm.) of color status.

(supplementary text, section S2) supported the inference that vaccine efficacy was generally higher for the upper versus lower tertile subgroup (table S8), which suggests that vaccine efficacy would have still increased with each antibody marker level even if additional (hypothetical) unmeasured confounders had been present.

Given the overlap of marker distributions in vaccine breakthrough cases and vaccine recipient noncases (Fig. 1 and fig. S7), our results do not support that a day 29 or day 57 antibody marker could be highly effective in guiding individual decisions of whether to be revaccinated or boosted. However, if a vaccinated person has negative IgG response or undetectable neutralization response, on the basis of our results, it would be rational for this person to be concerned about relatively weak protection and to therefore prompt the seeking out of other means of protection.

nAbs mediate about two-thirds of the mRNA-1273 vaccine efficacy

For bAbs at both time points, and for nAbs at day 57, a challenging issue is understanding vaccine efficacy for vaccine recipients with negative or undetectable antibody levels, given that <2% of vaccine recipients had negative or undetectable antibodies. Consequently, the 95% CIs about vaccine efficacy for these subgroups were wide, and assessment of mediation through these markers was not technically possible because of insufficient overlap of marker values in placebo and vaccine recipients. However, day 29 ID_{50} and ID_{80} titers could be assessed as mediators of vaccine efficacy by the Benkeser et al. method (30), given that 18 and 36% of vaccine recipients had undetectable titer, respectively, providing enhanced precision [e.g., estimated vaccine efficacy 79% (95% CI, 62 to 90%) at undetectable ID_{50}]. The quantitative ID_{50} and ID_{80} variables were studied. An estimated 68.5% (58.5 to 78.4%) of vaccine efficacy was mediated by day 29 ID_{50} titer and 48.5% (34.5 to 62.4%) by day 29 ID_{80} titer (table S9).

This result of positive vaccine efficacy for the undetectable subgroup implies a lack of full mediation of vaccine efficacy through the day 29 antibody marker (28), with an estimated 68% of the overall vaccine efficacy mediated through day 29 ID_{50} titer. Therefore, if nAbs circulating on day 29 could be removed but the other consequences of vaccination remained, overall vaccine efficacy would be expected to be reduced by 68% (on the log scale), from 92 to 56%. However, because >98% of vaccine recipients achieved detectable nAbs by day 57, these day 29 mediation results do not reflect a complete deactivation of the nAb response to the level at both day 29 and day 57 (undetectable) that would have been obtained without vaccination. Yet, under the reasonable assumption that the vaccine's effect on the risk of COVID-19 operating through the day 57 ID_{50} marker is nonnegative, 68% is a lower bound for the proportion of vaccine efficacy that is mediated through ID_{50} at both day 29 and day 57 (see conditions in supplementary text, section S2). In comparison, hemagglutination inhibition titer against the B/Brisbane/60/ 2008-like (Victoria lineage) strain of influenza virus (included in the trivalent inactivated influenza vaccine) mediated an estimated 57% of vaccine efficacy against virologically confirmed influenza B/Victoria illness (31). As hemagglutination inhibition titer has been used to guide influenza vaccine strain selection and approval, this defines a potential benchmark for influencing COVID-19 vaccine approval decisions (32).

A possible interpretation also consistent with our results is that neutralization as a biological function mediated a large proportion of the vaccine efficacy, but the specific day 29 ID_{50} and ID_{80} immune markers studied—measured with a particular immunoassay—had inadequate sensitivity to quantify low-level neutralization below the positivity cutoff that could be present and functionally important. Passive transfer of purified IgG from mRNA-1273–immunized rhesus macaques protected golden Syrian hamsters from disease after SARS-CoV-2 challenge, which suggests that functionally active antibodies can mediate protection (24). However, additional immune markers are likely needed to fully explain the observed vaccine efficacy in COVE—for example, markers measuring additional immune functions beyond neutralization (e.g., Fc effector functions or functional T cells), markers not measured fully in serum (e.g., mucosal), and/or anamnestic responses not fully represented by a single time point measurement.

Further clarification of functional mediation of protection may be provided by future correlates analyses that study antibody markers
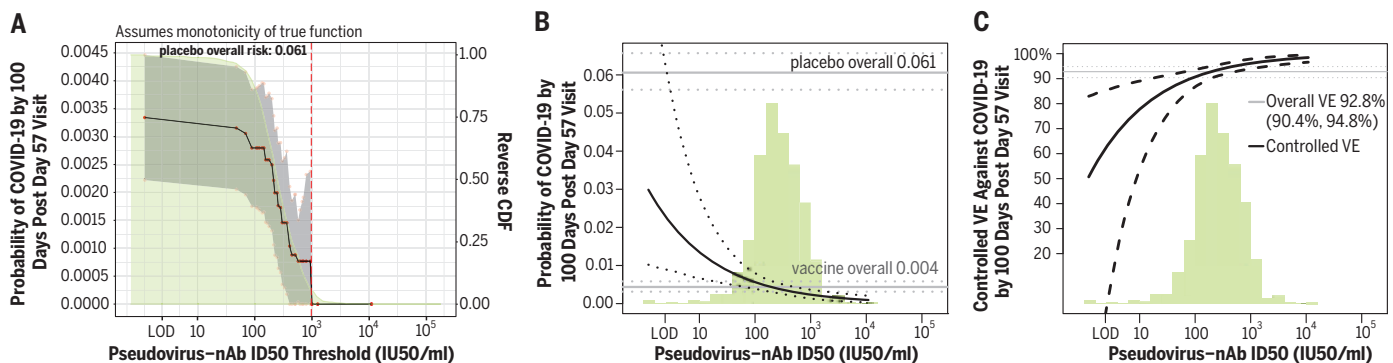
Fig. 4. Further analyses of day 57 ID_{50} level as a correlate of risk and as a correlate of protection. (**A**) Covariate-adjusted cumulative incidence of COVID-19 by 100 days after day 57 by vaccinated baseline SARS-CoV-2–negative per-protocol subgroups defined by day 57 ID_{50} level above a threshold, with reverse cumulative distribution function (CDF) of day 57 ID_{50} level overlaid in green. The red dots are point estimates at 35 threshold values equally spaced over quantiles of the observed marker values, linearly interpolated by solid black lines. The gray shaded area is pointwise 95% CIs. The upper boundary of the green shaded area is the estimate of the reverse CDF of day 57 ID_{50} level in baseline SARS-CoV-2–negative per-protocol vaccine recipients. The vertical red dashed line is the day 57 ID_{50} threshold above which no post–day 57 COVID-19 end points occurred. (**B**) Covariate-adjusted cumulative incidence of COVID-19 by 100 days after day 57 by day 57 ID_{50} level. The dotted black lines indicate bootstrap pointwise 95% CIs. The upper and lower horizontal gray lines are the overall cumulative incidence of COVID-19 from 7 to 100 days after day 57 in placebo and vaccine recipients, respectively. (**C**) Vaccine efficacy (VE) (solid black line) by day 57 ID_{50} level, estimated using the method of Gilbert, Fong, and Carone (28). The dashed black lines indicate bootstrap pointwise 95% CIs. The horizontal gray line is the overall vaccine efficacy from 7 to 100 days after day 57, with the dotted gray lines indicating the 95% CIs [this number, 92.8%, differs from the 94.1% reported in (6), which was based on counting COVID-19 end points starting 14 days after day 29]. In (B) and (C), the green histograms are an estimate of the density of day 57 ID_{50} level in baseline-negative per-protocol vaccine recipients. Baseline covariates are adjusted for baseline risk score, at risk status, and community of color status.

over time in relation to the timing of breakthrough infections of variants with variable neutralization sensitivity, with the antibodies measured against the variants of concern as well as against the ancestral strain. In particular, future research in COVE aims to measure bAbs and nAbs to the Delta variant in the same immunogenicity subcohort as examined in the current study and in all additional vaccine breakthrough cases that occur during follow-up. This should enable analyses to assess the consistency of an ancestral-strain correlate of protection for ancestral-strain COVID-19 compared with a Delta-variant correlate of protection against Delta-variant COVID-19.

Similar results are seen in a cross-trial–cross-platform comparison

Our use of validated assays, with all results reported in WHO international units (IU) or calibrated to WHO international standards, enables comparison with other studies and vaccine platforms. Immune correlates results for the COV002 trial (33), which is testing the AZD1222 chimpanzee adenoviral-vectored vaccine (also called ChAdOx1 nCoV-19), are available (19). The COV002 correlates results for spike IgG and RBD IgG can be quantitatively compared with the COVE results by virtue of the same MSD assay platform, conversion of IgG concentration to WHO international units per milliliter, and the same antibody measurement time—4 weeks after the second dose. Estimated AZD1222 vaccine efficacy was 70 and 90% at spike IgG levels of 113 [95% CI < limit of detection (LOD) = 0.31 to 245] and 899 (369 to NC) BAU/ml, respectively, and at RBD IgG levels of 165 (<LOD = 1.59 to 452) and 2360 (723 to NC) BAU/ml, respectively (where NC means not calculated) (19). For COVE, there is low precision at 70% vaccine efficacy because few vaccine recipients had IgG < 100 BAU/ml, such that we only compare results at 90% vaccine efficacy. Estimated mRNA-1273 vaccine efficacy was 90% at day 57 spike IgG level 298 (1 to 1786) BAU/ml and at day 57 RBD IgG level of 775 (29 to 2819) BAU/ml. Although the point estimates of IgG levels at 90% efficacy were about three times as high for COV002 compared with those for COVE, the overlapping CIs are consistent with similar results across the two trials.

Pseudovirus neutralization results can also be compared between the trials using ID_{50} titers calibrated to the international standard, where estimated AZD1222 vaccine efficacy was 70 and 90% at ID_{50} titer of 8 (<LOD = 2.42 to 26) and 140 (43 to NC) IU_{50}/ml, compared with COVE results at ID_{50} titer of 4 (<LOD = 2.42 to 22) and 83 (16 to 188) IU_{50}/ml. These results support that nAb titers have a similar quantitative relationship with vaccine efficacy for the two vaccine platforms, which is promising for potential applications of a neutralization biomarker. The materials and methods provide a sensitivity analysis comparing correlate of protection results between COV002 and COVE.

With the caveats of different study end points and hosts, the COVE results are also consistent with results on spike IgG and nAb titers as correlates of protection against SARS-CoV-2 replication in mRNA-1273–vaccinated rhesus macaques. For instance, all macaques with spike IgG > 336 IU/ml at 4 weeks after second dose were protected from >10,000 subgenomic RNA copies per milliliter in bronchoalveolar lavages (24), and in COVE, day 57 spike IgG of 336 IU/ml corresponded to 90% vaccine efficacy against COVID-19 (fig. S24).

Conclusions

Our findings that all evaluated bAb and nAb markers strongly inversely correlated with COVID-19 risk and directly correlated with

---

Table 1. Anti-spike and anti-RBD IgG response rates and geometric mean concentrations (GMCs) and pseudovirus neutralization titer ID_{50} and ID_{80} response rates and geometric mean titers (GMTs) by COVID-19 outcome status. Analysis based on baseline-negative per-protocol vaccine recipients in the day 29 marker or day 57 marker case-cohort sets. Median (interquartile range) number of days from dose one to day 29 was 28 (28 to 30) and from day 29 to day 57 was 28 (28 to 30). The N category under "Noncases in immunogenicity subcohort" indicates the number of noncases in the immunogenicity subcohort and hence with day 1, day 29, and day 57 antibody marker data, included in both the day 29 and day 57 marker correlates analyses. The N category under "COVID-19 cases" indicates either the number of vaccine breakthrough cases with day 1 and day 29 antibody marker data included (for day 29 marker analyses) or the number of vaccine breakthrough cases with day 1, day 29, and day 57 antibody data included (for day 57 marker analyses). See fig. S2. GM, geometric mean.

Visit for marker	Marker	COVID-19 cases*			Noncases in immunogenicity subcohort			Comparison	
		N	Response rate (95% CI)	GMC or GMT (95% CI)	N	Response rate (95% CI)	GMC or GMT (95% CI)	Response rate difference (95% CI)	Ratio of GM (cases/noncases) (95% CI)
Day 29	Anti-spike IgG (BAU/ml)	46	97.8% (85.4 to 99.7%)	183 (126 to 266)	1005	98.6% (97.4 to 99.2%)	318 (292 to 347)	−1% (−13 to 1%)	0.57 (0.39 to 0.84)
Day 29	Anti-RBD IgG (BAU/ml)	46	97.8% (85.4 to 99.7%)	207 (147 to 293)	1005	98.4% (97.2 to 99.1%)	327 (302 to 354)	−1% (−13 to 2%)	0.63 (0.44 to 0.90)
Day 29	Pseudovirus nAb ID_{50} (IU_{50}/ml)	46	65.2% (50.1 to 77.8%)	7.6 (5.4 to 10.8)	1005	81.7% (78.8 to 84.3%)	13.0 (11.9 to 14.1)	−17% (−32 to −4%)	0.59 (0.41 to 0.84)
Day 29	Pseudovirus nAb ID_{80} (IU_{80}/ml)	46	43.5% (29.7 to 58.4%)	18.0 (13.3 to 24.2)	1005	63.9% (60.4 to 67.3%)	29.0 (27.1 to 31.0)	−20% (−35 to −5%)	0.62 (0.46 to 0.84)
Day 57	Anti-spike IgG (BAU/ml)	36	100.0% (100.0 to 100.0%)	1890 (1449 to 2465)	1005	99.4% (98.2 to 99.8%)	2652 (2457 to 2863)	1% (0 to 2%)	0.71 (0.54 to 0.94)
Day 57	Anti-RBD IgG (BAU/ml)	36	100.0% (100.0 to 100.0%)	2744 (2056 to 3664)	1005	99.4% (98.3 to 99.8%)	3937 (3668 to 4227)	1% (0 to 2%)	0.70 (0.52 to 0.94)
Day 57	Pseudovirus nAb ID_{50} (IU_{50}/ml)	36	100.0% (100.0 to 100.0%)	160 (117 to 220)	1005	98.7% (97.6 to 99.3%)	247 (231 to 264)	1% (1 to 2%)	0.65 (0.47 to 0.90)
Day 57	Pseudovirus nAb ID_{80} (IU_{80}/ml)	36	97.2% (81.6 to 99.6%)	332 (248 to 444)	1005	98.3% (97.1 to 99.1%)	478 (450 to 508)	−1% (−17 to 2%)	0.69 (0.52 to 0.93)

*Cases for day 29 marker correlates analyses (intercurrent cases + post–day 57 cases) are baseline SARS-CoV-2–negative per-protocol vaccine recipients with the symptomatic infection COVID-19 primary end point diagnosed starting 7 days after day 29 through the end of the blinded phase. Cases for day 57 marker correlates analyses (post–day 57 cases) are baseline SARS-CoV-2–negative per-protocol vaccine recipients with the symptomatic infection COVID-19 primary end point diagnosed starting 7 days after day 57 through the end of the blinded phase. The last COVID-19 end point within the blinded phase occurred 100 days after day 57.

vaccine efficacy add evidence toward establishing an immune marker surrogate end point for mRNA COVID-19 vaccines. Moreover, the prespecification of the analyses and the absence of post hoc modifications bolsters the credibility of our conclusions.

For per-protocol recipients of two doses of mRNA-1273 COVID-19 vaccine in the COVE clinical trial, all four antibody markers at day 29 and at day 57 were inverse correlates of risk of COVID-19 occurrence through ~4 months after the second dose. Based on any of the antibody markers, estimated COVID-19 risk was about 10 times as high for vaccine recipients with negative or undetectable values compared with the estimated risk for those with antibodies in the top 10% of values. The nonparametric threshold analyses (Fig. 4A) suggested a continuum model where COVID-19 risk decreased incrementally with increasing increments in antibody level rather than a threshold model where an antibody cut-point sharply discriminated risk.

Together with evidence from other studies, the current results support that neutralization titer is a potential surrogate marker for mRNA-1273 vaccination against COVID-19 that can be considered as a primary end point for basing certain provisional approval decisions. For example, an immunogenicity noninferiority approach has been proposed for adding vaccine spike variants and boosters (34). An advantage of a noninferiority approach is avoiding the need to specify an absolute antibody benchmark for approval, such as one based on the percentage of vaccine recipients with ID_{50} titer above a threshold and geometric mean titer above a threshold. However, some applications may be aided by an absolute benchmark if data allowing head-to-head noninferiority evaluation are unavailable. Such a benchmark based on ID_{50} values from vaccinated individuals in a bridging study could be based on predicted vaccine efficacy being sufficiently high, where, for example, predicted vaccine efficacy could be calculated on the basis of the COVE correlates of protection results (Fig. 4C) and averaging over the distribution of ID_{50} values.

The evidence level for justifying various bridging applications differs across applications. Currently, confidence is greatest for bridging short-term vaccine efficacy (i.e., over 4 to 6 months) against COVID-19 to new subgroups for the same vaccine (e.g., to young children) or for bridging to a modified dose or schedule for the same vaccine (e.g., completing the primary series with a third dose). Less evidence is available to buttress the use of a humoral immune marker to predict long-term protection, to bridge to a new vaccine within the same vaccine platform, or to bridge to new spike variant inserts for the same vaccine. An open question challenging the latter application is whether higher nAb responses to emergent SARS-CoV-2 variants, such as

Delta, will be needed to achieve similar levels of vaccine efficacy, although modeling data are beginning to support the ability to make cross-variant predictions (16). Less evidence still is available for justifying bridging to a new candidate vaccine in a different vaccine platform. When immune correlates results are available from several COVID-19 phase 3 vaccine efficacy trials covering a multiplicity of vaccine platforms, it will be possible to conduct validation analyses of how well antibody markers can be used to predict vaccine efficacy across platforms (35). Uncertainties in bridging predictions can also be addressed by animal models that characterize immunological mechanisms of vaccine protection and by postauthorization or postapproval vaccine effectiveness studies (36). Notably, immune marker–based provisional approval mechanisms require post-approval studies verifying that the vaccine provides direct clinical benefit, such that the rigorous design and analysis of such studies is a critical component of the decision-making process for use of immune markers to accelerate the approval and distribution of vaccines.

Limitations of this immune correlates study include the inability to control for SARS-CoV-2 exposure factors (e.g., virus magnitude) and a lack of experimental assignment of antibody levels, which implies that the study could evaluate statistical correlates of protection or surrogate end points but not mechanistic correlates of protection (10). Additionally, scope limitations include the following: (i) the lack of data for assessing correlates against other outcomes besides COVID-19 (e.g., severe COVID-19, asymptomatic SARS-CoV-2 infection, infection regardless of symptomology, and viral shedding); (ii) the lack of assessment of non–antibody-based correlates (e.g., spike-specific functional T cell responses, which were not feasible to assess in the context of this study); (iii) the relatively short follow-up time of 4 months that precluded the assessment of immune correlate durability; (iv) the relatively small number of COVID-19 cases; (v) the lack of assessment of correlates for recipients of only one mRNA-1273 dose; (vi) the inability to assess the effects of boosting (homologous or heterologous) because this study pre-dated the addition of a third dose; (vii) the lack of data for assessing the potential contribution of anamnestic responses to the immune correlates; and (viii) the fact that almost all COVID-19 cases resulted from infections with viruses with a spike sequence similar to that of the vaccine strain, which precluded the assessment of robustness of correlates to SARS-CoV-2 variants of concern. However, the relative uniformity in circulating virus is also a strength in affording a clear interpretation as correlates against COVID-19 caused by variants genetically close to the vaccine. An additional strength is the racial and ethnic diversity of the trial

participants and the large number of diverse participants sampled for immunogenicity measurements (37).

REFERENCES AND NOTES

1. World Health Organization, "Coronavirus Disease (COVID-19): COVID-19 vaccine EUL issued" (2021); https://extranet.who.int/pqweb/vaccines/covid-19-vaccines.
2. US Food and Drug Administration (FDA), "COVID-19 Vaccines: COVID-19 Vaccines Authorized for Emergency Use" (2021); www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines#news [last updated 11 March 2021].
3. US Food and Drug Administration, "FDA Approves First COVID-19 Vaccine: Approval Signifies Key Achievement for Public Health" (FDA news release, 23 August 2021); www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine.
4. C. Weller, "Four reasons why we need multiple vaccines for Covid-19" (Wellcome Opinion, 6 June 2021); https://wellcome.org/news/four-reasons-why-we-need-multiple-vaccines-covid-19.
5. O. J. Wouters et al., Lancet 397, 1023–1034 (2021).
6. L. R. Baden et al., N. Engl. J. Med. 384, 403–416 (2021).
7. US Food and Drug Administration, "Moderna COVID-19 Vaccine," (updated 1 April 2021); www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine.
8. M. G. Thompson et al., MMWR Morb. Mortal. Wkly. Rep. 70, 495–500 (2021).
9. Moderna, "Moderna Announces TeenCOVE Study of its COVID-19 Vaccine in Adolescents Meets Primary Endpoint and Plans to Submit Data to Regulators in Early June" (press release, 25 May 2021); https://investors.modernatx.com/news-releases/news-release-details/moderna-announces-teencove-study-its-covid-19-vaccine.
10. S. A. Plotkin, P. B. Gilbert, Clin. Infect. Dis. 54, 1615–1617 (2012).
11. S. A. Plotkin, Clin. Vaccine Immunol. 17, 1055–1065 (2010).
12. S. A. Plotkin, P. B. Gilbert, in Plotkin's Vaccines, S. A. Plotkin, W. A. Orenstein, P. A. Offit, K. M. Edwards, Eds. (Elsevier, seventh ed., 2018), chap. 3.
13. J. Yu et al., Science 369, 806–811 (2020).
14. X. He et al., Cell 184, 3467–3473.e11 (2021).
15. K. McMahan et al., Nature 590, 630–634 (2021).
16. D. S. Khoury et al., Nat. Med. 27, 1205–1211 (2021).
17. K. A. Earle et al., Vaccine 39, 4423–4428 (2021).
18. D. Cromer et al., medRxiv 2021.06.29.21259504 [Preprint] (2021). https://doi.org/10.1101/2021.06.29.21259504.
19. S. Feng et al., Nat. Med. 27, 2032–2040 (2021).
20. A. Addetia et al., J. Clin. Microbiol. 58, e02107-20 (2020).
21. S. S. Abdool Karim, T. de Oliveira, N. Engl. J. Med. 384, 1866–1868 (2021).
22. M. S. Cohen et al., JAMA 326, 46–55 (2021).
23. Regeneron, "Phase 3 Prevention Trial Showed 81% Reduced Risk of Symptomatic SARS-CoV-2 Infections with Subcutaneous Administration of REGEN-COV™ (casirivimab with imdevimab)" (press release, 12 April 2021); https://investor.regeneron.com/news-releases/news-release-details/phase-3-prevention-trial-showed-81-reduced-risk-symptomatic-sars.
24. K. S. Corbett et al., Science 373, eabj0299 (2021).
25. M. Bergwerk et al., N. Engl. J. Med. 385, 1474–1484 (2021).
26. P. R. Krause et al., Lancet 398, 1377–1380 (2021).
27. G. R. Siber et al., Vaccine 25, 3816–3826 (2007).
28. P. B. Gilbert, Y. Fong, M. Carone, Assessment of Immune Correlates of Protection via Controlled Vaccine Efficacy and Controlled Risk. arXiv:2107.05734 [stat.ME] (2021).
29. T. J. VanderWeele, P. Ding, Ann. Intern. Med. 167, 268–274 (2017).
30. D. Benkeser, I. Díaz, J. Ran, Inference for natural mediation effects under case-cohort sampling with applications in identifying COVID-19 vaccine correlates of protection. arXiv:2103.02643 [stat.ME] (2021).
31. B. J. Cowling et al., Clin. Infect. Dis. 68, 1713–1717 (2019).
32. T. R. Fleming, J. H. Powers, Stat. Med. 31, 2973–2984 (2012).
33. M. Voysey et al., Lancet 397, 99–111 (2021).
34. US Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research, "Emergency Use Authorization for Vaccines to Prevent COVID-19: Guidance for Industry," appendix 2, pp. 18–22 (25 May 2021); www.fda.gov/regulatory-information/search-fda-guidance-documents/emergency-use-authorization-vaccines-prevent-covid-19.

35. R. A. Koup et al., Nat. Med. 27, 1493–1494 (2021).
36. J. Kertes et al., medRxiv 2021.09.01.21262957 [Preprint] (2021). https://doi.org/10.1101/2021.09.01.21262957.
37. M. P. Andrasik et al., PLOS ONE 16, e0258858 (2021).
38. P. B. Gilbert et al., COVID-19 Prevention Network Immune Correlates Analyses, version 1.1.1, Zenodo (2021); https://doi.org/10.5281/zenodo.5593130.

SPLICEOSOME

Structural basis of branch site recognition by the human spliceosome

Jonas Tholen[1,2]†, Michal Razew[1], Felix Weis[3], Wojciech P. Galej[1]*

Recognition of the intron branch site (BS) by the U2 small nuclear ribonucleoprotein (snRNP) is a critical event during spliceosome assembly. In mammals, BS sequences are poorly conserved, and unambiguous intron recognition cannot be achieved solely through a base-pairing mechanism. We isolated human 17S U2 snRNP and reconstituted in vitro its adenosine 5´-triphosphate (ATP)–dependent remodeling and binding to the pre–messenger RNA substrate. We determined a series of high-resolution (2.0 to 2.2 angstrom) structures providing snapshots of the BS selection process. The substrate-bound U2 snRNP shows that SF3B6 stabilizes the BS:U2 snRNA duplex, which could aid binding of introns with poor sequence complementarity. ATP-dependent remodeling uncoupled from substrate binding captures U2 snRNA in a conformation that competes with BS recognition, providing a selection mechanism based on branch helix stability.

Removal of introns from pre-mRNA is catalyzed by a large and dynamic RNA-protein complex known as the spliceosome. The spliceosome is assembled de novo on each pre-mRNA substrate from five small nuclear ribonucleoprotein (snRNP) particles and several dozen protein factors. During spliceosome assembly, three conserved positions in the pre-mRNA—the 5′ splice site (5′-SS), branch site (BS), and 3′ splice site (3′-SS)—are specifically recognized by the components of the spliceosome, allowing a two-step transesterification reaction to occur.

In mammalian cells, the BS is initially recognized by SF1 (mBBP) in cooperation with U2AF2 (U2AF65), which binds the polypyrimidine tract (PPT) sequence (1). Concomitantly, U1 snRNP binds to the 5′-SS and together they form the first, adenosine 5´-triphosphate (ATP)–independent spliceosome assembly intermediate known as complex E (2). The U2 snRNP is loosely associated with complex E (3), and its stable incorporation into the prespliceosome (complex A) requires ATP and formation of base-pairing interactions between the BS and the U2 snRNA (4).

In yeast, several factors have been shown to facilitate complex A formation. Among them are Cus2 (human HTATSF1) (5) and the RNA-dependent DEAD-box ATPase, Prp5 (human DDX46), whose activity is required for Cus2 displacement (6–8) and fidelity control of BS recognition (9, 10).

During BS recognition, an evolutionarily conserved branchpoint-interacting stem loop (BSL) presents U2 nucleotides to the intron BS for base pairing (11, 12). Although branch helix formation is subject to a fidelity checkpoint, it is unclear how branch helix stability is sensed by the splicing machinery (9, 10).

Upon engagement of U2 snRNA with the pre-mRNA substrate, a 15–nucleotide (nt)–long branch helix is formed, adopting helical geometry even in the absence of full complementarity. The length of the branch helix is conserved between yeast and human spliceosomes and maintained throughout different stages of splicing (13–15). In early splicing complexes, the branch helix is accommodated within a cavity formed by the heteroheptameric SF3b complex (7), which contacts the pre-mRNA around the BS and stabilizes the U2 snRNA:BS base-pairing interaction (16, 17). The branchpoint adenosine (BP-A) is bulged out of the branch helix and binds into a pocket formed by SF3B1 and PHF5A (13, 14). Mutations in SF3B1 associated with myelodysplastic syndromes have been shown to modulate BS selection (18, 19).

[1]European Molecular Biology Laboratory, 71 Avenue des Martyrs, 38042 Grenoble, France. [2]Heidelberg University, Faculty of Biosciences, Heidelberg, Germany. [3]European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany.
*Corresponding author. Email: wgalej@embl.fr
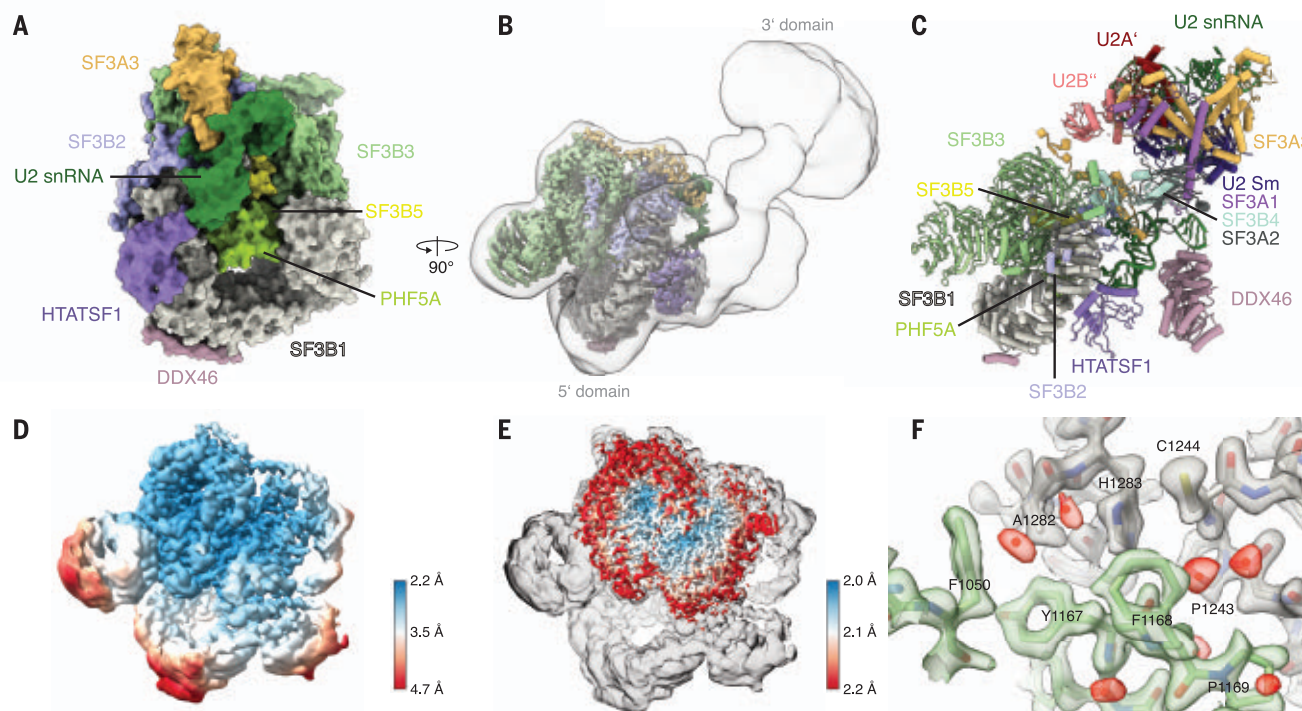†Candidate for Joint PhD degree from EMBL and Heidelberg University.

Fig. 1. High-resolution structure of the human 17S U2 snRNP. (**A**) Surface representation of the 5′ domain of the 17S U2 snRNP model. (**B**) Experimental cryo-EM map for the 17S U2 snRNP showing the high-resolution 5′ domain (colored by chain identity) embedded in a low-pass–filtered map showing the position of the 3′ domain. (**C**) Pseudo-atomic model for the fully assembled 17S U2 snRNP. The 3′ domain was modeled by rigid-body docking of the previously reported coordinates (PDB: 6Y5Q). (**D**) Cryo-EM map of the 17S U2 snRNP filtered and colored by local resolution. (**E**) The cryo-EM map obtained by merging several U2 snRNP datasets overlaid with the map of the 17S U2 snRNP. (**F**) Atomic modeling into the highest-resolution region at an interface of SF3B1 and SF3B3. The map was colored by chain identity; water molecules are colored red. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; F, Phe; H, His; P, Pro; and Y, Tyr.

SF3B6 (p14), which has no homolog in *Saccharomyces cerevisiae*, was shown to cross-link to the BP-A in HeLa nuclear extract, indicating a potential role in BS recognition (*20*). However, the position of SF3B6 in human Bact spliceosomes (*21*) does not explain the cross-linking data or its role in splicing.

Although parts of the U2 snRNP structure have been determined as a component of yeast and mammalian spliceosomes, there is no high-resolution structural information for the 17S U2 snRNP and early splicing complexes in humans (i.e., E and A). This is of particular interest as sequence conservation and base-pairing potential of human branch sites are weak compared with those of yeast (*22*), and the mechanism of BS selection remains elusive.

Here, we isolated human 17S U2 snRNP and reconstituted in vitro its binding to a model BS and the remodeling leading to the dissociation of HTATSF1 from the complex. We determined a series of cryo–electron microscopy (EM) structures of the U2 snRNP in different conformational states, including two previously unknown assembly intermediates. Our high-resolution reconstructions provide insight into the architecture and dynamics of the human U2 snRNP and pre-spliceosome

formation. These new data point at the critical roles of HTATSF1 and SF3B6 in facilitating pre-mRNA recognition.

Results

Purification of the 17S U2 snRNP complex

Existing methods for purification of the U2 snRNP use antibodies against SF3a or SF3b components (*7, 12*), which can in principle capture particles in multiple states. To specifically select a subset of U2 snRNPs representing a single functional state, we used CRISPR-Cas9–mediated genome editing to introduce a green fluorescent protein (GFP) tag into the HTATSF1 genomic locus of human embryonic kidney (HEK) 293F cells (fig. S1). Affinity chromatography with anti-GFP nanobodies allowed isolation of an intact 17S U2 snRNP, containing U2 snRNA and 22 proteins that accounted for a total estimated molecular weight of 1.08 MDa (fig. S1).

High-resolution structure of the 17S U2 snRNP

We determined the cryo-EM structure of the 5′ domain of the human 17S U2 snRNP at 2.2-Å resolution, which allowed accurate atomic modeling of the SF3b complex, SF3A3, HTATSF1^RRM, and the 5′ end of the U2 snRNA (Fig. 1 and figs. S2 to S4). The overall architecture of the com-

plex agrees well with that of previous studies (*23, 24*), including a recent low-resolution cryo-EM reconstruction (*12*). The low-pass–filtered map reveals an unresolved density at the periphery, which likely corresponds to the U2 snRNP core (3′ domain) (Fig. 1, B and C). This domain appears flexible relative to the resolved 5′ domain and could not be improved by further data processing.

Parts of SF3B2 and SF3A3 have been previously observed in cryo-EM maps of mammalian snRNPs and spliceosomes (*21, 25*), but owing to limited resolution, they were not interpreted with atomic coordinates. The high-resolution reconstruction provides atomic insights into several interfaces, including HTATSF1^RRM:SF3B1 (Fig. 1, A and C) and SF3B2:SF3A3 (Fig. 1 and fig. S4), consistent with previous lower-resolution structures (*12, 26*).

In vitro reconstitution of branch site recognition by the U2 snRNP

To obtain mechanistic insights into BS recognition by cryo-EM analysis, we reconstituted BS recognition in vitro with purified 17S U2 snRNP and a model BS oligonucleotide (BPS oligo). The BPS oligo is complementary to the positions 27 to 42 of the U2 snRNA and includes a bulged-out adenosine, which mimics
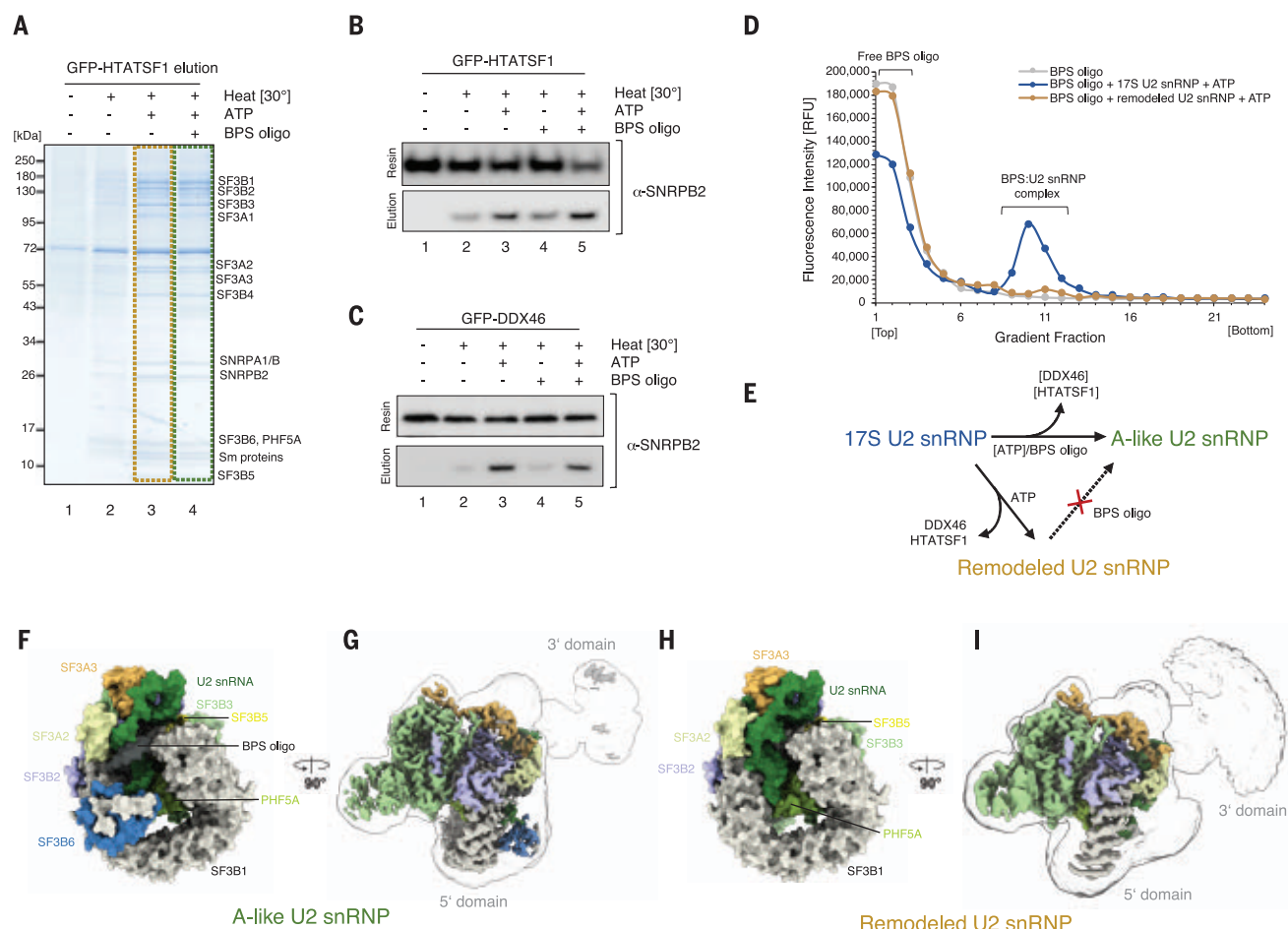
Fig. 2. Sample preparation and in vitro reconstitution of the branch site recognition by the U2 snRNP. (**A**) SDS–polyacrylamide gel electrophoresis analysis of the eluates from the GFP-HTATSF1–tagged 17S U2 snRNP immobilized on the GFP nanobody resin and incubated under various conditions. (**B**) Western blot analysis of the reconstitution reaction performed as in (A). Elution and resin fractions were probed with antibodies against SNRPB2, a core U2 snRNP component. (**C**) The same as in (B), but the 17S U2 snRNP sample was immobilized using GFP tag attached to DDX46. (**D**) Analysis of the Cy5-labeled BPS oligonucleotide binding to the 17S U2 snRNP or remodeled U2 snRNP by glycerol gradients. RFU, relative fluorescence units. (**E**) Schematic summarizing the outcome of the in vitro remodeling and substrate binding experiments. (**F** and **H**) Surface representation of the 5′ domains of the A-like and Remodeled U2 snRNPs models. (**G** and **I**) Experimental cryo-EM maps of A-like and remodeled U2 snRNPs showing the high-resolution 5′ domain (colored by chain identity) embedded in a low-pass–filtered (5 Å) map, showing positioning of their 3′ domains.

the BP-A. Similar minimal substrates have been used previously (27, 28).

We immobilized 17S U2 snRNP on anti-GFP nanobody resin by using a GFP tag on HTATSF1 or DDX46 and incubated it under various conditions. In the presence of ATP and the BPS oligo, U2 snRNP is released from the resin (Fig. 2A, lane 4) and remains bound to the BPS oligo when analyzed by glycerol gradient centrifugation. These BPS oligonucleotide–bound complexes likely resemble the substrate-bound U2 snRNP within complex A. Thus, we refer to these complexes as "A-like U2 snRNPs" hereafter. Addition of ATP alone, without the BPS oligo, also induces HTATSF1 dissociation at increased temperature (Fig. 2A, lane 3), suggesting that this remodeling can be functionally uncoupled from substrate binding, leading to the formation of a "remodeled U2

snRNP." We further investigated both reactions biochemically by probing eluates with antibodies specific to the core U2 snRNP component, SNRPB2 (Fig. 2, B and C). Western blotting revealed that some HTATSF1 dissociation can occur spontaneously at increased temperature (Fig. 2B, lane 2), but it is greatly stimulated by the presence of ATP (Fig. 2B, lanes 3 and 5), consistent with previous experiments (8). The same results were obtained regardless of whether the sample was tethered to the resin via HTATSF1 (Fig. 2B) or DDX46 (Fig. 2C).

Next, we investigated the requirements for U2 snRNP engagement with a model substrate. 17S U2 snRNP engages efficiently with the BPS oligo (Fig. 2D), and the binding occurs in a wide range of conditions without requiring ATP (fig. S5). No binding was observed when

the remodeled (ATP-treated) U2 snRNP variant was used in this assay (Fig. 2D). This indicates that displacement of HTATSF1 and DDX46 uncoupled from substrate binding leads to the formation of an inhibited conformation of the U2 snRNP.

We determined high-resolution cryo-EM structures of these two newly identified U2 snRNP complexes (Fig. 2, E to I, and figs. S2 to S4).

Structures of the minimal A-like and remodeled U2 snRNP complexes

The overall architecture of the A-like U2 snRNP is in good agreement with the lower-resolution descriptions of the U2 snRNP embedded within the fully assembled human B and Bact spliceosomes (21, 29) (Fig. 2, F and G). However, the structure exhibits features that are
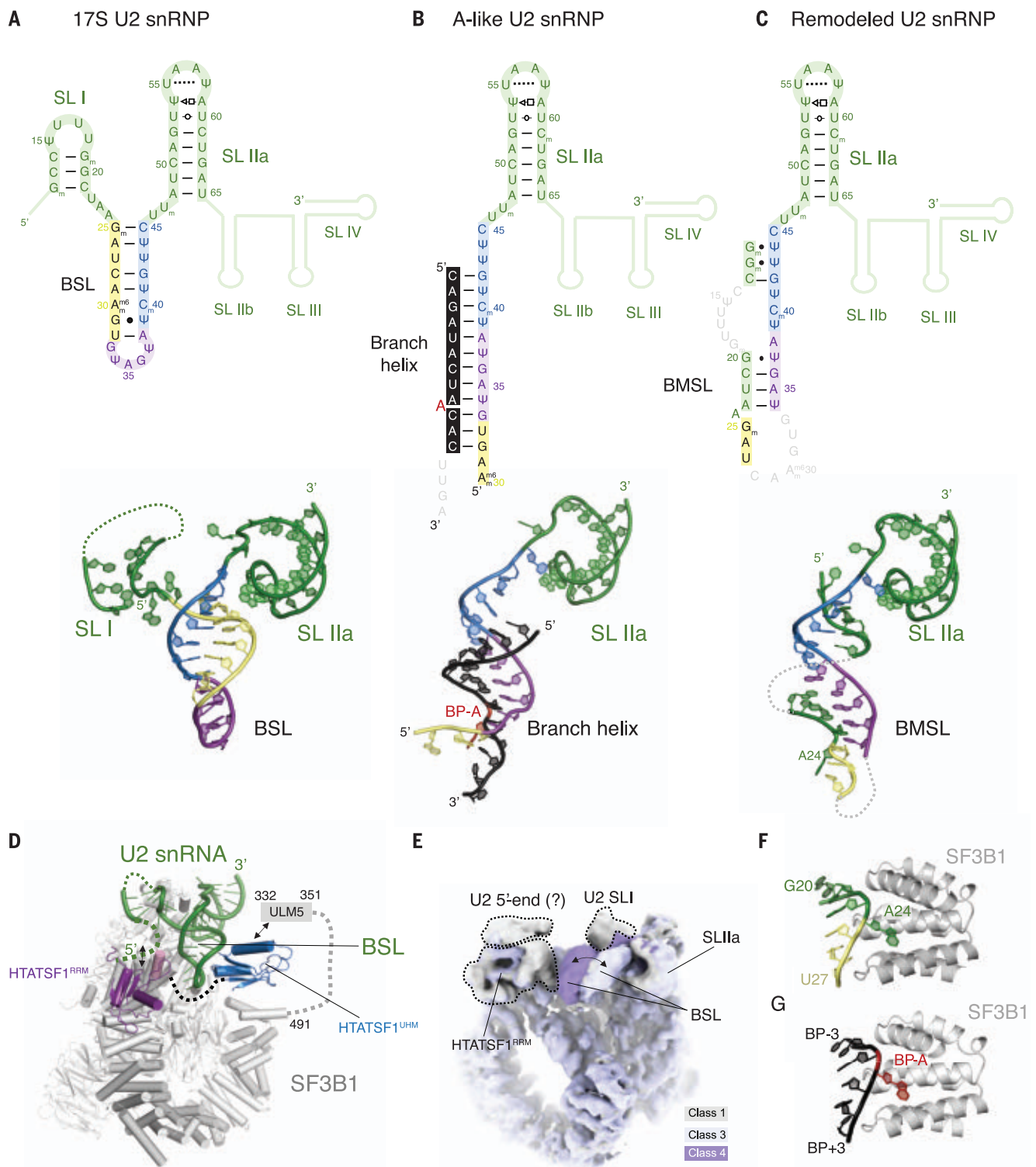
Fig. 3. Structure and dynamics of the U2 snRNA during branch site recognition. Secondary and tertiary structure of the U2 snRNA in the (**A**) 17S, (**B**) A-like, and (**C**) remodeled U2 snRNPs. (**D**) BSL is stabilized directly by the two domains of HTATSF1 in the 17S U2 snRNP complex. (**E**) Structural dynamics of the BSL visualized directly in the cryo-EM map (see also movie S2). (**F**) Adenosine 24 of the U2 snRNA mimics BP-A in the remodeled U2 snRNP complex. (**G**) Environment of the BP-A in the A-like U2 snRNP in the same orientation as in (F).

incompatible with those later splicing complexes, indicating that the A-like U2 snRNP represents a distinct splicing intermediate. The remodeled U2 snRNP closely resembles A-like U2 snRNPs, except for the major differences in the 5′ end of the U2 snRNA and missing SF3B6 (Fig. 2, F and H). Similar to the 17S U2 snRNP, both complexes can be divided into a well-resolved 5′ domain and a 3′ domain (Fig. 2, G and I). The 3′ domains remain flexible but occupy different positions compared to the 17S complex. HTATSF1 and
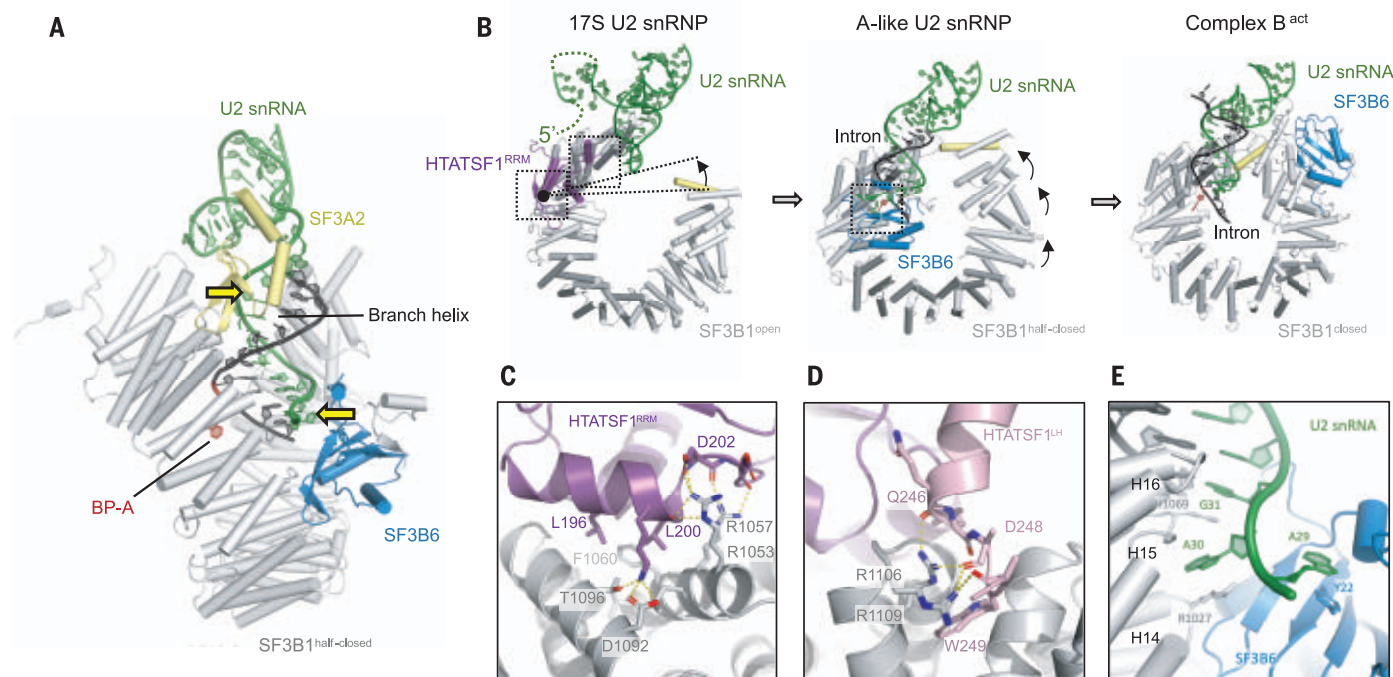
Fig. 4. SF3B6 stabilizes branch helix in the A-like U2 snRNP while SF3B1 HEAT repeats adopt a half-closed conformation. (**A**) Side view of the A-like U2 snRNP showing positions of the branch helix and its stabilization by the SF3A2 and SF3B6; yellow arrows indicate U2 snRNA contact points enforcing helical geometry of the branch helix. (**B**) Structure of the RNA and HEAT repeats in the 17S U2 snRNP, the A-like U2 snRNP, and the Bact complex (PDB: 6FF7), showing two-step transition from open to closed SF3B1 conformation. (**C** and **D**) Atomic model of the interfaces between SF3B1HEAT and HTATSF1RRM/LH. (**E**) Atomic model of the interface of SF3B6 with SF3B1HEAT and U2 snRNA. Abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; F, Phe; G, Gly; H, His; K, Lys; L, Leu; Q, Gln; R, Arg; T, Thr; W, Trp; and Y, Tyr.

the DDX46 helix are missing from these two reconstructions, consistent with the biochemical data and the sample preparation protocol.

Structure and dynamics of the U2 snRNA during branch site recognition

SLIIa of the U2 snRNA is resolved to nearly 2 Å in all our reconstructions, which allowed modeling of three additional, noncanonical base pairs within this stem-loop and its interactions with the components of the SF3a and SF3b complexes (figs. S4 and S6). These interactions remain unchanged during the transition from the 17S U2 snRNP to the A-like and remodeled U2 snRNP complexes.

In the 17S U2 snRNP reconstruction, a helical density emerges from the 5′ end of SLIIa and points toward SF3B1 and PHF5A along the side of HTATSF1RRM. This density was interpreted as the BSL and modeled by rigid-body docking of an idealized RNA helix with a base-pairing pattern based on previous predictions and the low-resolution reconstruction of the 17S U2 snRNP (12) (Fig. 3A). In contrast to the neighboring U2 SLIIa, the BSL density is not well resolved in our map, pointing at the intrinsically dynamic nature of this structure. Indeed, a three-dimensional (3D) classification focused on this region allows the separation of the ensemble structure into at least three distinct

conformational states of the BSL (Fig. 3, D and E), suggesting a dynamic probing mechanism for BS recognition.

During transition from 17S to A-like U2 snRNP, the BSL sequence engages with the BPS oligo, forming a 12-nt U2 snRNA:BS duplex (Fig. 3B). This duplex forms interactions with SF3B1, PHF5A, SF3A2, and SF3A3. The SF3A2 zinc finger domain binds the branch helix as previously described (14, 21, 26). SF3A3 becomes more ordered in the A-like complex, likely stabilized by SF3A2 and its interaction with the branch helix (Fig. 4A). Most prominent of the contacts formed during this transition are the interactions of charged amino acids of SF3B1 (K1071, R1106, N1107, R1109, K1149) with the BPS RNA phosphate backbone, consistent with previous studies (14, 21). The SF3a complex is less well resolved in the A-like complex assembled in the presence of AMP-PCP, suggesting that the ATP-dependent remodeling may play a role in facilitating SF3a docking to the complex A (fig. S7).

5′ end of the U2 snRNA mimics pre-mRNA substrate in the absence of HTATSF1

Upon ATP-dependent remodeling and HTATSF1 dissociation, U2 snRNA nucleotides 11 to 44 form a distinct bulged stem-loop structure in the remodeled U2 snRNP complex (Fig. 3C).

Comparison with the A-like U2 snRNP reveals that this newly formed stem-loop closely mimics the branch duplex and its interactions with the U2 snRNP proteins (Fig. 3, F and G); therefore, we refer to this stem-loop as the branch helix–mimicking stem-loop (BMSL). Formation of the BMSL is mutually exclusive with the pre-mRNA binding by the U2 snRNP, suggesting that the two structures could compete with one another during BS recognition. This finding provides a potential mechanism for the selection of branch helix stability by the spliceosome and could represent a previously unknown BS fidelity checkpoint.

SF3B6 stabilizes the branch helix in the A-like U2 snRNP

Upon BPS oligo binding to the 17S U2 snRNP, an additional density appears near H14 and H15 of SF3B1. This density could be unambiguously interpreted by rigid-body docking of the SF3B6:SF3B1 crystal structure [Protein Data Bank (PDB): 3LQV (30), fig. S4]. Although it is a stable component of the SF3b complex, SF3B6 has not been observed in any previously reported structure of the SF3b complex or U2 snRNP (12, 23) and it differs substantially from the SF3B6:SF3B1 interface in the Bact spliceosome (21). SF3B6 binds to the U2 snRNA at the 5′ end of the branch helix (Fig. 4A);
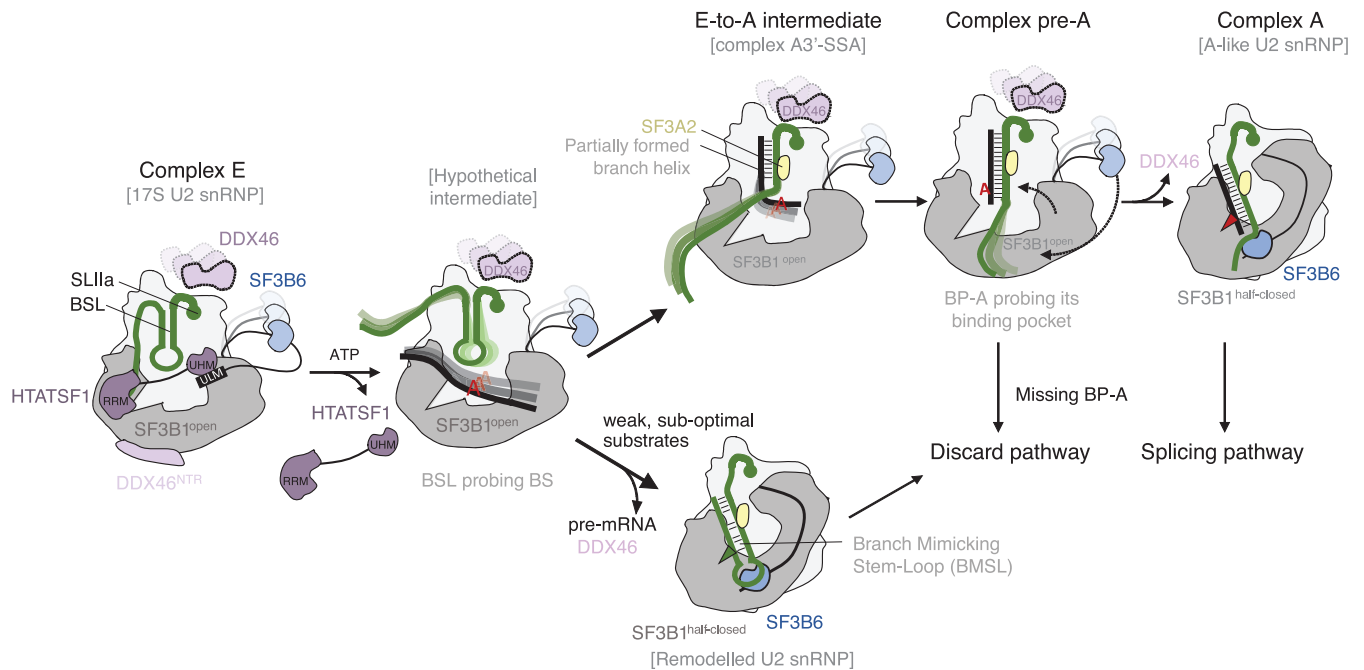
Fig. 5. Schematic model of branch site recognition by the U2 snRNP based on recent structural data. U2 snRNP associated with spliceosomal complex E is likely structurally similar to the 17S U2 snRNP described by (12) and in this work. Dissociation of HTATSF1 creates competition between the formation of a branch helix and the BMSL. Rejection of weak, suboptimal substrates results in the remodeled U2 snRNP, which is targeted to a discard pathway (this work). Stable substrates gradually form the branch helix as shown in the E-to-A (41) and pre-A (42) intermediates. In the absence of properly positioned, bulged out BP-A, the pre-A complex is targeted to a discard pathway. Productive engagement of the branch helix leads to the formation of complex A, wherein U2 snRNP is structurally similar to the A-like U2 snRNP (this work).

therefore, it defines the exact position of the bulged BP-A relative to the end of the branch helix. Such an interaction is supported by previous RNA-protein cross-linking (31, 32) and cross-linking mass spectrometry data (12, 23). A29 of the U2 snRNA inserts into the same pocket where adenine was placed in the co-crystal structure (30) and stacks against residue Y22 of SF3B6 (Fig. 4E). Moreover, SF3B6 is oriented in such a way that its disordered N terminus points toward BP-A and is close enough to explain previous cross-linking data (20, 33). Our data show that SF3B6 plays a previously unknown role in stabilizing branch helix, which could be particularly relevant for branch sequences with poor complementarity to the U2 snRNA.

HTATSF1 stabilizes the BSL in 17S U2 snRNP

Comparison of the 17S and A-like U2 snRNP shows that binding of SF3B6 and HTATSF1^{RRM} to SF3B1^{HEAT} are mutually exclusive and HTATSF1^{RRM} needs to be displaced before stable docking of SF3B6 (Fig. 4B). Our reconstruction of the 17S U2 snRNP shows that HTATSF1^{RRM} binds in a hydrophobic groove formed by HEAT repeats H15 and H16 of SF3B1 (Fig. 4C). The neighboring H16 and H17 repeats form the interface for the C terminus of the HTATSF1 linker helix (HTATSF1^{LH}), comprising residues 239 to 251 (Fig. 4D). The C terminus of HTATSF1^{LH} points toward the BSL and a

globular density nearby, which likely belongs to the HTATSF1^{UHM} domain that is known to bind the SF3B1^{ULM} motif (Fig. 3D) (12, 34, 35). Therefore, the two domains of HTATSF1 form stable interfaces with SF3B1 and flank the U2 snRNA BSL from both sides, suggesting a direct stabilization mechanism for this transient RNA secondary structure.

Movement of the BSL correlates with the disappearance of the extra density on top of HTATSF1^{RRM} and presumably the short variant of the U2 stem-loop I structure (Fig. 3C and movie S1). Given the concerted movement with other U2 snRNA elements, we speculate that at least part of this density could belong to the 5′ end of the U2 snRNA, especially that it occupies the surface that is typically involved in RRM-RNA binding. Indeed, the Y48D mutation (Y136 in HTATSF1) in the yeast homolog Cus2 abolishes U2 snRNA binding (5). Recombinant HTATSF1^{RRM} exhibits some nonspecific affinity for RNA (fig. S8). This supports the hypothesis that interaction between HTATSF1^{RRM} and the 5′ end of the U2 snRNA could additionally stabilize the BSL in an indirect manner by preventing BMSL formation.

Two-step conformational change in SF3B1 upon pre-mRNA binding

SF3B1 was previously reported to transition from an open to a closed conformation around a hinge between HEAT repeats H15 and H16

(36). Similar remodeling occurs in our in vitro system, with no extra factors needed, even when the BPS oligo is incubated on ice with the 17S U2 snRNP in the presence of AMP-PCP, a nonhydrolysable ATP analog (table S1 and fig. S7). This indicates that branch helix formation is the only driving force for the rearrangement around the first hinge region and that it does not depend on ATP hydrolysis.

Although the hinge-like movement of SF3B1 is reconstituted in our system, the conformation of the N-terminal part of the HEAT repeat differs appreciably from any of the previously reported states. In the closed conformation, SF3B1 helix H1 (residues 509 to 523) inserts into the major groove of the branch helix, providing additional stabilization for the branch helix, whereas in the A-like complex it remains ~20 Å away from this binding site (Fig. 4B). We refer to this new SF3B1 conformation as half-closed, in accordance with the previous convention. The movement from half-closed to closed is different from the hinge-like closure and involves multiple small changes in the curvature of the HEAT repeats in its N-terminal part (H1-H12) (Fig. 4B). It is possible that binding of the intron sequence downstream from the BS could facilitate complete closure of the SF3B1.

Discussion

Recognition of the branch point sequences by the U2 snRNP is a critical step of spliceosome

assembly. In this work, we used a minimal in vitro system to analyze the structure of the human U2 snRNP and its conformational changes upon ATP-dependent remodeling and engagement with the pre-mRNA substrate.

The 17S U2 snRNP structure shows that HTATSF1 and BSL stabilize each other in two distinct ways: directly, through interactions between BSL and HTATSF1[LH/UHM], and indirectly, via a possible association of the 5′ end of the U2 snRNA with the HTATSF1[RRM], which would prevent formation of RNA structures that compete with the BSL (i.e., the long variant of the SLI or the BMSL).

Our data show that a model BPS oligonucleotide can engage in base-pairing interaction with the U2 snRNP without a requirement for prior remodeling. Although formation of complex A has been shown to be ATP dependent in HeLa cell nuclear extract, Amin complex can form without ATP when the sequence upstream of the BS is missing (27, 37). This could be due to the absence of certain BS binding proteins (e.g., SF1) or lack of topological restraints for branch helix formation. To form the branch helix, U2 snRNA has to wind around the long pre-mRNA substrate, and it is possible that ATP is required to liberate the 5′ end of the U2 snRNA from HTATSF1 to allow that. Indeed, in yeast, the ATPase activity of the DDX46 homolog Prp5 is required for complex A formation, but deletion of the HTATSF1 homolog Cus2 removes this dependence (8, 34). However, the ATP-dependent BS fidelity control by Prp5 remains unchanged in the absence of Cus2, suggesting a more complex function of this protein.

The structure of the A-like U2 snRNP captured SF3B6 interacting with the branch helix, which has two major implications. First, it provides a specific binding site for the U2 snRNA in addition to SLIIa and SF3A2[ZnF], which imposes helical geometry on the U2 snRNA within the branch helix binding pocket. This provides a mechanism for the stabilization of weak branch point sequences, such as those found in mammals, even in the absence of extensive complementarity. Second, SF3B6 binds at the junction of the branch helix duplex and the single-stranded region of the U2 snRNA; therefore, it defines the length of the branch helix and the exact position of the bulged BP-A relative to its end. It has been previously shown in an orthogonal yeast system that the position of the BP-A within the branch helix is critical for productive splicing (38). In budding yeast, the BS has evolved to be highly conserved, and sequence complementarity between BS and U2 snRNA ensures proper positioning of the BP-A (39). Weak BS sequence conservation and base-pairing potential in other organisms, including mammals and fission yeast (22), require an additional BP-A positioning mechanism, which is fulfilled by SF3B6. Consequently, SF3B6 is conserved in

many species with low BS conservation, but not in *S. cerevisiae* (fig. S9).

The emerging data suggest that the transition from E to A complex requires ATP-dependent displacement of HTATSF1, which destabilizes the BSL and allows it to probe BS sequences (40) (Fig. 5). The absence of HTATSF1 creates competition between the branch helix and the BMSL structure within U2 snRNA, providing a mechanism for the selection of the branch helix stability. Formation of the BMSL would mean rejection of the potential BS sequences. Therefore, the structure of the remodeled U2 snRNP likely represents an intermediate on the discard pathway after suboptimal substrate rejection. Such a state was predicted to exist in the framework of the kinetic proofreading model (9).

BS sequences that withstand competition with the BMSL would continue to progressively form the branch helix through a recently proposed toehold strand invasion mechanism (41). An intermediate state in this process (A3′-SSA complex) was captured by blocking spliceosome assembly with spliceostatin A (SSA) (41), which trapped U2 snRNP with a partially formed branch helix, missing bulged BP-A. Consequently, the branch helix was not accommodated in its pocket and SF3B1 remains in the open conformation, resembling that found in the 17S U2 snRNP (Fig. 5).

Without inhibition by SSA, BS sequences would continue to fully form the branch helix. At this point another checkpoint would be reached. If a bulged-out BP-A is present, it will bind the pocket in SF3B1, causing transition to the half-closed conformation and dissociation of DDX46, as shown in the A-like U2 snRNP. However, in the absence of a properly positioned BP-A, SF3B1[HEAT] remains in the open conformation and the spliceosome is stalled, as shown in the structure of the yeast pre-A complex (42). In this complex, Prp5 provides steric hindrance for the next step of spliceosome assembly, recruitment of the tri-snRNP (10). A prolonged block by Prp5 will likely initiate a discard pathway.

The remodeled U2 snRNP described in this work and the pre-A complex (42) are two distinct intermediates that direct suboptimal substrates to the discard pathway. They represent different checkpoints in BS fidelity control, ensuring both the formation of a stable branch helix and the presence of properly positioned BP-A (Fig. 5).

Only properly positioned bulged out BP-A can bind the SF3B1-PHF5A pocket and cause the transition to the half-closed conformation of SF3B1, as observed in the A-like U2 snRNP. During this transition, an extensive interaction surface forms between the branch helix, including the bulged BP-A and the HEAT repeats H15 to H19. Our minimal system shows that this interaction is the sole driving force

for the SF3B1 hinge-like movement. It is not clear which factors are needed for the second phase of the transition from half-closed to closed SF3B1 conformation and when this conformational change occurs.

Upon A complex formation, poorly defined branch sequences would benefit from stabilization by SF3B6, which enforces helical geometry of the U2 snRNA, even in the absence of extensive branch site complementarity. During subsequent steps of the spliceosome assembly, SF3B6 has to relocate to its binding site observed in the Bact complex (21), as its position in the A-like U2 snRNP would clash with Prp8 and prevent early Bact formation.

Our data provide several high-resolution snapshots of the complex process of BS recognition by the U2 snRNP and contribute to a better understanding of the mechanism of pre-mRNA splicing in humans.

REFERENCES AND NOTES

1. J. Valcárcel, R. K. Gaur, R. Singh, M. R. Green, *Science* **273**, 1706–1709 (1996).
2. S. Michaud, R. Reed, *Genes Dev.* **5** (12b), 2534–2546 (1991).
3. R. Das, Z. Zhou, R. Reed, *Mol. Cell* **5**, 779–787 (2000).
4. J. Wu, J. L. Manley, *Genes Dev.* **3**, 1553–1561 (1989).
5. D. Yan et al., *Mol. Cell. Biol.* **18**, 5000–5009 (1998).
6. C. L. O'Day, G. Dalbadie-McFarland, J. Abelson, *J. Biol. Chem.* **271**, 33261–33267 (1996).
7. C. L. Will et al., *EMBO J.* **21**, 4978–4988 (2002).
8. R. Perriman, I. Barta, G. K. Voeltz, J. Abelson, M. Ares Jr., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13857–13862 (2003).
9. Y.-Z. Xu, C. C. Query, *Mol. Cell* **28**, 838–849 (2007).
10. W.-W. Liang, S.-C. Cheng, *Genes Dev.* **29**, 81–93 (2015).
11. R. Perriman, M. Ares Jr., *Mol. Cell* **38**, 416–427 (2010).
12. Z. Zhang et al., *Nature* **583**, 310–313 (2020).
13. R. Rauhut et al., *Science* **353**, 1399–1405 (2016).
14. C. Yan, R. Wan, R. Bai, G. Huang, Y. Shi, *Science* **353**, 904–911 (2016).
15. W. P. Galej et al., *Nature* **537**, 197–201 (2016).
16. O. Gozani, R. Feld, R. Reed, *Genes Dev.* **10**, 233–243 (1996).
17. C. L. Will et al., *EMBO J.* **20**, 4536–4546 (2001).
18. Q. Tang et al., *Genes Dev.* **30**, 2710–2723 (2016).
19. T. J. Carrocci, D. M. Zoerner, J. C. Paulson, A. A. Hoskins, *Nucleic Acids Res.* **45**, 4837–4852 (2017).
20. A. M. MacMillan et al., *Genes Dev.* **8**, 3008–3020 (1994).
21. D. Haselbach et al., *Cell* **172**, 454–464.e11 (2018).
22. A. J. Taggart et al., *Genome Res.* **27**, 639–649 (2017).
23. C. Cretu et al., *Mol. Cell* **64**, 307–319 (2016).
24. A. Krämer, P. Grüter, K. Gröning, B. Kastner, *J. Cell Biol.* **145**, 1355–1368 (1999).
25. X. Zhang et al., *Cell Res.* **28**, 307–322 (2018).
26. C. Plaschka, P.-C. Lin, K. Nagai, *Nature* **546**, 617–621 (2017).
27. C. C. Query, P. S. McCaw, P. A. Sharp, *Mol. Cell. Biol.* **17**, 2944–2953 (1997).
28. E. G. Folco, K. E. Coil, R. Reed, *Genes Dev.* **25**, 440–444 (2011).
29. X. Zhan, C. Yan, X. Zhang, J. Lei, Y. Shi, *Cell Res.* **28**, 1129–1140 (2018).
30. M. J. Schellenberg, E. L. Dul, A. M. MacMillan, *RNA* **17**, 155–165 (2011).
31. O. Dybkov et al., *Mol. Cell. Biol.* **26**, 2803–2816 (2006).
32. R. Spadaccini et al., *RNA* **12**, 410–425 (2006).
33. M. J. Schellenberg et al., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1266–1271 (2006).
34. J. Talkish et al., *RNA* **25**, 1020–1037 (2019).
35. S. Loerch et al., *J. Biol. Chem.* **294**, 2892–2902 (2019).
36. C. Cretu et al., *Mol. Cell* **70**, 265–273.e8 (2018).
37. C. M. Newnham, C. C. Query, *RNA* **7**, 1298–1309 (2001).
38. D. J. Smith, M. M. Konarska, C. C. Query, *Mol. Cell* **34**, 333–343 (2009).
39. J. Sales-Lee et al., *Curr. Biol.* **31**, 1–13 (2021).

40. C.-Y. Kao, E.-C. Cao, H. L. Wai, S.-C. Cheng, *Nucleic Acids Res.* **49**, 9965–9977 (2021).
41. C. Cretu *et al.*, *Nat. Commun.* **12**, 4491–15 (2021).
42. Z. Zhang *et al.*, *Nature* **596**, 296–300 (2021).

REPORTS

GEOPHYSICS

On the relative temperatures of Earth's volcanic hotspots and mid-ocean ridges

Xiyuan Bao[1]*, Carolina R. Lithgow-Bertelloni[1], Matthew G. Jackson[2], Barbara Romanowicz[3]

Volcanic hotspots are thought to be fed by hot, active upwellings from the deep mantle, with excess temperatures (T_{ex}) ~100° to 300°C higher than those of mid-ocean ridges. However, T_{ex} estimates are limited in geographical coverage and often inconsistent for individual hotspots. We infer the temperature of oceanic hotspots and ridges simultaneously by converting seismic velocity to temperature. We show that while ~45% of plume-fed hotspots are hot ($T_{ex} \geq 155$°C), ~15% are cold ($T_{ex} \leq 36$°C) and ~40% are not hot enough to actively upwell (50°C ≤ T_{ex} ≤ 136°C). Hot hotspots have an extremely high helium-3/helium-4 ratio and buoyancy flux, but cold hotspots do not. The latter may originate at upper mantle depths. Alternatively, the deep plumes that feed them may be entrained and cooled by small-scale convection.

O n Earth's surface, two types of volcanism are observed. The dominant type, observed at tectonic plate boundaries, manifests the large-scale global circulation in Earth's mantle. In contrast, isolated intraplate volcanoes, such as those of Hawaii, Iceland, or the Galápagos, do not fit with classical plate tectonics theory and are thought to reflect dynamical processes rooted in the deep mantle. The lavas found at these "hotspot" volcanoes provide a singular window into the thermochemical dynamics of Earth's interior. These hotspots usually appear as chains of intraplate volcanoes that may reflect relative movement between plates and the underlying mantle (*1*). The presence of active, hot, upwelling plumes rising from the core-mantle boundary (CMB) to the bottom of the lithosphere under hotspot volcanoes has been intensely debated (*2, 3*) since it was proposed 50 years ago (*1*). Nonetheless, the idea of hot mantle plumes originating in the deep mantle that sample sources distinct from those that give rise to mid-ocean ridge volcanism reconciles many geophysical and geochemical observations. For example, seismic studies showing an expected thinner mantle transition zone under hotspots (*4, 5*) and low-velocity columns in tomographic models that extend from the surface to the CMB beneath most hotspots (*6*) suggest that plumes can indeed extend well into the deep mantle. Geochemically distinct signals can be observed between ocean island basalts (OIBs) at hotspots and mid-ocean ridge basalts (MORBs), reflecting their source regions. MORBs have relatively uniform ^{3}He/^{4}He ratios, ~8 ± 1 Ra (atmospheric ratio) (*7*) (1σ), whereas OIBs have a much wider range, with ratios up to 43 Ra in Iceland (*8*) [and up to 50 Ra at the Baffin Island large igneous province (*9*)]. The high ^{3}He/^{4}He at hotspots might reflect a deep reservoir that preserves ancient Hadean material (*10*). Hotspots with high ^{3}He/^{4}He are also those with the highest buoyancy flux, a measure of plume strength (*11*). High ^{3}He/^{4}He signals and high buoyancy flux suggest a deep origin and active upwellings, which are further confirmed by the correlation between low shear wave velocities (V_S) at 200 km and high ^{3}He/^{4}He signals at hotspots (*12*). If the low seismic velocities are dominantly thermal, it implies that hotspots with higher ^{3}He/^{4}He anomalies are hotter and sufficiently buoyant to entrain the possibly denser high ^{3}He/^{4}He domain (*12, 13*) from the deep mantle.

Classical plume theory predicts focused thermal anomalies beneath hotspots. Directly measuring the excess temperature of the subhotspot mantle relative to the mantle upwelling beneath ridges may therefore allow us to constrain the origin and dynamics of plumes that feed them. Previous temperature estimates using petrological thermometers suggest that the subhotspot mantle is typically 100° to 300°C hotter than the subridge mantle (*4, 14, 15*). This implies an excess temperature (T_{ex}) for hotspots compared with ridges, where T_{ex} is the difference between the potential temperature (T_p)—the temperature a parcel of mantle will have at Earth's surface if extrapolated along an adiabat—beneath an individual hotspot and the average ridge temperature $\overline{T_p}$ over all ridge segments. Dynamically, a T_{ex} of between 100° and 150°C beneath the lithosphere is needed for pure thermal plumes of ~100 km radius to rise fast enough (~10 cm/year) in the upper mantle (*16*) to avoid cooling too much by diffusion, still create excess melt, and have enough buoyancy to continue rising (*17*). This estimate provides a dynamical limit for the minimum T_{ex} and is coincidentally the same as the lower bound of typical petrological estimates (i.e., $T_{ex} = 100$°C). The question of whether all oceanic hotspots exceed the average ridge temperature and the minimum dynamic limit is hard to answer from petrological thermometers alone, given the limited geographical coverage and inconsistent estimates. Only a subset of hotspots (≤28) have petrological temperature estimates, and values vary substantially between studies (*4, 14, 15*). Estimates of the average temperature at mid-ocean ridges range from 1280° to 1400°C (*4, 14, 15, 18, 19*).

[1]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA 90095, USA. [2]Department of Earth Science, University of California, Santa Barbara, CA 93106, USA. [3]Department of Earth and Planetary Science, University of California, Berkeley, CA 94720, USA.
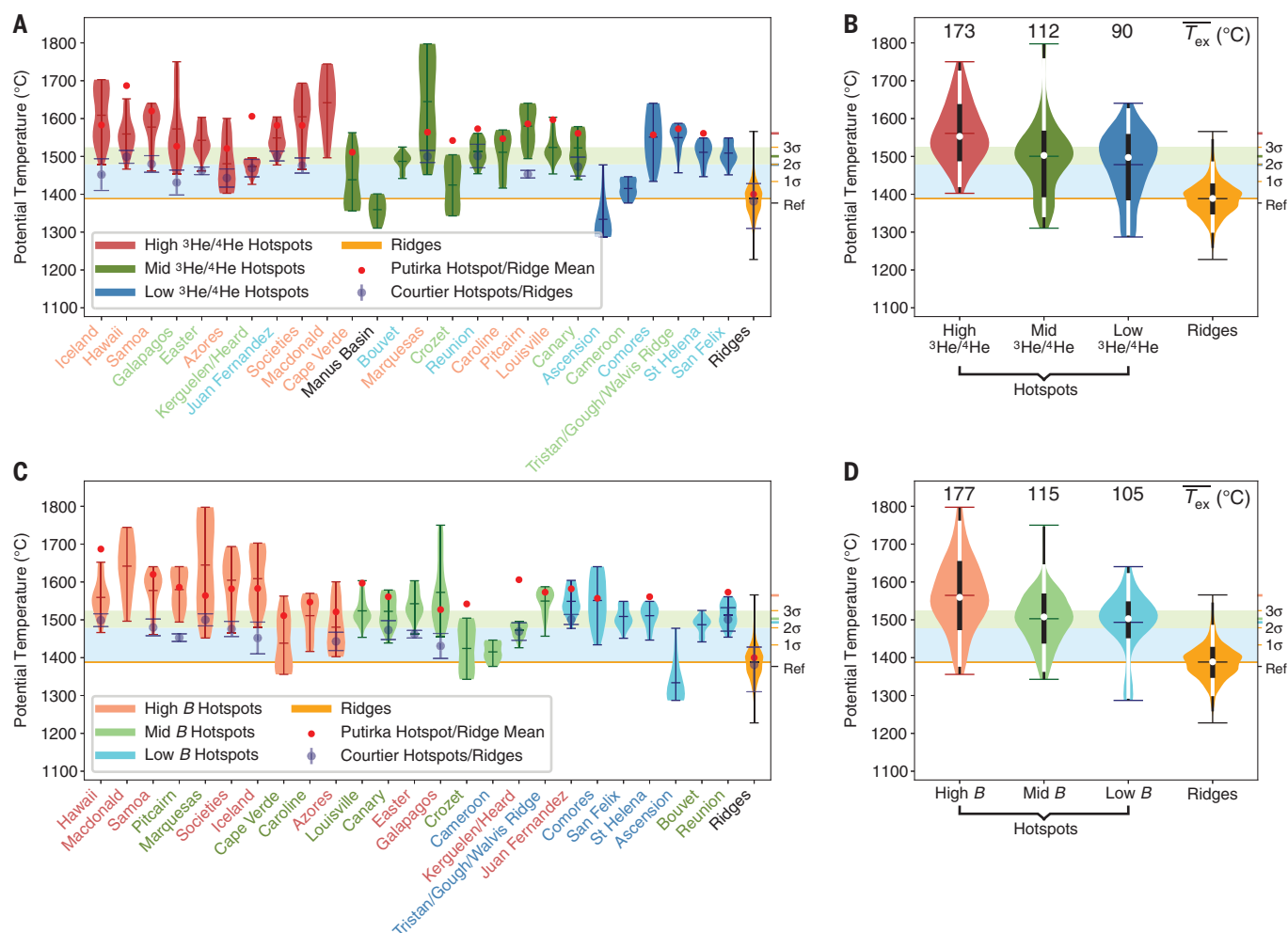*Corresponding author. Email: xiyuanbao@g.ucla.edu

Fig. 1. Map of global hotspots and mid-ocean ridges and schematic of velocity extraction. (**A**) Red, green, and blue squares represent hot, warm, and cold oceanic hotspots, respectively. Squares with black outlines have resolved plumes (6). The numbers correspond to the hotspot IDs in data S1. Orange triangles show mid-ocean ridges. The −1% $\delta \ln V_S$ contour at the CMB (22) represents LLSVPs in white dashed lines. (**B**) Sketch illustrating our search strategy. Radius R decreases with depth to account for the mantle's decreasing volume. Distance A defines the ridge-hotspot proximity threshold, used in (21). We use $A = 1000$ km and $R = 500$ km (21).

Fig. 2. Potential temperature of oceanic hotspots and ridges. Violin plots of the distribution of potential temperature (T_p). (**A**) On the right side, the 1377°C reference adiabat (21) is marked by a black tick. Three yellow ticks indicate $\overline{T_p}$ + 1σ, 2σ, and 3σ for ridges. The width of the violin represents the probability density of temperature estimates from all depths (260 to 600 km). In every violin, horizontal bars, from top to bottom, are maximum, mean, and minimum values. Black outlines are plume-fed hotspots. From left to right, hotspots are arranged in order of decreasing T_p. Red, green, and blue violins are hot, warm, and cold hotspots, respectively, classified by cluster analysis, with ticks on the right showing their $\overline{T_{ex}}$. The blue shaded region ($T_{ex} < 2σ$) covers plumes that cannot actively upwell, green ($2σ < T_{ex} < 3σ$) covers those that barely can. Olivine thermometry estimates from Courtier et al. (4) for hotspots and ridges in purple [average and 1σ, see (21)] and Putirka's (14) in solid red circles. Ridge estimates from this study and Dalton et al. (19) (with purple outlines) shown by yellow violins on the right. (**B**) Violin stacks for three clusters of hotspots compared with ridges (yellow). $\overline{T_{ex}}$ for each group shown on top. White circles are the median values, black and white bars away from the center indicate the central 68th, 95th, and 99.7th percentiles.

In this study, we reexamined whether oceanic hotspots are hotter than ridges using complete ridge segment ($N = 424$) (20) and oceanic hotspot ($N = 46$) (5, 12) catalogs (data S1 and Fig. 1A). We inferred temperature directly from seismic tomography in the upper mantle by converting velocity to temperature, similarly to Dalton et al. (19) for ridges. Dalton et al. (19) converted velocity to temperature at 300 km depth below ridges, but not hotspots, and Jackson et al. (12) did not convert to temperatures and focused only on hotspots at 200 km depth. Examining both hotspots and ridges simultaneously is key to both computing excess temperatures and comparison with previous studies using petrological thermometers (14). We choose to compute T_{ex} with respect to the ridge $\overline{T_p}$ in keeping with previous work (14, 15) and understanding of MORB formation (21). For both ridges and hotspots, we inferred T_p in 20-km intervals from 260 to 600 km depth, below the depths with the highest seismic attenuation (21). Hereafter we refer to potential temperature as temperature. We converted velocity anomalies from the global shear wave tomography model SEMUCB-WM1 (22) to temperature using HeFESTo (23),

Fig. 3. Potential temperature of hotspots with different ³He/⁴He ratios or geometrical buoyancy flux (B). Violin plots of the inferred T_p, for plume hotspots sorted by ³He/⁴He and B. Red, green, and blue ticks on the right are the $\overline{T_{ex}}$ for the corresponding group. Ticks and shaded regions as in Fig. 2. $\overline{T_p}$ of ridges in the yellow violins. (**A**) From left to right: dark red, high (>15.7 Ra); dark green, mid; and dark blue, low (≤9 Ra) ³He/⁴He. Purple and red circles as in Fig. 2. Hotspot names colored after groups from B. (**B**) Hotspots stacked by high, mid, and low (red, green, and blue) ³He/⁴He and compared with ridges (yellow). Black and white bars as in Fig. 2B. (**C**) From left to right: light red, high (>0.66 Mg/s); light green, mid; and light blue, low (≤0.19 Mg/s) B. Purple and red circles as in (A). Hotspot names colored after groups from ³He/⁴He. (**D**) As in (B), but stacking B. For both ³He/⁴He and B, when comparing temperature of the "high" group and the "mid" plus "low" group hotspots (plume hotspots or all hotspots), Welch's t test gives a P value of ≤0.04. Note that 9 Ra corresponds to 1σ above MORB average (see text for more details).

a self-consistent thermodynamic model of the phase assemblage of the mantle and its physical properties (*21*). SEMUCB-WM1 (*22*) uses full waveform inversions and accurate forward wavefield computations and—as is crucial for inferring temperature—has some of the most robust estimates of absolute velocities in the upper mantle (*24*). SEMUCB-WM1 also shows 26 oceanic hotspots with well-resolved plumes (*6*) in our catalog. The reference adiabat, extracted velocities for ridges and hotspots, velocity-to-temperature conversions, and inferred temperatures are shown in fig. S1, and our compiled dataset with the temperature inferred from all seismic models (*22*, *25–28*) can be found in data S1.

Whereas Dalton *et al.* (*19*) found temperature to be the primary driver of MORB major-

element compositional variability and seismic velocity variations, we went a step further and started with the null hypothesis that seismic anomalies in both the subridge and the sub-hotspot mantle are due to temperature alone. Further, we assumed that the major element source of mantle melting beneath hotspots and ridges is identical.

Our reference mantle composition was depleted MORB mantle (DMM) (*29*) with a reference adiabat of 1377°C (1650 K) (*21*), consistent with previous studies (*4*, *19*) and MORB formation (*30*). At each hotspot, and every depth interval, we searched for the local minimum V_S anomaly ($\delta\ln V_S$), which translates to the highest mantle T_p, within a search radius R centered around the hotspot. We set R to 500 km at Earth's surface to account for the

possible lateral deflection of the plume conduit by the "mantle wind" (Fig. 1B) (*21*). For ridges, we took the average seismic velocity as representative of the ambient mantle that melts to form MORB. We expect the average temperature to be more representative of their dynamical origin, as ridges are largely the result of passive return flow away from slabs and have no undisputed slow anomalies throughout the entire upper mantle (*21*). Hence, at each ridge segment, for every depth interval, we averaged all $\delta\ln V_S$ values in a disk of radius R, centered on the ridge segment. By using the local highest temperature for hotspots and the local average temperature beneath each ridge segment, we maximized the inferred T_{ex} between ridges and hotspots (*21*). Even in this conservative case, optimal for making hotspots

hotter than ridges, we found that >10% of oceanic hotspots [~15% of plume-fed hotspots (6)] are not resolvably hotter than ridges (T_{ex} < 50°C). Had we chosen the local highest temperature for ridges, as we did for hotspots, then ~70% of oceanic hotspots (~80% of plume-fed hotspots) would not be hotter than ridges.

The effects of limited resolution, smoothing, and damping in seismic tomography models broaden and reduce the amplitude of $\delta \ln V_S$, leading to lower inferred temperatures. This may be especially true for the local minimum $\delta \ln V_S$ extracted beneath hotspots compared with the broad average $\delta \ln V_S$ beneath midocean ridges. We attempted to correct for these effects beneath hotspots by scaling up $\delta \ln V_S$ by a constant scaling factor f (21). For SEMUCB-WM1 (22), three-dimensional (3D) synthetic resolution tests (6, 22) suggest a conservative scaling factor of $f = 2$ for plumes that have a core radius [cut off at half peak $\delta \ln V_S$ (21)] of 150 km in the upper mantle. To assess both the resolving power of SEMUCB-WM1 for narrower conduits and the appropriate corresponding f value, we performed two additional 3D synthetic resolution tests (fig. S2) for 100 km radius conduits under Ascension (T_{ex} = −10°C) and Cameroon (T_{ex} = 27°C), two cold hotspots. This complements existing resolution tests on Hawaii and Iceland (6). For narrower plumes, f may, depending on resolution, be greater (figs. S2 and S3) (21), but not by enough to alter our main conclusions.

In Fig. 2, we show violin plots (21) of the stacked inferred temperature for all oceanic hotspots and ridge segments at all depths (260 to 600 km). The width of each violin column represents the number density of T_p. The $\overline{T_p}$ for all ridges is 1388° ± 45°C (1σ), very close to the 1377°C reference adiabat. The $\overline{T_p}$ of all oceanic hotspots is 1527° ± 95°C, resulting in an average T_{ex} ($\overline{T_{ex}}$) of 139°C, which spreads over a large range of T_{ex} for individual hotspots. Cluster analysis (Fig. 2) (21) suggests three distinct groups of hotspots: hot ($\overline{T_{ex}}$ = 199°C), warm ($\overline{T_{ex}}$ = 104°C), and cold ($\overline{T_{ex}}$ = −10°C). For 26 oceanic hotspots associated with seismically resolved plumes [i.e., plume-fed hotspots as defined in (6), where we expect the inferred T_p to be most reliable, as the seismic anomaly is visible through much of the mantle], the $\overline{T_p}$ (1519° ± 93° versus 1527° ± 95°C) and $\overline{T_{ex}}$ (131° versus 139°C) are essentially unchanged compared with all oceanic hotspots. We thus focused only on plume-fed hotspots (Fig. 2A, black outlines).

The T_{ex} of the four hotspots (~15% of the 26 hotspots considered) that are cold (−54° ≤ T_{ex} ≤ 36°C) all fall within 1σ (45°C) of the ridge $\overline{T_p}$. For the 10 hotspots (~40%) that are warm (50° ≤ T_{ex} ≤ 136°C), the T_{ex} falls within the ridge 3σ (135°C). Only 6 of the 10 warm hot-

spots have sufficiently high T_{ex} to match or exceed the minimum dynamical limit for mantle plumes with a 100 km radius to rise at 10 cm/year (138°C) (table S1) (21). Thus, even warm hotspots are barely hot and buoyant enough to actively upwell. The remaining 12 hot hotspots have T_{ex} ≥ 155°C, beyond the ridge 3σ and well above the dynamical limit. Taken at face value, our results suggest that nearly a third (N= 8) of plume-fed hotspots are either not resolvably hotter than ridges or not beyond the minimum dynamical limit (T_{ex} ≤ 100°C). The presence of three classes of hotspots, including cold hotspots, is robust and independent of our choice of seismic model (fig. S4) or reference profile (fig. S5). It is of interest that the cutoff between hot and warm hotspots from the cluster analysis (136°C) matches the T_{ex} needed for a 100 km radius plume to rise at 10 cm/year (138°C) (21) (table S1). We use this plume radius and terminal speed in the main text for all calculations, unless otherwise noted.

We found that the $\overline{T_p}$ for ridges (1388° ± 45°C) is consistent with Courtier et al.'s (4) and Putirka's (14) petrological estimates (1381°C and 1400° ± 35°C, respectively) and Dalton et al.'s (19) hybrid thermometer (1385° ± 40°C). Our estimate of the $\overline{T_{ex}}$ for plume-fed hotspots (131° ± 77°C) lies between those of Courtier et al. (4) and Putirka (14) (91° ± 24°C and 177° ± 57°C, respectively) (Fig. 2 and figs. S6 and S7) (21). We note that Courtier et al. (4) also found hotspots with T_{ex} as low as 50°C, below the typical lower bound of 100°C of other petrological studies (14, 31) but in agreement with our inferences.

We need to evaluate whether our null compositional hypothesis is valid by allowing for the presence of a substantial fraction of recycled crust in the plume source (32). We expect the addition of recycled crust (eclogite) to increase the inferred T_{ex} because eclogite is seismically faster than DMM (33). Mechanically mixing 10 and 25% normal N-MORB (normal MORB) (29) with DMM in the hotspot mantle source (21) increases the inferred plume-fed hotspot $\overline{T_p}$ by 14°C (1533° ± 97°C) and 35°C (1554° ± 107°C), respectively, compared with the eclogite-free results ($\overline{T_p}$ of 1519° ± 93°C). Even after adding 25% eclogite, 8% of plume-fed hotspots are still cold (T_{ex} ≤ 45°C) (fig. S8).

However, the addition of a recycled crustal component comes at a substantial dynamical cost. Any additional eclogitic fraction increases density substantially and requires higher T_{ex} to overcome the negative buoyancy. For example, plumes with 25% recycled crust would need a T_{ex} of 368°C (142°C for 10%) just to be neutrally buoyant in the upper mantle (table S1) (21), but they gain only an additional 35°C from the presence of eclogite in the seismic velocity–to–temperature conversion. A T_{ex} of 368°C is consistent with prior estimates (34)

for the upper temperature bound for eclogite-bearing plumes to remain neutrally buoyant at the bottom of the lithosphere. At such a T_{ex} threshold, all hotspots with 25% added eclogite have insufficient T_{ex} to remain neutrally buoyant, let alone rise actively. At 10% eclogitic fraction, >40% of all plume-fed hotspots remain below the neutral buoyancy threshold (142°C). Therefore, the addition of eclogite cannot explain cold hotspots.

The existence of cold hotspots appears robust both from our results and recent work. A petrological study at Cameroon (35) and a seismological study at Ascension (36) show lower T_{ex} (~50°C or less; T_p of 1400° to 1430°C), in agreement with our inferences for these hotspots. Low petrological T_{ex} values have been found for Juan Fernández (37) and Pitcairn (Gambier Island) (38), although these appear as hot hotspots in our estimates. This opens the possibility that, in some cases, we may have overestimated T_{ex} for some plumes, which would strengthen our conclusions.

Low T_{ex} values would seem to make it hard for plumes to rise rapidly enough without losing buoyancy. A minimum T_{ex} of ~135°C is needed for pure DMM plumes with a 100 km radius to rise at ~10 cm/year (16) in the upper mantle (21), consistent with the dynamical limit (17) and simulations (39). Plumes of the same radius containing eclogite require much higher T_{ex} to rise at the same velocity (>300°C for 10% fraction) (table S1). Envisioning cold, or even warm, hotspots as the same dynamical entities as hot hotspots is difficult.

One hypothesis proposed by Courtillot et al. (40) suggests that cold and warm hotspots are fed by passive upwellings or have a shallow source (3). A shallow source for pure DMM hotspots may be possible if they are generated by edge-driven or small-scale convection (41) or by sublithospheric drainage (42). Alternatively, cold and warm hotspots may still be fed by deep plumes that become trapped and cooled by small-scale convection in the upper mantle (43). In either case, these plumes will have small or even negligible T_{ex} in the deeper upper mantle. This could be the case for some cold hotspots without clear age progressions, such as Cameroon (42) (T_{ex} = 27°C) near the West African passive margin. Or these plumes may be weak (T_{ex} ~ 50°C), as suggested by geophysical (36) and geochemical (44) studies beneath Ascension. For narrower plumes (<50 km core radius), it may be that global tomography is unable to resolve them in the upper mantle. This may be the case for Cape Verde (45), which has a relatively high ^{3}He/^{4}He signal [15.7 Ra (46)] but low inferred T_{ex} in our model and is well resolved in the lower mantle (6). But it is not the case for Ascension and Cameroon, two cold plumes that are resolved even at core radii of 50 km (fig. S2) (21). A core radius of <35 km would be required for the hottest cold

plume (T_{ex} = 36°C) to equal the $\overline{T_{ex}}$ for hot hotspots (199°C), as our synthetic resolution tests suggest (21). Narrower plumes, however, would need to rise faster, and thus be hotter than the minimum dynamical limit, so as not to lose their buoyancy and be buffeted by the mantle wind.

For hotspots with a small fraction of recycled oceanic crust (e.g., ~5%), a T_{ex} of ~50°C is enough to overcome the excess density of eclogite. Such low T_{ex} is enough to allow a thermochemical plume to rise with the help of broad passive upward return flow that occurs above large low shear wave velocity provinces (LLSVPs), a return flow that complements downward flow at global subduction zones (21). Broad passive flow can contribute half [~1.5 to 3 cm/year (47)] of the velocity needed by low T_{ex} plumes to rise fast enough to keep their buoyancy (21), the other half arising from their T_{ex}. We cannot rule out the possibility that the cold hotspots are wet, which lowers the melting temperature at shallow depths and leads to decompression melting (48). Whatever the explanation, these hotspots do not fit the classical plume model.

On the other hand, hotspots with the highest T_{ex} do fit the classical model, as they are more than hot enough to rise actively. These hot plumes are also associated with many of the largest buoyancy fluxes and the highest ^{3}He/^{4}He. Figure 3 shows a series of violin plots of the stacked T_p of plume-fed hotspots, colored and sorted by ^{3}He/^{4}He (12) (Fig. 3, A and B) or geometrical buoyancy flux B (5, 21) (Fig. 3, C and D). We use ^{3}He/^{4}He ≤ 9 Ra and B ≤ 0.19 Mg/s as the threshold criterion for hotspots with low flux and He ratios, where 9 Ra is the 1σ deviation from the MORB mean (8 Ra) (7), and both correspond to the lowest 30th percentile. We further divide the remaining hotspots into high (^{3}He/^{4}He > 15.7 Ra; B > 0.66 Mg/s), the top 70th percentile, and intermediate ($9 < ^3$He/^{4}He ≤ 15.7 Ra; $0.19 < B$ ≤ 0.66 Mg/s) categories. We find that the $\overline{T_{ex}}$ of hotspots reduces as a function of decreasing ^{3}He/^{4}He or B (Fig. 3 and figs. S9 and S10). These results directly confirm the relationships between higher plume temperature and extreme ^{3}He/^{4}He and buoyancy flux proposed by Jackson et al. (12) using seismic velocity at 200 km depth, and by Putirka (14) using petrologically derived temperatures.

For Iceland and Hawaii, the two hotspots with the highest ^{3}He/^{4}He, we find a T_p of 1609°C (T_{ex} = 221°C) and 1559°C (T_{ex} = 171°C), respectively, firmly in the hot hotspot cluster (Fig. 2). Iceland's higher T_{ex} compared with Hawaii is more compatible with its higher ^{3}He/^{4}He signal [up to 47.5 Ra (49) versus 35.3 Ra (50)] and estimates of its buoyancy flux (21, 51) but is not compatible with the greater lithospheric thickness beneath Hawaii (48).

If high ^{3}He/^{4}He domains are associated with primordial and denser material, it may be that only the more buoyant (hotter) plumes can entrain it and rise to the surface (12). High ^{3}He/^{4}He in OIBs is geographically correlated to the two LLSVPs (Fig. 1A) in the lowermost mantle (10). LLSVPs may be denser and chemically distinct (52), compatible with a reservoir of dense oceanic crust or primordial material with high ^{3}He/^{4}He (53). Our results suggest that hot hotspots are indeed thermochemical in nature but are hot and buoyant enough to entrain LLSVP material with high ^{3}He/^{4}He.

We find a $\overline{T_{ex}}$ of 205°C for five hotspots overlying large ultra low velocity zones ("mega-ULVZs") at the CMB: Iceland, Hawaii, Samoa, and Marquesas (54) as well as Galápagos (55) (fig. S11A). These hotspots also have the highest ^{3}He/^{4}He, with the exception of Marquesas (54), which has a moderately-high ^{3}He/^{4}He but high B. If these mega-ULVZs are broad regions of partial melt they may provide roots for some hot, strong plumes (54). They may also represent a core–mantle interaction zone, which may be the source of ancient isotopic anomalies (56) in hotspot lavas. Two other classes of plumes appear to be cooler and less thermally buoyant and give rise to low-flux hotspots. They do not appear to entrain deep-seated primordial domains—possibly because they are too cold and thus insufficiently buoyant to entrain a deep and dense high ^{3}He/^{4}He domain (12)—but they may provide important clues to shallower mantle processes such as slow-rising plumes and small-scale convection.

REFERENCES AND NOTES

1. W. J. Morgan, *Nature* **230**, 42–43 (1971).
2. N. H. Sleep, *Spec. Pap. Geol. Soc. Am.* **430**, 29–45 (2007).
3. D. L. Anderson, *Aust. J. Earth Sci.* **60**, 657–673 (2013).
4. A. M. Courtier et al., *Earth Planet. Sci. Lett.* **264**, 308–316 (2007).
5. S. D. King, C. Adam, *Phys. Earth Planet. Inter.* **235**, 66–83 (2014).
6. S. W. French, B. Romanowicz, *Nature* **525**, 95–99 (2015).
7. C. D. Graham, *Rev. Mineral. Geochem.* **47**, 247–317 (2002).
8. M. G. Jackson et al., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30993–31001 (2020).
9. F. M. Stuart, S. Lass-Evans, J. G. Fitton, R. M. Ellam, *Nature* **424**, 57–59 (2003).
10. C. D. Williams, S. Mukhopadhyay, M. L. Rudolph, B. Romanowicz, *Geochem. Geophys. Geosyst.* **20**, 4130–4145 (2019).
11. A. M. Jellinek, M. Manga, *Rev. Geophys.* **42**, RG3002 (2004).
12. M. G. Jackson, J. G. Konter, T. W. Becker, *Nature* **542**, 340–343 (2017).
13. H. Samuel, C. G. Farnetani, *Earth Planet. Sci. Lett.* **207**, 39–56 (2003).
14. K. Putirka, *Geology* **36**, 283 (2008).
15. C. Herzberg, P. D. Asimow, *Geochem. Geophys. Geosyst.* **9**, Q09001 (2008).
16. B. Steinberger, M. Antretter, *Geochem. Geophys. Geosyst.* **7**, Q11018 (2006).
17. M. Albers, U. R. Christensen, *Geophys. Res. Lett.* **23**, 3567–3570 (1996).
18. D. Mckenzie, M. J. Bickle, *J. Petrol.* **29**, 625–679 (1988).
19. C. A. Dalton, C. H. Langmuir, A. Gale, *Science* **344**, 80–83 (2014).
20. A. Gale, C. A. Dalton, C. H. Langmuir, Y. Su, J. G. Schilling, *Geochem. Geophys. Geosyst.* **14**, 489–518 (2013).
21. Materials and methods are available as supplementary materials.
22. S. W. French, B. A. Romanowicz, *Geophys. J. Int.* **199**, 1303–1327 (2014).
23. L. Stixrude, C. Lithgow-Bertelloni, *Geophys. J. Int.* **184**, 1180–1213 (2011).
24. M. Meschede, B. Romanowicz, *Geophys. J. Int.* **200**, 1078–1095 (2015).
25. N. A. Simmons, A. M. Forte, L. Boschi, S. P. Grand, *J. Geophys. Res.* **115**, B12310 (2010).
26. J. Ritsema, A. Deuss, H. J. Van Heijst, J. H. Woodhouse, *Geophys. J. Int.* **184**, 1223–1236 (2011).
27. A. J. Schaeffer, S. Lebedev, *Geophys. J. Int.* **194**, 417–449 (2013).
28. E. Debayle, F. Dubuffet, S. Durand, *Geophys. Res. Lett.* **43**, 674–682 (2016).
29. R. K. Workman, S. R. Hart, *Earth Planet. Sci. Lett.* **231**, 53–72 (2005).
30. R. S. White, T. A. Minshull, M. J. Bickle, C. J. Robinson, *J. Petrol.* **42**, 1171–1196 (2001).
31. C. Herzberg et al., *Geochem. Geophys. Geosyst.* **8**, Q02006 (2007).
32. S. Pugh, J. Jenkins, A. Boyce, S. Cottaar, *Earth Planet. Sci. Lett.* **561**, 116813 (2021).
33. W. Xu, C. Lithgow-Bertelloni, L. Stixrude, J. Ritsema, *Earth Planet. Sci. Lett.* **275**, 70–79 (2008).
34. M. Pertermann, M. M. Hirschmann, *J. Petrol.* **44**, 2173–2201 (2003).
35. C. N. Ngwa, T. H. Hansteen, C. W. Devey, F. M. van der Zwan, C. E. Suh, *Lithos* **288–289**, 326–337 (2017).
36. T. A. Minshull, N. J. Bruguier, J. M. Brozena, *Geology* **26**, 115–118 (1998).
37. J. Reyes et al., *Geochim. Cosmochim. Acta* **257**, 311–335 (2019).
38. H. Delavault, C. Chauvel, A. Sobolev, V. Batanova, *Earth Planet. Sci. Lett.* **426**, 23–35 (2015).
39. M. Arnould, N. Coltice, N. Flament, C. Mallard, *Earth Planet. Sci. Lett.* **547**, 116439 (2020).
40. V. Courtillot, A. Davaille, J. Besse, J. Stock, *Earth Planet. Sci. Lett.* **205**, 295–308 (2003).
41. S. D. King, J. Ritsema, *Science* **290**, 1137–1140 (2000).
42. A. R. Guimarães, J. G. Fitton, L. A. Kirstein, D. N. Barfod, *Earth Planet. Sci. Lett.* **536**, 116147 (2020).
43. S. French, V. Lekic, B. Romanowicz, *Science* **342**, 227–230 (2013).
44. E. Bourdon, C. Hémond, *Mineral. Petrol.* **71**, 127–138 (2001).
45. X. Liu, D. Zhao, *Geophys. J. Int.* **225**, 872–886 (2021).
46. R. Doucelance, S. Escrig, M. Moreira, C. Gariépy, M. D. Kurz, *Geochim. Cosmochim. Acta* **67**, 3717–3733 (2003).
47. A. M. Forte et al., *Earth Planet. Sci. Lett.* **295**, 329–341 (2010).
48. J. G. Fitton, R. Williams, T. L. Barry, A. D. Saunders, *J. Petrol.* **61**, egaa111 (2020).
49. S. Harðardóttir, S. A. Halldórsson, D. R. Hilton, *Chem. Geol.* **480**, 12–27 (2018).
50. P. J. Valbracht, T. Staudacher, A. Malahoff, C. J. Allègre, *Earth Planet. Sci. Lett.* **150**, 399–411 (1997).
51. R. Parnell-Turner et al., *Nat. Geosci.* **7**, 914–919 (2014).
52. M. Ishii, J. Tromp, *Science* **285**, 1231–1236 (1999).
53. T. D. Jones, R. Maguire, P. van Keken, J. Ritsema, P. Koelemeijer, EarthArXiv [Preprint] (2021). https://doi.org/10.31223/osf.io/k98p7.
54. D. Kim, V. Lekić, B. Ménard, D. Baron, M. Taghizadeh-Popp, *Science* **368**, 1223–1228 (2020).
55. S. Cottaar, Z. Li, *AGU Fall Meeting Abstracts* **2019**, DI33A-04 (2019).
56. A. Mundl-Petermeier et al., *Geochim. Cosmochim. Acta* **271**, 194–211 (2020).

NANOMATERIALS

Single-walled zeolitic nanotubes

Akshay Korde[1], Byunghyun Min[1], Elina Kapaca[2], Omar Knio[1], Iman Nezam[1], Ziyuan Wang[1], Johannes Leisen[3], Xinyang Yin[4], Xueyi Zhang[4], David S. Sholl[1], Xiaodong Zou[2], Tom Willhammar[2]*, Christopher W. Jones[1,3]*, Sankar Nair[1]*

We report the synthesis and structure of single-walled aluminosilicate nanotubes with microporous zeolitic walls. This quasi-one-dimensional zeolite is assembled by a bolaform structure-directing agent (SDA) containing a central biphenyl group connected by C_{10} alkyl chains to quinuclidinium end groups. High-resolution electron microscopy and diffraction, along with other supporting methods, revealed a unique wall structure that is a hybrid of characteristic building layers from two zeolite structure types, beta and MFI. This hybrid structure arises from minimization of strain energy during the formation of a curved nanotube wall. Nanotube formation involves the early appearance of a mesostructure due to self-assembly of the SDA molecules. The biphenyl core groups of the SDA molecules show evidence of π stacking, whereas the peripheral quinuclidinium groups direct the microporous wall structure.

Zeolites are widely used as size- and shape-selective catalysts and adsorbents because of their ordered microporous structure (1–3). There has been considerable interest in the synthesis of zeolites with hierarchical porosity (4–16) that allow access to a wider range of molecules. Early approaches (10–12) included postsynthesis treatments to etch mesopores into zeolite crystals. More recently, new structure-directing agents (SDAs) have been used to create two-dimensional (2D) zeolite nanosheets interspersed by mesoporous regions (6–8, 13–15), yielding nanosheets of several zeolitic topologies such as MFI, MWW, FAU, AEL, and others (6–13, 15–23). This is usually achieved with di-quaternary ammonium surfactant SDAs, in which the quaternary ammonium groups direct zeolite formation in two dimensions, whereas the long hydrocarbon moieties prevent zeolite crystallization in the third dimension. Interactions such as π stacking between the SDA molecules (15, 18–23) can also enhance their self-assembly into lamellar structures that allow 2D zeolite formation.

We report the first synthesis and structural characterization of a quasi-1D hierarchical zeolite, specifically a single-walled nanotube that has a microporous zeolitic wall enclosing a central mesoporous channel. We synthesized a bolaform SDA (BCPh10Qui; Fig. 1) that is capable of π stacking because of the central biphenyl moiety and has bulky quinuclidinium SDA head groups linked to the biphenyl moiety by C_{10} alkyl chains. This SDA was used for hydrothermal synthesis at 423 K in an alkaline

aluminosilicate medium with an Si/Al ratio of ~30 (see the supplementary materials, including fig. S1, for SDA and zeolite synthesis and characterization methods). Although "rational" design of SDAs for zeolite synthesis remains difficult and unreliable, we speculated that a long-chain SDA containing an aromatic (π-stacking) species at its center might also template a nanotubular zeolite because many conventional surfactants can form lamellar and rodlike micelles. Further, the attachment of bulky quaternary ammonium head groups using sufficiently long and flexible alkyl chain connectors could direct zeolite formation away from lamellar (2D) to tubular (1D) materials and allow the formation of a cylindrical zeolitic wall.

The formation of nanotubes was apparent from transmission electron microscopy (TEM) images showing individual nanotubes and nanotube bundles in the as-made material (fig. S2, A and B) and after SDA removal by calcination at 823 K (Fig. 2A and fig. S2, C and D). Other materials such as 3D crystals or 2D nanosheets were not observed. The typical nanotube yield (see the materials and methods) was >28% based on Si and >60% based on Al. High-resolution N_2 physisorption at 77 K (24) clearly revealed mesopores (the nanotube channels that form the bulk of the total porosity) and micropores (indicating zeolitic nanotube walls) (fig. S3). The mesopore size distribution (BJH method) shows a narrow peak at ~2.5 nm, suggesting a quite monodisperse channel diameter. The micropore size distribution (HK method) shows a peak at ~0.5 nm in the range of a medium-pore zeolite. Because of the large mesoporosity, the nanotubes have a very high BET surface area of 980 m^2/g compared with 410 m^2/g for a conventional MFI material. Ar adsorption measurements (fig. S4, A and B) allow greater microporosity resolution. The mesopore size distribution (fig. S4C) exhibits a sharp peak at 3 nm, in good agreement with N_2 physisorp-

tion. The micropore size is in the same range (5.6 to 6.2 Å) as those of conventional MFI and beta zeolites (fig. S4D). An artifact peak at 8 to 12 Å in all three materials is caused by a known phase transition of adsorbed Ar (7).

Low-angle and wide-angle powder x-ray diffraction (PXRD) patterns of the calcined nanotubes are shown in Fig. 2, B and C (also see fig. S5). As shown earlier for imogolite nanotubes (25, 26), the low-angle PXRD patterns are dominated by the scattering form factors of individual nanotubes and small nanotube bundles, and the primary peak position approximately corresponds to the outer diameter of individual nanotubes. This peak (~4.2 nm in Fig. 2B) is representative of the nanotube diameter, and the subsequent peaks (2 and 1.1 nm) are higher-order scattering peaks. The peaks (0.58 and 0.39 nm) in the wide-angle PXRD pattern indicate periodicity within the nanotube walls (Fig. 2C). The curvature of thin (~1-nm) nanotube walls into a closed cylinder rather than an extended 3D crystal or 2D sheet results in broad PXRD peaks unsuited for structure determination (27–29). Figure S6A shows the Fourier transform infrared (FTIR) spectra of the as-made and calcined nanotubes and the pure SDA. Peaks from the SDA are visible in as-made nanotubes and disappear upon calcination. Peaks in the ~1225 and ~550 cm^{-1} regions are clearly present in the nanotubes and indicate pentasil [silicate five-membered rings (5MRs)] structural units (30–33). Figure
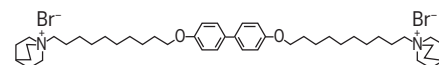
Fig. 1. Structure-directing agent BCPh10Qui-1,1'-(((([1,1'-biphenyl]-4,4'-diylbis(oxy))bis(decane-10,1-diyl))bis(quinuclidin-1-ium) bromide.

S6B compares the FTIR spectra of calcined nanotubes with three pentasil-rich zeolites: 3D BEA, 3D MFI, and 2D MFI. The BEA and MFI spectra show well-known and distinct 5MR signatures at ~1225 cm^{-1} (external asymmetric stretching of 5MR chains) and 525 to 580 cm^{-1} (double 5MRs) (30–34). The ^{29}Si nuclear magnetic resonance (NMR) spectrum of the as-made nanotubes (fig. S7A) shows three peaks at –99 ppm (Q^3), –106.6 ppm (Q^4 3Si,1Al), and –113.3 ppm (Q^4 4Si). The Q^3 signals are from Si atoms on the wall surface that are presumably terminated by Si-OH groups, and Q^4 signals are from interior Si atoms in the wall. The Al-bonding environment (^{27}Al NMR peak at 54 ppm; fig. S8) in as-made and calcined nanotubes corresponds to tetrahedral Al, with no evidence of octahedral or extraframework Al. On the basis of the peak areas (35), the Si/Al ratio was calculated as 16. The fraction of Q^3 Si atoms is 0.15, similar to 2D zeolite sheets with nearly single-unit cell thicknesses (36, 37).

[1]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]Department of Materials and Environmental Chemistry, Stockholm University, 10691 Stockholm, Sweden. [3]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA. [4]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA.
*Corresponding author. Email: sankar.nair@chbe.gatech.edu (S.N.); cjones@chbe.gatech.edu (C.W.J.); tom.willhammar@mmk.su.se (T.W.)

The calcined nanotubes (fig. S7B) show peaks at −102 ppm (Q^3) and −110 ppm (Q^4), with a Q^3 fraction of 0.17. This value is similar to the as-made nanotubes and indicates no significant condensation of surface silanols after calcination. Normalized FTIR spectra for the nanotubes and a beta zeolite of similar Si/Al ratio (fig. S9) show a similar nature of silanol peaks in both materials, with higher silanol peak intensity in the nanotubes. Isolated (3745 cm^{-1}), terminal (3710 cm^{-1}), internal (3670 cm^{-1}), Al-bridged (3610 cm^{-1}), and H-bonded (3520 cm^{-1}) silanols (38, 39) are recognizable, but the presence of multiple broad O–H stretch bands at lower wave numbers precludes the identification of any other peaks (40).

The acid site densities were estimated by temperature-programmed desorption of ammonia (NH$_3$-TPD) and FTIR measurements of pyridine adsorption on the nanotubes in their proton-exchanged form. Figure S10A shows NH$_3$-TPD traces and acid site concentrations for the proton-exchanged nanotubes and, for comparison, a conventional MFI (ZSM-5) catalyst with a similar Si/Al ratio (20). The strength and density of weak acid sites for the two materials are quite comparable, but ZSM-5 has more strong acid sites. The ZSM-5 catalyst would have a theoretical acid site density of 794 μmol/g, close to the measured sum of weak and strong acid sites (721 μmol/g) shown in fig. S10A. The nanotubes have a measured total acid site density of 479 μmol/g (the theoretical value is 988 μmol/g based on the Al content). The sum of Brønsted (B) and Lewis (L) acid site densities measured by pyridine adsorption is 151 μmol/g (fig. S10B), much lower than the NH$_3$-TPD result. The nanotube material thus has an acid site accessibility factor (AF = pyridine acid site density/NH$_3$ acid site density) of 0.31 and a B/L site ratio of 0.93. Although ^{27}Al NMR provides no evidence for extra-framework Al (fig. S8) that is sometimes hypothesized to provide Lewis acid sites, it has been shown (41) that there is no correlation between such Al sites and Lewis acid site densities in zeolites. The moderate AF may relate to the high aggregation tendency of the nanotubes, making a considerable fraction of acid sites inaccessible to the larger pyridine molecules. Discrepancies between NH$_3$-TPD and pyridine infrared are well known in zeolites and other materials (42, 43), but detailed analysis of the nanotube acid site behavior should be performed in the future.

The crystal structures of most polycrystalline 3D periodic zeolites have been determined either from 3D electron diffraction or PXRD. For materials lacking 3D periodicity, such as 2D zeolites, high-resolution TEM imaging has been key to structure elucidation. These techniques are particularly challenging in the present context because of the reduction of the zeolite to a hollow cylindrical 1D form with a
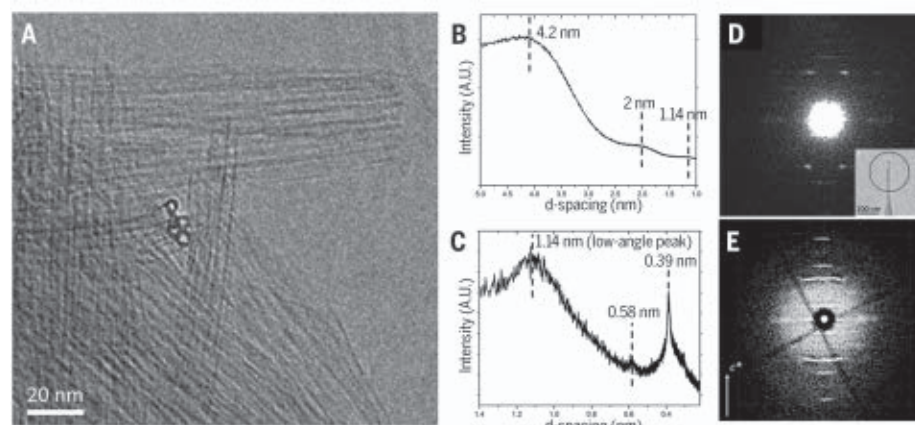
Fig. 2. Zeolite nanotube morphology and diffraction patterns. (**A**) TEM image showing the tubular morphology of the calcined zeolitic nanotube material. (**B** and **C**) PXRD patterns of the calcined nanotube material showing low-angle (B) and wide-angle (C) regions. (**D**) Selected area electron diffraction pattern from a nanotube (marked in the inset) showing typical tubular features with periodicity along the nanotube direction and characteristic diffraction streaks perpendicular to the nanotube direction. (**E**) Reconstructed 3D reciprocal lattice from cRED data collected from a bundle of zeolitic nanotubes with the nanotube direction c* marked.

very thin (~1-nm) wall. A region of the sample consisting of one or two aligned nanotubes was used for selected area electron diffraction (SAED) and 3D continuous rotation electron diffraction (cRED) (Fig. 2, D and E). The patterns show characteristic features similar to those of carbon nanotubes (44) and imogolite nanotubes (45). Both the SAED pattern and reconstructed 3D reciprocal space based on cRED data in Fig. 2, D and E, reveal a distinct periodicity of 12.5 Å along the nanotube direction (denoted c*), with no apparent periodicity observed perpendicular to the c* axis.

High-resolution annular dark-field scanning TEM (ADF-STEM) and integrated differential phase contrast (iDPC) images were obtained both perpendicular to and along the nanotube direction after sectioning the nanotubes by ultramicrotomy (Fig. 3). Images acquired along the nanotube direction of individual and fused nanotubes (Fig. 3, A and B, and fig. S11) confirm a tubular structure with an ~5-nm outer diameter and an ~3-nm inner diameter. Ten identical repeating units with square-like features are frequently observed around the circumference of the nanotubes, and the distance between adjacent units is 12 to 13 Å. Occasionally, nanodomains with micropores of ~6-Å diameter and an arrangement resembling 3D zeolite beta (*BEA) (27) are observed (figs. S12 and S13, A to E). The square-like feature is found in both the nanotubes and the zeolite beta-like domains. A structural model of the circumferential building unit of the nanotube (fig. S13E) could be deduced from the image of an incomplete nanotube (fig. S13, A to C) and the beta structure. Images acquired perpendicular to the nanotube direction reveal the projected wall structure in

more detail (Fig. 3, C and D). The Fourier transform of the image (inset in Fig. 3C) confirms the periodicity of ~12.5 Å along the nanotube direction and a lack of long-range periodicity perpendicular to the nanotube direction, consistent with the electron diffraction data. Isolated dark features of ~6-Å diameter are observed on the nanotube wall surface (Fig. 3D), revealing the presence of micropores on the nanotube wall. The pore size range corresponds to 10MRs to 12MRs perforating the wall. The micropores are arranged at an oblique angle of ~108° (fig. S14) with respect to the nanotube channel axis at a distance of 12 Å, similar to that in zeolite beta.

On the basis of the iDPC STEM images and the axial periodicity from cRED, the structural model of the nanotube is deduced (Fig. 4). The circumferential building unit (Fig. 3 and fig. S13) is depicted in fig. S15A. Repetition of 10 such building units leads to the circumferential cross-section of the nanotube (Fig. 3A and fig. S15B). In the nanotube circumference, these building units are connected through a 5MR (figs. S13 and S15) rather than through a 6MR as in zeolite beta. Although the connection through a 6MR in zeolite beta retains the orientation of the building units, the connection through a 5MR in the nanotube enforces a ~36° rotation of the building unit relative to its neighbors (fig. S15, B and C). This leads to closure of the cylindrical sheet (nanotube) with 10 building units (fig. S15B). The terminal T sites in the walls can act as branching points to form fused nanotubes. Branching occurs between two circumferential building units, as observed in the ADF and iDPC-STEM images (Fig. 3B and fig. S16). The Q^3 fraction of T atoms in the structural model is 0.23, which
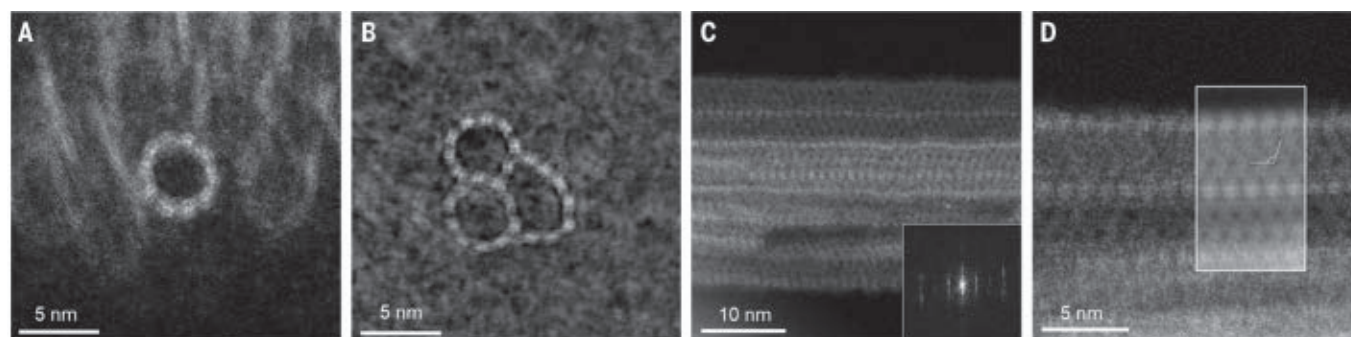
Fig. 3. Zeolite nanotube structure. (**A**) ADF-STEM imaging of an individual single-walled zeolitic nanotube viewed along the nanotube direction. (**B**) Three fused nanotubes imaged with iDPC STEM imaging. The two circular nanotubes (left) each display 10 identical building units around the circumference, and the third tube (right) contains 11 such building units and is no longer circular. (**C**) ADF-STEM image viewed perpendicular to the nanotube direction, with the Fourier transform (inset) showing a periodicity of 12.5 Å along the nanotube direction. (**D**) enlarged ADF-STEM image revealing the fine structure of the nanotube, with the Fourier-filtered image shown in the inset. Micropores with a diameter of ~6 Å are visible as isolated dark features in (C) and (D), corresponding to 10MRs to 12MRs.

is slightly higher than the value of 0.17 obtained by ^{29}Si NMR for calcined nanotubes. The branching of the nanotubes and presence of zeolite beta-like nanoregions will reduce the Q^3 fraction of T atoms, leading to a lower Q^3 fraction compared with the ideal nanotube model. The pentasil-rich structure is consistent with the FTIR spectra (fig. S6), showing characteristic 5MR vibrations.

Geometrical optimization of the pure-silica (no Al) nanotube structure was performed after termination of the Q^3 Si atoms with hydroxyl groups. The model (Fig. 4, A and B, and structural model in the supplementary materials) converged to a structure with reasonable bond geometries (table S1). The geometry-optimized structure has a periodicity of 12.65 Å along the nanotube axis, which agrees well with the SAED, ADF, and iDPC results. Its outer diameter (based on the outermost Si atoms) and wall thickness are 4.6 and 0.5 nm, respectively, in agreement with the STEM images (fig. S16). The wall structure allows for polytypic structural disorder (fig. S17) similar to 3D zeolite beta (46, 47). This stacking disorder is based on allowed ±1/3 translations of the 12.65-Å periodicity along the extended *c* axis. To close the nanotube, the sum of all translation vectors should be an integer ($\pm n*c$). This restriction might account for the observation of some incomplete nanotubes in the microscope images. The simulated and experimental PXRD patterns of the individual nanotubes are in very good agreement (fig. S18). A notable feature emerging naturally from the arrangement of the building units is the presence of 10MR and 12MR micropores on the inner and outer wall surfaces, respectively (Fig. 4, C to E). Because of the nanotube curvature, the two surfaces have different topological structures. The outer surface is built from 4MRs, 5MRs, and 6MRs, leading to 12MR micropores, whereas the inner surface is built from only 5MRs and 6MRs, leading to 10MR micropores. The

outer surface is topologically identical to a layer of zeolite beta. For the case of strictly consecutive stacking (+1/3, +1/3... or –1/3, –1/3... translations), the inner surface is topologically identical to a building layer in the *ac* plane of zeolite MFI. The nanotube wall can thus be considered a unique "atomic-scale" hybrid of zeolites beta (polymorph B) and MFI. Such a hybrid cannot be formed in a 3D or 2D structure, and instead requires curvature into a cylindrical nanotubular morphology.

To study the energetics of the nanotube diameter, structural models were constructed from six, eight, 10, 12, and 14 building units (fig. S19) and geometrically optimized. The nanotube built from 10 units (which is also the experimentally observed nanotube) has the most favorable geometry in terms of Si–O distances as well as O–Si–O and Si–O–Si angles (table S1 and fig. S20). This nanotube also exhibits a clear minimum in the computed surface energy (fig. S21) caused by optimal balance of bond geometries in the inner and outer surfaces. Because of the curvature-induced strain, the major and minor dimensions of the 12MRs (7.91 × 6.44 Å, after subtraction of two oxygen radii of 1.35 Å) and 10MRs (5.89 × 5.63 Å) in the optimized nanotube structure are distorted relative to the 12MRs in the 3D *BEA [100] projection (7.17 × 6.33 Å) and MFI (6.23 × 4.98 Å) optimized with the same force field. This may also influence the effective pore size distributions obtained from Ar adsorption (fig. S4).

To obtain initial observations of the nanotube formation process, synthesis products from 1 to 7 days of hydrothermal synthesis at 423 K were analyzed by PXRD and TEM (figs. S22 and S23). Small-angle PXRD patterns show early development of mesopore domains with a characteristic scale that does not change significantly with time. The wide-angle PXRD patterns show evolution of the nanotube wall structure from amorphous to an ordered zeolitic form. These observations are consistent

with the TEM images, in which the mesoporosity of the material is clearly visible at an early stage. Proto-nanotubes are visible at 3 days and distinct nanotubes at 5 to 7 days. Thus, the overall growth mechanism of the zeolite nanotubes appears to have some similarities to the growth of 2D zeolite nanosheets, i.e., the initial formation of a mesophase followed by transformation to an ordered zeolitic material (48). A key difference is in the morphology-directing effect of the bolaform SDA used in this work, which creates a 1D nanotubular morphology rather than 2D nanosheets. Other bolaform molecules with aromatic rings in their hydrophobic core are known to π-stack and form stable cylindrical or rod-like micellar assemblies (15, 16, 18–20, 49). Thus, we hypothesized that our bolaform SDA (BCPh10Qui) might π-stack sufficiently to direct the formation of nanotubular zeolites. Figure S24 shows ultraviolet-visible diffuse reflectance absorption spectra of the solid SDA, dilute aqueous SDA solution, and as-made nanotubes. In the dilute solution, the SDA molecules are isolated, and a single absorption is observed at 265 nm (π-HOMO → π*-LUMO transition). In the solid SDA, this transition is red-shifted to a double peak beyond 300 nm because of π stacking. The as-made nanotubes also show a double peak that is evidence of significant π stacking, albeit not as extensive as in the solid SDA. Figure S25 compares ^{13}C CPMAS NMR spectra of as-made nanotubes and solid SDA and confirms that the SDA is intact in the nanotubes. Elemental analysis (table S2) reveals a C/N atomic ratio of 25 in the as-made nanotubes (agreeing with C/N = 23 in the SDA; Fig. 1) and an Si/Al ratio of 15 (in agreement with the Si/Al ratio of 16 from NMR). Thermogravimetric analysis of the as-made nanotubes shows that the SDA accounts for 51% of total mass (fig. S26), in agreement with elemental analysis in which C, H, and N account for 48% of total mass. The
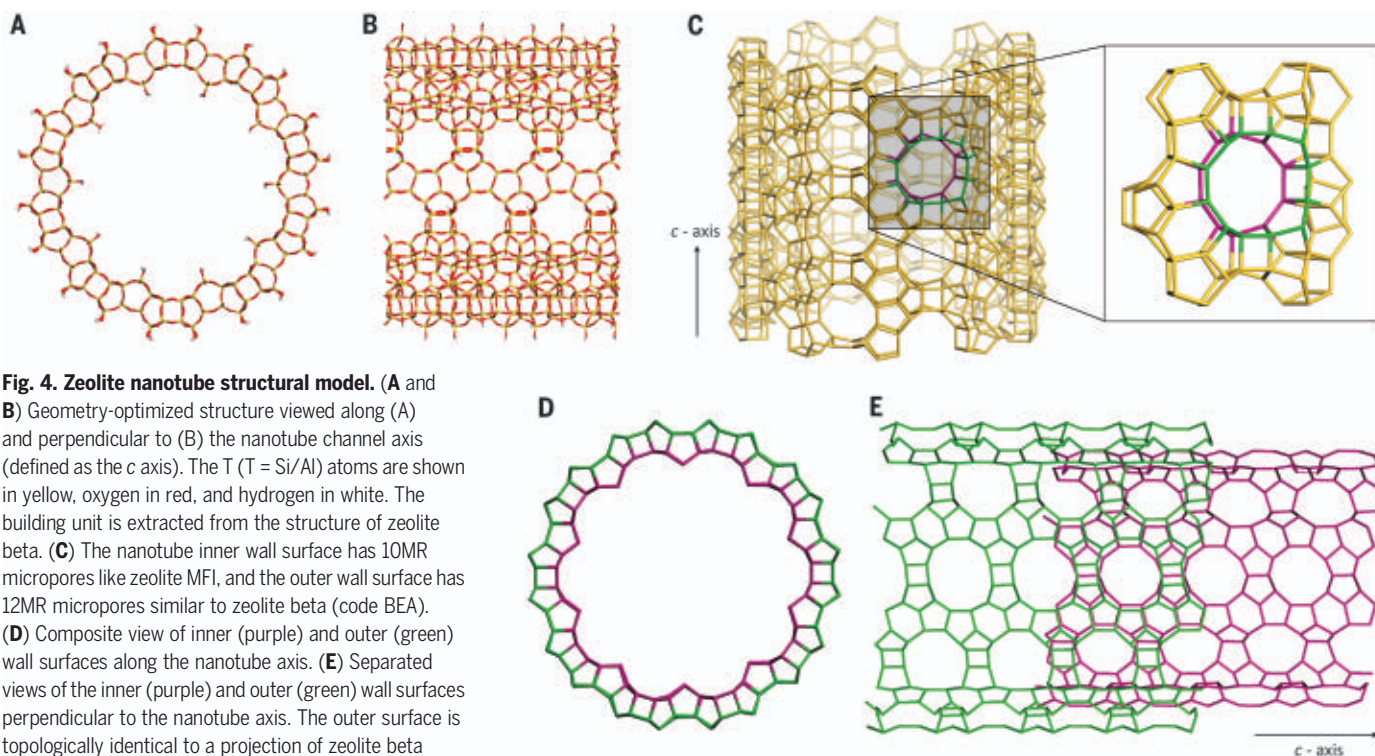
Fig. 4. Zeolite nanotube structural model. (**A** and **B**) Geometry-optimized structure viewed along (A) and perpendicular to (B) the nanotube channel axis (defined as the *c* axis). The T (T = Si/Al) atoms are shown in yellow, oxygen in red, and hydrogen in white. The building unit is extracted from the structure of zeolite beta. (**C**) The nanotube inner wall surface has 10MR micropores like zeolite MFI, and the outer wall surface has 12MR micropores similar to zeolite beta (code BEA). (**D**) Composite view of inner (purple) and outer (green) wall surfaces along the nanotube axis. (**E**) Separated views of the inner (purple) and outer (green) wall surfaces perpendicular to the nanotube axis. The outer surface is topologically identical to a projection of zeolite beta (in this case, polymorph B) and the inner surface to a projection of zeolite MFI.

hydrophobic core of the long-chain SDA is the main contributor to this high organic content, which is also seen in hierarchical 2D zeolitic materials synthesized with long-chain SDAs (*14*, *15*, *19*, *20*),

In conclusion, a quasi-1D zeolite in the form of single-walled nanotubes with zeolitic walls has been synthesized for the first time and its structure revealed. The concept of directing zeolite nanotube synthesis using bolaform SDAs capable of π-stacking of the hydrocarbon core is introduced. Closure of a thin zeolitic sheet into a nanotube is shown to result in a nanotube wall with structurally different inner and outer surfaces, in the present case, a hybrid of zeolite beta and MFI layers. The exact arrangement of the SDA molecules in the as-made nanotubes is not currently known. Our experimental observations suggest that the biphenyl rings of the SDA molecules may form a stable π-stacked hydrophobic core along the nanotube axis, whereas the flexible alkyl chains with the quinuclidinium head groups stretch out along the radius of the nanotube in different directions, reaching into the microporous walls that are templated by the head groups. A number of different 1D zeolite nanotubes could potentially be synthesized by the above concept using a wide range of bolaform SDAs and reaction conditions. More detailed studies of the formation mechanisms and the effects of synthesis conditions are also

desirable to better guide such strategies. The zeolitic nanotubes are stable under high-temperature calcination, like 2D and 3D zeolites. New functional properties could result from the ability to transport molecules axially within catalytically active nanotubular zeolitic channels while also allowing radial molecular transport, exchange, and catalytic conversion through the ultrathin (~1-nm) microporous wall. These phenomena cannot be realized with conventional nanotubes.

REFERENCES AND NOTES

1. C. S. Cundy, P. A. Cox, *Chem. Rev.* **103**, 663–702 (2003).
2. A. Corma, *J. Catal.* **216**, 298–312 (2003).
3. M. E. Davis, *Nature* **417**, 813–821 (2002).
4. Y. Tao, H. Kanoh, L. Abrams, K. Kaneko, *Chem. Rev.* **106**, 896–910 (2006).
5. J. Čejka, S. Mintova, *Catal. Rev., Sci. Eng.* **49**, 457–509 (2007).
6. X. Zhang *et al.*, *Science* **336**, 1684–1687 (2012).
7. K. Na *et al.*, *Science* **333**, 328–332 (2011).
8. V. J. Margarit, M. E. Martínez-Armero, M. T. Navarro, C. Martínez, A. Corma, *Angew. Chem. Int. Ed.* **54**, 13724–13728 (2015).
9. K. Möller, T. Bein, *Chem. Soc. Rev.* **42**, 3689–3707 (2013).
10. D. Verboekend, J. Pérez-Ramírez, *Catal. Sci. Technol.* **1**, 879–890 (2011).
11. R. A. Beyerlein, C. Choi-feng, J. B. Hall, B. J. Huggins, G. J. Ray, *Top. Catal.* **4**, 27–42 (1997).
12. A. Feliczak-Guzik, *Microporous Mesoporous Mater.* **259**, 33–45 (2018).
13. W. Fan *et al.*, *Nat. Mater.* **7**, 984–991 (2008).
14. M. Choi *et al.*, *Nature* **461**, 246–249 (2009).
15. D. Xu *et al.*, *Nat. Commun.* **5**, 4262 (2014).
16. D. Xu, Z. Jing, F. Cao, H. Sun, S. Che, *Chem. Mater.* **26**, 4612–4619 (2014).
17. D. P. Serrano, J. M. Escola, P. Pizarro, *Chem. Soc. Rev.* **42**, 4004–4035 (2013).
18. B. K. Singh *et al.*, *Chem. Mater.* **26**, 7183–7188 (2014).
19. X. Shen *et al.*, *Angew. Chem. Int. Ed.* **57**, 724–728 (2018).
20. Y. Zhang *et al.*, *Chemistry* **25**, 738–742 (2019).
21. Y. Zhang, Y. Ma, S. Che, *Chem. Mater.* **30**, 1839–1843 (2018).
22. Y. Zhang, S. Che, *Angew. Chem. Int. Ed.* **59**, 50–60 (2020).
23. Y. Seo, S. Lee, C. Jo, R. Ryoo, *J. Am. Chem. Soc.* **135**, 8806–8809 (2013).
24. P. I. Ravikovitch, S. C. O. Domhnaill, A. V. Neimark, F. Schüth, K. K. Unger, *Langmuir* **11**, 4765–4772 (1995).
25. D. Y. Kang *et al.*, *Nat. Commun.* **5**, 3342 (2014).
26. D. Y. Kang *et al.*, *ACS Nano* **4**, 4897–4907 (2010).
27. Database of Zeolite Structures, http://www.iza-structure.org/databases/.
28. R. Pophale, P. A. Cheeseman, M. W. Deem, *Phys. Chem. Chem. Phys.* **13**, 12407–12412 (2011).
29. O. Knio, A. J. Medford, S. Nair, D. S. Sholl, *Chem. Mater.* **31**, 353–364 (2019).
30. J. C. Jansen, F. J. van der Gaag, H. van Bekkum, *Zeolites* **4**, 369–372 (1984).
31. G. Coudurier, C. Naccache, J. C. Vedrine, *J. Chem. Soc. Chem. Commun.* (24): 1413–1415 (1982).
32. P. A. Jacobs, H. K. Beyer, J. Valyon, *Zeolites* **1**, 161–168 (1981).
33. D. Lesthaeghe *et al.*, *J. Phys. Chem. C* **112**, 9186–9191 (2008).
34. S. R. Tomlinson, T. McGown, J. R. Schlup, J. L. Anthony, *Int. J. Spectrosc.* **2013**, 1–7 (2013).
35. J. Klinowski, S. Ramdas, J. M. Thomas, C. A. Fyfe, J. S. Hartman, *J. Chem. Soc. Faraday Trans. 2 Mol. Chem. Phys.* **78**, 1025–1050 (1982).
36. H. Zhang *et al.*, *Angew. Chem. Int. Ed.* **55**, 7184–7187 (2016).
37. J. E. Schmidt, D. Xie, M. E. Davis, *Chem. Sci.* **6**, 5955–5963 (2015).
38. J.-P. Gallas *et al.*, *Langmuir* **25**, 5825–5834 (2009).
39. A. A. Gabrienko *et al.*, *J. Phys. Chem. C* **122**, 25386–25395 (2018).
40. F. Eder, M. Stockenhuber, J. A. Lercher, *J. Phys. Chem. B* **101**, 5414–5419 (1997).
41. M. Ravi, V. L. Sushkevich, J. A. van Bokhoven, *Nat. Mater.* **19**, 1047–1056 (2020).
42. P. M. Kester, J. T. Miller, R. Gounder, *Ind. Eng. Chem. Res.* **57**, 6673–6683 (2018).
43. J. S. Yoon, M. B. Park, Y. Kim, D. W. Hwang, H.-J. Chae, *Catalysts* **9**, 933 (2019).

44. J. C. Meyer, M. Paillet, G. S. Duesberg, S. Roth, *Ultramicroscopy* **106**, 176–190 (2006).
45. P. D. G. Cradwick *et al.*, *Nat. Phys. Sci (Lond.)* **240**, 187–189 (1972).
46. M. M. J. Treacy, J. M. Newsam, *Nature* **332**, 249–251 (1988).
47. F. Taborda, T. Willhammar, Z. Wang, C. Montes, X. Zou, *Microporous Mesoporous Mater.* **143**, 196–205 (2011).
48. K. Na *et al.*, *J. Am. Chem. Soc.* **132**, 4169–4177 (2010).
49. C. Wang, Z. Wang, X. Zhang, *Acc. Chem. Res.* **45**, 608–618 (2012).

BATTERIES

Capturing the swelling of solid-electrolyte interphase in lithium metal batteries

Zewen Zhang[1], Yuzhang Li[1,2], Rong Xu[1], Weijiang Zhou[3], Yanbin Li[1], Solomon T. Oyakhire[4], Yecun Wu[5], Jinwei Xu[1], Hansen Wang[1], Zhiao Yu[4,6], David T. Boyle[6], William Huang[1], Yusheng Ye[1], Hao Chen[1], Jiayu Wan[1], Zhenan Bao[4], Wah Chiu[3,7,8]*, Yi Cui[1,9]*

Although liquid-solid interfaces are foundational in broad areas of science, characterizing this delicate interface remains inherently difficult because of shortcomings in existing tools to access liquid and solid phases simultaneously at the nanoscale. This leads to substantial gaps in our understanding of the structure and chemistry of key interfaces in battery systems. We adopt and modify a thin film vitrification method to preserve the sensitive yet critical interfaces in batteries at native liquid electrolyte environments to enable cryo–electron microscopy and spectroscopy. We report substantial swelling of the solid-electrolyte interphase (SEI) on lithium metal anode in various electrolytes. The swelling behavior is dependent on electrolyte chemistry and is highly correlated to battery performance. Higher degrees of SEI swelling tend to exhibit poor electrochemical cycling.

E lectrode-electrolyte interfaces are important to technologies ranging from electrical energy generation and storage to the synthesis of chemicals and materials (*1*, *2*). These electrochemical interfaces are complex and experimentally difficult to study, in part as the result of a lack of effective tools to characterize with high resolution. This gap in understanding has contributed to insufficient experimental control over interfacial structure and reactivity. For example, the solid-electrolyte interphase (SEI)—an interfacial layer formed at the electrode-electrolyte interface because of the electrochemical and chemical decomposition of electrolytes—is a key component responsible for the reversible operation of Li-ion and Li metal batteries (*3–5*). Thus, efforts have been made to engineer the SEI to enable battery chemistries with higher energy densities and longer cycles (*6–9*). However, fundamental understanding of the interfacial phenomena in these battery chemistries is still limited. Elucidating the nanoscale structures and chemistries at the electrode-electrolyte interface is therefore critical for developing high–energy density batteries (*10–13*).

Conventional characterization techniques with high spatial resolution, such as high-resolution transmission electron microscopy (HRTEM), are incompatible with volatile liquid electrolytes and sensitive solid electrodes, like Li metal anode. Moreover, both electrodes and electrolytes are highly reactive and easily subject to contamination or damage during sample preparation and transfer. Cryogenic temperatures can help stabilize sensitive battery materials and interfaces during sample preparation and enable high-resolution characterization in TEM (*14–18*). Nonetheless, the nanoscale structure of SEI in the layer that is closely interfaced with the electrode revealed with cryo–electron microscopy (cryo-EM) in many state-of-the-art electrolytes is often amorphous (*6*, *7*). Thus, it is hard to correlate the difference in battery performance with the SEI nanostructure and chemistry.

The experiments referenced in the previous paragraph were performed in the absence of liquid electrolyte; however, ideally one would want to preserve the solid-liquid interface in the "wet" state with liquid electrolyte. A cryo–scanning transmission electron microscopy (cryo-STEM) method, combined with cryo–focused ion beam (cryo-FIB), was reported to access the buried interface in batteries with solid and liquid phases together (*19*). However, high-resolution imaging of SEI in the electrolyte is difficult because of the technical challenge in preparing thin enough lamellae suitable for HRTEM. Additionally, the effect of ion milling on SEI nanostructure and chemistry is also a concern.

We adapt the original thin film vitrification method (*20*) to preserve the electrode-electrolyte interface of batteries in its native organic liquid electrolyte environment. Such samples can be characterized with cryo-(S)TEM to investigate the intact structure and chemistry of the interphase in Li metal batteries. The key is to directly obtain thin film specimens of organic liquid electrolyte interfaced with the solid battery material while avoiding any mechanical or chemical artifacts from extra sample preparation steps.

Figure 1, A and B, shows a schematic of the thin film vitrification method developed for batteries and the cross-sectional view of the vitrified specimen. Such a process yields uniform thin films inside the holes throughout the grid (fig. S1) and generates the electron-transparency of the specimen (fig. S2, A to C). There are two crucial factors to ensure that the vitrification of organic electrolytes is a practical method. (i) Organic solvent molecules often require substantially slower cooling rates for vitrification than aqueous solutions of biological samples (*21*), so the original method

[1]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Chemical and Biomolecular Engineering, University of California Los Angeles, Los Angeles, CA 90095, USA. [3]Biophysics Program, School of Medicine, Stanford University, Stanford, CA 94305, USA. [4]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [5]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [6]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [7]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [8]Division of Cryo-EM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [9]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.

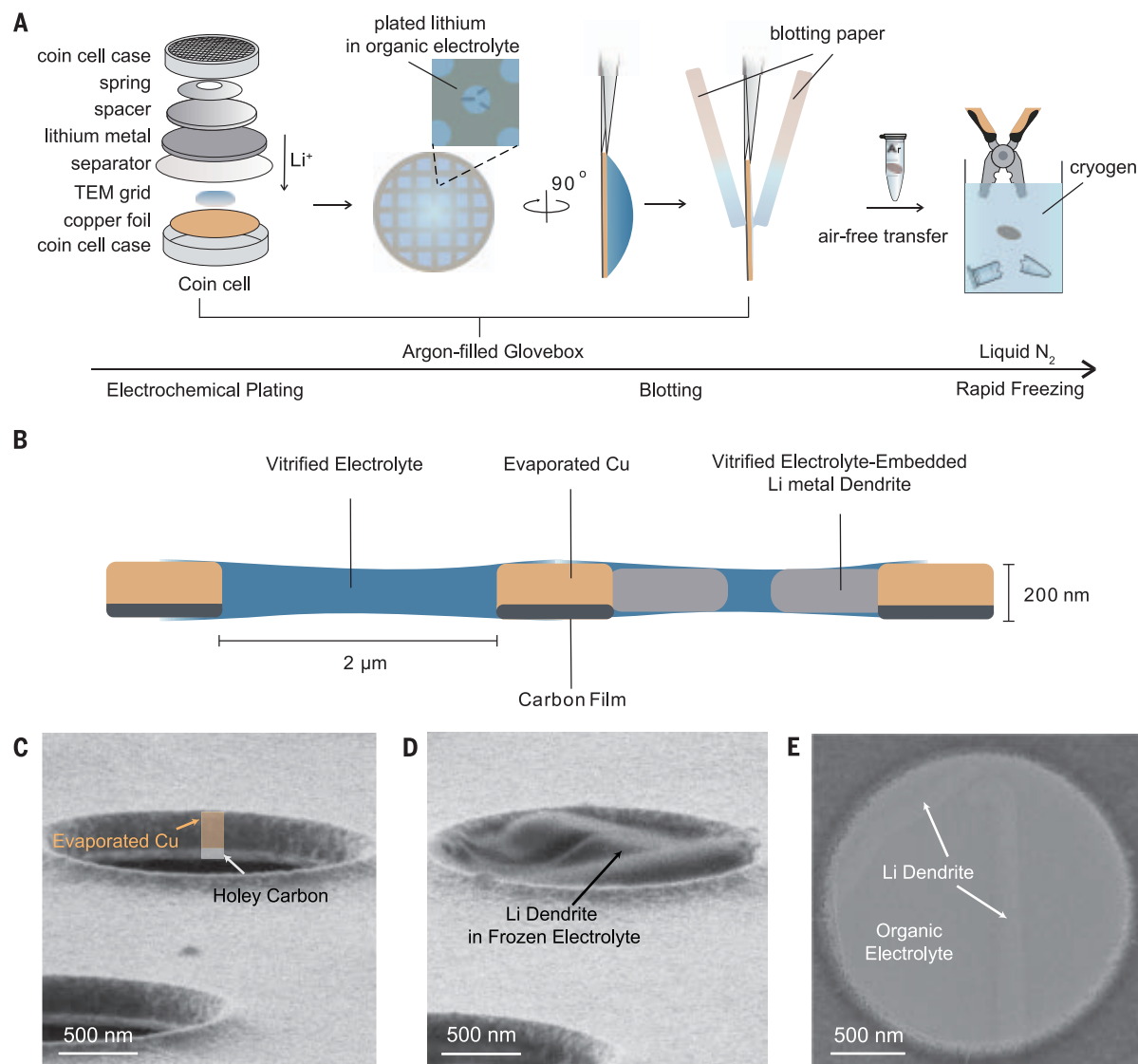*Corresponding author. Email: wahc@stanford.edu (W.C.); yicui@stanford.edu (Y.C.)

Fig. 1. Sample preparation of dendrite in vitrified organic electrolyte. (**A**) Schematic process of sample preparation for vitrified specimens. Cu-evaporated commercial holey carbon TEM grids were used as the working electrode for Li metal plating in the coin cell setup. Upon coin cell disassembly after Li metal deposition, excess electrolyte on the TEM grid is removed with double-sided blotting (movie S1) in an Ar-filled glove box and vitrified by liquid nitrogen without air exposure. (**B**) Schematic cross section of vitrified specimens. (**C**) Cryo-SEM image of Cu-evaporated TEM grid. (**D**) Cryo-SEM image of frozen Li metal dendrite along with electrolyte. (**E**) Cryo-TEM image of Li dendrite in frozen electrolyte. The light-contrast rodlike region represents the Li metal dendrite.

of freezing biological specimens directly in liquid nitrogen was used. (ii) Although lower in surface tension, organic electrolytes can still form a self-supporting thin film of sub-micron thickness by themselves and can remain for seconds before breaking as aqueous solutions. The amorphous diffraction pattern of pure frozen electrolyte without any salt or solvent crystallization confirms the successful vitrification process (fig. S2, D and E). Cryo–scanning electron microscopy (cryo-SEM) revealed rod-shape morphologies covered by a thin film in the TEM grid hole, corresponding to Li metal dendrites covered with a thin layer of vitrified electrolyte (Fig. 1, D and E, and fig. S2, F and G).

Li metal plated in commercial carbonate electrolyte—1 M LiPF$_6$ in ethylene carbonate/diethyl carbonate (EC/DEC)—was used as an example to reveal the SEI in the electrolyte. In Fig. 2A, Li metal dendrites show a lighter contrast compared with that of the organic electrolyte as a result of a lower average atomic number. A thick layer of ~20 nm with slightly darker contrast in the vitrified electrolyte was identified as the SEI layer (Fig. 2C). However, the SEI characterized in the absence of liquid electrolyte is ~10 nm thick (Fig. 2D). There is a visible thickness difference between these two samples (Fig. 2E) that can be observed across multiple experiments (Fig. 2F). A video recorded after electrolyte removal but without drying shows that the SEI shrinks under electron beam exposure as a result of the evaporation of volatile solvent species (movie S2). Thus, this change in thickness should be ascribed to the loss of electrolyte species during washing and drying in preparing dry-state samples, which indicates a swollen SEI in the electrolyte environment. In the following discussion, the SEI in the electrolyte is denoted as w-SEI to indicate that the SEI is in a vitrified (also referred to as a wet or w-) state, and the SEI in the absence of electrolyte is denoted as d-SEI to indicate that the SEI is in a dry state.

We used cryo-STEM and electron energy-loss spectroscopy (EELS) to explore the chemistry of Li metal and its SEI in vitrified electrolyte.
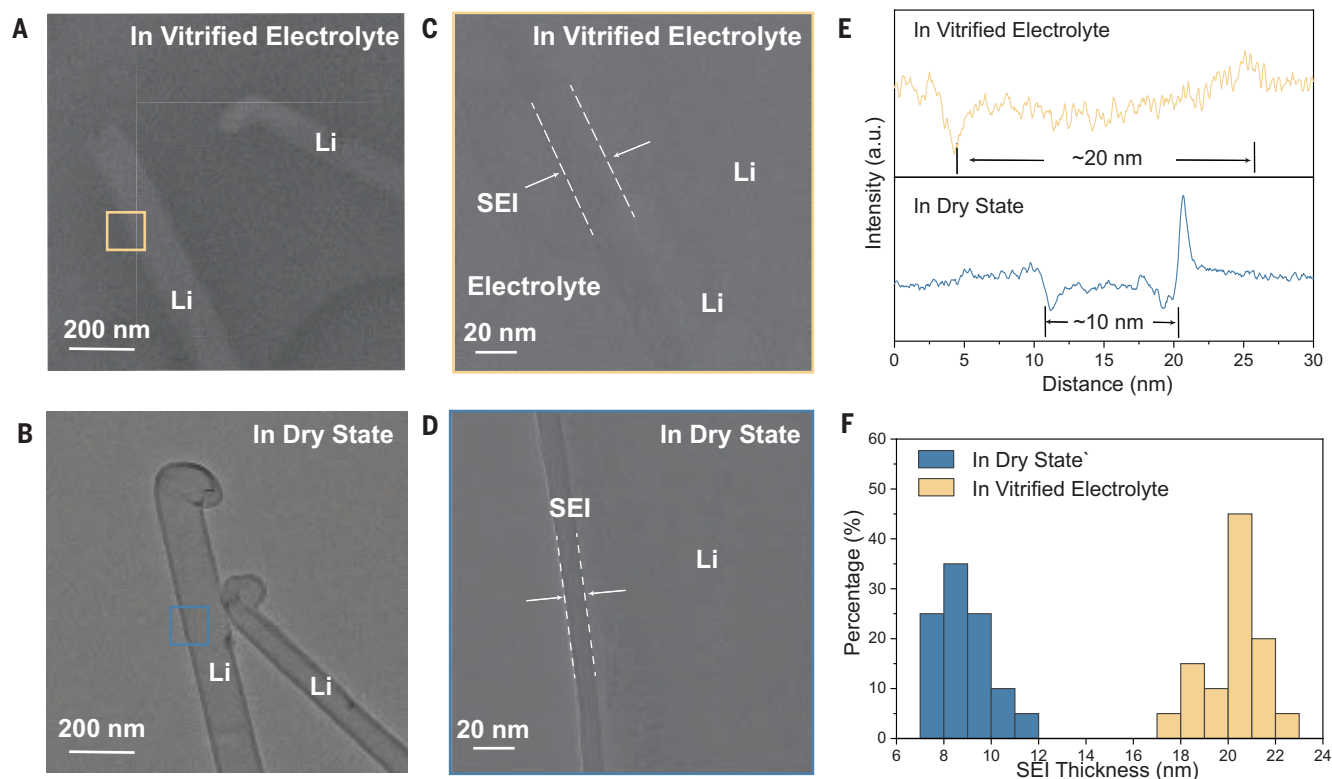
Fig. 2. SEI on Li dendrite in dry state and vitrified organic electrolyte imaged with cryo-TEM. (**A** and **B**) Li metal dendrites in vitrified electrolyte (A) and in dry state (B). (**C** and **D**) HRTEM of SEI on Li metal dendrite in vitrified electrolyte (C) and in dry state (D). (**E**) Representative line profiles of intensity across the interfaces on Li metal dendrite deposited in 1 M LiPF$_6$ in EC/DEC electrolyte. a.u., arbitrary units. (**F**) The histogram of SEI thickness in vitrified electrolyte and in dry state across multiple Li metal dendrites with 20 measurements for both in vitrified electrolyte and in dry state.
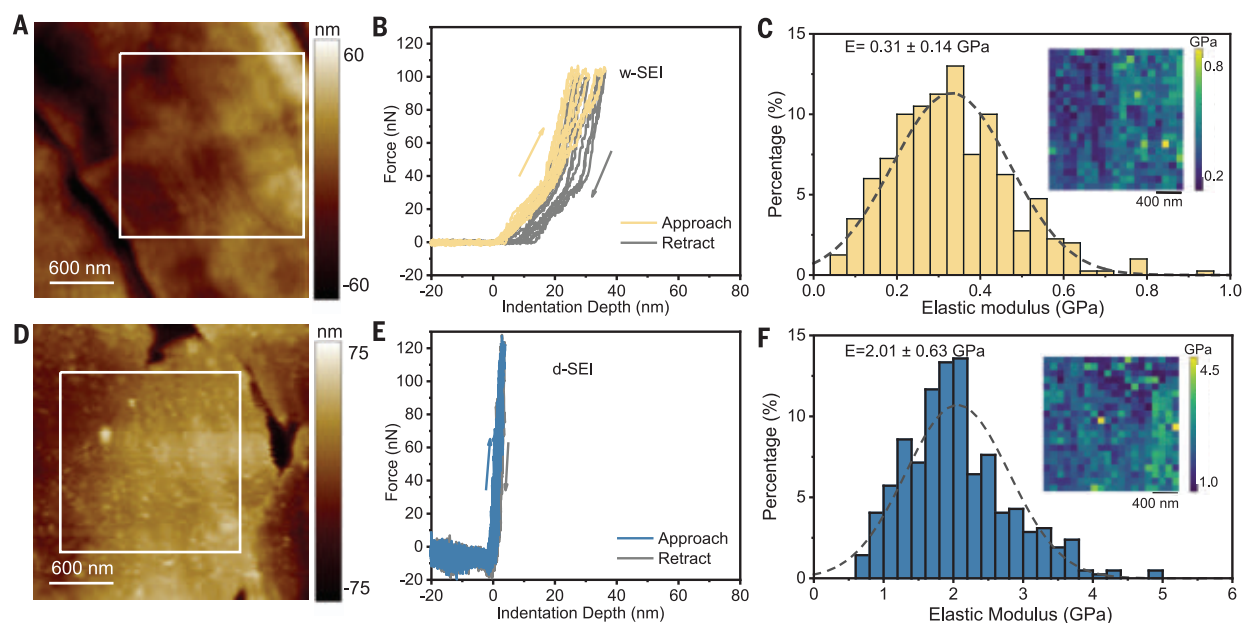
Fig. 3. AFM nanoindentation analysis of SEI in liquid electrolyte. (**A** and **D**) AFM height images of deposited Li metal in liquid electrolyte (A) and in dry state (D). The white box indicates the region for nanoindentation mapping. (**B** and **E**) Representative force-displacement curves for nanoindentation experiments for both w-SEI (B) and d-SEI (E). (**C** and **F**) Histograms of the elastic modulus of w-SEI (C) and of d-SEI (F). The insets are corresponding two-dimensional maps of elastic moduli in the regions of interest. Indentation displacement curves were offset for clarity. The scales of the x axes are different in (C) and (F).

Spectroscopic mapping of EELS shows that the SEI layer is tens of nanometers thick (fig. S4). We observe distinct carbon-bonding environments in the d-SEI, the w-SEI, and the electrolyte in the carbon K-edge fine structures (fig. S5). The peaks at 288 and 290 eV correspond to C–H and carbonate C=O bonds present in all three regions, consistent with evidence that SEI is mainly composed of alkyl carbonates in carbonate-based electrolyte (22). The increase in the relative intensity of C–H bonds from the w-SEI compared with the d-SEI correlates well with the observed swelling behavior. More carbonate-based organic molecules are present in the SEI layer in the wet state. Thus, the average carbon and oxygen bonding environment in the w-SEI more closely resembles that in the electrolyte as compared with the d-SEI.

Local mechanical properties of SEI were measured by nanoindentation with atomic force microscopy (AFM). The measurements were carried out in an inert environment to prevent undesired side reactions, and for w-SEI particularly, a closed liquid cell for AFM was used to further keep the electrode in the liquid electrolyte environment (fig. S6A). Typical force-displacement curves for nanoindentation experiments on both d-SEI and w-SEI are shown in Fig. 3. w-SEI showed an elastic-plastic deformation, where the force-displacement curves during loading and unloading are not fully reversible. However, under similar force loading, d-SEI only exhibited elastic deformation with small displacement (<5 nm) (fig. S6, C and D). The elastic modulus of w-SEI is 0.31 ± 0.14 GPa, whereas that of d-SEI is 2.01 ± 0.63 GPa. This difference can be explained by the swelling behavior of SEI in liquid electrolyte because swelling can cause polymers to soften (23). Additionally, swelling has been shown to increase the spatial heterogeneity of polymer materials (24), which corresponds to a more diverse distribution of elastic modulus from w-SEI.

Our result suggests that SEI is in a swollen state in liquid electrolyte. This is important in part because it suggests that SEI may not be a dense layer and that there is a nontrivial amount of electrolyte in this region. This is different from previous understanding, where SEI was thought to be a mixing layer of solid inorganic species (such as Li_2O, Li_2CO_3, etc.) and polymers and thus was electrolyte blocking and surface passivating to make the electrode-electrolyte interface metastable. Our results indicate that the electrolyte is in close contact with the electrode at the solid-liquid interface in batteries. Several fundamental yet critical aspects about this interface, for example the Li-ion desolvation process and Li-ion transport mechanism through SEI, need to be reconsidered to better understand the key processes during battery cycling. Notably, after calendar aging or cycling, SEI can become much thicker, where the swelling might become more substantial and eventually lead to the drying out of the electrolyte (25, 26).

Furthermore, the swelling of SEI sheds light on the mechanism of SEI growth after the formation of the initial SEI layer. Previously, the decrease in the rate of SEI formation was projected to be caused by the need for the reactants to diffuse through the already-existing layer (27). However, whether it is solvent diffusion through SEI inward to electrode surface or electron conduction through SEI outward toward electrolyte is still subject to debate. On the basis of our observation, it is highly likely that solvent diffusion plays a more significant role in the continuous growth of the SEI, particularly because the presence of solvents within the SEI reduces the distance required for electron tunneling during the decomposition of electrolytes. The reaction hotspot is now at or near the electrode-SEI interface.
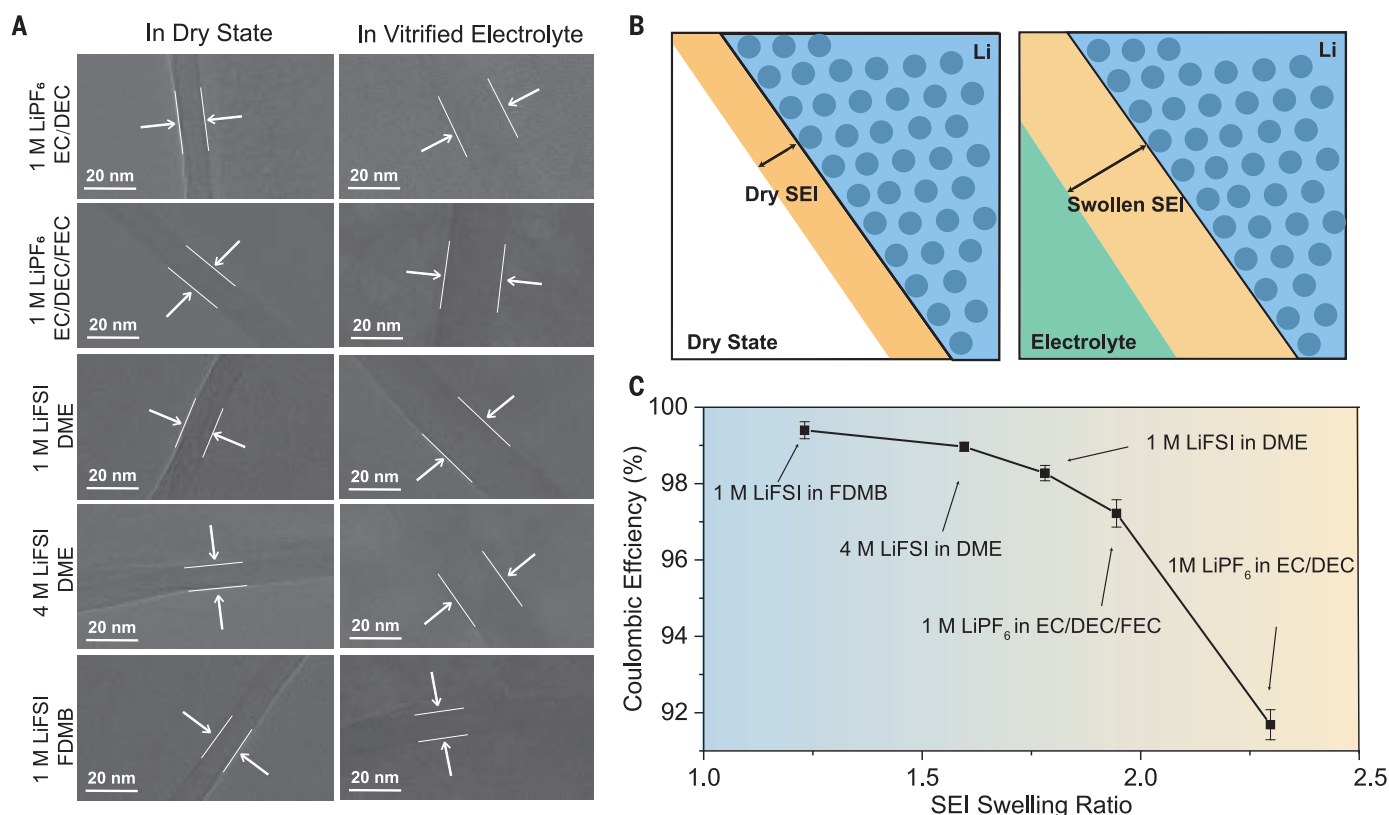
Fig. 4. The correlation of Li metal anode performance and swelling ratio of SEI in different electrolytes. (**A**) Representative comparison of SEI thickness in d-SEI and w-SEI with high-resolution cryo-TEM for various electrolyte systems. (**B**) d-SEI is thinner compared with w-SEI in vitrified electrolyte for all five systems. (**C**) SEI swelling ratio (w-SEI thickness versus d-SEI thickness) as a function of CE.

This observation provides a practical method to quantitatively measure the electrolyte uptake of SEI in the liquid environment. Alternatively, this can also be viewed as SEI porosity at (sub)-nanoscale, as projected in earlier SEI models (27–30). By measuring the swelling ratio, defined as the thickness ratio of w-SEI and d-SEI, we can estimate the amount of electrolyte in the SEI region. For example, in 1 M LiPF$_6$ in EC/DEC, the w-SEI has an average thickness of 20 nm, whereas the d-SEI is ~10 nm. This indicates that ~50% of the SEI volume is composed of the electrolyte. Further questions, like nanoscale pore or electrolyte distribution in the SEI, need to be addressed for better understanding of the transport mechanism of Li across the interface.

SEI is the key determinant for Li metal anode performance, and its properties vary with electrolyte systems, where solvent chemistry and salt composition largely determine SEI composition and structure (31). Even changing salt concentration would alter the solvation chemistry and the derived SEI (9). Generally, a mechanically robust, spatially uniform, and chemically passivating SEI is desirable (32). The swelling of SEI directly contradicts these design principles. We hypothesize that a better SEI should swell less with the electrolyte.

As a test of the above hypothesis, we performed electrochemical experiments and extended this cryo-TEM analysis to four other electrolytes with a variety of salts, solvents, and additives from the literature: 1 M LiPF$_6$ in EC/DEC with 10% fluoroethylene carbonate (EC/DEC, 10% FEC), 1 M lithium bis(fluorosulfonyl)imide (LiFSI) in 1,2-dimethoxyethane (DME), 4 M LiFSI in DME, and 1 M LiFSI in fluorinated 1,4-dimethoxylbutane (FDMB) (7). These electrolytes have different Coulombic efficiencies (CEs) measured with the Aurbach method, ranging from 97.2 to 99.4%. Despite their differences in surface tension and viscosity, we could obtain high-quality thin film vitrified specimens for all these electrolytes (fig. S7). No salt precipitation was observed even for the highly concentrated electrolyte (fig. S8).

We find the swelling of SEI in the electrolyte to be a universal phenomenon across all these electrolyte systems, regardless of solvent chemistry (Fig. 4 and table S1). This swelling behavior is dependent on electrolyte chemistry and highly correlated to battery performance, where higher degrees of SEI swelling tend to exhibit poor electrochemical cycling. The average d-SEI thicknesses in these five electrolytes are 8.8, 10.2, 9.9, 9.8, and 8.8 nm, whereas the average thicknesses of corresponding w-SEI are 20.1, 19.8, 17.6, 15.7, and 10.9 nm, respectively (fig. S9 and table S1). We correlate the cycling performance of Li metal anode represented by CE with SEI swelling behaviors.

Among the five electrolytes examined here, 1 M LiPF$_6$ in EC/DEC—the electrolyte with the lowest CE or worst cycling performance—has the largest swelling ratio, ~2.3. For one of the best performing electrolytes, 1 M LiFSI in FDMB, this ratio is the smallest, ~1.2. Overall, an increased swelling ratio correlates to a decreased CE (i.e., cycle life) (table S1).

We also find that the increase of elements associated with salt decomposition in d-SEI is accompanied by a decrease in swelling ratio. These elements most likely form inorganic domains in the SEI, and inorganic species in SEI have less affinity toward organic solvents compared with organic species. This results in a less–electrolyte-philic SEI with a smaller swelling ratio. In the 0.1 M LiPF$_6$ in EC/DEC electrolyte, where the salt concentration is much lower than that in commercial carbonate electrolytes, we observed a swelling ratio of ~2.6, which is higher than that of 1 M LiPF$_6$ in EC/DEC (fig. S10 and table S2). The elastic moduli of both d-SEI and w-SEI are lower than those of 1 M LiPF$_6$ in EC/DEC, respectively (fig. S11), corresponding to a more polymeric composition, as expected. Such analysis is also valid in ether-based electrolytes (table S3). The highly concentrated electrolyte, 4 M LiFSI in DME, exhibited a smaller swelling ratio as well as a higher elastic modulus for both d-SEI and w-SEI compared with 1 M LiFSI in DME (fig. S12), in accord with the account that SEI from 4 M LiFSI in DME is highly anion derived (9). Such observation of smaller swelling ratios in more–inorganic-rich SEI provides a possible explanation for the pursuit of more–anion-derived SEI in the community. The better anion-derived SEI has a higher ratio of elements from the decomposition products of the salt instead of solvents, which means that the SEI swells less with the electrolyte to remain mechanically robust and chemically passivating. This relationship between SEI swelling and battery performance can be a potential design principle in conjunction with other electrochemical and mechanical properties, such as ionic conductivity, elasticity, and uniformity. Because current density plays a critical role in controlling the structure of SEI (33), this analysis could be further extended to understand current density effect on SEI composition and nanostructure (fig. S13 and table S4). Beyond that, given the similarities in chemical composition of SEI, we also expect this swelling behavior in SEI on other negative electrodes. Furthermore, such insights also highlight the importance of preserving both the liquid and solid phases for studying complex interfacial phenomena with high resolution using cryo-EM methods.

REFERENCES AND NOTES

1. A. J. Bard et al., J. Phys. Chem. 97, 7147–7173 (1993).
2. V. R. Stamenkovic, D. Strmcnik, P. P. Lopes, N. M. Markovic, Nat. Mater. 16, 57–69 (2017).
3. K. Xu, Chem. Rev. 114, 11503–11618 (2014).
4. E. Peled, S. Menkin, J. Electrochem. Soc. 164, A1703–A1719 (2017).
5. M. B. Pinson, M. Z. Bazant, J. Electrochem. Soc. 160, A243–A250 (2013).
6. X. Cao et al., Nat. Energy 4, 796–805 (2019).
7. Z. Yu et al., Nat. Energy 5, 526–533 (2020).
8. R. Weber et al., Nat. Energy 4, 683–689 (2019).
9. J. Qian et al., Nat. Commun. 6, 6362 (2015).
10. M. Gauthier et al., J. Phys. Chem. Lett. 6, 4653–4672 (2015).
11. X. Yu, A. Manthiram, Energy Environ. Sci. 11, 527–543 (2018).
12. X.-B. Cheng, R. Zhang, C.-Z. Zhao, Q. Zhang, Chem. Rev. 117, 10403–10473 (2017).
13. J. W. Choi, D. Aurbach, Nat. Rev. Mater. 1, 16013 (2016).
14. Y. Li et al., Science 358, 506–510 (2017).
15. X. Wang et al., Nano Lett. 17, 7606–7612 (2017).
16. W. Huang et al., Nano Lett. 19, 5140–5148 (2019).
17. W. Huang et al., Matter 1, 1232–1245 (2019).
18. Z. Zhang et al., Matter 4, 302–312 (2021).
19. M. J. Zachman, Z. Tu, S. Choudhury, L. A. Archer, L. F. Kourkoutis, Nature 560, 345–349 (2018).
20. K. A. Taylor, R. M. Glaeser, Science 186, 1036–1037 (1974).
21. D. Danino, R. Gupta, J. Satyavolu, Y. Talmon, J. Colloid Interface Sci. 249, 180–186 (2002).
22. L. Wang et al., Nat. Chem. 11, 789–796 (2019).
23. M. Rubinstein, R. H. Colby, Polymer Physics (Oxford Univ. Press, 2003).
24. R. Subramani et al., Front. Mater. 7, 212 (2020).
25. S. Kranz, T. Kranz, A. G. Jaegermann, B. Roling, J. Power Sources 418, 138–146 (2019).
26. M. Nojabaee, K. Küster, U. Starke, J. Popovic, J. Maier, Small 16, e2000756 (2020).
27. F. Single, B. Horstmann, A. Latz, Phys. Chem. Chem. Phys. 18, 17810–17814 (2016).
28. M. Garreau, J. Power Sources 20, 9–17 (1987).
29. P. Guan, L. Liu, X. Lin, J. Electrochem. Soc. 162, A1798–A1808 (2015).
30. J. Popovic, Energy Technol. 9, 2001056 (2021).
31. X. Ren et al., Proc. Natl. Acad. Sci. U.S.A. 117, 28603–28613 (2020).
32. X.-B. Cheng et al., Adv. Sci. 3, 1500213 (2016).
33. Y. Xu et al., ACS Nano 14, 8766–8775 (2020).

SOLAR CELLS

Metastable Dion-Jacobson 2D structure enables efficient and stable perovskite solar cells

Fei Zhang[1]*†, So Yeon Park[1]†, Canglang Yao[2,3]†, Haipeng Lu[1], Sean P. Dunfield[4,5,6], Chuanxiao Xiao[4], Soňa Uličná[7], Xiaoming Zhao[8], Linze Du Hill[9], Xihan Chen[1], Xiaoming Wang[2,3], Laura E. Mundt[7], Kevin H. Stone[7], Laura T. Schelhas[1,7], Glenn Teeter[4], Sean Parkin[10], Erin L. Ratcliff[9,11,12], Yueh-Lin Loo[8], Joseph J. Berry[4,5,13], Matthew C. Beard[1], Yanfa Yan[2,3]*, Bryon W. Larson[1]*, Kai Zhu[1]*

The performance of three-dimensional (3D) organic-inorganic halide perovskite solar cells (PSCs) can be enhanced through surface treatment with 2D layered perovskites that have efficient charge transport. We maximized hole transport across the layers of a metastable Dion-Jacobson (DJ) 2D perovskite that tuned the orientational arrangements of asymmetric bulky organic molecules. The reduced energy barrier for hole transport increased out-of-plane transport rates by a factor of 4 to 5, and the power conversion efficiency (PCE) for the 2D PSC was 4.9%. With the metastable DJ 2D surface layer, the PCE of three common 3D PSCs was enhanced by approximately 12 to 16% and could reach approximately 24.7%. For a triple-cation–mixed-halide PSC, 90% of the initial PCE was retained after 1000 hours of 1-sun operation at ~40°C in nitrogen.

P erovskite solar cells (PSCs) are a promising photovoltaic (PV) technology, and certified power conversion efficiencies (PCEs) as high as 25.5% have been reported (1). Despite this high performance, device stability hinders their commercialization. Efforts to improve device stability include defect passivation, contact layer modification, and encapsulation (2–5). The use of two-dimensional (2D) perovskite as the interfacial modification layer has great potential for addressing surface defects, in particular to improve the stability and efficiency of PSCs (6–8). The Ruddlesden-Popper (RP) 2D layered perovskites that are based on bulky cations, such as phenethylammonium (PEA⁺) or butylammonium (BA⁺), have been widely applied to the surface of 3D perovskite thin films to decrease defect densities and enhance device stability (8–11). Such bulky organic cations often self-assemble into a barrier layer

that protects against surface water adsorption or ingress. However, bulky-cation–based 2D structures often exhibit anisotropic and poor charge transport across the organic layer and are susceptible to charge-extraction barrier formation that inhibits efficient device operation (12–14).

We show a rational design strategy to maximize the out-of-plane hole transport based on a metastable Dion-Jacobson (DJ) 2D perovskite surface layer with a reduced transport energy barrier by using asymmetric bulky organic molecules, leading to highly efficient and stable perovskite solar cells. Our general design strategy to maximize the out-of-plane charge transport in 2D perovskites is illustrated in Fig. 1. Because the free electrons and holes are localized in the conduction band minimum (CBM) and valence band maximum (VBM) of the [PbI₆] planes, respectively, and because of

the long distance between two adjacent [PbI₆] planes, the out-of-plane charge transport must traverse the bulky cationic organic layers. Thus, it is mainly limited by two factors: (i) the low carrier mobility within the organic layer and (ii) the energy barrier between the [PbI₆] planes and the bulky organic cations. To mitigate the first limit, DJ 2D structures based on a short and single layer of divalent organoammonium cations (15–18) are generally more preferable than the RP 2D structures based on double layers of monovalent organoammonium cations (19). To mitigate the second limiting factor, the band offsets between the [PbI₆] planes and the bulky cationic organic layers need to be optimized.

The coupling (interaction) between [PbI₆] planes and the organic cations is through hydrogen bonding, and the change in the bonding strength can affect the band offsets (20). For a weaker hydrogen bonding configuration, the bonding states of the bulky organic layers are normally at a higher energy position, which brings them nearer the VBM of the [PbI₆] planes (Fig. 1A). This effect leads to a smaller band offset or barrier for hole transport between the [PbI₆] inorganic planes and organic cations. Because of the spin-orbital coupling of Pb 6p orbitals, the antibonding states of the organic layers are much higher than the CBM of the [PbI₆] planes. Thus, a DJ structure with weaker hydrogen bonding should improve hole transport. Yet, a weaker hydrogen bonding (or H-bonding) configuration generally means a less stable structure. Thus, a metastable DJ 2D structure with short cationic organic layers could in principle facilitate out-of-plane hole transport.

A rational strategy to induce the desired metastable H-bonding motifs in DJ 2D structures is to use asymmetric diammonium cations in lieu of symmetric straight chain divalent

[1]Chemistry and Nanoscience Center, National Renewable Energy Laboratory, Golden, CO 80401, USA. [2]Department of Physics and Astronomy, University of Toledo, Toledo, OH 43606, USA. [3]Wright Center for Photovoltaics Innovation and Commercialization, University of Toledo, Toledo, OH 43606, USA. [4]Materials Science Center, National Renewable Energy Laboratory, Golden, CO 80401, USA. [5]Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309, USA. [6]Materials Science and Engineering Program, University of Colorado, Boulder, CO 80309, USA. [7]SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [8]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA. [9]Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ 85721, USA. [10]Department of Chemistry, University of Kentucky, Lexington, KY 40506, USA. [11]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA. [12]Department of Materials Science and Engineering, University of Arizona, Tucson, AZ 85721, USA. [13]Department of Physics, University of Colorado, Boulder, CO 80309, USA.
*Corresponding author. Email: fei.zhang@nrel.gov (F.Z.); yanfa.yan@utoledo.edu (Y.Y.); bryon.larson@nrel.gov (B.W.L.); kai.zhu@nrel.gov (K.Z.)
†These authors contributed equally to this work.

Table 1. PV parameters of PSCs based on control and DMePDAI₂-modified perovskite thin films by using different perovskite compositions. V_{oc}, open-circuit voltage; FF, fill factor.

Device	Scan	J_{sc} (mA/cm²)	V_{oc} (V)	FF	PCE (%)	SPO (%)
FA₀.₈₅MA₀.₁Cs₀.₀₅PbI₂.₉Br₀.₁	Forward	24.35	1.111	0.773	20.9	20.4
	Reverse	24.32	1.099	0.764	20.4	
FA₀.₈₅MA₀.₁Cs₀.₀₅PbI₂.₉Br₀.₁ /DMePDAI₂	Forward	24.97	1.167	0.822	24.0	23.7
	Reverse	24.93	1.167	0.814	23.7	
FA₀.₉₇MA₀.₀₃PbI₂.₉₁Br₀.₀₉	Forward	25.21	1.103	0.791	22.0	21.7
	Reverse	25.15	1.108	0.781	21.8	
FA₀.₉₇MA₀.₀₃PbI₂.₉₁Br₀.₀₉ /DMePDAI₂	Forward	25.25	1.158	0.843	24.7	24.3
	Reverse	25.26	1.158	0.839	24.5	
MAPbI₃	Forward	23.09	1.090	0.742	18.7	18.2
	Reverse	23.09	1.080	0.729	18.2	
MAPbI₃/DMePDAI₂	Forward	23.19	1.131	0.797	20.9	20.8
	Reverse	23.19	1.132	0.794	20.8	

cations. For example, both *N,N*-dimethyl-1,3-propane diammonium (DMePDA^{2+}) and 1,4-butane diammonium (BDA^{2+}) form DJ 2D structures with short interlayer distances (*19*). Whereas BDA^{2+} is symmetric and features two terminal primary ammonium ions on the butyl (C4) chain, DMePDA^{2+} is asymmetric, with a primary ammonium on one end and a dimethyl-substituted tertiary ammonium on the other end of the propyl (C3) chain. The "head or tail" H-bonding options for the DMePDA^{2+} molecules are asymmetric, giving rise to different possible relative orientations of the adjacent molecules, and the different H-bonding interactions possible within the [PbI$_6$] planes could lead to both stable and metastable energy polymorphs of the 2D structure (Fig. 1B). The alternating relative head-to-tail alignment of adjacent DMePDA^{2+} cations (most stable orientation configuration) provides a larger compensation for overall structural relaxation than those of other orientation arrangements. By contrast, the

symmetric BDA^{2+} molecule has only one possible orientation configuration (Fig. 1B) and cannot form metastable polymorphs.

We further examined single-crystal 2D DJ structures from BDAI$_2$ and DMePDAI$_2$ and conducted first-principle calculations to verify our design strategy. We found that 1,3-propane diammonium diiodine (PDAI$_2$)—which is often assumed the shortest diamine (*15, 21*) to form DJ 2D perovskites—templated Pb-I to a non-perovskite structure (empirical formula: [PDAPbI$_4$]$_{15}$•[PDAI$_2$]) (fig. S1 and table S1). Thus, BDA^{2+} represents the shortest linear-alkyl-chain diamine that forms an iodide-based 2D DJ structure (BDAPbI$_4$) (fig. S2A and table S2). C3-based DMePDAI$_2$ with two methyl groups attached to one side of PDA can form 2D DJ structures with two polymorphs, which we refer to as DMePDAPbI$_4$-1 (fig. S2B and table S3) and DMePDAPbI$_4$-2, respectively (fig. S2C) (*22*).

We grew the DMePDAPbI$_4$-1 single crystal, which was based on the most stable DMePDA^{2+} orientation alignment, from a

concentrated hydroiodic acid solution using a slow-crystallization process, as adapted from our previous report (*23*) and consistent with a previous theoretical predication (*24*). By contrast, the DMePDAPbI$_4$-2 single crystal, which was based on a metastable orientational alignment, was formed from either a fast cooling (*22*) or antisolvent quenching during single-crystal growth (*25*), both of which represent a fast-crystallization process. In comparison with DMePDAPbI$_4$-1, DMePDAPbI$_4$-2 had an emission wavelength that was red-shifted ~25 nm, which is consistent with the corresponding absorption data (fig. S3). The average interlayer distances were comparable among these 2D structures (~10.10 to 10.39 Å), with that of BDAPbI$_4$ being the shortest. The corresponding hydrogen-bonding configurations for these three single-crystal structures (figs. S4, S5, and S6) were consistent with the analysis in Fig. 1B.

We confirmed our design strategy by means of density functional theory (DFT) calculation.
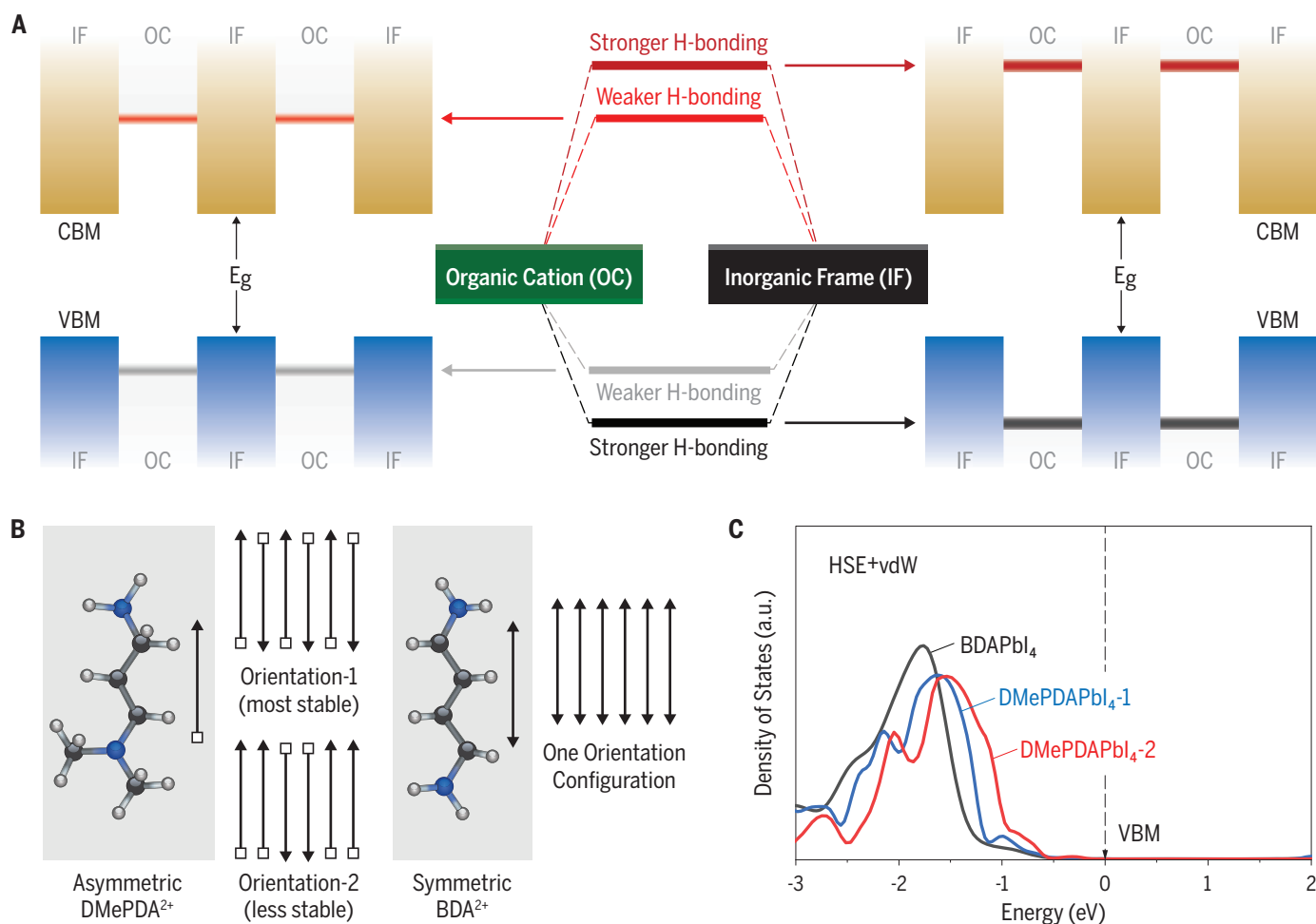
Fig. 1. Design concept. (**A**) Illustration of band offsets between [PbI$_6$] planes and bulky organic cations with a weaker and stronger degree of H-bonding. For clarity, the inorganic framework orbital diagram is omitted in the middle of the panel. (**B**) Two possible arrangements of asymmetric DMePDA^{2+} cations and the sole arrangement of symmetric BDA^{2+} cations. (**C**) HSE+vdW calculated total DOSs of the organic cations in BDAPbI$_4$, DMePDAPbI$_4$-1 [with orientation-1 in (B)], and DMePDAPbI$_4$-2 [with orientation-2 in (B)]. The VBMs were set to 0.0 eV.

We calculated the effect of organic molecules using the screened hybrid functional and van der Waals (vdW) interaction (HSE+vdW) (26, 27). The DMePDAPbI$_4$-2 structure was indeed less stable than the DMePDAPbI$_4$-1 structure. The energy level differences of the organic cations in BDAPbI$_4$, DMePDAPbI$_4$-1, and DMePDAPbI$_4$-2 could be seen in the total density of states (DOSs) of the organic cations (the sum of states of C, N, and H atoms) (Fig. 1C). The total DOS of BDA^{2+} cations in BDAPbI$_4$ was lower in energy (farther from VBM) than that of DMePDA^{2+} cations in DMePDAPbI$_4$-1, which in turn was lower in energy compared with the total DOS of DMePDA^{2+} cations in DMePDAPbI$_4$-2. Thus, we expected the out-of-plane hole transport to improve from BDAPbI$_4$ to DMePDAPbI$_4$-2.

Rapid perovskite film growth conditions from standard solution deposition also led to the formation of the metastable DMePDAPbI$_4$-2 structure. The XRD patterns of the DMePDAPbI$_4$ thin film prepared by means of spin coating are shown in Fig. 2A. The powder XRD pattern measured from DMePDAPbI$_4$-1 and DMePDAPbI$_4$-2 single-crystal samples, along with the calculated powder XRD patterns shown in Fig. 2A for comparison, revealed the differences of XRD patterns between these two single crystals. The XRD pattern of

the thin-film sample matched that of the DMePDAPbI$_4$-2 structure. A metastable polymorph does not mean it is unstable under synthetic or ambient conditions. The phase transformation between polymorphs requires 180° rotation of the alkyl chain, which is highly energetically unfavorable (Fig. 2B) (figs. S7 and S8). A wide range of thin-film growth conditions from solution all formed DMePDAPbI$_4$-2 thin films (fig. S9).

To test our hypothesis that the reduced energy barrier from the asymmetric bulky organic cation layer could facilitate charge transport between inorganic [PbI$_6$] sheets, we conducted time-resolved microwave conductivity (TRMC) measurements along the out-of-plane direction (28). In Fig. 2C, we compare the normalized TRMC results between several n = 1 2D perovskite thin films calibrated by their corresponding internal quantum yield of charges measured in devices. The out-of-plane transport for DMePDAPbI$_4$ (or more specifically, DMePDAPbI$_4$-2) is about a factor of 4 to 5 faster than that of BDAPbI$_4$, despite the slightly longer interlayer distance. Space-charge–limited current (SCLC) measurements further verified that the DMePDAPbI$_4$-2 structure had faster out-of-plane hole transport than that of the DMePDAPbI$_4$-1 structure (fig. S10). These results confirmed

the role of reducing the energy barrier for improving out-of-plane charge transport. The out-of-plane transport for the two 2D DJ structures (DMePDAPbI$_4$ and BDAPbI$_4$) was faster than those of the two 2D RP structures (BA$_2$PbI$_4$ and PEA$_2$PbI$_4$). These TRMC results were consistent with the current density-voltage (J–V) results of PSCs on the basis of the corresponding n = 1 2D structures (Fig. 2D and table S4). The DMePDAPbI$_4$-based PSC reached a PCE of 4.90% (forward scan) and 4.33% (reverse scan), which is among the highest obtained thus far for any n = 1 2D lead iodide–based PSCs (6); the corresponding external quantum efficiency (EQE) spectrum is shown in fig. S11.

The use of 2D systems to passivate defects and enhance performance has recently been used in many polycrystalline PV technologies (29). We validated the impact of this metastable design motif with the use of DMePDAPbI$_4$ as a surface layer to improve the quality of 3D perovskite absorbers. We spin-coated the corresponding bulky organic halide salt in isopropanol (IPA) solution on top of a 3D perovskite absorber layer (6). Specifically, the DMePDAI$_2$/IPA solution was coated atop (FAPbI$_3$)$_{0.85}$(MAPbI$_2$Br)$_{0.1}$(CsPbI$_3$)$_{0.05}$ (or FA$_{0.85}$MA$_{0.1}$Cs$_{0.05}$PbI$_{2.9}$Br$_{0.1}$) followed by annealing, where FA is formamidinium and MA is methylammonium. The thin-film XRD

Fig. 2. 2D thin-film structure, transport, and device characteristics. (**A**) XRD patterns of a solution-grown DMePDAPbI$_4$ thin film and the powder XRD patterns (measured and calculated) from DMePDAPbI$_4$-1 and DMePDAPbI$_4$-2 single-crystal structures. X-ray source, Cu Kα radiation. The peak labeled with an asterisk is from the fluorine tin oxide (FTO) substrates. (**B**) Energy profile along the transition path between DMePDAPbI$_4$-1 and DMePDAPbI$_4$-2. (**C**) TRMC comparison of out-of-plane charge transport across the layers of n = 1 2D perovskites. (**D**) J–V characteristics of PSCs based on n = 1 2D perovskite thin films using a device stack of glass/FTO/compact-TiO$_2$/2D-perovskite/2,2',7,7'-Tetrakis[N,N-di(4-methoxyphenyl)amino]-9,9'-spirobifluorene (spiro-OMeTAD)/Au.
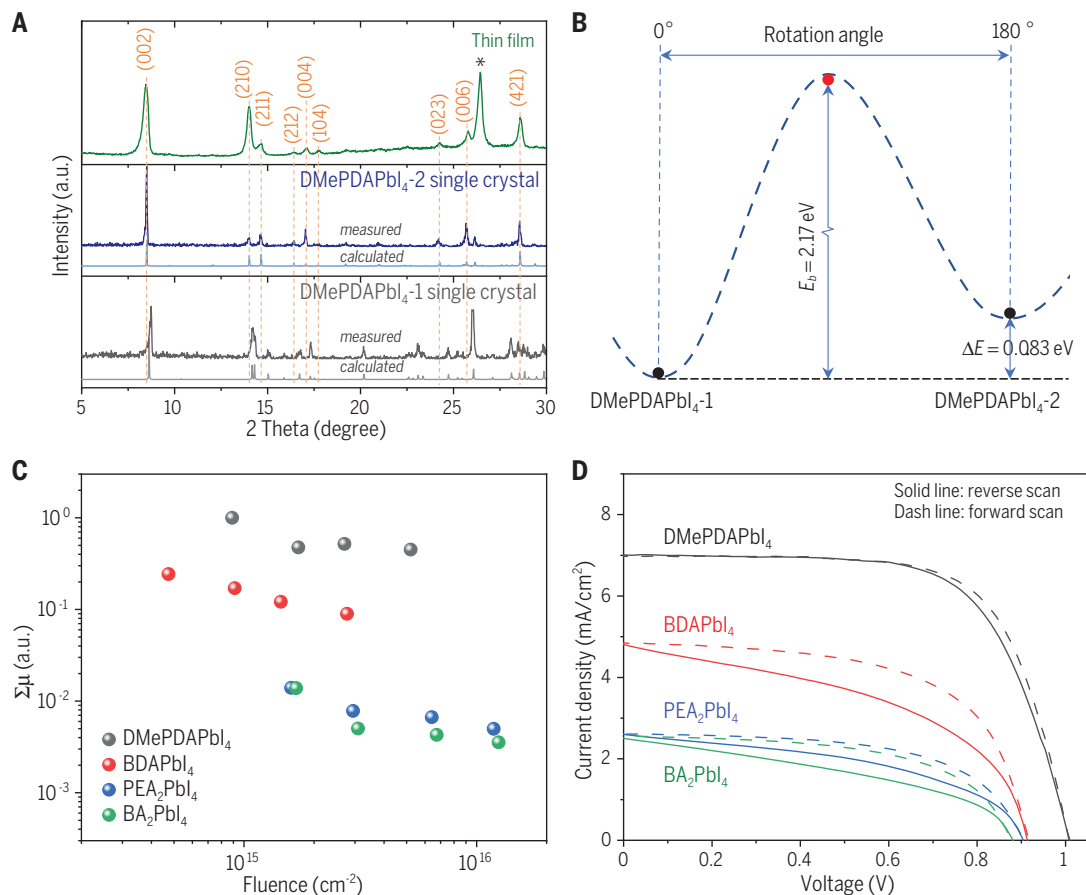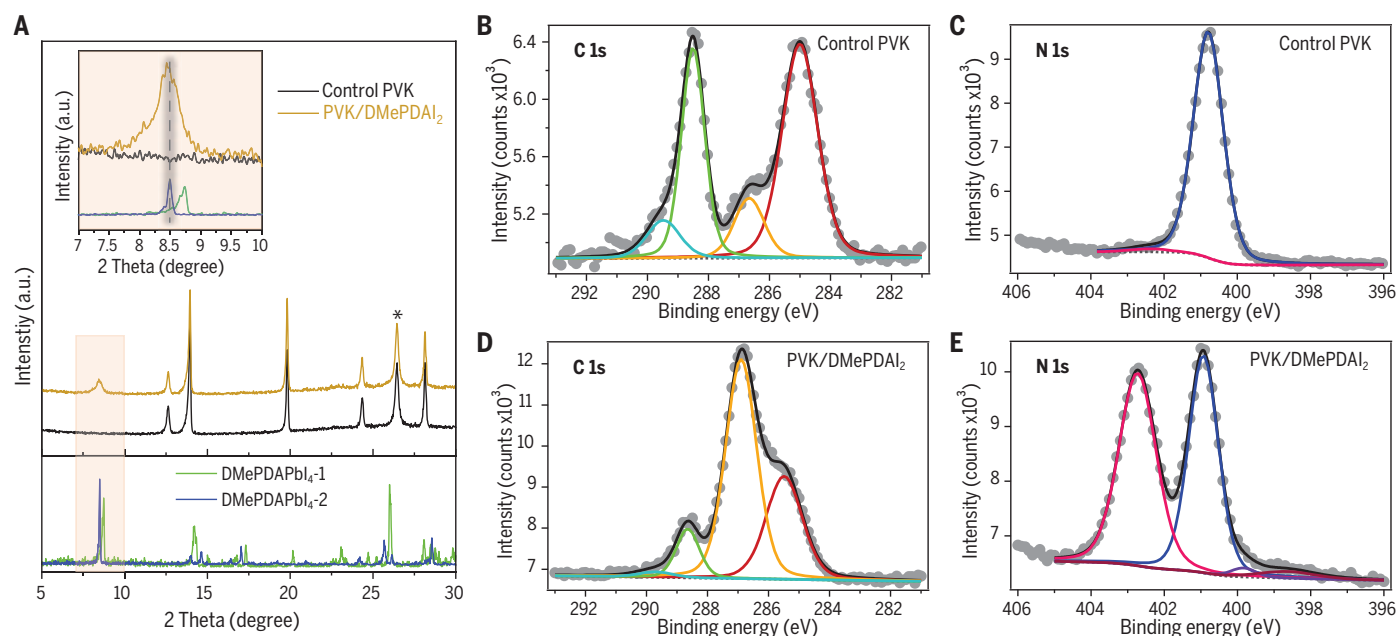
Fig. 3. Surface layer treatment. (**A**) Comparison of grazing incident XRD (GIXRD) patterns of thin films of DMePDAPbI$_4$ and perovskites without (control PVK) and with DMePDAI$_2$ surface treatment (PVK/DMePDAI$_2$). (Inset) Zoom-in view of the GIXRD pattern from 7° to 10°. The peak labeled with an asterisk is from the FTO substrate. X-ray source, Cu Kα radiation. (**B** to **E**) Comparison of the XPS spectra of N1s and C1s for [(B) and (C)] the control and [(D) and (E)] the DMePDAI$_2$-modified perovskite thin film.

results suggested that the DMePDAPbI$_4$-2 structure formed, as evidenced by the characteristic low-angle diffraction peak, at ~8.5° for DMePDAPbI$_4$-2, rather than ~8.7° for DMePDAPbI$_4$-1 (Fig. 3A).

We also checked the 2D structures on top of three other common perovskite compositions of Cs$_{0.05}$FA$_{0.95}$PbI$_3$, (FAPbI$_3$)$_{0.95}$(MAPbBr$_3$)$_{0.05}$, and FAPbI$_3$ (fig. S12). For these compositions, the characteristics peaks at (002), (004), and (006) matched well to DMePDAPbI$_4$-2, which were absent in the DMePDAPbI$_4$-1 spectrum. Last, the low-angle diffraction peak associated with the 2D structure from the thin-film XRD results were further confirmed with grazing-incidence wide-angle x-ray scattering (GIWAXS) measurements (fig. S13). In terms of 2D surface-layer topology and coverage, the scanning electron microscopy (SEM) measurements indicated that the treatment induced formation of a thin surface layer with small apparent grain sizes (figs. S14 and S15). The conductive-atomic force microscopy (C-AFM) measurements show that the current of the treated film is much more uniform and lower than that of the control film, which is consistent with the formation of a capping layer over the 3D perovskite layer (fig. S16).

To gain more insight into how the DMeP-DAI$_2$ modification affects the optoelectronic properties in perovskite films, we conducted steady-state photoluminescence (PL), time-resolved photoluminescence (TRPL), and TRMC studies on these samples. The DMePDAI$_2$ treatment led to enhanced PL intensity (fig.

S17), longer TRPL lifetime (fig. S18 and table S5), and improved TRMC mobility and lifetime (fig. S19) that were consistent with the improved surface properties (8, 30). In addition, the ultraviolet photoelectron spectroscopy (UPS) measurements showed that the 2D surface treatment improved the energetics for hole transport from the 3D perovskite to the 2D surface layer (fig. S20).

The impact of the DMePDAI$_2$ treatment on the perovskite surface chemistry was investigated with x-ray photoelectron spectroscopy (XPS). Normalized core levels from key elements identified on the sample surface are included in figs. S21 and S22. The spectral shapes of most core levels showed minimal change between the two samples, indicating similar bonding environments, but surface treatment caused change in the C 1s and N 1s core levels. We fit the core levels (Fig. 3, B to E) using constrained fitting procedures (summarized in tables S6 and S7). The control sample had a N 1s region whose relative peak areas were dominated by a C=NH$_2^+$ (FA) peak (~401 eV) with a small shoulder to higher binding energy (~403 eV) that corresponded to C–NH$_3$ (MA). The DMePDAI$_2$ treatment increased the area of the C–NH$_3$ peak and also led to two additional peaks at a lower binding energy consistent with that of C–NH$_2$ (~400 eV) and the tertiary amine in DMePDAI$_2$ (~398 eV). Concomitant with these changes, redistribution occurred in the features in the C 1s spectra comprising four main peaks that are consistent with primarily C–C or C–H (~285 eV),

N–CH$_3$ (~287 eV), HC(NH$_2$)$_2$ (~289 eV), and C–O or C=O bonds (~290 eV). The surface treatment decreased the concentration of HC(NH$_2$)$_2$ bonds from FA on the surface while simultaneously increasing the amount of N–CH$_3$ and C–C or C–H bonds. In addition, XPS revealed that surface treatment increased the amount of halide on the surface, from about 2.6 halide-to-lead ratio for the control to 3.1 for the DMePDAI$_2$-treated film. Collectively, these results suggest that both organic and halide components of the additive incorporated into the top surface of the treated films. Undercoordinated lead can cause donor defects on the surface, resulting in downward band bending and increased recombination centers (31), so the increase in the halide-to-lead ratio associated with the formation of 2D interfacial component upon surface treatment was consistent with a less defective surface.

We investigated the impact of DMePDAI$_2$ surface treatment on the PV performance by fabricating PSCs using the standard n-i-p device architecture, glass/FTO/electron transport layer (ETL)/perovskite/hole transport layer (HTL)/Au, where ETL is TiO$_2$ or SnO$_2$ and HTL is spiro-OMeTAD, with more details in the supplementary materials (23). Typical cross-section SEM images of devices are shown in fig. S23. In Fig. 4A, we compare the J–V curves of the PSCs on the basis of triple-cation–mixed-halide FA$_{0.85}$MA$_{0.1}$Cs$_{0.05}$PbI$_{2.9}$Br$_{0.1}$ without and with DMePDAI$_2$ treatment under simulated 100-mW/cm^2 air mass coefficient (AM) 1.5 G
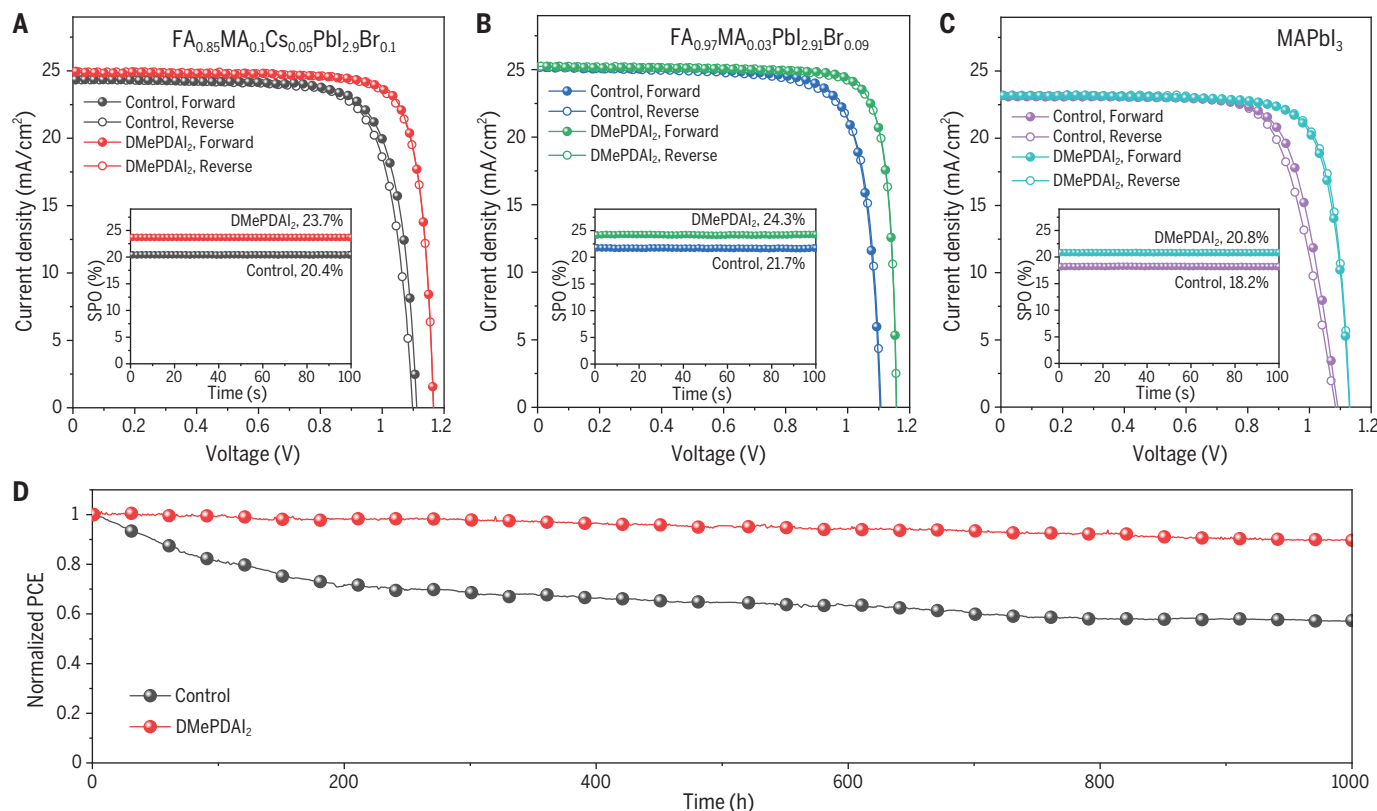
Fig. 4. Device characteristics. (**A** to **C**) *J–V* characteristics of PSCs based on different perovskite compositions. (A) $FA_{0.85}MA_{0.1}Cs_{0.05}PbI_{2.9}Br_{0.1}$. (B) $FA_{0.97}MA_{0.03}PbI_{2.91}Br_{0.09}$. (C) $MAPbI_3$. (Insets) SPOs of the corresponding devices. (**D**) Operation ISOS-L-1 stability (maximum power point tracking, in N_2, continuous one-sun illumination at ~40°C) of unencapsulated PSC based on $FA_{0.85}MA_{0.1}Cs_{0.05}PbI_{2.9}Br_{0.1}$. The initial PCE was 20.5% for the control and 23.1% for the $DMePDAI_2$-treated device.

illumination (Table 1). With the surface treatment, the device PCE increased from about 20.9 to 24.0% from forward scan and from 20.4 to 23.7% from reverse scan. The PCE improvement is also consistent with a better perovskite-HTL junction on the basis of the cross-sectional Kelvin probe force microscopy (KPFM) measurements (fig. S24) (*32*). The optimum concentration for $DMePDAI_2$-surface treatment was found at 0.5 mg/mL (fig. S25).

In addition to the $FA_{0.85}MA_{0.1}Cs_{0.05}PbI_{2.9}Br_{0.1}$ perovskite composition, we also examined the impact of $DMePDAI_2$ surface treatment on PSCs on the basis of double-cation–mixed-halide ($FA_{0.97}MA_{0.03}PbI_{2.91}Br_{0.09}$) and single-cation–single-halide ($MAPbI_3$) using ETL of SnO_2 and TiO_2, respectively, and found PCE improvements for both compositions (Fig. 4, B and C). Noteworthy for PSCs based on $FA_{0.97}MA_{0.03}PbI_{2.91}Br_{0.09}$, the PCE was improved from 22.0 to 24.7% from forward scan and from 21.8 to 24.5% from reverse scan, with short-circuit current density (J_{sc}) > 25 mA/cm², which is in agreement with the EQE spectrum (fig. S26). For all three perovskite compositions, the stabilized power outputs (SPOs) for PSCs based on the control and $DMePDAI_2$-modified perovskite thin films matched well with the *J–V* mea-

surements (Fig. 4, A to C, insets, and Table 1). The PCE improvement for all three perovskite compositions was reproducible on the basis of the statistical comparison (fig. S27). The devices with this treatment also exhibited higher PCE than that of devices based on other surface treatments with similar length of bulky organic salts for either RP or DJ 2D perovskites (fig. S28).

Last, we checked the operation stability of unencapsulated $FA_{0.85}MA_{0.1}Cs_{0.05}PbI_{2.9}Br_{0.1}$–based PSCs using maximum power point (MPP) tracking at ~40°C in N_2, following the ISOS-L-1 stability protocol (*33*). The $DMePDAI_2$-modified PSC (Fig. 4D) showed only 10% relative efficiency drop after 1000 hours of continuous operation, whereas the PCE of the control device decreased by ~43%. The stability improvement with $DMePDAI_2$ surface treatment was also observed when the devices were tested at high-moisture (>85% relative humidity) or high-temperature (85°C) conditions (figs. S29 and S30). These results suggest that the $DMePDAI_2$-modification to form a 2D DJ phase surface layer is a general way to improve PSC performance. Our use of the metastable 2D DJ structure through hydrogen bonding tuning based on asymmetric bulky organic molecules

represents a promising chemical design element for perovskite interfacial engineering to enhance PSC efficiency and stability.

REFERENCES AND NOTES

1. National Renewable Energy Laboratory, Best research-cell efficiency chart (August, 2020); www.nrel.gov/pv/cell-efficiency.html.
2. D. Bi *et al.*, *Nat. Energy* **1**, 16142 (2016).
3. X. Zheng *et al.*, *Nat. Energy* **2**, 17102 (2017).
4. F. Zhang *et al.*, *Adv. Mater.* **29**, 1606806 (2017).
5. X. Li *et al.*, *Nat. Chem.* **7**, 703–711 (2015).
6. F. Zhang *et al.*, *Energy Environ. Sci.* **13**, 1154–1186 (2020).
7. E. H. Jung *et al.*, *Nature* **567**, 511–515 (2019).
8. H. Min *et al.*, *Science* **366**, 749–753 (2019).
9. L. Liang, H. Luo, J. Hu, H. Li, P. Gao, *Adv. Energy Mater.* **10**, 2000197 (2020).
10. Y.-W. Jang *et al.*, *Nat. Energy* **6**, 63–71 (2021).
11. J. Xue *et al.*, *Science* **371**, 636–640 (2021).
12. D. Thrithamarassery Gangadharan, D. Ma, *Energy Environ. Sci.* **12**, 2860–2889 (2019).
13. E. Shi *et al.*, *Nature* **580**, 614–620 (2020).
14. Y. Gao *et al.*, *Nat. Chem.* **11**, 1151–1157 (2019).
15. S. Ahmad *et al.*, *Joule* **3**, 794–806 (2019).
16. L. Mao *et al.*, *J. Am. Chem. Soc.* **140**, 3775–3783 (2018).
17. Y. Lv *et al.*, *J. Mater. Chem. A Mater. Energy Sustain.* **8**, 10283–10290 (2020).
18. X. Jiang *et al.*, *Nano Energy* **75**, 104892 (2020).
19. X. Li, J. M. Hoffman, M. G. Kanatzidis, *Chem. Rev.* **121**, 2230–2291 (2021).
20. G. Miessler, P. Fischer, D. Tarr, *Inorganic Chemistry* (Prentice Hall, ed. 5, 2014).
21. C. Ma, D. Shen, T.-W. Ng, M.-F. Lo, C.-S. Lee, *Adv. Mater.* **30**, 1800710 (2018).

22. T. Yu, L. Zhang, J. Shen, Y. Fu, Y. Fu, *Dalton Trans.* **43**, 13115–13121 (2014).

23. F. Zhang et al., *Chem* **7**, 774–785 (2021).

24. S. Silver, S. Xun, H. Li, J.-L. Brédas, A. Kahn, *Adv. Energy Mater.* **10**, 1903900 (2020).

25. W. Zhao et al., *J. Mater. Chem. A Mater. Energy Sustain.* **8**, 9919–9926 (2020).

26. A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, G. E. Scuseria, *J. Chem. Phys.* **125**, 224106 (2006).

27. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).

28. F. Zhang et al., *J. Am. Chem. Soc.* **141**, 5972–5979 (2019).

29. D. L. McGott et al., *Joule* **5**, 1057–1073 (2021).

30. H. Zhu et al., *J. Am. Chem. Soc.* **143**, 3231–3237 (2021).

31. S. P. Dunfield et al., *Adv. Energy Mater.* **10**, 1904054 (2020).

32. Y. Hou et al., *Science* **367**, 1135–1140 (2020).

33. M. V. Khenkin et al., *Nat. Energy* **5**, 35–49 (2020).

HEAVY FERMIONS

Evidence for a delocalization quantum phase transition without symmetry breaking in CeCoIn₅

Nikola Maksimovic[1,2]*, Daniel H. Eilbott[1,2], Tessa Cookmeyer[1,2], Fanghui Wan[1,2], Jan Rusz[3], Vikram Nagarajan[1,2], Shannon C. Haley[1,2], Eran Maniv[1,2], Amanda Gong[1,2], Stefano Faubel[1,2], Ian M. Hayes[1,2], Ali Bangura[4], John Singleton[5], Johanna C. Palmstrom[5], Laurel Winter[5], Ross McDonald[5], Sooyoung Jang[1,2], Ping Ai[2], Yi Lin[2], Samuel Ciocys[1,2], Jacob Gobbo[1,2], Yochai Werman[1,2], Peter M. Oppeneer[3], Ehud Altman[1,2], Alessandra Lanzara[1,2], James G. Analytis[1,2]*

The study of quantum phase transitions that are not clearly associated with broken symmetry is a major effort in condensed matter physics, particularly in regard to the problem of high-temperature superconductivity, for which such transitions are thought to underlie the mechanism of superconductivity itself. Here we argue that the putative quantum critical point in the prototypical unconventional superconductor CeCoIn₅ is characterized by the delocalization of electrons in a transition that connects two Fermi surfaces of different volumes, with no apparent broken symmetry. Drawing on established theory of f-electron metals, we discuss an interpretation for such a transition that involves the fractionalization of spin and charge, a model that effectively describes the anomalous transport behavior we measured for the Hall effect.

C eCoIn₅ is an f-electron metal with notable similarities to high-temperature superconducting copper oxides, for example, in crystal structure, transport properties, and unconventional superconductivity (*1–10*). Both CeCoIn₅ and this category of copper oxides also exhibit signatures of a quantum phase transition (QPT), a phase transition induced by a nonthermal parameter, underlying the superconducting state. However, in many unconventional superconductors, it is unclear whether the underlying QPT can be understood in the conventional sense as a process of separating phases with different symmetries. Unconventional types of QPTs, such as non–symmetry breaking (*11*) or weakly symmetry breaking (*12, 13*) QPTs, have therefore become a subject of intense study. In this work, we provide evidence that CeCoIn₅ is proximate to a QPT in which the density of itinerant electrons changes, apparently without the breaking of symmetry. Established

theory of f-electron metals provides a means to interpret such a transition.

At the microscopic level, f-electron metals such as CeCoIn₅ are described by a Kondo lattice model. Each Ce atom hosts a single f-level valence electron, which contributes a localized spin-½ moment. These local moments coexist with a sea of itinerant conduction electrons. In the conventional metallic ground state of the Kondo lattice, the f-electrons appear to become an integral part of the itinerant metal. In particular, they join the conduction electrons, contributing their full share to the total Fermi volume as prescribed by Luttinger's theorem (*14*). This phenomenon occurs through the formation of Kondo singlet correlations between the local f moments and the conduction electrons, which effectively hybridize the f level with the conduction bands.

A long-standing challenge has been to characterize a QPT in which the f electrons recover their localized character and withdraw from the itinerant Fermi volume. Superficially, the remaining Fermi volume without f electrons is in apparent violation of Luttinger's theorem. The loss of Fermi volume when f electrons localize is therefore conventionally accompanied by a transition to a spin-density wave state, whereby Luttinger's theorem is recovered in the appropriately folded Brillouin zone associated with translational symmetry breaking (*15–19*). In this paper, we present

¹Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA. ²Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ³Department of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden. ⁴National High Magnetic Field Laboratory, Tallahassee, FL 32310, USA. ⁵National High Magnetic Field Laboratory, Los Alamos, NM 97545, USA.

*Corresponding author. Email: nikola_maksimovic@berkeley.edu (N.M.); analytis@berkeley.edu (J.G.A.)
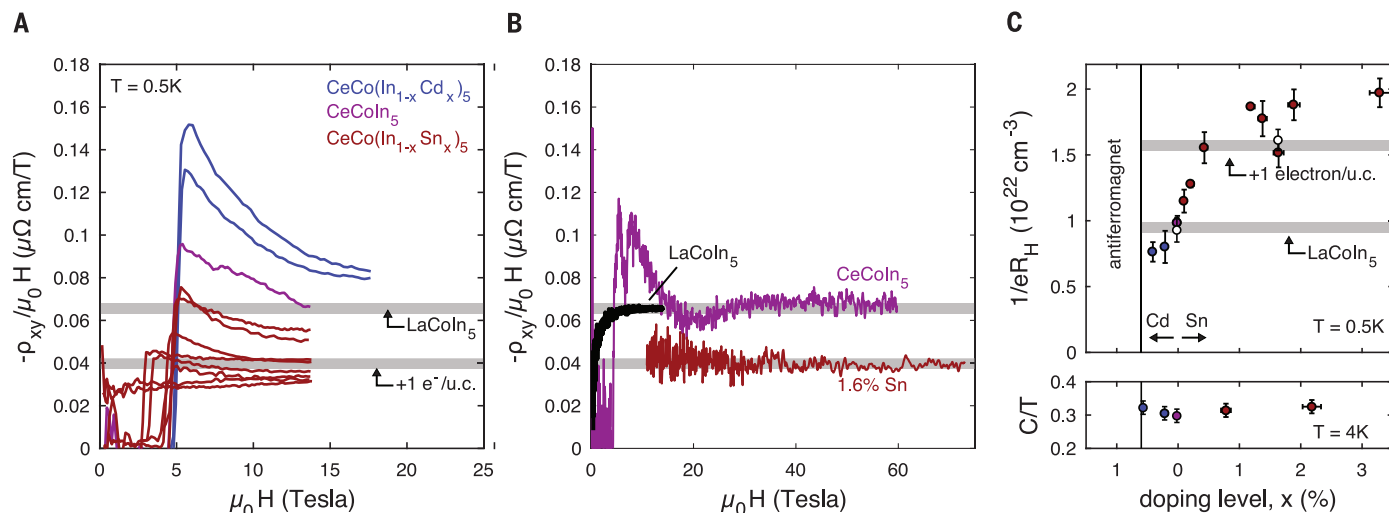
Fig. 1. Carrier density measurements in doped CeCoIn$_5$. (A) Hall coefficient as a function of field in doped CeCoIn$_5$ with Cd concentrations 0.2 and 0.4% and Sn concentrations 0.11, 0.22, 0.33, 0.44, 1.2, 1.39, 1.65, 1.9, and 3.3%. As discussed in the main text, the inverse of the Hall coefficient $\left(\rho_{xy}/\mu_0 H\right)$ in the high-field limit can be used to approximate the net carrier density [see also section 4 of (21)]. Gray lines denote the high-field Hall coefficient of the non-f analog LaCoIn$_5$ and the calculated value including one additional electron (e$^-$) per unit cell (u.c.). μΩ, microhm. **(B)** Pulsed-field Hall resistivity of CeCoIn$_5$ ($T = 0.66$ K) and Sn-doped CeCoIn$_5$ ($T = 0.5$ K) overlaid on the continuous-field Hall resistivity of LaCoIn$_5$ (1.8 K). **(C)** Inverse high-field Hall coefficient of CeCoIn$_5$ at 0.5 K as a function of doping level, including measurements in continuous field up to 14 or 18 T (filled circles) and pulsed field up to 73 T (open circles). With Sn substitution, the apparent carrier density of CeCoIn$_5$ increases by about one electron per unit cell above that of LaCoIn$_5$. This trend provides evidence that Sn substitution delocalizes the single Ce f electron per unit cell in CeCoIn$_5$. The value of $1/eR_H$ in some Sn-doped samples lies above the calculated +1 electron line, likely because the Hall coefficient has not completely saturated in these samples at 14 T. At higher fields, the value of $1/eR_H$ seems to saturate at the +1 electron value, as seen in the 1.6% Sn-doped sample at 70 T. The lower panel shows the 4 K heat capacity (units of millijoules per mole kelvin squared) across this doping series. Error bars indicate uncertainties in measurement of geometrical factors (in the case of the Hall resistivity) and sample mass (in the case of heat capacity).

Hall effect, quantum oscillation, and angle-resolved photoemission spectroscopy (ARPES) measurements of CeCoIn$_5$ with small levels of chemical substitution and compare the experimental data to ab initio calculations. We find evidence for an f-electron delocalization QPT without symmetry breaking.

Figure 1A presents low-temperature measurements of the Hall resistivity, ρ_{xy}, versus magnetic field, $\mu_0 H$, for CeCoIn$_5$ samples with varying levels of Cd (hole doping) or Sn (electron doping), both of which substitute In. The Hall coefficient, $R_H = \rho_{xy}/\mu_0 H$, can be used to estimate the net carrier density enclosed by the Fermi surface according to the formula (20)

$$n_{\text{net}} = \frac{1}{eR_H(H\to\infty)} \quad (1)$$

where n_{net} is the net carrier density—i.e., the carrier density of electrons minus that of holes. In multiple-band metals such as CeCoIn$_5$, Eq. 1 applies only in the limit where high fields eliminate the effects of carrier mobility imbalances and R_H becomes field independent [see section 3 of (21) for more details on the high-field limit]. For each sample, we measured the high-field value of R_H at 0.5 K to approximate the net carrier density. Many of the traces shown in Fig. 1A appear to saturate at high fields, and fig. S4 shows that the high-field

slope of ρ_{xy} is in good agreement with the high-field value of $\rho_{xy}/\mu_0 H$, which suggests that, at these temperatures and fields, the Hall coefficient is nearly field independent. In addition, select samples were measured in pulsed magnetic fields up to 75 T, as shown in Fig. 1B, where the Hall coefficient is field independent over an extended field range; the extracted Hall coefficients from pulsed and continuous fields are in good agreement for these samples (Fig. 1C). Finally, our Hall coefficient measurements on pure CeCoIn$_5$ are consistent with measurements at 20 mK, at which the Hall resistivity is completely linear in field (22). Together, these findings provide evidence that our extracted Hall coefficient values can be interpreted as an approximate measurement of the net carrier density as described by Eq. 1.

Figure 1C shows the value of $1/eR_H$, approximating the net carrier density, extracted for samples with different levels of chemical substitution in continuous and pulsed magnetic fields. The carrier density of these material, excluding the f electron, can be established by using Hall resistivity measurements of LaCoIn$_5$ (Fig. 1B) (its Hall coefficient is field independent above 5 T at 1.8 K; see also fig. S3)—LaCoIn$_5$ can be thought of as CeCoIn$_5$ without the f electron. We find that the Hall coefficient of CeCoIn$_5$, evaluated either up to

60 T or up to 14 T at 0.5 K, is close to that of LaCoIn$_5$ (Fig. 1C). This suggests that the two materials have similar net carrier densities, implying that the f electrons are nearly localized in CeCoIn$_5$. With Cd substitution, $1/eR_H$ remains close to that of LaCoIn$_5$, but with Sn substitution it increases to a value consistent with the addition of one itinerant electron per unit cell. Identifying the additional electron as the single Ce f electron suggests that Sn substitution induces a delocalization transition of the f electrons. None of these samples show a finite-temperature phase transition other than that associated with superconductivity. Only when Cd-substitution levels exceed 0.6% is an antiferromagnetic phase observed (fig. S1) (23). In addition, the specific heat capacity at moderate temperature remains constant across this substitution series (Fig. 1C); we will comment more on this later.

When the f electrons delocalize, the Fermi surfaces are expected to reconstruct and increase in volume. The results of our density functional theory (DFT) calculations of the three Fermi surfaces to compare the (de)localized f-electron models are visualized in Fig. 2A [DFT calculation details are provided in section 1 of (21)]. According to the calculations, f-electron delocalization causes the extended γ surface to disconnect into small
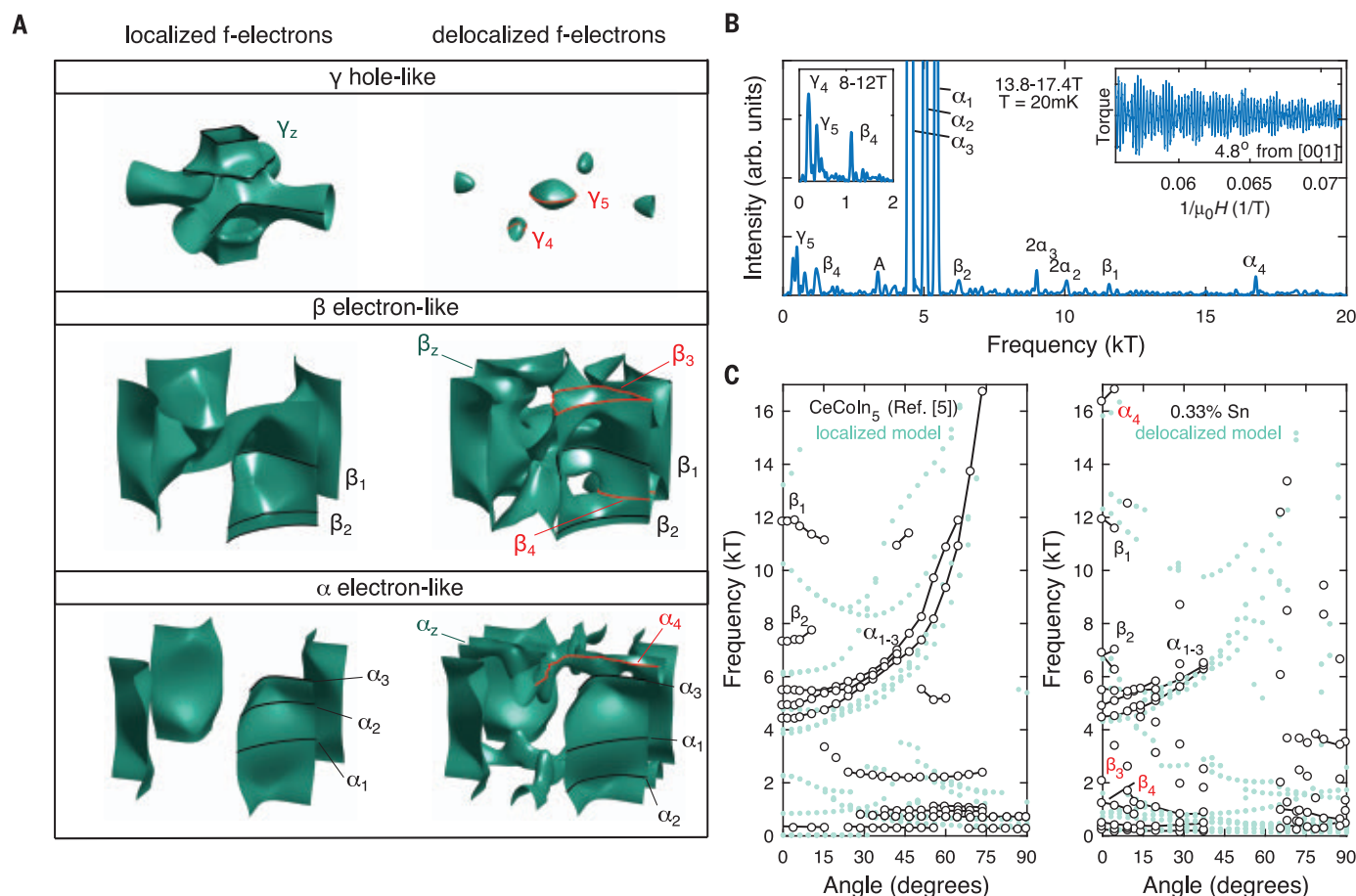
Fig. 2. De Haas–van Alphen oscillations in Sn-doped CeCoIn₅ and comparison to DFT calculations. (A) DFT-calculated Fermi surface sheets of CeCoIn₅ with localized and delocalized f-electron models. Predicted dHvA orbits for $H \parallel [001]$ are drawn in black and red; red orbits are specific to the delocalized f-electron model. **(B)** Characteristic dHvA spectrum with the magnetic field 4.8° away from [001] of a crystal of 0.33% Sn-doped CeCoIn₅. The right inset shows oscillations in the magnetic torque after background subtraction. **(C)** dHvA oscillation frequencies as a function of the angle that tilts the magnetic field from the crystallographic [001] to [100] directions in pure CeCoIn₅ (5) and 0.33% Sn-doped CeCoIn₅. Light green points are DFT-calculated frequencies of the localized and delocalized f-electron models, respectively.

ellipsoidal pockets at the Brillouin zone center and edge and results in the disappearance of the γ pocket at the zone top (γ_Z). Also, large extended surfaces α_Z and β_Z appear at the zone top, and the α and β cylinders expand slightly. In pure CeCoIn₅, previous ARPES data at 10 to 20 K are in better qualitative agreement with the localized f-electron model, as α_Z and β_Z are absent and γ_Z is present (24, 25). However, the volumes of the α and β cylinders are slightly larger than those of the localized model (24–26), and the smaller γ Fermi surface seems to exhibit features of both the delocalized and localized models: They are potentially disconnected (which suggests delocalization), but they retain γ_Z (indicative of localization) (24, 25, 27, 28). These characteristics may point to a partially delocalized f-electron character in pure CeCoIn₅. This interpretation is also promoted by previous magnetic resonance (29) and photoemission (24, 30, 31) studies. Notably, our Hall effect measurements suggest that the f elec-

trons only weakly contribute to the Fermi volume of CeCoIn₅, even at 0.5 K, consistent with the presence of partially localized f electrons in the low-temperature limit.

De Haas–van Alphen (dHvA) oscillations measure extremal areas of the Fermi surface perpendicular to the field direction, thus enabling examination of the Fermi surface structure at extremely low temperature. Here we compare our dHvA measurements of Sn-doped CeCoIn₅ and published data on pure CeCoIn₅ (5). As seen in Table 1 and Fig. 1C, the sizes of the α and β cylinder orbits in pure CeCoIn₅ are more consistent with the delocalized model, implying that f electrons incorporate into these Fermi surface sheets. However, there do not appear to be additional frequencies associated with the α_Z and β_Z sheets of the delocalized model, and orbit β₂ increases as a function of tilt angle away from [001] (Fig. 2C), further suggesting that the β cylinder is fully connected in better qualitative agreement

with the localized model. In the Sn-substituted sample, the sizes of the α and β cylinders change slightly compared with their sizes in pure CeCoIn₅ (Table 1). In addition, an oscillation of ~16 kT appears for two field angles near [001]. This oscillation does not appear to be harmonically related to the α_{1–3} branches, and its frequency and angle dependence are consistent with a predicted orbit on α_Z of our delocalized model calculations. Furthermore, 1.2- and 2-kT frequencies for field angles near [001] are indicative of holes in the β cylinder (Fig. 2A), and a branch of the β₂ cylinder orbit appears to decrease as a function of tilt angle from [001], in better agreement with the delocalized model (Fig. 1C) and further suggesting that holes develop in the β cylinder. Finally, possible low-frequency oscillations <800 T at several angles, which seem to be present in pure CeCoIn₅ over certain angular ranges as well, are most naturally assigned to small γ ellipsoids (Fig. 2C) but could also
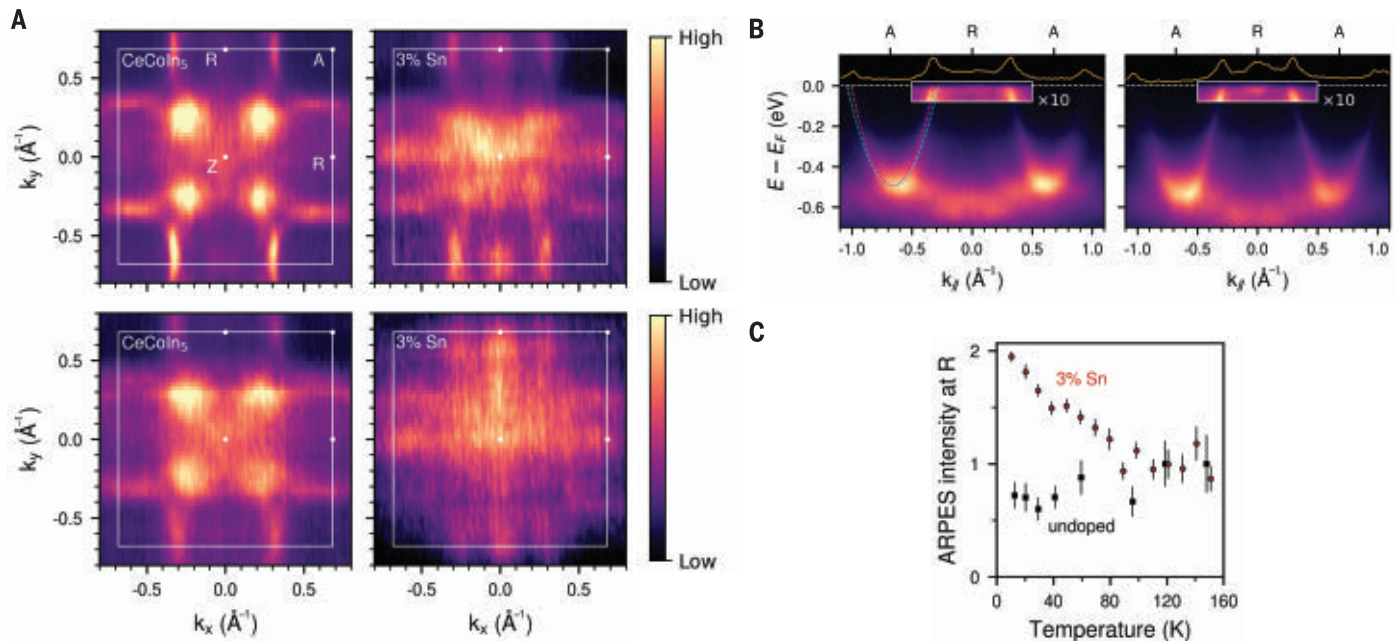
Fig. 3. ARPES measurements of CeCoIn$_5$ and Sn-doped CeCoIn$_5$. (**A**) Fermi surface maps in pure and 3% Sn-substituted CeCoIn$_5$ at the Brillouin zone top (RZA plane). A new Fermi surface sheet appears at the zone top in the Sn-substituted sample. Each of the four subpanels represents measurements on a different cleave. (**B**) A-R-A dispersion cuts. Parabolic α and β bands are labeled by red and blue dotted lines, respectively. The new Fermi surface in the Sn-substituted sample is observed as an increase in spectral intensity at the Fermi level at R. The spectral intensity within the white box has been enhanced by a factor of 10 for clarity. E is the energy, E_F is the Fermi energy, and $E - E_F$ denotes the energy relative to the Fermi energy. (**C**) Comparison of temperature-dependent intensity at the R point normalized to the average value between 120 and 160 K. Error bars indicate estimates of the standard error of the intensity in the fits to the spectral feature at R shown in fig. S15 and an additional 100% margin.

Table 1. De Haas–van Alphen extremal orbit assignments. Values are given in units of kilotesla, $H \parallel [001]$, as determined from experiments and DFT calculations. Each orbit is labeled according to the assigned Fermi surface sheet (orbits are visualized on the calculated Fermi surface sheets in Fig. 2A). Values in parentheses are low-frequency orbits that are consistent in size with the gamma ellipsoids but are not present at all angles as expected for ellipsoidal pockets. –, not present.

Fermi surface	dHvA orbit label	Localized f-electron model	CeCoIn$_5$ (5)	0.33% Sn-doped CeCoIn$_5$	Delocalized f-electron model
γ_Z	γ_1	0.8	–	–	–
γ_Z	γ_2	2.3	–	–	–
γ Cross	γ_3	13.2	–	–	–
γ Ellipsoid	γ_4	–	–	(0.46)	0.7
γ Ellipsoid	γ_5	–	(0.24)	(0.2)	0.22
α Cylinder	α_1	4.8	5.6	5.4	5.6
α Cylinder	α_2	4.0	4.5	4.8	4.4
α Cylinder	α_3	3.9	4.2	4.4	4.3
α_Z	α_4	–	–	16.3	15.8
β Cylinder	β_1	10.3	12.0	11.9	12.3
β Cylinder	β_2	6.1	7.5	6.8	6.7
β_Z/cylinder	β_3	–	–	2.0	1.6
β_Z/cylinder	β_4	–	–	1.2	0.9

originate from the γ_Z sheet. Table 1 summarizes the frequency assignments based on comparison to DFT calculations, suggesting that the α_Z and β_Z sheets are present in the Fermi surface of the Sn-substituted sample. From dHvA, it is not possible to conclusively say whether these sheets are absent in pure CeCoIn$_5$ at low temperature because the orbit frequencies on α_Z and β_Z are sensitive to the precise structure of these Fermi surfaces. Nevertheless, the comparison provided in Table 1 is indicative of a Fermi surface reconstruction induced by Sn substitution.

Our ARPES measurements corroborate the dHvA evidence for a Fermi surface reconstruction. Figure 3 compares Fermi surface maps at the Brillouin zone top in pure CeCoIn$_5$ and 3% Sn-substituted CeCoIn$_5$ at 10 K [additional data are provided in section 9 of (21)]. Our data on pure CeCoIn$_5$ agree well with earlier reports. The cylindrical Fermi surfaces centered at the zone corners are visible. Bright spots near the Z point are probably signatures of the γ_Z Fermi surface, as discussed previously
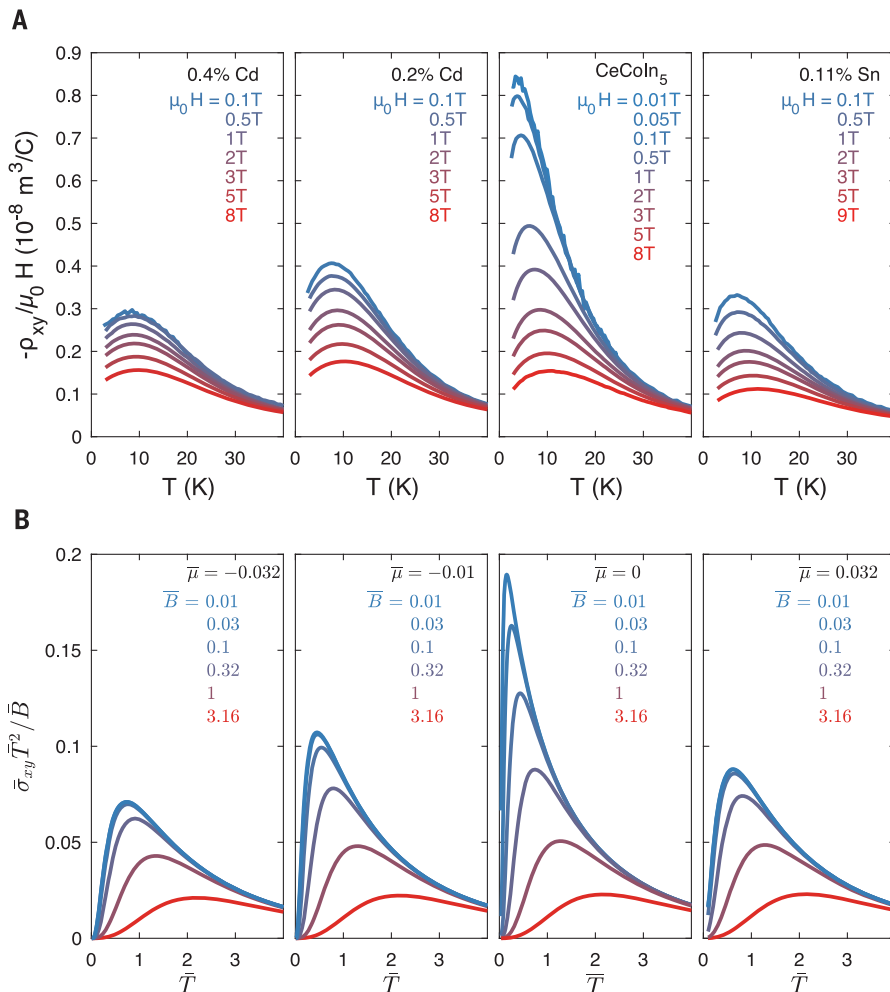
Fig. 4. Comparison of experimental data and theoretical calculations of the conductivity of critical valence fluctuations around an f-electron delocalization transition. (**A**) Experimentally measured Hall resistivity, divided by the applied magnetic field, for samples with different compositions. (**B**) Theoretically predicted Hall effect from bosonic valence fluctuations of the fractionalized Fermi liquid model. Each panel is labeled by the chemical potential in the theory corresponding to the doping level in the experiment, where $\mu < 0$ corresponds to hole doping and $\mu > 0$ corresponds to electron doping. Curves are labeled by the normalized magnetic field value, and all theory data include a parametrization of impurity scattering, $\bar{C} = 4$. See section 7 of (21) for the details of the calculation and relevant parameter normalizations.

(27, 31). In the 3% Sn-substituted sample, we observed enhanced intensity at the R point of the Brillouin zone relative to the pure material, as well as a qualitative change in structure near the Z point. Overall, the electronic structure appears to change with Sn substitution, with a sharp cross-shaped structure emerging in the RZA plane, which resembles α_Z or β_Z of our delocalized model calculations (α and β bands nearly overlap along this cut; as such, they may be difficult to distinguish from one another in ARPES). Weak features appear at the R point in pure CeCoIn$_5$ as well, potentially indicating that incoherent states exist at the R point—these states may exist because of the partially delocalized f-electron character in the pure material. In Fig. 3C, we explore the temperature

dependence of these Fermi surface sheets via the ARPES intensity at the R point. The relative intensity at R increases in the Sn-substituted sample when the temperature decreases below ~90 K with the onset of f/conduction hybridization (fig. S16). In the pure material, the R point spectral weight is relatively constant down to 10 K. This comparison suggests that the Fermi surface sheet in 3% Sn-doped CeCoIn$_5$ emerges, or is made relatively more coherent, because of enhanced f/conduction electron hybridization induced by Sn substitution.

One way to view f-electron delocalization is as a result of Kondo hybridization between the f level and conduction electrons. Although there are reports of hybridization developing below ~45 K in pure CeCoIn$_5$ (27) and Cd-

doped CeCoIn$_5$ (32), resulting in a detectable f-electron contribution to the Fermi surface, we find that the low-temperature carrier density of these materials is consistent with predominantly localized f electrons (Fig. 1). In contrast to that of the pure material, the net carrier density of Sn-substituted samples appears to include the f electrons (Fig. 1C). This change coincides with signatures of new Fermi surface sheets (Fig. 3 and Table 1), which seem to agree well with predicted Fermi surfaces that are specific to the delocalized f-electron DFT model (Fig. 2A). Taken together, these data suggest that Sn substitution of CeCoIn$_5$ induces a Fermi volume–changing transition between a phase with predominantly localized f electrons to one with a delocalized character. This transition could be attributed to an enhancement of the Kondo coupling induced by electron doping (25, 33, 34). High magnetic fields may compete with the Kondo coupling by polarizing the f electrons, but notably the Hall resistivity remains linear up to 73 T (Fig. 1B), so it seems likely that higher fields are required to induce a complete breakdown of Kondo hybridization.

A delocalization transition is a reasonable scenario from the perspective of doping-tuned Kondo coupling. Because of the constraints imposed by Luttinger's theorem, the reduction in Fermi volume in the more localized f-electron regime is expected to coincide with antiferromagnetic order where the Brillouin zone is reduced (15). It is, however, difficult to reconcile this scenario with the data because the transition to antiferromagnetism is seen only around a Cd doping of 0.6% (23), considerably removed from the suggested delocalization transition induced by Sn substitution (Fig. 1C). Furthermore, magnetic order has never been observed in native CeCoIn$_5$ or Sn-substituted CeCoIn$_5$ (6, 25, 33, 34), and the ARPES and dHvA data suggest that the Brillouin zone is essentially unchanged by Sn substitution. An alternative possibility is the formation of a fractionalized phase in the more localized f-electron regime (11). In this theoretically predicted phase, the f-electron charge localizes to the Ce site, reducing the Fermi volume, whereas the spin excitations of the f moments remain itinerant and form a charge-neutral Fermi surface (11). We can speculate that the specific heat remains constant across the substitution series (Fig. 1C) owing to the presence of such a neutral Fermi surface, which conserves the fermionic degrees of freedom of the system even when the density of itinerant electrons appears to increase. One may also expect quantum fluctuations associated with a delocalization transition to enhance the specific heat coefficient. Such an enhancement has been observed as a function of decreasing temperature below 2 K in pure CeCoIn$_5$ (2). The confinement of these effects to <2 K temperatures

could explain why we do not detect singular behavior in heat capacity (C)/temperature (T) at 4 K across the substitution series.

Our calculations of the Hall conductivity of such a fractionalized phase capture several distinctive aspects of the low-field Hall coefficient in this material. In the simplest description of the fractionalized Fermi liquid, the f electron separates into a fermionic spinon carrying its spin and a gapped bosonic mode, in this case a valence fluctuation, carrying its charge. Delocalization of f electrons can be identified with the closing of the boson gap. Near this transition, the electrical conductivity has contributions from the fermionic spinons, the charged bosons, and the conduction electrons. The spinon and the bosons should be added in series (35). The bosons' resistivity will then dominate, owing to their much smaller number, and we therefore neglect the spinon contribution. Adding to this the resistivity of the conduction band in parallel gives

$$R_\mathrm{H} = R_\mathrm{H}^c \frac{\sigma_c^2}{(\sigma_{tot})^2} + \frac{1}{\mu_0 H}\frac{\sigma_{xy}^b}{(\sigma_{tot})^2} \quad (2)$$

where σ_c and R_H^c are the longitudinal conductivity and Hall coefficient of the conduction electrons, respectively; σ_{xy}^b is the Hall conductivity of the bosonic valence fluctuations; and σ_{tot} is the total conductivity. In our calculation, we consider two processes that contribute to the scattering rate of the valence fluctuations. One process is provided by the internal gauge field (11). The other mechanism is scattering on the doped ions, which grows linearly with the doping level (fig. S5). One may expect an enhancement of the low-field Hall coefficient stemming from the second term in Eq. 2, caused by the singular behavior of the valence fluctuations when the boson gap closes. This expectation is corroborated by a semiclassical Boltzmann analysis, the details of which are given in section 7 of (21). As shown in Fig. 4, the results of the calculation of the conductivity in this model are in good agreement with the measured Hall coefficient as a function of temperature, doping level, and magnetic field, with the assumption that pure CeCoIn$_5$ is the sample closest to the delocalization transition. The results shown in Fig. 4B are obtained from a calculation of σ_{xy}^b and are converted to a Hall coefficient using the physical resistivity of the system $1/\sigma_{tot} = \rho_{xx} \sim T$, as observed in the experiment over the relevant temperature range. A more complete description of the longitudinal resistivity in this model will be the subject of future work.

We emphasize that the experimental observations illustrated in Fig. 4 are difficult to reconcile with more conventional transport models. From the point of view of band theory, the low-field R_H is proportional to the carrier density of the most-mobile carriers (20), so it is surprising that R_H has such a strong temperature dependence with a peak at finite temperature and retains the same sign and uniformly decreases with either hole or electron doping. In addition, the observed symmetric-in-doping Hall coefficient cannot be readily attributed to disorder scattering induced by substitution, as we find that disordering the material by other means, substituting La for Ce, has a relatively small effect on the low-field R_H (fig. S6). These key features of the experimental transport data are captured by the valence fluctuation model described above.

The present study provides evidence that CeCoIn$_5$ exists near a QPT associated with the delocalization of f-electron charge. The absence of evidence for symmetry breaking around this transition opens the possibility for the fractionalization of f electrons into separate spin and charge degrees of freedom. Although our conductivity calculations support this theoretical picture, direct evidence for such fractionalized electrons is desirable and may be possible with inelastic neutron measurements (36) or Josephson tunneling experiments (37). On a final note, recent experiments on cuprate high-temperature superconductors find evidence for a Fermi surface reconstruction in which the localized charge of the Mott insulator gradually delocalizes over a certain oxygen doping range near the end point of the pseudogap phase [sometimes referred to as a p to $1 + p$ transition, where p denotes the doped hole concentration (38)]. We have presented evidence for an analogous transition in an f-electron metal. It is possible that such a QPT underlies some of the similarities between CeCoIn$_5$ and cuprate superconductors (1–9), and perhaps our work may help guide interpretation of these recent results for cuprates.

REFERENCES AND NOTES

1. C. Petrovic et al., J. Phys. Condens. Matter 13, L337 (2001).
2. A. Bianchi, R. Movshovich, I. Vekhter, P. G. Pagliuso, J. L. Sarrao, Phys. Rev. Lett. 91, 257001 (2003).
3. J. Paglione et al., Phys. Rev. Lett. 91, 246405 (2003).
4. Y. Nakajima et al., Phys. C 460-462, 680–681 (2007).
5. R. Settai et al., J. Phys. Condens. Matter 13, L627 (2001).
6. Y. Kohori et al., Phys. Rev. B 64, 134526 (2001).
7. V. A. Sidorov et al., Phys. Rev. Lett. 89, 157004 (2002).
8. B. B. Zhou et al., Nat. Phys. 9, 474–479 (2013).
9. C. Stock, C. Broholm, J. Hudis, H. J. Kang, C. Petrovic, Phys. Rev. Lett. 100, 087001 (2008).
10. Y. Tokiwa, E. D. Bauer, P. Gegenwart, Phys. Rev. Lett. 111, 107003 (2013).
11. T. Senthil, M. Vojta, S. Sachdev, Phys. Rev. B 69, 035111 (2004).
12. C. M. Varma, Phys. Rev. B 73, 155113 (2006).
13. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, Proc. Natl. Acad. Sci. U.S.A. 114, 4905–4910 (2017).
14. M. Oshikawa, Phys. Rev. Lett. 84, 3370–3373 (2000).
15. Q. Si, F. Steglich, Science 329, 1161–1166 (2010).
16. S. Paschen et al., Nature 432, 881–885 (2004).
17. A. Schröder et al., Nature 407, 351–355 (2000).
18. P. Gegenwart et al., Phys. Rev. Lett. 81, 1501–1504 (1998).
19. J. Custers et al., Nature 424, 524–527 (2003).
20. A. B. Pippard, Magnetoresistance in Metals (Cambridge Univ. Press, 2009).
21. Supplementary materials.
22. S. Singh et al., Phys. Rev. Lett. 98, 057001 (2007).
23. L. D. Pham, T. Park, S. Maquilon, J. D. Thompson, Z. Fisk, Phys. Rev. Lett. 97, 056404 (2006).
24. Q. Y. Chen et al., Phys. Rev. B 100, 035117 (2019).
25. K. Chen et al., Phys. Rev. B 97, 045134 (2018).
26. A. Koitzsch et al., Phys. Rev. B 77, 155128 (2008).
27. S. Jang et al., Proc. Natl. Acad. Sci. U.S.A. 117, 23467–23476 (2020).
28. N. Gauthier et al., Phys. Rev. B 102, 125111 (2020).
29. N. J. Curro et al., Phys. Rev. B 64, 180514(R) (2001).
30. S. Fujimori, J. Phys. Condens. Matter 28, 153002 (2016).
31. Q. Y. Chen et al., Phys. Rev. B 96, 045107 (2017).
32. Q. Y. Chen et al., Phys. Rev. B 100, 235148 (2019).
33. K. Gofryk et al., Phys. Rev. Lett. 109, 186402 (2012).
34. H. Sakai et al., Phys. Rev. B 92, 121105 (2015).
35. L. B. Ioffe, A. I. Larkin, Phys. Rev. B 39, 8988–8999 (1989).
36. A. Banerjee et al., NPJ Quant. Mater. 3, 8 (2018).
37. T. Senthil, M. P. A. Fisher, Phys. Rev. B 64, 214511 (2001).
38. S. Badoux et al., Nature 531, 210–214 (2016).
39. N. Maksimovic, J. Rusz, T. Cookmeyer, D. Eilbott, J. Gobbo, Ce115, OSF (2021).

NEUROSCIENCE

Voltage compartmentalization in dendritic spines in vivo

Victor Hugo Cornejo*, Netanel Ofer, Rafael Yuste

Dendritic spines mediate most excitatory neurotransmission in the nervous system, so their function must be critical for the brain. Spines are biochemical compartments but might also electrically modify synaptic potentials. Using two-photon microscopy and a genetically encoded voltage indicator, we measured membrane potentials in spines and dendrites from pyramidal neurons in the somatosensory cortex of mice during spontaneous activity and sensory stimulation. Spines and dendrites were depolarized together during action potentials, but, during subthreshold and resting potentials, spines often experienced different voltages than parent dendrites, even activating independently. Spine voltages remained compartmentalized after two-photon optogenetic activation of individual spine heads. We conclude that spines are elementary voltage compartments. The regulation of voltage compartmentalization could be important for synaptic function and plasticity, dendritic integration, and disease states.

Dendritic spines are small protrusions that cover the dendrites of neurons (1) and mediate most excitatory connections in the brain (2–4). Aside from connecting neurons, they must play an additional role because excitatory axons specifically target spines, avoiding dendritic shafts (5). Spine morphologies, with a small (~1 μm in diameter) head connected to the dendrite by a thin (~140 nm in diameter) neck, suggest that they isolate synapses from the dendrite. Indeed, spines are calcium compartments (6, 7) that biochemically isolate synaptic inputs, en-

Neurotechnology Center, Department of Biological Sciences, Columbia University, New York, NY 10027, USA.
*Corresponding author. Email: victor.c@columbia.edu

abling input-specific plasticity (8, 9). However, calcium compartmentalization also occurs in aspiny dendrites (10, 11), so spines may implement additional functions. One possibility is that spines are also electrical compartments (12–15). Although dendritic voltages should fully invade the spine, synaptic potentials could attenuate as they propagate from the spine toward the parent dendrite (13, 16, 17). If spines were electrical compartments, the local depolarization at the spine caused by synaptic potentials could be substantially higher than what is measured at the dendrite or soma (18).

The introduction of optical methods has enabled measurements of the membrane potential of spines in vitro using organic dyes

or genetically encoded voltage indicators (GEVIs) (19). Although the invasion of action potentials (APs) into spines has been demonstrated (20–23), optical measurements of the electrical responses of spines to synaptic inputs have been inconsistent, with some studies finding similar potentials in spines and dendrites (21) and others finding larger depolarizations in spines (22, 23). Experiments with two-photon photobleaching or glutamate uncaging have been used to estimate the degree of electrical compartmentalization of spines, with a variety of results (24–31). Consistent with voltage compartmentalization, nanopipette recordings from spines in brain slices reveal large amplitude synaptic potentials (32). However, these experiments examined spines in vitro, so the electrical behavior of spines during physiological states remains unexplored.

To investigate the electrical function of spines in vivo, we developed a GEVI that could be efficiently excited with two-photon illumination and expressed it in pyramidal neurons from layer 2/3 mouse somatosensory cortex. This sensor, postASAP (short for "postsynaptic ASAP"), used an ASAP (accelerated sensor of APs) GEVI backbone (33), modified with mutations for enhanced sensitivity and a PSD95.FingR nanobody domain to enrich its expression in spines (34, 35) (Fig. 1A; see supplementary text). To express postASAP in vivo, we used in utero electroporation to achieve sparse, yet robust, expression and measured the coexpression of postASAP and red fluorescent protein (RFP) (Fig. 1B and fig. S2A). The soluble RFP showed higher fluorescence

Fig. 1. Characterization of postASAP. (**A**) Construction of postASAP. D, Asp; G, Gly; H, His; L, Leu; N, Asn; R, Arg; S, Ser; T, Thr. (**B**) Two-photon imaging of neurons expressing RFP-p2a-postASAP. Somatic expression (scale bar, 15 μm) is shown on the left. Dendritic expression (scale bar, 3 μm) is shown on the right. (**C**) Sensitivity of postASAP in ND7/23 cells to 500-ms voltage steps (mean ± SD; n = 5 cells). The red area corresponds to a linear range: y = 0.6x, coefficient of determination (R^2) = 0.9973, and p = 0.0014. (**D**) Experimental design. (**E**) On the left, a patched neuron expressing postASAP is shown. The red lines show the pipette outline, and the white dotted square shows the region of interest (ROI) for fluorescence measurement (scale bar, 5 μm). On the right, a representative optical trace of postASAP (light green, raw fluorescence; black, 10-Hz low-pass filtered) and electrical recording (red) are shown. (**F**) Average somatic electrical subthreshold signals (red, six events) and simultaneous fluorescence changes (black and gray, mean ± SD). (**G**) Correlation of peak postASAP fluorescent changes and subthreshold electrical amplitude [mean ± SD; n = 317 subthreshold events, 14 cells, and 8 animals; linear regression (red): y = 0.58x, confidence interval = 0.55 to 0.61, R^2 = 0.7156, and p = 0.0003].
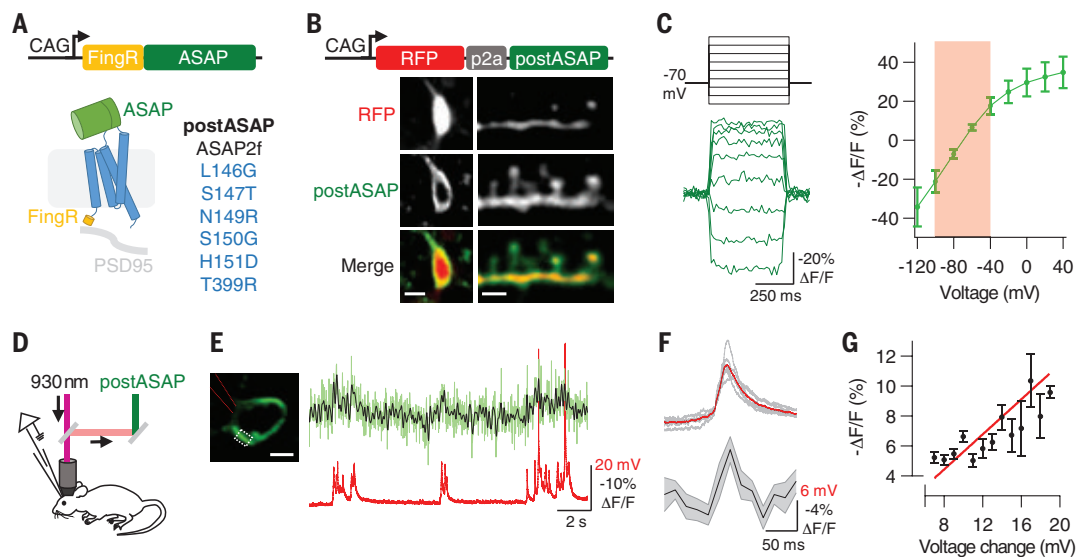
in larger volume compartments [fluorescence ratio soma/spine = 3.02 ± 1.89, soma/dendrite = 2.09 ± 1.84, dendrite/spine = 1.45 ± 1.21 (mean ± SD); $p < 0.0001$, Kruskal-Wallis test; fig. S2B]. Meanwhile, postASAP fluorescence levels were evenly distributed among neuronal compartments [fluorescence ratio soma/spine = 0.98 ± 0.59, soma/dendrite = 1.04 ± 0.58, dendrite/spine = 0.95 ± 0.62 (mean ± SD); $p = 0.40$, Kruskal-Wallis test], indicating membrane targeting, with robust labeling of dendrites and spines. To investigate the biophysical properties of postASAP, we first analyzed the fluorescence-voltage (F-V) relation with two-photon excitation in cultured ND7/23 cells (Fig. 1C). F-V curves were fitted with a Boltzmann function and displayed linearity in the range of −100 and −40 mV. To calibrate postASAP signals in vivo, we used whole-cell patch-clamp recordings and, simultaneously, two-photon imaging of postASAP in neuronal somata (Fig. 1D), finding a clear correspondence between electrical activity and fluorescence (Fig. 1E). Changes in fluorescence correlated with subthreshold depolarizations, rather than with APs (Fig. 1, E and F; see supplementary text), as expected from the low-pass behavior of the voltage sensor and imaging rate (~16 ms per frame). Within subthreshold depolarizations, the relation between voltage and changes in fluorescence was fitted by a linear function with a slope of 1.71 ± 0.02 mV per percent −ΔF/F [mV/(−ΔF/F%)] (slope ± SEM, $p < 0.0001$) (Fig. 1G), similar to the linear range of the F-V curve in cultured cells [1.67 mV/(−ΔF/F%)].

We then used two-photon imaging in vivo to measure the voltage dynamics experienced by basal dendrites and their spines from neurons expressing postASAP while performing simultaneous somatic whole-cell recordings (Fig. 2A; see supplementary text). During spontaneous activity, we found three spatiotemporal patterns of depolarizations in dendritic trees (Fig. 2B and movie S1). In the first spatial pattern ("AP"), which occurred during trains of APs, dendrites and spines were synchronously depolarized (Fig. 2, B and C). Peak depolarizations during APs were similar in spines and adjacent dendrites, confirming the functional expression of postASAP in spines and AP invasion into spines without failures or decrement (6, 7, 20, 36, 37) [Fig. 2, D and E; spine = 20.0 ± 9.5 mV, dendrite = 21.3 ± 9.9 mV (mean ± SD); $p = 0.24$, Mann-Whitney test]. In the absence of APs, during subthreshold potentials or even in the absence of pronounced somatic depolarization in the whole-cell recordings, two other spatial types of depolarization were observed (Fig. 2, B and C; fig. S8; and supplementary text). In a second spatial pattern ("Dendrite+Spines"), localized segments of dendrites with associated spines were depolarized together [Fig. 2F; spine = 15.9 ± 5.8 mV,

Fig. 2. Spine and dendritic voltage dynamics in vivo during spontaneous activity. (**A**) In vivo two-photon imaging and somatic whole-cell recording of a neuron expressing postASAP is shown at the top (red lines, pipette outline). Imaged dendrites (43 μm from center of the image to cell body) of a patched cell are shown at the bottom (scale bars, 5 μm). (**B**) A somatic electrical recording of the neuron in (A) is shown at the top. AP, train of APs; Sub, subthreshold depolarization; RMP, resting membrane potential. Simultaneous fluorescence changes of numbered spines and adjacent dendrites in (A) are shown at the bottom. (**C**) Representative image with peak fluorescence changes in dendrites and spines during three conditions in (B). (**D**) Depolarization during APs, generated by three 100-ms current pulses (300 pA). Somatic imaging and electrophysiological recording are shown at the top (scale bar, 5 μm). Representative fluorescence changes in a dendrite are shown at the bottom [average three trials; spine at 48 μm from cell body; scale bar, 5 μm; color scale same as (B)]. (**E**) Peak spine and dendrite fluorescence changes during AP trains ($n = 125$ spines, 37 dendrite segments, 5 cells, and 4 animals; linear regression: $y = 0.93x$, $R^2 = 0.823$, and $p < 0.0001$). Box and whiskers represent median (line), 25th to 75th percentiles (box), range (whiskers), and mean as a "+." (**F**) Examples of Dendrite +Spines patterns are shown on the left (average 10 events; scale bar, 5 μm). Peak fluorescence changes in spine heads and adjacent dendrites are shown on the right ($n = 221$ spines, 90 dendritic segments, 13 cells, and 7 animals). n.s., not significant. (**G**) Same as (F) for Spine-only pattern ($n = 116$ spines, 90 dendritic segments, 13 cells, and 7 animals; scale bar, 5 μm). ****$p < 0.0001$.
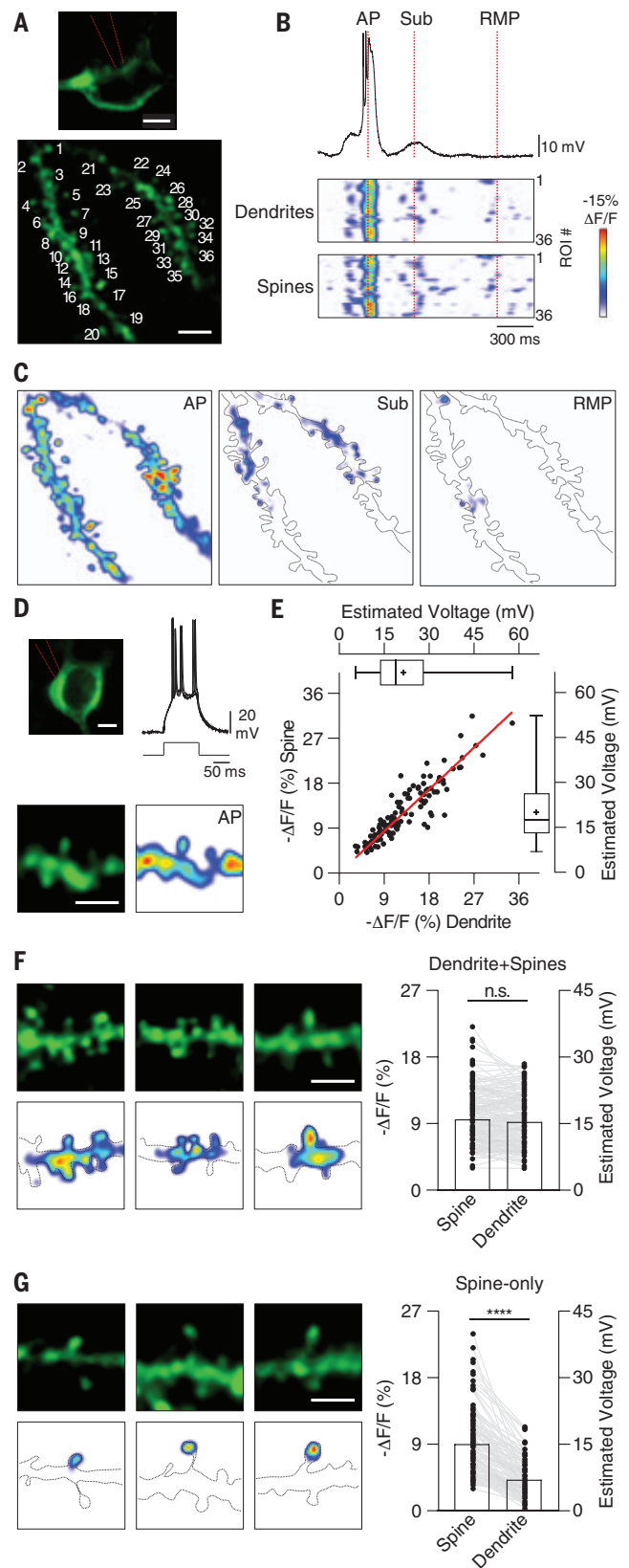
Fig. 3. Spine and dendritic voltage dynamics in vivo after sensory stimulation. (**A**) Experimental design. (**B**) Representative image of dendrites and spines expressing postASAP (scale bar, 5 μm). (**C**) Simultaneous fluorescence changes of numbered spines and adjacent dendrites in (B). (**D**) Example image with peak fluorescence changes in dendrites and spines in the time point indicated in (C) is shown on the left. Peak fluorescence changes in spine heads and adjacent dendrites for the Dendrite+Spines pattern are shown on the right (n = 255 spines, 49 dendrite segments, and 5 animals). (**E**) Same as (D) for Spine-only events (n = 181 spines, 49 dendrites, and 5 animals). ****p < 0.0001.
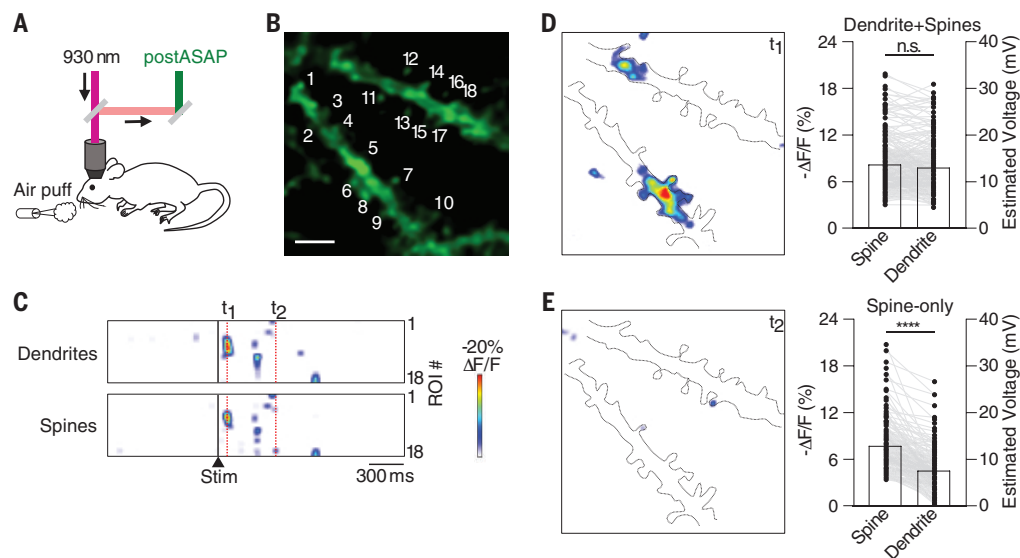
dendrite = 15.3 ± 5.6 mV (mean ± SD); p = 0.3767, Mann-Whitney test]. Interestingly, we also observed a third pattern ("Spine-only"), where individual spines became depolarized independently, or with a reduced depolarization of the parent dendrite, presumably reflecting isolated synaptic events [Fig. 2G; spine = 15.0 ± 7.2 mV, dendrite = 6.97 ± 4.2 mV (mean ± SD); p < 0.0001, Mann-Whitney test]. Spine voltage peaks reached by Spine-only events ranged from 5.0 to 39.8 mV, although this might represent an overestimate owing to the difficulty of measuring smaller depolarizations with sufficient signal-to-noise ratio (fig. S4A and supplementary text).

Next, we wondered what spatial patterns of depolarizations were found after sensory stimulation. We used air puffs to activate the whiskers while simultaneously imaging postASAP in dendrites and spines (Fig. 3A; see supplementary text) and also found AP (fig. S7, C and D, and movie S2), Dendrite +Spines, and Spine-only patterns (Fig. 3, B and C). After sensory stimulation, Dendrite +Spines events [Fig. 3D; spine = 13.7 ± 5.5 mV, dendrite = 13.0 ± 5.4 mV (mean ± SD); p = 0.1299, Mann-Whitney test] were detected more frequently, whereas Spine-only events did not change significantly (fig. S8 and supplementary text). Spine-only patterns [Fig. 3E; spine = 12.8 ± 5.4 mV, dendrite = 7.5 ± 4.5 mV (mean ± SD); p < 0.0001, Mann-Whitney test] had similar peak amplitudes to those during spontaneous activity. All dendritic and spine depolarizations were reversibly blocked by synaptic antagonists (fig. S9 and supplementary text), confirming their physiological nature.

We then turned our attention to the Spine-only patterns, where spines activated independently. To explore this, we activated spine heads or dendritic segments using two-photon optogenetics (38) with ChrimsonR (39) (Fig. 4A; see supplementary text) while simultaneously measuring their voltages with postASAP (Fig. 4B). For calibration, we imaged postASAP in the soma during two-photon optogenetic activation (Fig. 4C) and found a gradual fluorescence response, proportional to laser power (fig. S12A). In vivo patch-clamp recordings indicated reliable optogenetic responses and allowed us to measure the currents generated by the photostimulation (Fig. 4D). We then photostimulated small segments of dendrites while measuring postASAP fluorescence in dendrites and spines and found similar depolarizations in dendrites and adjacent spines [Fig. 4E; spine = 8.1 ± 4.2 mV, dendrite = 7.6 ± 4.0 mV (mean ± SD); p = 0.55, Mann-Whitney test]. Finally, we photostimulated spine heads to mimic synaptic potentials and found depolarizations that decreased as they spread into the parent dendrites [Fig. 4F; spine = 11.2 ± 3.6 mV, dendrite = 6.0 ± 2.2 mV (mean ± SD); p < 0.0001, Mann-Whitney test]. These optogenetic results were in agreement with our measurements during spontaneous or sensory-evoked activity (Figs. 2 and 3), confirming voltage attenuation from spine to dendrite but no attenuation from dendrite to spine. This asymmetric electrical behavior is expected from cable properties (13, 16).

To estimate the electrical properties of the spines, we used a simplified steady-state electrical equivalent circuit (30, 40) (Fig. 4G), with a very small (<0.01 pF) spine capacitance, so that synaptic currents flow through the spine neck resistance (R_n). By defining the spine input resistance (R_{sp}) as the sum of R_n and the dendritic input resistance (R_{den}), we obtain

$$R_{sp} = R_n + R_{den} \qquad (1)$$

According to Ohm's law, R_{sp} depends on the current flowing across the synapse (I_{syn}) and the voltage in the spine (V_{sp}):

$$R_{sp} = \frac{V_{sp}}{I_{syn}} \qquad (2)$$

Applying the voltage divider equation, one finds

$$\frac{R_n}{R_{den}} = \frac{V_{sp}}{V_{den}} - 1 \qquad (3)$$

Because we obtained a similar average resistance ratio (R_n/R_{den}) for Spine-only events during spontaneous activity and sensory stimuli, and also for optogenetically stimulated spines [Fig. 4H; 1.7 ± 2.8, 2.1 ± 5.9, and 1.1 ± 0.8, respectively (mean ± SD); p = 0.34, Kruskal-Wallis test], R_n would be equivalent to R_{den} for most spines (29). To estimate R_n, we solved Eqs. 1 to 3, obtaining

$$R_n = \frac{V_{sp} - V_{den}}{I_{syn}} \qquad (4)$$

and assuming an I_{syn} = 22.7 pA for photostimulated spines (Fig. 4D), we obtained an average R_n of 226.6 ± 128.8 megohm (mean ± SD), ranging from ~0 to 530.8 megohm (Fig. 4I; see supplementary text), in line with previous estimates with different methods (29, 30, 32).

Our results demonstrate that measurements of spine depolarization in vivo are feasible with two-photon GEVI imaging, providing an initial glimpse into the rich spatiotemporal patterns of depolarizations of dendrites and spines during spontaneous activity or sensory stimulation. Consistent with previous reports (6, 7, 20–22), we found synchronous depolarization of dendrites and spines during backpropagation of axonal APs into dendrites and spines (AP pattern). In addition, we detected
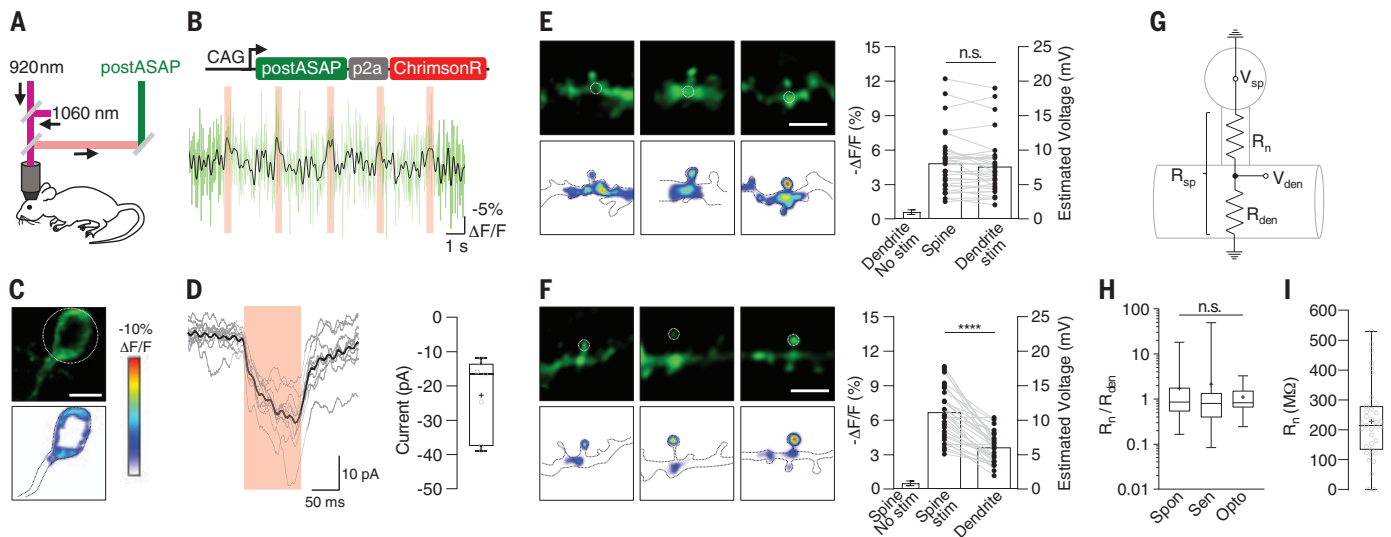
Fig. 4. Two-photon optogenetics and voltage imaging in vivo. (**A**) Experimental design. (**B**) Construct and representative postASAP fluorescence changes in soma (light green, raw fluorescence; black, 10-Hz low-pass filtered) during 500-ms stimulation trials (red, 100 mW power). (**C**) Representative soma (top) and peak fluorescence response (bottom) during stimulation trials (×10, 500 ms, 100 mW). The dotted circle shows the stimulation area (scale bar, 10 μm). (**D**) Representative in vivo voltage-clamp recordings during optogenetic stimulation of proximal dendrites (100 mW, 100 ms) are shown on the left. Peak currents are shown on the right; −22.7 ± 11.3 pA (mean ± SD), 10 trials ($n = 7$ cells and 4 animals). (**E**) Representative peak fluorescence changes during optogenetic activation of dendritic shafts are shown on the left [stimulation ROIs are indicated by white dotted circles; 10 trials, 100 ms, 100 mW; color bar same as (C); scale bar, 5 μm]. Peak fluorescence changes in stimulated dendritic shaft (dendrite stim), adjacent dendritic spine (spine), and unstimulated dendritic shaft (dendrite no stim) are shown on the right ($n = 34$ dendrites and 9 animals). (**F**) Same as (E) during optogenetic activation of spines ($n = 35$ spines and 12 animals; scale bar, 5 μm). ****$p < 0.0001$. (**G**) Simplified electrical model. (**H**) Resistance ratio (R_n to R_{den}) of Spine-only events during spontaneous (Spon, $n = 116$), sensory stimuli (Sen, $n = 181$), and optogenetic spine stimulation (Opto, $n = 35$). (**I**) Values of spine neck resistance for stimulated spines; median = 213.7 megohms ($n = 35$ spines). In (D), (H), and (I), boxes and whiskers represent median (line), 25th to 75th percentiles (box), range (whiskers), and mean as a "+".

local and transient depolarizations of dendritic segments and their spines (Dendrite+Spines pattern), likely corresponding to a combinatory of synaptic potentials and dendritic spikes (*41–44*). Importantly, we also found individual spines activating independently (Spine-only pattern), even in the absence of dendritic or somatic activity. These spine-independent depolarizations likely represent individual synaptic potentials, owing to their occurrence during subthreshold synaptic inputs, their spatial restriction, and their sensitivity to synaptic blockers (*6, 7, 45, 46*). These synaptic depolarizations of spines can be large in amplitude, at least of tens of millivolts (*16, 32, 47–49*). The presence of spine-independent depolarization during spontaneous and sensory-evoked activity implies that spines can compartmentalize voltage in physiological states in vivo. Indeed, using two-photon optogenetic activation of individual spines, we demonstrated this directly while also revealing that dendritic potentials propagated into spines faithfully. Thus, our data indicate that the spines act asymmetrically, with no attenuation of dendritic potentials or AP but significant attenuation of synaptic potentials (*13, 16*), maintaining a voltage gradient with the dendrite (*50, 51*). The mechanisms underlying spine voltage compartmentalization could involve geometrical and structural factors, passive cable properties,

voltage-sensitive ion channels, or synaptic receptors (*18*). Regardless of the mechanism, our results show that dendritic spines, in addition to serving as biochemical compartments, are also elementary electrical compartments for synapses. The regulation of the voltage compartmentalization of synaptic inputs by spines could be important for synaptic function (*52*), synaptic plasticity (*8, 53*), and dendritic integration (*54*) and be affected in mental and neurological diseases (*55*).

REFERENCES AND NOTES

1. S. Ramon y Cajal, *Rev. Trim. Histol. Norm. Patol.* **1**, 1–10 (1888).
2. E. G. Gray, *Nature* **183**, 1592–1593 (1959).
3. K. M. Harris, S. B. Kater, *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
4. R. Yuste, *Dendritic Spines* (MIT Press, 2010).
5. A. Peters, I. R. Kaiserman-Abramof, *Am. J. Anat.* **127**, 321–355 (1970).
6. R. Yuste, W. Denk, *Nature* **375**, 682–684 (1995).
7. X. Chen, U. Leischner, N. L. Rochefort, I. Nelken, A. Konnerth, *Nature* **475**, 501–505 (2011).
8. J. C. Magee, D. Johnston, *Science* **275**, 209–213 (1997).
9. S. J. Lee, Y. Escobedo-Lozoya, E. M. Szatmari, R. Yasuda, *Nature* **458**, 299–304 (2009).
10. J. H. Goldberg, G. Tamas, D. Aronov, R. Yuste, *Neuron* **40**, 807–821 (2003).
11. G. J. Soler-Llavina, B. L. Sabatini, *Nat. Neurosci.* **9**, 798–806 (2006).
12. W. Rall, in *Excitatory Synaptic Mechanisms, Proceedings of the Fifth International Meeting of Neurobiologists*, P. Andersen, J. Jansen, Eds. (Universitets Forlaget, Oslo, 1970), pp. 175–187.
13. J. J. B. Jack, D. Noble, R. W. Tsien, *Electric Current Flow in Excitable Cells* (Oxford Univ. Press, 1975).
14. C. Koch, T. Poggio, *Proc. R. Soc. London Ser. B* **218**, 455–477 (1983).
15. C. Koch, A. Zador, T. H. Brown, *Science* **256**, 973–974 (1992).
16. D. Johnston, S. M.-S. Wu, *Foundations of Cellular Neurophysiology* (MIT Press, 1995).
17. D. Tsay, R. Yuste, *Trends Neurosci.* **27**, 77–83 (2004).
18. R. Yuste, *Annu. Rev. Neurosci.* **36**, 429–449 (2013).
19. Y. Bando, C. Grimm, V. H. Cornejo, R. Yuste, *BMC Biol.* **17**, 71 (2019).
20. M. Nuriya, J. Jiang, B. Nemet, K. B. Eisenthal, R. Yuste, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 786–790 (2006).
21. M. A. Popovic, X. Gao, N. T. Carnevale, D. Zecevic, *Cereb. Cortex* **24**, 385–395 (2014).
22. T. Kwon, M. Sakamoto, D. S. Peterka, R. Yuste, *Cell Rep.* **20**, 1100–1110 (2017).
23. C. D. Acker, E. Hoyos, L. M. Loew, *eNeuro* **3**, ENEURO.0050-15.2016 (2016).
24. K. Svoboda, D. W. Tank, W. Denk, *Science* **272**, 716–719 (1996).
25. B. L. Bloodgood, B. L. Sabatini, *Science* **310**, 866–869 (2005).
26. R. Araya, K. B. Eisenthal, R. Yuste, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18799–18804 (2006).
27. R. Araya, J. Jiang, K. B. Eisenthal, R. Yuste, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17961–17966 (2006).
28. B. L. Bloodgood, A. J. Giessel, B. L. Sabatini, *PLOS Biol.* **7**, e1000190 (2009).
29. M. T. Harnett, J. K. Makara, N. Spruston, W. L. Kath, J. C. Magee, *Nature* **491**, 599–602 (2012).
30. J. Tønnesen, G. Katona, B. Rózsa, U. V. Nägerl, *Nat. Neurosci.* **17**, 678–685 (2014).
31. R. Araya, T. P. Vogels, R. Yuste, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2895–E2904 (2014).
32. K. Jayant et al., *Nat. Nanotechnol.* **12**, 335–342 (2017).
33. F. St-Pierre et al., *Nat. Neurosci.* **17**, 884–889 (2014).
34. V. Villette et al., *Cell* **179**, 1590–1608.e23 (2019).
35. G. G. Gross et al., *Neuron* **78**, 971–985 (2013).
36. C. R. Rose, Y. Kovalchuk, J. Eilers, A. Konnerth, *Pflugers Arch.* **439**, 201–207 (1999).

37. L. M. Palmer, G. J. Stuart, *J. Neurosci.* **29**, 6897–6903 (2009).
38. A. M. Packer *et al.*, *Nat. Methods* **9**, 1202–1205 (2012).
39. N. C. Klapoetke *et al.*, *Nat. Methods* **11**, 338–346 (2014).
40. C. Koch, A. Zador, *J. Neurosci.* **13**, 413–422 (1993).
41. R. Yuste, M. J. Gutnick, D. Saar, K. D. Delaney, D. W. Tank, *Neuron* **13**, 23–43 (1994).
42. J. Schiller, Y. Schiller, G. Stuart, B. Sakmann, *J. Physiol.* **505**, 605–616 (1997).
43. K. Holthoff, Y. Kovalchuk, R. Yuste, A. Konnerth, *J. Physiol.* **560**, 27–36 (2004).
44. H. Jia, N. L. Rochefort, X. Chen, A. Konnerth, *Nature* **464**, 1307–1312 (2010).
45. T. W. Chen *et al.*, *Nature* **499**, 295–300 (2013).
46. D. E. Wilson, D. E. Whitney, B. Scholl, D. Fitzpatrick, *Nat. Neurosci.* **19**, 1003–1009 (2016).
47. I. Segev, W. Rall, *J. Neurophysiol.* **60**, 499–523 (1988).
48. R. Araya, V. Nikolenko, K. B. Eisenthal, R. Yuste, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12347–12352 (2007).
49. N. L. Golding, N. Spruston, *Neuron* **21**, 1189–1200 (1998).
50. D. Holcman, Z. Schuss, *J. Math. Neurosci.* **1**, 10 (2011).
51. T. Lagache, K. Jayant, R. Yuste, *J. Comput. Neurosci.* **47**, 77–89 (2019).
52. R. Yuste, *Neuron* **71**, 772–781 (2011).
53. W. Rall, in *Studies in Neurophysiology*, R. Porter, Ed. (Cambridge Univ. Press, 1974), pp. 203–209.
54. M. London, M. Häusser, *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
55. D. P. Purpura, *Science* **186**, 1126–1128 (1974).

STRUCTURAL BIOLOGY

Cryo-EM structure of human GPR158 receptor coupled to the RGS7-Gβ5 signaling complex

Dipak N. Patil[1]†, Shikha Singh[2]†, Thibaut Laboute[1], Timothy S. Strutzenberg[3], Xingyu Qiu[4,5], Di Wu[4,5], Scott J. Novick[3], Carol V. Robinson[4,5], Patrick R. Griffin[3], John F. Hunt[2], Tina Izard[6], Appu K. Singh[7,8]*, Kirill A. Martemyanov[1]*

GPR158 is an orphan G protein–coupled receptor (GPCR) highly expressed in the brain, where it controls synapse formation and function. GPR158 has also been implicated in depression, carcinogenesis, and cognition. However, the structural organization and signaling mechanisms of GPR158 are largely unknown. We used single-particle cryo–electron microscopy (cryo-EM) to determine the structures of human GPR158 alone and bound to an RGS signaling complex. The structures reveal a homodimeric organization stabilized by a pair of phospholipids and the presence of an extracellular Cache domain, an unusual ligand-binding domain in GPCRs. We further demonstrate the structural basis of GPR158 coupling to RGS7-Gβ5. Together, these results provide insights into the unusual biology of orphan receptors and the formation of GPCR-RGS complexes.

G protein–coupled receptors (GPCRs), which form the largest family of proteins encoded in mammalian genomes, detect extracellular signals to program cellular response. GPCRs are essential to understanding physiology, disease, and drug development (*1*, *2*). The canonical model posits that GPCRs transduce their signals by recruitment and activation of heterotrimeric G proteins (*3*). This model was subsequently updated to accommodate alternative signal propagation by recruitment of β-arrestin scaffolds (*4*). Termination of GPCR signaling requires the action of RGS (regulator of G protein signaling) proteins, which directly deactivate G proteins (*5*, *6*). GPCRs and RGS are thus classically considered as opposing forces in controlling cellular responses. However, they have long been reported to form complexes, which suggests the existence of additional signaling mechanisms (*7*, *8*).

Orphan GPCRs are attractive drug targets with important roles in physiology and disease (*9*, *10*). Yet in many cases, their mechanisms, ligands, and signaling reactions are poorly understood. An example is the orphan receptor GPR158. It is one of the most abundant GPCRs in the brain, well documented for its pivotal role in regulating mood and cognition, and is implicated in a range of diseases (*11–15*). It shapes synaptic organization and function by regulating ion channels and second messengers (*16*, *17*). GPR158 features a large ex-tracellular domain with distinctive sequence suggesting unique ligand recognition principles. The central feature of GPR158 is its association with the neuronal RGS7-Gβ5 protein complex (*11*). Binding with GPR158 potentiates RGS activity (*18*), and both proteins act together to regulate homeostasis of the second messenger cyclic adenosine monophosphate (cAMP) to control neuronal activity with marked impact on brain physiology (*19*). However, its signaling mechanisms and structural organization remain elusive.

We used single-particle cryo–electron microscopy (cryo-EM) to obtain structures of GPR158 in the apo state and in complex with RGS7-Gβ5 at an average resolution of 3.4 Å and 3.3 Å, respectively (figs. S1 to S3 and table S1). The structure of GPR158 reveals a homodimer assembly (Fig. 1A) where the dimerization interface involves the extracellular domain (ECD), the transmembrane (TM) region, and cytoplasmic elements (Fig. 1A and fig. S4A). Each protomer features prominent extracellular and transmembrane domains linked by a flexible "stalk" domain. The N-terminal portion of the ECD adopts a characteristic Cache domain fold (the name derives from "calcium channels and chemotaxis receptors"). The TM region of protomers contains well-resolved helices. We observed continuous density for TM1 through TM7 including all extra- and intracellular loops (ECLs and ICLs), except for ICL2. We further detected two phospholipids in the cavity generated by the dimeric interface and several cholesterol-like molecules packed against hydrophobic residues of the TM domain, including the dimeric interface. The density at the ECD is limited, and many of the side-chain densities are not visible. The overall *B* factor and average side-chain *B* factor for ectodomain are high (>50 Å) (fig. S4C), suggesting greater conformational flexibility of the ECD. Nonetheless, the key organizational features of ECD are clearly distinguishable.

[1]Department of Neuroscience, The Scripps Research Institute, Jupiter, FL 33458, USA. [2]Department of Biological Sciences, Columbia University, New York, NY 10027, USA. [3]Department of Molecular Medicine, The Scripps Research Institute, Jupiter, FL 33458, USA. [4]Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK. [5]Kavli Institute for Nanoscience Discovery, Oxford OX1 3QU, UK. [6]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, Jupiter, FL 33458, USA. [7]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India. [8]Mehta Family Centre for Engineering in Medicine, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India.
*Corresponding author. Email: kirill@scripps.edu (K.A.M.); singhappu@iitk.ac.in (A.K.S.)
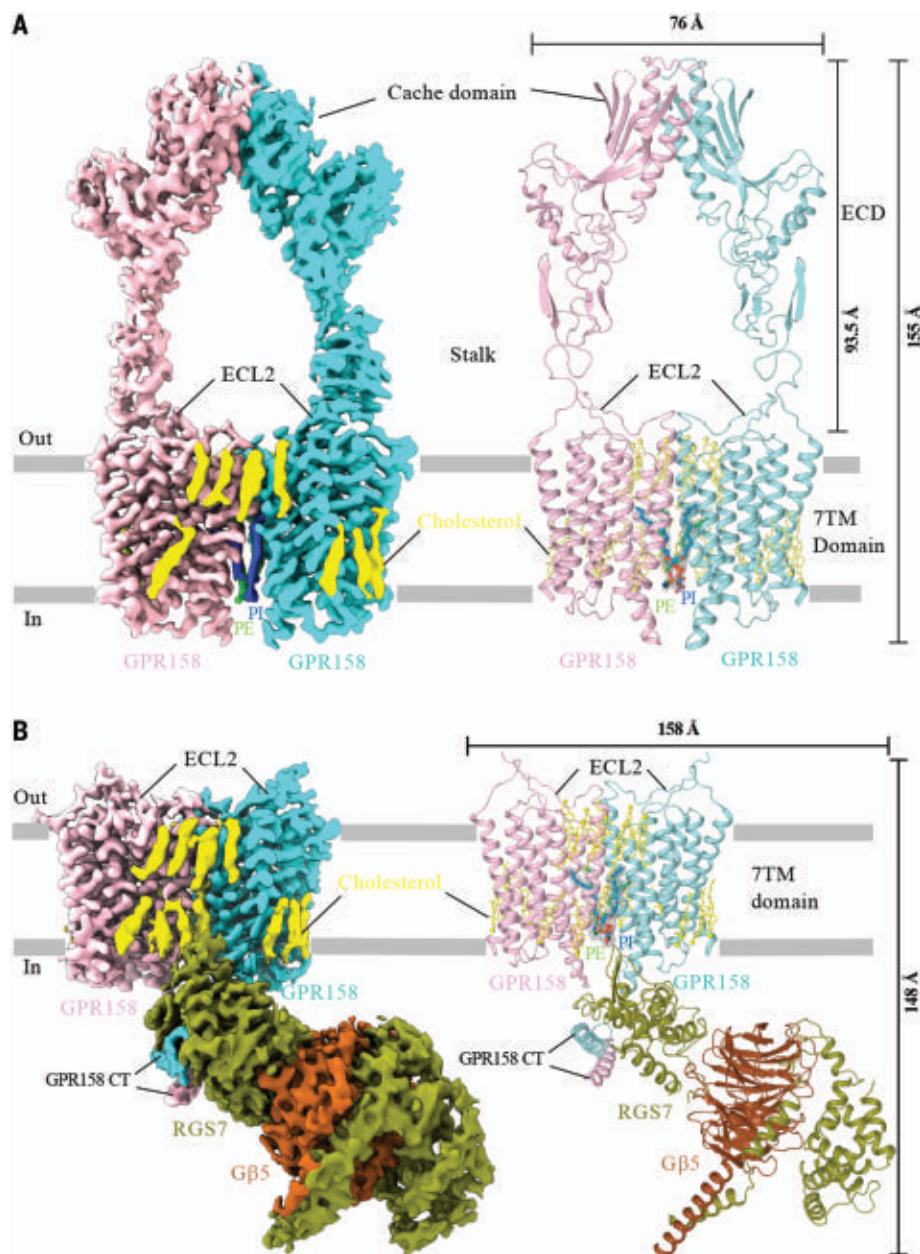†These authors contributed equally to this work.

Fig. 1. Cryo-EM structures of GPR158 in its apo and RGS7-Gβ5–bound states. (**A**) Cryo-EM map (left) and model (right) of GPR158 homodimer in ribbon representation with protomers colored in cyan and pink. Phospholipids PE and PI are shown in green and blue, respectively; cholesterols are yellow. (**B**) Cryo-EM map (left) and model (right) of GPR158 homodimer complexed with RGS7-Gβ5 in ribbon representation colored as in (A); RGS7 and Gβ5 are shown in olive and brown.

The structure of the GPR158 homodimer in complex with RGS7-Gβ5 shows that one RGS7-Gβ5 heterodimer interacts with two sites on GPR158 (Fig. 1B and fig. S4B). The first interface is a two-helical bundle comprising a C-terminal helix (CT-CC) from each protomer. The second contact surface is at the dimer interface between the seven-transmembrane (7TM) regions of the GPR158 protomers. We found that in the RGS7-bound state, the density for the GPR158 ectodomain becomes diffuse. Although the density is observable at low contour, the model could not be built. This higher conformational flexibility at the ectodomain, possibly induced by RGS7-Gβ5 binding, suggests that extracellular and intracellular elements of the complex could be allosterically connected. This notion is further supported by a three-dimensional variability analysis, which showed that ECD undulations coincide with RGS7-Gβ5 binding (movies S1 and S2). Along with that, the GPR158 CT-CC and

the RGS domain of RGS7 show a higher B factor, reflecting conformational dynamics at these regions (fig. S4D). Cross-linking mass spectroscopy (XL-MS) confirmed the main contact points in the complex at the single amino acid level (fig. S5).

Detailed analysis of the 7TM domain of GPR158 reveals a compact dimeric interface (Fig. 2A and fig. S6A). Superposition of the two protomer 7TM regions shows a similar arrangement of all the elements including extra- and intracellular loops with RMSD value of 0.35 Å (fig. S6B). Within each protomer, ECL2 Cys[573] forms a disulfide bond with TM3 Cys[481], an interaction conserved throughout GPCRs. The β hairpin of ECL2 interacts with the stalk-TM linker through hydrophobic interactions, bridging the ectodomain and TM domain of each protomer. ECL2 also interacts with ECL1 and TM3 hydrophobically and with ECL3 through polar interactions (Fig. 2B).

The dimer interface at the 7TM domain can be separated into two parts. The first is formed at the extracellular end by TM4, TM5, and ECL2 of each protomer. The helices assemble in an inverted V-shape, creating a cavity. The second interaction at the cytoplasmic end closes the cavity (Fig. 2C and fig. S6C). This type of dimeric architecture has not been seen in other GPCRs, including class C receptors (fig. S6D).

The extracellular portion of the TM interface formed by TM4, TM5, and ECL2 features a series of hydrophilic interactions and hydrophobic contacts (Fig. 2C). The cytoplasmic portion of the TM interface is formed by TM3 and ICL2 of both the protomers. The ICL2 regions interact with each other, and residues in TM3 form a basic patch that engages phospholipids (fig. S6, C and E). The polar contacts between Gln[516] across protomers also form the lid that covers the cavity from the intracellular end (II in Fig. 2C). The cavity formed at the dimer interface is further shielded at both its intra- and extracellular ends by several cholesterol molecules from both the front and back sides (fig. S6A). The TM domain interaction with cholesterols may stabilize the dimeric interface.

We observed densities for two phospholipids at the cavity formed by the TM dimeric interface and performed mass spectrometry to identify lipids copurified with the GPR158 complex (fig. S6F). Although not exhaustive, this lipidomics analysis identified several phospholipid species, most notably phosphatidylethanolamine (PE) and phosphatidylinositol (PI). Both molecules are well accommodated into the respective cryo-EM densities (Fig. 2A and fig. S3). The phospholipids are arranged in the cavity in such a way that one hydrophobic tail of each lipid is inserted into the shallow hydrophobic pocket
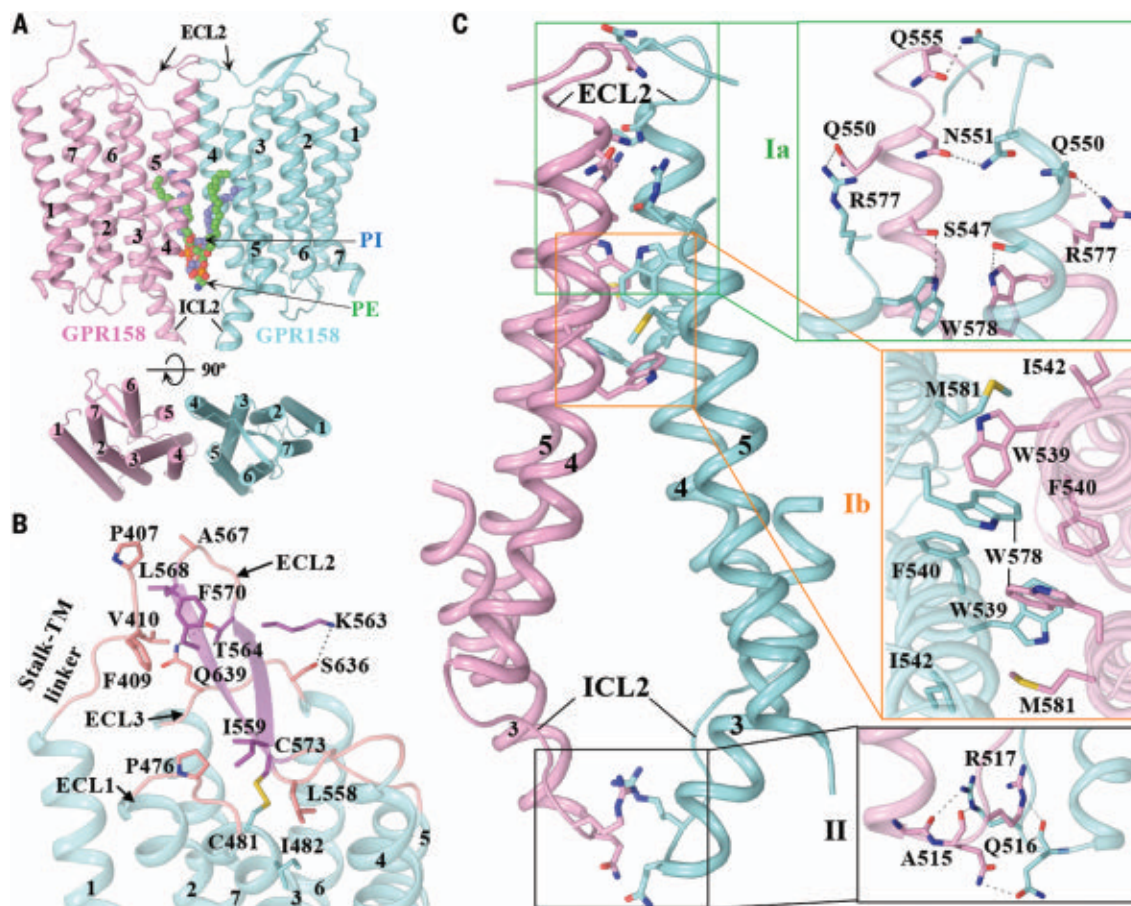
Fig. 2. Organization of GPR158 transmembrane domain and its homodimer interface. (**A**) Overall arrangement of the 7TM region of GPR158 protomers, in side and top views. Phospholipids PE and PI are identified at the cavity formed by the TM dimeric interface are shown as spheres; 7TM protomers and phospholipids are colored as in Fig. 1. (**B**) Close-up view of the extracellular loop region. ECL2 caps the extracellular pocket by interacting with TM3, ECL1, and ECL3 residues. ECL2 Cys[573] preserves the conserved disulfide bond with TM3 Cys[481]. ECL2 is also stabilized by interaction with a stalk-TM linker that connects the ectodomain with 7TM. (**C**) The 7TM dimer interface is formed at two sites (I and II), the extra- and intracellular sides. Direct contacts at the extracellular side are formed by TM4, TM5, and ECL2 of both protomers; contacts are shown at right. The intracellular-side interface is formed by TM3 and ICL2 of both protomers; contacts formed by ICL2 are shown at right. Amino acid abbreviations: A, Ala; C, Cys; E, Glu; F, Phe; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp.

created by TM3, TM4, and TM5, stabilizing the dimer interface. The hydrophobic tails of PE and PI are stacked against the side chains of bulky hydrophobic residues, and head groups occupy basic clusters at the interface (fig. S6, G and H). Comparisons with other class C GPCR structures show both similarities and differences in 7TM organization (fig. S7).

Overall, the distinct dimeric arrangement of the TM domain that forms an extensive web of interactions, and the additional stability conferred by phospholipid and cholesterol molecule interactions, make the TM dimeric interface more compact and unlikely to be compatible with G protein activation. Indeed, our functional studies of GPR158 show no constitutive activity in any of the G protein signaling assays (fig. S8). Furthermore, mutagenesis at the dimerization interface aimed at increasing the dynamics by mimicking sub-

stitutions that activate other class C GPCRs (20, 21) also failed to unlock the constitutive activity of GPR158 (fig. S8 and table S2). We conclude that the GPR158 7TM domain has an organization of the dimeric interface that locks it in a conformation that prevents constitutive G protein activation.

The ECDs of two protomers interact with each other side-by-side, forming cross-subunit contacts. The ectodomain of each protomer is composed of the N-terminal α-helical region, a prominently folded central domain, and a C-terminal stalk region (Fig. 3A). Superposition of the two ECDs showed substantial structural similarity with an RMSD of 1.37 Å (fig. S9A). The largest portion of the GPR158 ectodomain, comprising 412 amino acids, lacks sequence similarity with the ectodomains of other GPCRs. The GPR158 structure reveals a well-defined globular domain consisting of six antiparallel

β sheets flanked by two α helices (Fig. 3B). A structural homology search revealed that it shares a similar fold to Cache domains present in numerous proteins but not previously found in GPCRs. The β sheets of the Cache domain are curved and form an amphipathic pocket analogous to a ligand-binding site in the Cache domains of other proteins (Fig. 3C). The helices α1 and α2 assemble at the back of the β sheets' core structure, stabilizing the cavity by hydrophobic interactions with β-sheet residues. The density for the third helix, α3, at its respective position is not well resolved to correctly interpret. The α3 likely caps the putative ligand-binding pocket from one side while being connected by flexible loops (β3α3 and α3β4) (Fig. 3B). The density surrounding the binding pocket is not uniformly resolved, likely reflecting its dynamic nature. It possibly becomes ordered upon ligand binding. The Cache
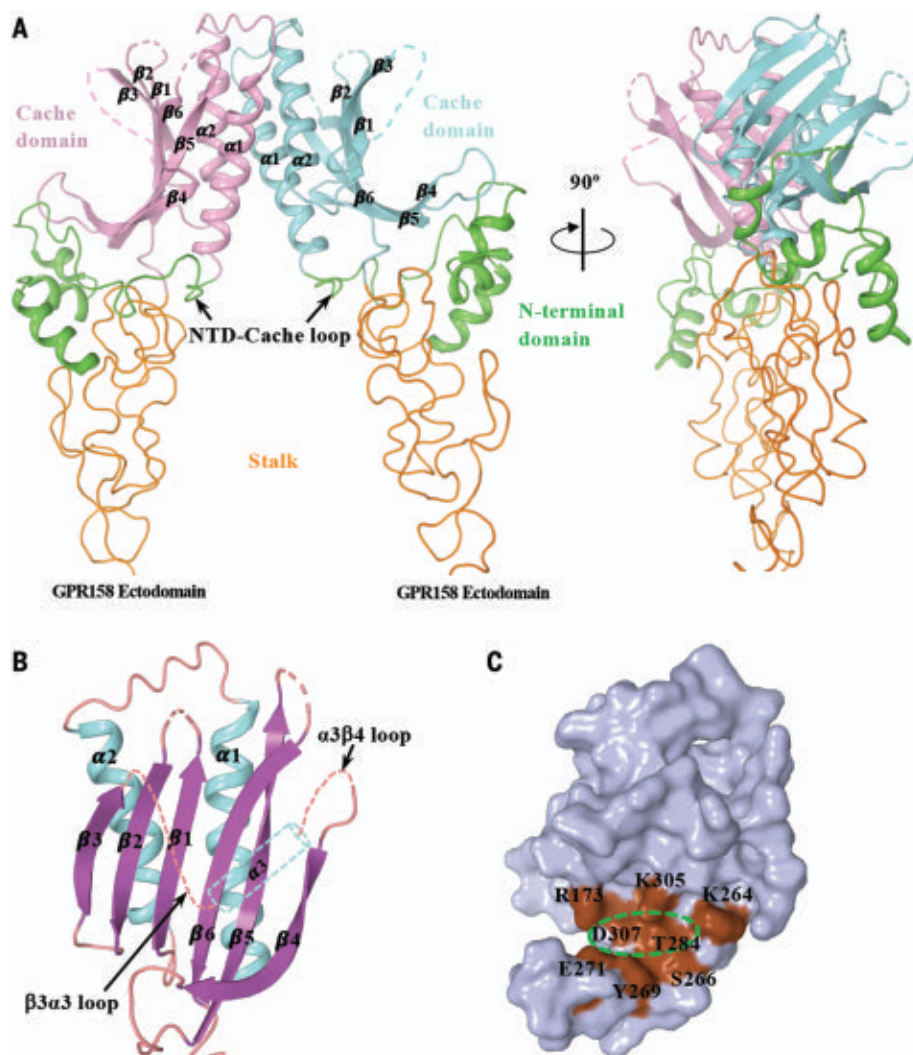
Fig. 3. A distinctive organization of the GPR158 ectodomain featuring the Cache domain. (**A**) Side view of GPR158 ectodomain consisting of N-terminal domain (NTD), Cache domain, and stalk region. The GPR158 ectodomain forms a dimeric interface with the Cache domain. (**B**) The Cache domain is composed of six antiparallel β sheets flanked by α helices. The density for the α3 helix is not well resolved and is represented as a dashed cylinder at the respective position. The missing flexible loops in the model are shown as other dashed lines. (**C**) Cache domain putative ligand-binding pocket. The curved sheets form a putative ligand-binding pocket (dashed oval), equivalent to the prokaryotes' extracellular Cache domain ligand-binding site, which generates an amphipathic environment. The putative ligand-interacting residues are shown in brown. However, densities for most of the side chains of pocket residues are not well resolved and have a high *B* factor. The putative ligand-binding pocket is possibly capped by the dynamic α3 helix from one side.

domain of GPR158 has the highest structural homology with the ligand-sensing extracellular Cache domains of the bacterial chemoreceptors and histidine kinase receptors (fig. S9B).

The Cache domain is connected through the NTD-Cache loop to the N-terminal domain (NTD), which has three helices. The NTD-Cache loop contains Cys[99], which forms a disulfide bond with Cys[272] of the Cache domain β4β5 loop. Interactions with the NTD-Cache loop and N-terminal region likely provide additional stability to the Cache domain (fig. S9C).

The C terminus of the Cache domain is connected to the stalk region that contains flexible loops and has a very weak density map. The stalk region is cysteine-rich, and its 10 cysteines may engage in intradomain disulfide bridges (fig. S9D). A structurally similar cysteine-rich domain (CRD) is found in class C GPCRs, which suggests that it likely plays a similar role in receptor activation (*21, 22*). Following the stalk region is a stalk-TM linker that hydrophobically interacts with ECL2 and connects to TM1 of the transmembrane domain.

One of the most prominent features of the GPR158 ectodomain is the dimerization of the Cache domains; this occurs through helices α1 and α2 of each Cache domain, which cross at an angle to create a four-helix bundle at the dimer interface (fig. S9E). The loops connecting helices α1 and α2 are also likely involved in intersubunit interaction. The dimer interface is stabilized by an extensive network of hydrophobic and hydrophilic interactions, with a buried surface area of 2178.3 Å^2 (fig. S9F).

The structure of the GPR158 homodimer bound to the RGS7-Gβ5 complex reveals several key insights into RGS docking and regulation by GPCRs. Overall, we observe an asymmetric assembly involving two GPR158 protomers interacting with one RGS7-Gβ5 complex (Fig. 1B). The two docking sites for RGS7-Gβ5 on GPR158 are both created by the dimerization of GPR158. Upon RGS7 binding to GPR158, the intracellular C-terminal helices of both GPR158 protomers, which were disordered in the GPR158-apo structure, stack into a coiled-coil configuration (CT-CC). The side-chain densities are not well resolved for CT-CC residues and have a high *B* factor (fig. S4D). The CT-CC domain contributes a third contact point for the dimerization of GPR158 in the complex, which is predicted to be held together mainly by hydrophobic contacts and stabilized by ionic and polar bonding (fig. S10, A to E). The contacts with GPR158 are made exclusively by RGS7, with no direct interactions involving Gβ5.

The primary binding site (site I) is formed by CT-CC (Fig. 4A), which potentially engages in an extensive web of hydrophobic and hydrophilic interactions with the RGS7 DEP-DHEX domain (fig. S10, F to H). One α helix is engaged in the interactions with the Eα1, Eα3, and Eα4 helices and the Eα3Eα4 and DEP-DHEX loops, whereas the other α helix interacts with the Eα3 helix and the Eα2Eα3 and DEP-DHEX loops. In addition, we observe insertion of the C-terminal loop in one of the GPR158 protomers into the pocket created by the DEP Dα1 and DHEX Eα4 helices and the β-hairpin loop (fig. S10H). The second docking site (site II) for RGS7-Gβ5 is provided by the intracellular portion of the 7TM dimerization interface. However, the contacts at this site are made only with one GPR158 protomer. The interaction involves TM3, TM5, and ICL3 of the GPR158 7TM domain and the Eα1, Eα2, and Eα3 helices and the Eα1Eα2 loop of the DHEX domain (Fig. 4A). The interactions at this interface are mainly hydrophobic, stabilized by hydrogen bonding at the periphery (fig. S10I). Comparison with the apo structure shows that RGS7-Gβ5 binding results in pulling the GPR158 protomers apart and remodeling the interface to accommodate RGS7 (Fig. 4B).
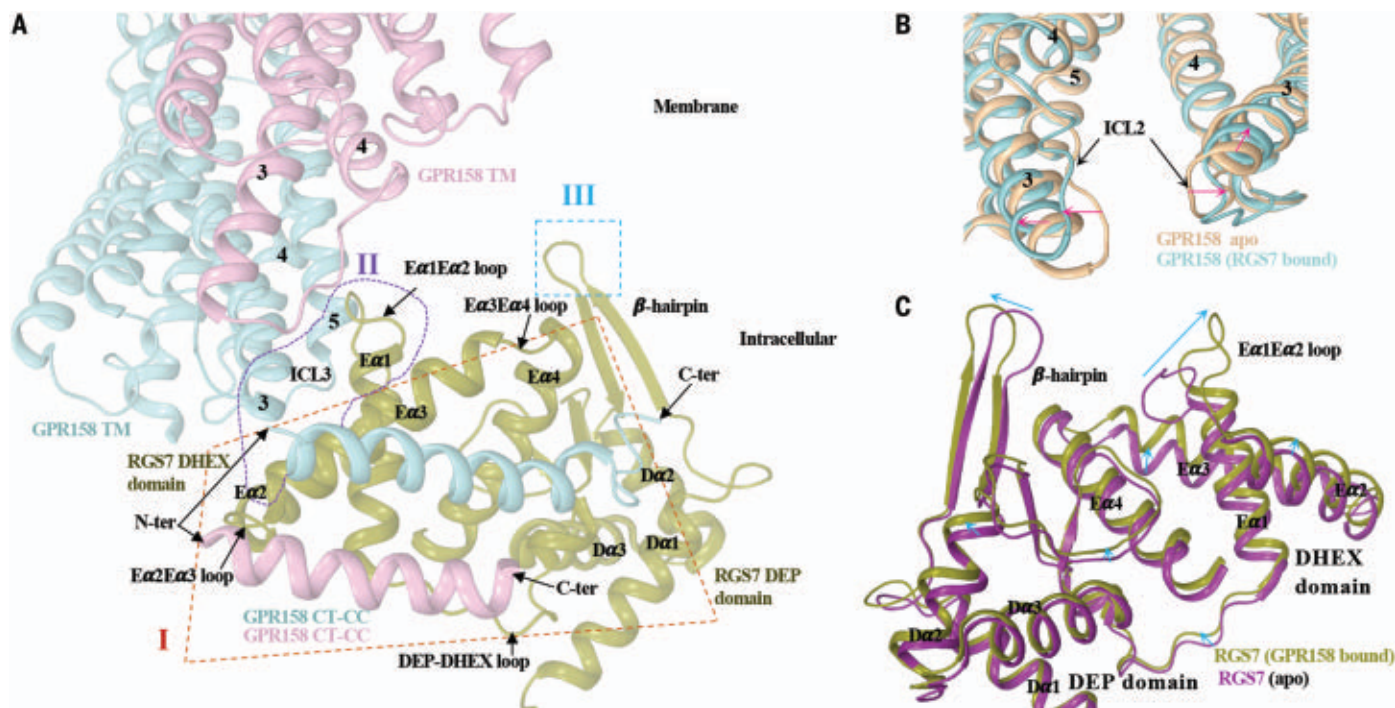
Fig. 4. Mechanism of GPR158 interaction with RGS7. (A) RGS7 forms two distinct binding sites (I and II) on GPR158, both created by the dimerization of GPR158. The first binding interface is formed between GPR158 CT-CC and the RGS7 DEP-DHEX domain. The second interface is formed by GPR158 7TM and the RGS7 DHEX domain. In addition, the β-hairpin loop of RGS7 is inserted into the membrane (shown as III) and could facilitate the orientation of RGS7 toward the membrane. **(B and C)** Conformational rearrangement on the GPR158 TM dimeric interface and RGS7 upon complex formation. The TM3 of one protomer shifts toward the 7TM core while dissociating it from another protomer to accommodate RGS7 at the interface (B). A large conformation shift at the Eα1Eα2 loop of the RGS7 DHEX domain follows, along with rearrangement of the β-hairpin loop and helices that shift up toward the membrane (C).

The GPR158-RGS7 interactions at the second site are weaker than the binding to CT-CC, are likely transient in nature, and involve remodeling of both GPR158 and RGS7-Gβ5. Comparison of the apo GPR158 structure with the structure of the GPR158-RGS7-Gβ5 complex reveals rearrangement of the cytoplasmic end of TM3 and ICL2 of one 7TM, shifting inward toward its 7TM core upon RGS7 binding. This is accompanied by destabilization of the TM3 C terminus and ICL2 at the other 7TM protomer affecting the 7TM dimerization interface (Fig. 4B). Comparison of the apo RGS7-Gβ5 crystal structure (*23*) with the complex structure also shows substantial reorganization of the RGS7 DEP-DHEX domain upon binding to GPR158 (Fig. 4C). Most prominently, the dynamic loop Eα1Eα2 moves upward and is stabilized by interacting with TM3, TM5, and ICL3. Along with this, DHEX helices Eα1, Eα2, and Eα3 move upward to interact with the TM domain and the DEP-DHEX loops move upward to interact with GPR158 CT. The β-hairpin loop of RGS7 moves upward so that its hydrophobic tip is inserted into the membrane, allowing Dα2 helix rearrangement in the DEP domain (Fig. 4B). These rearrangements collectively generate favorable complementary electrostatic surfaces on the GPR158-RGS7 interface. This likely serves to orient the RGS7-Gβ5 complex toward the membrane (fig. S10, J and K).

RGS7 recruitment is reminiscent of GPCR interactions with signal transducers. Indeed, the RGS binding surface on GPR158 substantially overlaps with the GPCR surface that binds heterotrimeric G proteins and β-arrestin (fig. S11, A and B). Our modeling, using structures of Gα complexes with diverse GPCRs, shows that the RGS7 DHEX domain occludes the G protein binding site from the TM3 and TM5 side where the α5 helix of Gα inserts into the 7TM central cavity and creates steric clashes with the Ras domain of Gα subunits (fig. S11, C to E). Thus, recruitment of RGS7-Gβ5 would preclude GPR158 from productively interacting with G proteins, supporting lack of G protein activation (fig. S8). We further detect bidirectional allosteric effects resulting from the GPR158-RGS7 binding similar to what is observed upon GPCR-Gα interaction. These include inward shift of the cytoplasmic end of TM3, as seen in the GABA$_B$-G$_I$ structure and modulation of the ligand-binding ectodomain upon RGS7 binding. The interaction of GPR158 with the RGS7-Gβ5 complex is quite distinct, and mutagenesis at the GPR158 dimerization interface that constitutively activates class C GPCRs failed to change RGS activity in the absence of a ligand (fig. S12).

To further investigate the conformational dynamics resulting from RGS7-Gβ5 recruitment to GPR158, we performed biochemical experiments. First, we studied the impact of binding to a synthetic C-terminal peptide that comprises the CT-CC module by gel filtration. Complexing with this peptide was sufficient to induce a large change in the hydrodynamic behavior of the RGS7-Gβ5 complex, consistent with substantial conformational changes in RGS7-Gβ5 upon binding to GPR158 (fig. S13A). We further refined these investigations using hydrogen/deuterium exchange mass spectrometry (HDX-MS), which showed that the C-terminal peptide induced substantial changes in solvent accessibility within the DEP-DHEX domain, specifically in the Dα1, Eα3, and Eα4 helices and the β-hairpin, Eα3Eα4, and DEP-DHEX loops of RGS7 (fig. S13, B to D).

In this work, we present high-resolution structures of an unusual receptor assembly that involves an orphan GPCR complexed with a signaling regulator, RGS protein. The RGS protein binds the same elements that GPCRs use for engaging their signal transducers: G proteins and β-arrestins. In the present

structure, we observe constitutive engagement of the RGS7-Gβ5 complex by GPR158 in the absence of a G protein. We speculate that binding of a ligand to the ECD would activate GPR158 rearrangement of the cytoplasmic domains that engage RGS to alter its activity. Given that RGS binding precludes GPR158 from canonical activation of G proteins, one can describe it as an RGS-coupled receptor.

In addition to providing information on the GPCR-RGS structure, we show the role of two phospholipids in organizing the dimerization interface of GPCRs. These lipids staple the protomers and provide intriguing possibilities for GPCR modulation. We also identify a Cache domain, raising the possibility that GPR158 detects a small-molecule ligand that could regulate the RGS module—an avenue to be explored in future studies. We hope our findings will spur further progress in understanding the regulatory and signaling mechanisms of GPR158 by facilitating the structure-based discovery of its ligands and by guiding the exploration of GPR158-mediated control of RGS proteins in the endogenous neuronal setting.

REFERENCES AND NOTES

1. A. S. Hauser et al., Cell 172, 41–54.e19 (2018).
2. K. Sriram, P. A. Insel, Mol. Pharmacol. 93, 251–258 (2018).
3. A. G. Gilman, Annu. Rev. Biochem. 56, 615–649 (1987).
4. A. K. Shukla, K. Xiao, R. J. Lefkowitz, Trends Biochem. Sci. 36, 457–469 (2011).
5. S. Hollinger, J. R. Hepler, Pharmacol. Rev. 54, 527–559 (2002).
6. E. M. Ross, T. M. Wilkie, Annu. Rev. Biochem. 69, 795–827 (2000).
7. M. Abramow-Newerly, A. A. Roy, C. Nunn, P. Chidiac, Cell. Signal. 18, 579–591 (2006).
8. D. R. Ballon et al., Cell 126, 1079–1093 (2006).
9. O. Civelli, Neuron 76, 12–21 (2012).
10. P. A. Insel et al., Mol. Pharmacol. 88, 181–187 (2015).
11. C. Orlandi et al., J. Cell Biol. 197, 711–719 (2012).
12. N. Patel et al., PLOS ONE 10, e0117758 (2015).
13. L. Khrimian et al., J. Exp. Med. 214, 2859–2873 (2017).
14. L. P. Sutton et al., eLife 7, e33273 (2018).
15. H. Engqvist et al., BMC Cancer 19, 928 (2019).
16. G. Condomitti et al., Neuron 100, 201–215.e9 (2018).
17. C. Song, C. Orlandi, L. P. Sutton, K. A. Martemyanov, J. Biol. Chem. 294, 13145–13157 (2019).
18. C. Orlandi et al., J. Biol. Chem. 290, 13622–13639 (2015).
19. C. Orlandi, L. P. Sutton, B. S. Muntean, C. Song, K. A. Martemyanov, Neuropsychopharmacology 44, 642–653 (2019).
20. J. Park et al., Nature 584, 304–309 (2020).
21. A. Koehl et al., Nature 566, 79–84 (2019).
22. J. Du et al., Nature 594, 589–593 (2021).
23. D. N. Patil et al., eLife 7, e42150 (2018).
24. Y. Perez-Riverol et al., Nucleic Acids Res. 47, D442–D450 (2019).

CELL AND GENE THERAPY

CAR T cells produced in vivo to treat cardiac injury

Joel G. Rurik[1,2,3], István Tombácz[4]†, Amir Yadegari[4]†, Pedro O. Méndez Fernández[1,2,3], Swapnil V. Shewale[2], Li Li[1,2], Toru Kimura[4]‡, Ousamah Younoss Soliman[4], Tyler E. Papp[4], Ying K. Tam[5], Barbara L. Mui[5], Steven M. Albelda[4,6], Ellen Puré[7], Carl H. June[6], Haig Aghajanian[1,2,3]*, Drew Weissman[4]*, Hamideh Parhiz[4]*, Jonathan A. Epstein[1,2,3,4]*

Fibrosis affects millions of people with cardiac disease. We developed a therapeutic approach to generate transient antifibrotic chimeric antigen receptor (CAR) T cells in vivo by delivering modified messenger RNA (mRNA) in T cell–targeted lipid nanoparticles (LNPs). The efficacy of these in vivo–reprogrammed CAR T cells was evaluated by injecting CD5-targeted LNPs into a mouse model of heart failure. Efficient delivery of modified mRNA encoding the CAR to T lymphocytes was observed, which produced transient, effective CAR T cells in vivo. Antifibrotic CAR T cells exhibited trogocytosis and retained the target antigen as they accumulated in the spleen. Treatment with modified mRNA-targeted LNPs reduced fibrosis and restored cardiac function after injury. In vivo generation of CAR T cells may hold promise as a therapeutic platform to treat various diseases.

Cardiac fibroblasts become activated in response to various myocardial injuries through well-studied mechanisms including transforming growth factor β–SMAD2/3, interleukin-11, and other interactions with the immune system (1–6). In many chronic heart diseases, these fibroblasts fail to quiesce and secrete excessive extracellular matrix, resulting in fibrosis (7). Fibrosis both stiffens the myocardium and negatively affects cardiomyocyte health and function (8). Despite in-depth understanding of activated cardiac fibroblasts, clinical trials of antifibrotic therapeutics have only demonstrated a modest effect (5, 7) at best. Furthermore, these interventions aim to limit fibrotic progression and are not designed to remodel fibrosis once it is established. To address this substantial clinical problem, we recently demonstrated the use of chimeric antigen receptor (CAR) T cells to specifically eliminate activated fibroblasts as a therapy for heart failure (9). Elimination of activated fibroblasts in a mouse

[1]Department of Cell and Developmental Biology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [2]Penn Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [3]Institute for Regenerative Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [5]Acuitas Therapeutics, Vancouver, British Columbia V6T 1Z3, Canada. [6]Center for Cellular Immunotherapies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Biomedical Sciences, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA.
*Corresponding author. Email: haig@pennmedicine.upenn.edu (H.A.); dreww@pennmedicine.upenn.edu (D.W.); parhizh@pennmedicine.upenn.edu (H.P.); epsteinj@upenn.edu (J.A.E.)
†These authors contributed equally to this work.
‡Present address: Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan.
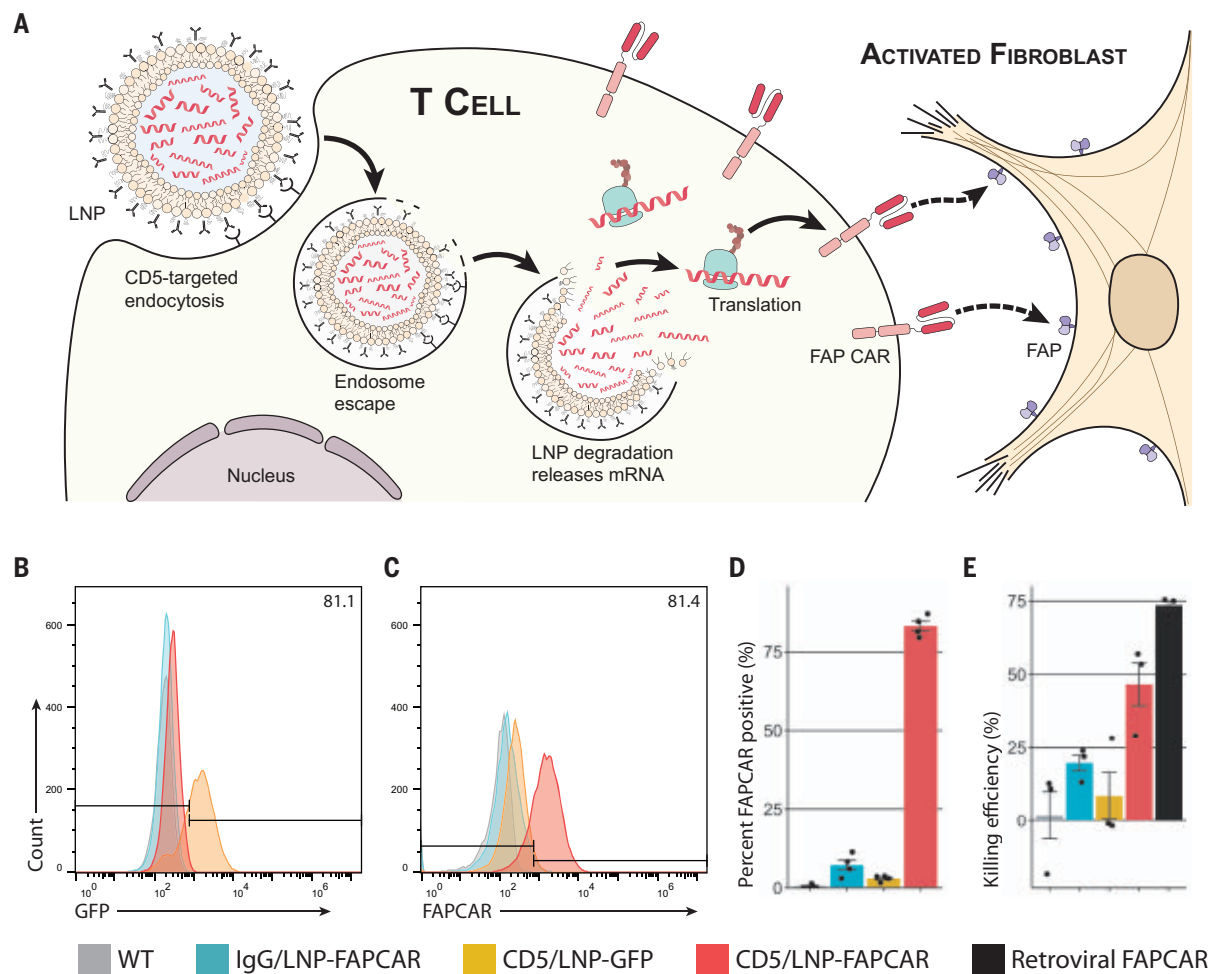
Fig. 1. CD5-targeted LNPs produce functional, mRNA-based FAPCAR T cells in vitro. (**A**) Schematic outlining the molecular process to create transient FAPCAR T cells using CD5-targeted LNPs. (**B** and **C**) Representative flow cytometry analysis of GFP (B) and FAPCAR (C) expression in murine T cells 48 hours after incubation with IgG/LNP-FAPCAR, CD5/LNP-GFP, or CD5/LNP-FAPCAR. (**D**) Quantification of murine T cells (percentage) staining positive for FAPCAR from biologically independent replicates ($n = 4$). (**E**) FAPCAR T cells were mixed with FAP-expressing target HEK293T cells overnight and assayed for killing efficiency in biologically independent replicates ($n = 3$). Data are shown as mean ± SEM.

model of heart disease resulted in a significant reduction of cardiac fibrosis and improved cardiac function (*9*). One caveat of that work is the indefinite persistence of engineered T cells similar to CAR T cell therapy currently used in the oncology clinical setting (*10*). Fibroblast activation is part of a normal wound-healing process in many tissues, and persistent antifibrotic CAR T cells could pose a risk in the setting of future injuries. Therefore, we leveraged the power of nucleoside-modified mRNA technology to develop a transient antifibrotic CAR T therapeutic.

Therapeutic mRNAs can be stabilized by the incorporation of modified nucleosides, synthetic capping, and the addition of lengthy poly-A tails, and can be enhanced with codon optimization (*11–13*). 1-Methylpseudouridine integration also boosts translation (*13, 14*). Direct introduction of mRNA into T cells

ex vivo by electroporation has been used successfully by our group and others to make CAR T cells (*15*); however, this process carries significant cost and risk and requires extensive infrastructure. Thus, we developed an approach that could be used to avoid removing T cells from the patient by packaging modified mRNAs in lipid nanoparticles (LNPs) capable of producing CAR T cells in vivo after injection. LNP-mRNA technology underlies recent successes in COVID-19 vaccine development and holds exceptional promise for additional therapeutic strategies (*16–20*). Once in the body, mRNA-loaded LNPs, absent of any specific targeting strategies, are endocytosed by various cell types (especially hepatocytes if injected intravenously) (*21, 22*). Shortly after cellular uptake, the mRNA escapes the endosome, releasing the mRNA into the cytoplasm, where it is transiently transcribed

before degrading (*11*). Targeting antibodies can be decorated on the surface of LNPs to direct uptake (and mRNA expression) to specific cell types (*23, 24*). We hypothesized that an LNP directed to T lymphocytes could deliver sufficient mRNAs to produce functional CAR T cells in vivo (Fig. 1A). Because mRNA is restricted to the cytoplasm and is incapable of genomic integration, intrinsically unstable, and diluted during cell division, these CAR T cells will be, by design, transient.

We generated modified nucleoside-containing mRNA encoding a CAR designed against fibroblast activation protein (FAP) (a marker of activated fibroblasts) and packaged it in CD5-targeted LNPs (referred to as "targeting antibody/LNP-mRNA cargo" or CD5/LNP-FAPCAR) (Fig. 1A) (*9, 25*). CD5 is naturally expressed by T cells and a small subset of

Fig. 2. CD5-targeted LNPs produce mRNA-based FAPCAR T cells in vivo. (**A**) Luciferase activity in CD3$^+$ splenocytes 24 hours after intravenous injection of 8 µg of control IgG/LNP-Luc or CD5/LNP-Luc. Bar graphs represent two biologically independent replicates. (**B**) Ai6 mice (Rosa26$^{CAG-LSL-ZsGreen}$) were injected with saline or 30 µg of either IgG/LNP-Cre or CD5/LNP-Cre. After 24 hours, ZsGreen expression was observed in 81.1% of CD4$^+$ splenocytes and in 75.6% of CD8$^+$ splenocytes, but in only 15.0% of CD3$^-$ splenocytes. Bar graphs represent two biologically independent replicates. (**C**) T cells were isolated from the spleens of AngII/PE–injured mice 48 hours after injection of 10 µg of LNPs. Representative flow cytometry analysis shows FAPCAR expression in animals injected with CD5/LNP-FAPCAR but not in control saline, IgG/LNP-FAPCAR, or CD5/LNP-GFP animals. (**D**) Quantification of murine T cells staining positive for FAPCAR in (C). *n* = 4 biologically independent mice in two separate cohorts. Data are shown as mean ± SEM.

(CD5/LNP-Luc) were found to express abundant luciferase activity in their splenic T cells, whereas mice injected with isotype control (nontargeting) IgG/LNP-Luc did not (Fig. 2A). Bioluminescence imaging demonstrated spleen targeting only in CD5/LNP-Luc–treated animals (fig. S2A). Liver expression of LNP-delivered mRNA was observed in both CD5/LNP-Luc– and IgG/LNP-Luc–treated animals, as expected mainly due to normal hepatic clearance of LNPs, as reported previously (*22, 24*). In another experiment, CD5/LNPs were loaded with mRNA encoding Cre recombinase (CD5/LNP-Cre) and injected into Ai6 Cre-reporter mice (Rosa26$^{CAG-LSL-ZsGreen}$). We found evidence of genetic recombination (ZsGreen expression) specifically in CD3$^+$ T cells (both CD4$^+$ and CD8$^+$ subsets) from CD5/LNP-Cre–injected animals but little evidence of Cre recombinase activity in CD3$^-$ (non-T) cells (mainly representing B cells, dendritic cells, and macrophages) or in IgG/LNP-Cre–injected mice (Fig. 2B). We next investigated whether targeted LNPs could deliver FAPCAR mRNA (CD5/LNP-FAPCAR) to T cells in an established murine hypertensive model of cardiac injury and fibrosis produced by constant infusion of angiotensin II/phenylephrine (AngII/PE) through implanted 28-day osmotic mini-pumps (*9, 28*). Mice were injured for 1 week to allow fibrosis to be established before injecting CD5/LNP-FAPCAR (*9*). Forty-eight hours after LNP injection, we found a consistent population of FAPCAR$^+$ T cells (17.5-24.7%) exclusively in mice that received CD5/LNP-FAPCAR (Fig. 2, C and D, and fig. S2B). By contrast, nontargeted (IgG/LNP-FAPCAR) and targeted LNPs containing GFP (CD5/LNP-GFP) did not produce FAPCAR T cells (Fig. 2, C and D, and fig. S2B). We observed FAPCAR expression in each major T cell subset with a slight enrichment in CD4$^+$ T cells above their prevalence in the spleen (of all FAPCAR T$^+$ cells, 87% were CD4$^+$ and 9 to 10% CD8$^+$, with most of both classes portraying a naïve phenotype; 25 to 37% of regulatory T cells were FAPCAR$^+$; fig. S2C and table S1). A mixture of CAR$^+$ T cell subtypes has been shown to benefit CAR effectiveness (*29*). We did not observe significant FAPCAR expression in splenic B cells or natural killer cells (fig. S2C). No FAPCAR expression was found in splenic T cells 1 week after injection, demonstrating the transient nature of FAPCAR expression in this model (table S1).

CAR T cell therapy has previously been associated with a process called trogocytosis, in which lymphocytes extract surface molecules through the immunological synapse from antigen-presenting cells (*30–32*) (Fig. 3A). We sought to determine whether FAPCAR T cells produced either in vivo with CD5/LNP-FAPCAR mRNA or adoptively transferred ex vivo virally engineered CAR T cells exhibit evidence of
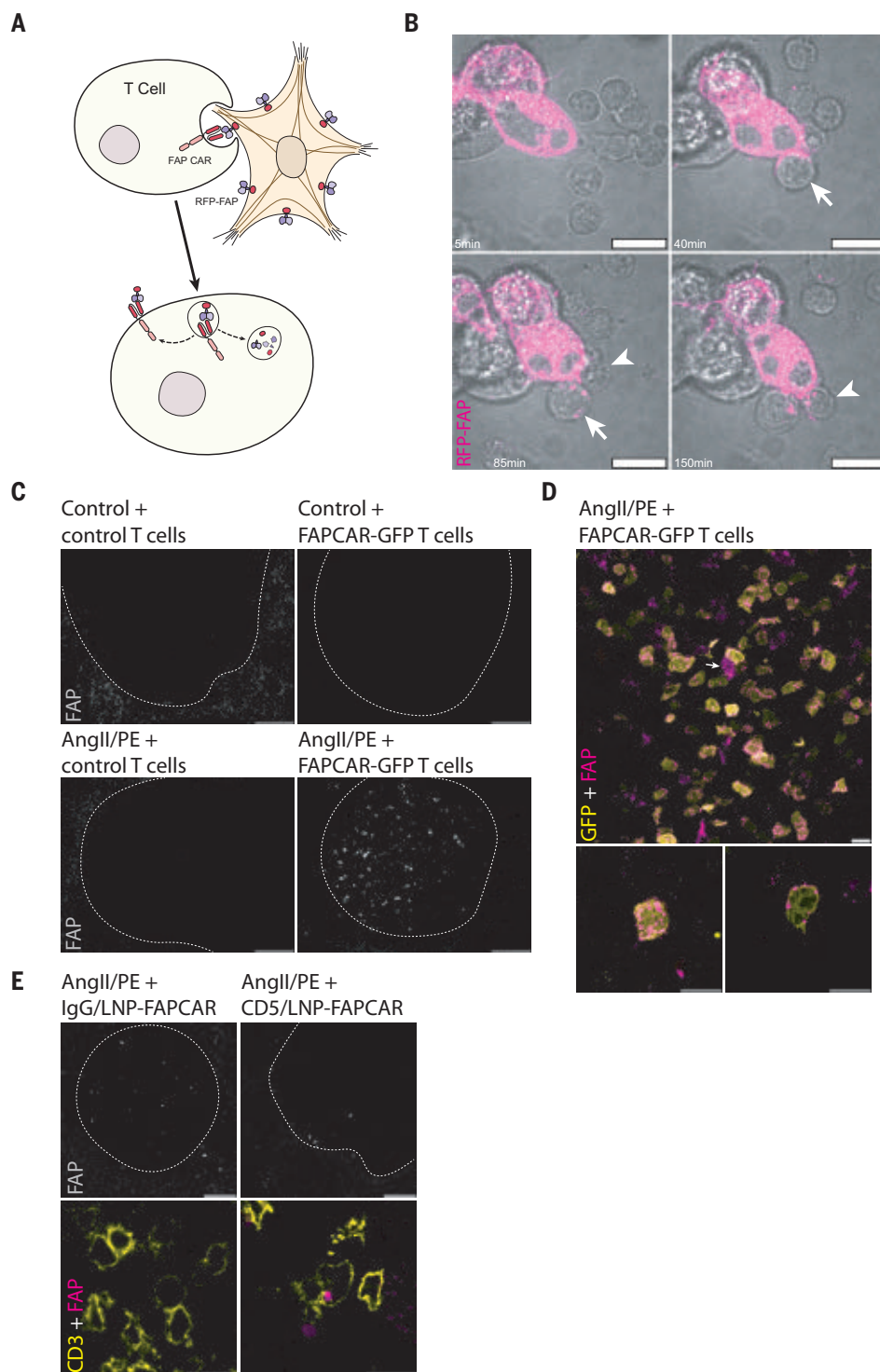
B cells and is not required for T cell effector function (*26, 27*). As a first proof-of-concept experiment, we incubated CD5/LNPs containing modified mRNA encoding either FAPCAR or green fluorescent protein (GFP) with freshly isolated, activated murine T cells in vitro for 48 hours. CD5-targeted LNPs delivered their mRNA cargo to most T cells in culture, where 81% expressed GFP after exposure to CD5/LNP-GFP (Fig. 1B) and 83% expressed FAPCAR after exposure to CD5/LNP-FAPCAR (Fig. 1, C and D), as measured by flow cytometry (fig. S1A). In vitro, CAR expression peaks at 24 hours and rapidly abates over the ensuing days (fig. S1B). LNPs decorated with isotype control [immunoglobulin G (IgG)] antibodies,

and thus not explicitly directed to lymphocytes, were only able to deliver mRNA to a small fraction (7%) of T cells in vitro (Fig. 1, C and D). These LNP-generated CAR T cells were able to effectively kill FAP-expressing target cells in vitro (Fig. 1E) in a dose-dependent manner (fig. S1C) similar to virally engineered FAPCAR T cells. Gene transfer through targeted LNPs in vitro is also possible and efficient (89 to 93%) in human T cells, as demonstrated by targeting ACH2 cells with CD5/LNP-GFP (fig. S1D).

We next assessed whether CD5-targeted LNP mRNA could also efficiently reprogram T cells in vivo. Mice that were intravenously injected with CD5/LNPs containing luciferase mRNA

Fig. 3. FAPCAR T cells trogocytose FAP from activated cardiac fibroblasts and return FAP to the spleen only in AngII/PE–injured, FAPCAR T cell–treated animals.
(**A**) Schematic representation of FAPCAR-expressing T cells trogocytosing FAP from activated fibroblasts. (**B**) Confocal time-lapse micrographs of two FAPCAR T cells first forming an immunological synapse at 40 min (arrow) and 85 min (arrowhead) and then trogocytosing RFP-FAP (magenta) from HEK293T cells [punctae can be seen at 85 min (arrow) and at 150 min (arrowhead) within FAPCAR T cells]. Scale bars, 10 μm. (**C**) Wide-field images of FAP-stained spleens (white pulp regions highlighted by the dashed line) of an uninjured animal 24 hours after adoptive transfer of 10^7 MigR1-control T cells, an uninjured animal 24 hours after adoptive transfer of 10^7 FAPCAR-GFP T cells, an AngII/PE–injured (7 days) animal 48 hours after adoptive transfer of 10^7 MigR1-control T cells, and an AngII/PE–injured (7 days) animal 48 hours after adoptive transfer of 10^7 FAPCAR-GFP T cells. Scale bars, 100 μm. (**D**) Confocal micrograph of FAP (magenta) and FAPCAR-GFP (yellow) in a white pulp region of the spleen of an AngII/PE–injured (7 days) animal 48 hours after adoptive transfer of 10^7 FAPCAR-GFP T cells. Shown are a maximum Z projection (lower left subpanel) and a single Z slice (lower right subpanel) of a representative FAP⁺/FAPCAR⁺ T cell. Scale bars, 10 μm. (**E**) Confocal micrographs of a white pulp region (dashed outline) of FAP-stained spleens from AngII/PE–injured (7 days) animals injected with 10 μg of IgG/LNP-FAPCAR or CD5/LNP-FAPCAR for 48 hours. FAP (gray and magenta) and CD3 (yellow) overlap specifically in the CD5/LNP-FAPCAR–treated condition. Scale bars, 100 μm (top row, grayscale) or 10 μm (bottom row, merged pseudocolored).

trogocytosis as further support that functional FAPCAR T cells are produced in situ. First, we mixed retrovirus-engineered FAPCAR T cells with human embryonic kidney (HEK) 293T cells overexpressing red fluorescence protein (RFP)–tagged FAP in vitro and observed trogocytosis with live-imaging confocal microscopy (Fig. 3B and movie S1). Immunofluorescence analysis of spleens from AngII/PE–injured animals treated with adoptively transferred, virally transduced GFP-tagged FAPCAR T cells revealed extensive FAP staining in the white pulp regions of the spleen, which was not seen in injured animals treated with control T cells or in uninjured animals (Fig. 3C and fig. S3). The FAP⁺ cells in the spleens of injured and treated animals co-stained for GFP, which indicates that they were transduced cells (Fig. 3D). Furthermore, the FAP staining appeared as cytoplasmic punctae consistent with trogocytosis (Fig. 3D). We observed some rare FAP⁺/GFP⁻ cells in the spleens of injured, treated animals that were not observed in the controls (Fig. 3D, arrow). CD3⁺ lymphocytes

Fig. 4. In vivo generation of transient FAPCAR T cells improves cardiac function after injury. Wild-type adult C57BL/6 mice were continuously dosed with saline or AngII/PE through an implanted 28-day osmotic minipump. After 1 week of cardiac pressure–overload injury, CD5-targeted LNPs were injected. Mice were analyzed after an additional 2 weeks. (**A**) Schematic representation of the experimental timeline. Echocardiograph measurements show improvements in LV volumes and diastolic and systolic function after a single injection of 10 μg of CD5/LNP-FAPCAR. (**B** and **C**) Measurements of end diastolic (B) and end systolic (C) volumes (in microliters). (**D**) M-mode estimate of weight-normalized LV mass (in milligrams per gram). (**E** to **G**) Diastolic function (E/e'), an estimate of LV filling pressure (E); ejection fraction [percent (F)]; and global longitudinal strain (G). (**H**) Representative M-mode echocardiography images. Echocardiograph data represent $n = 7$, $n = 7$, and $n = 8$ biologically independent mice per condition spread over three cohorts. (**I**) Picrosirius red staining highlights collagen (pink) in coronal cardiac sections of mock uninjured animals (3 weeks after saline pump implant + saline injection at week 1), injured control animals (AngII/PE + saline), and treated animals (AngII/PE + CD5/LNP-FAPCAR). Inset shows magnification of the LV myocardium. Scale bar, 100 μm. (**J**) Quantification of fibrosis (percentage) of the ventricles seen in (I). Histology data represent $n = 8$, $n = 11$, and $n = 12$ biologically independent mice per condition spread over five cohorts. Data are shown as mean ± SEM. P values shown are from Tukey's post hoc test after one-way ANOVA ($P < 0.05$).
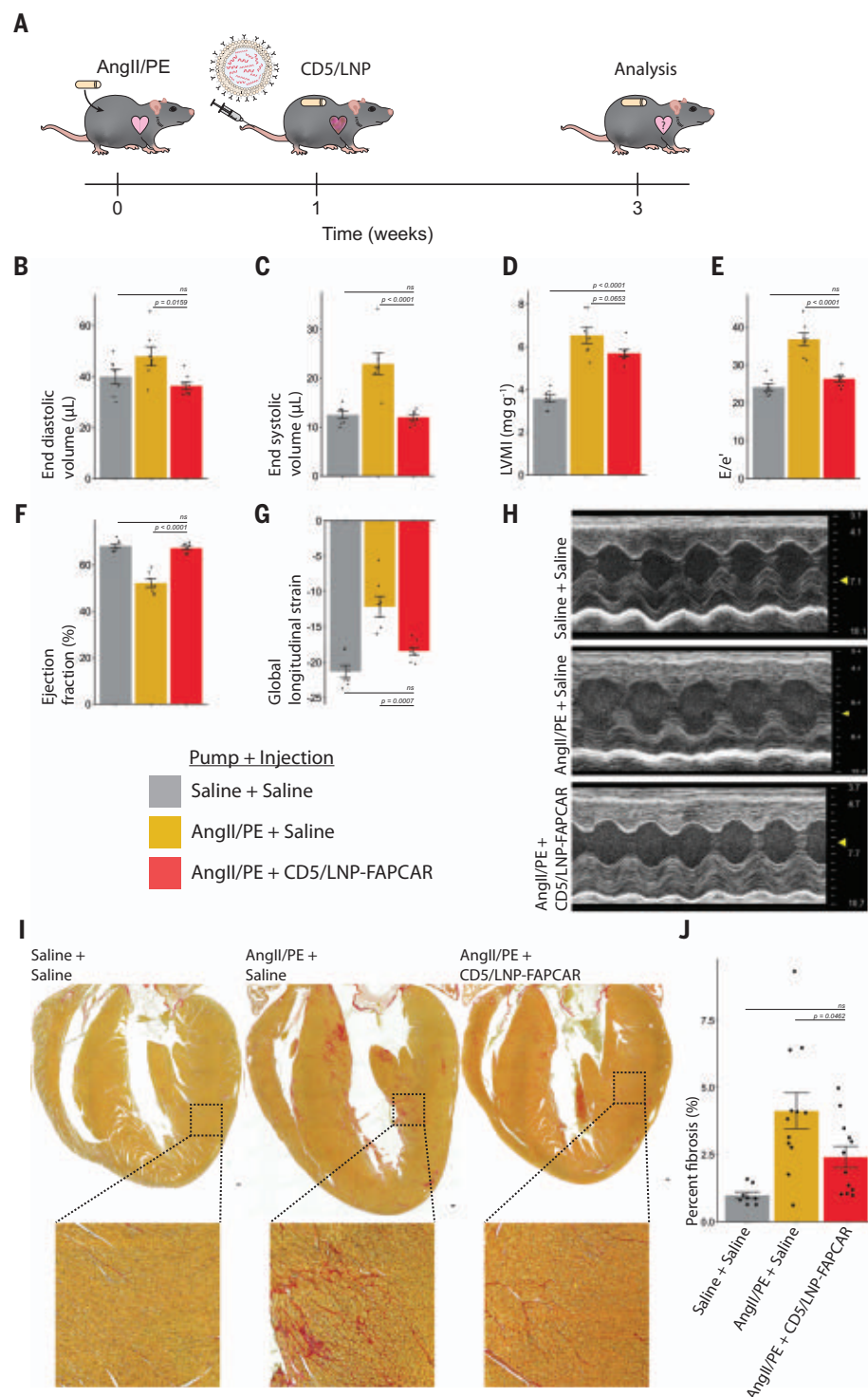
containing FAP^+ punctae were also seen in the spleens of injured animals treated with CD5/LNP-FAPCAR therapy but not in those treated with IgG/LNP-FAPCAR control (Fig. 3E). We are not aware of prior reports of CAR T cells exhibiting trogocytosis in the spleen after therapy, perhaps because prior studies have focused on CAR T cells directed against lymphocytic markers that would be difficult to distinguish from endogenous expression in the spleen. These findings are consistent with functional anti-FAP CAR T cells being produced in vivo after CD5/LNP-FAPCAR treatment.

We next assessed whether CD5/LNP-FAP-CAR treatment was able to improve cardiac function in injured mice, as was observed previously (9). To test this, we induced cardiac injury in mice with AngII/PE delivered through 28-day osmotic mini-pumps. After 1 week, when fibrosis is apparent (9), 10 μg of LNPs were injected intravenously. Two weeks after injection, cardiac function was analyzed by echocardiography (Fig. 4A and fig. S4, A and B).

We observed marked functional improvements in injured mice treated with in vivo–produced, transient FAPCAR T cells, consistent with our previous studies using adoptively transferred viral FAPCAR T cells (movies S2 to S5). AngII/ PE–injured mice treated with CD5/LNP-FAPCAR exhibited normalized left ventricular (LV) end diastolic and end systolic volumes (Fig. 4, B and C). Also consistent with our previous study (9), body weight–normalized LV mass (estimated in M-mode) did not show statistically significant differences after CD5/ LNP-FAPCAR injection, although a trend in improvement compared with control injured mice was noted (Fig. 4D). LV diastolic function (E/e') returned to uninjured levels (Fig. 4E). LV systolic function was also noticeably improved, as measured by ejection fraction (Fig. 4F) and global longitudinal strain (Fig. 4, G and H). Injection of nontargeting IgG/LNP-FAPCAR did not alter LV function (fig. S4C). In CD5/LNP-FAPCAR–injected animals, but not in controls, we observed an accumulation of CD3$^+$ T cells within regions occupied by FAP$^+$ fibroblasts (fig. S4D) (9). Furthermore, many of these CD3$^+$ T cells were FAPCAR$^+$ (80 of 137, or 58% of CD3$^+$ T cells observed in 25 highly magnified fields of view in five histologic sections), indicating that they had been transduced with FAPCAR mRNA, whereas CD3$^+$ T cells from control animals did not costain for the FAPCAR (fig. S4E). Consistent with our previous results (9), we observed a statistically significant improvement of the heart weight to body weight ratio (a measure of cardiac hypertrophy) in treated animals (fig. S5A).

Histologic analysis, as assessed by staining with picrosirius red, highlighted a significant improvement in the overall burden of extracellular matrix between injured mice treated with CD5/LNP-FAPCAR and those treated with saline or IgG/LNP-FAPCAR controls (Fig. 4, I and J, and fig. S5, B and C). Furthermore, a subset of treated animals (five of 12) was indistinguishable from uninjured controls, apart from the persistent perivascular fibrosis that results from activated fibroblasts that do not express FAP (9) (fig. S5D, arrows). Previous studies in which activated fibroblasts were eliminated by genetic ablation or treatment with virally transduced CAR T cells have also shown persistence of perivascular fibrosis (9, 28). Thus, CD5/LNP-FAPCAR treatment results in improved function and decreased interstitial fibrosis. We did not observe any gross histological changes in noncardiac organs or weight loss after CD5/LNP-FAPCAR injection (fig. S6, A and B).

These experimental results provide a proof of concept that modified mRNA encapsulated in targeted LNPs can be delivered intravenously to produce functional engineered T cells in vivo. The marked success and safety of modified mRNA/LNP severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines has stimulated broad efforts to extend this therapeutic platform to address numerous pathologies. By targeting LNPs to specific cell types, as we demonstrate here for lymphocytes, modified mRNA therapeutics are likely to have far-reaching applications. The generation of engineered T cells in vivo using mRNA is attractive for certain disorders because the transient nature of the produced CAR T cells is likely to limit toxicities, including risks incurred by lymphodepletion before injection, and allow for precise dosing. Unlike patients with cancer, those suffering from fibrotic disorders may not require a complete elimination of pathologic cells (activated fibroblasts) but may symptomatically benefit from an overall reduction in burden of disease. Furthermore, targeted LNP/mRNA technology affords the advantageous ability to titrate dosing and to re-dose as needed. Future studies will be needed to optimize the dosing strategy, LNP composition, and targeting approaches to further enhance therapeutic effects and limit potential toxicities. Nevertheless, the possibility of an "off-the-shelf" universal therapeutic capable of engineering specific immune functions provides promise for a scalable and affordable avenue to address the enormous medical burden of heart failure and other fibrotic disorders.

REFERENCES AND NOTES

1. H. Khalil et al., J. Clin. Invest. 127, 3770–3783 (2017).
2. S. Schafer et al., Nature 552, 110–115 (2017).
3. T. Moore-Morris et al., J. Clin. Invest. 124, 2921–2934 (2014).
4. T. Yokota et al., Cell 182, 545–562.e23 (2020).
5. J. G. Rurik, H. Aghajanian, J. A. Epstein, Circ. Res. 128, 1766–1779 (2021).
6. A. A. Widjaja et al., Front. Mol. Biosci. 8, 740650 (2021).
7. N. C. Henderson, F. Rieder, T. A. Wynn, Nature 587, 555–566 (2020).
8. A. González, E. B. Schelbert, J. Díez, J. Butler, J. Am. Coll. Cardiol. 71, 1696–1706 (2018).
9. H. Aghajanian et al., Nature 573, 430–433 (2019).
10. M. Kalos et al., Sci. Transl. Med. 3, 95ra73 (2011).
11. D. Weissman, Expert Rev. Vaccines 14, 265–281 (2015).
12. K. Karikó, M. Buckstein, H. Ni, D. Weissman, Immunity 23, 165–175 (2005).
13. K. Karikó et al., Mol. Ther. 16, 1833–1840 (2008).
14. O. Andries et al., J. Control. Release 217, 337–344 (2015).
15. Y. Zhao et al., Cancer Res. 70, 9053–9061 (2010).
16. N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, Nat. Rev. Drug Discov. 17, 261–279 (2018).
17. F. Rizvi et al., Nat. Commun. 12, 613 (2021).
18. C. Krienke et al., Science 371, 145–153 (2021).
19. J. Szőke et al., Nat. Commun. 12, 3460 (2021).
20. J. D. Gillmore et al., N. Engl. J. Med. 385, 493–502 (2021).
21. N. Pardi et al., J. Control. Release 217, 345–351 (2015).
22. A. Akinc et al., Mol. Ther. 18, 1357–1364 (2010).
23. H. Parhiz et al., J. Control. Release 291, 106–115 (2018).
24. I. Tombácz et al., Mol. Ther. 29, 3293–3304 (2021).
25. L. C. S. Wang et al., Cancer Immunol. Res. 2, 154–166 (2014).
26. L. Boumsell et al., J. Exp. Med. 152, 229–234 (1980).
27. G. Soldevila, C. Raman, F. Lozano, Curr. Opin. Immunol. 23, 310–318 (2011).
28. H. Kaur et al., Circ. Res. 118, 1906–1917 (2016).
29. D. Sommermeyer et al., Leukemia 30, 492–500 (2016).
30. M. Hamieh et al., Nature 568, 112–116 (2019).
31. E. Joly, D. Hudrisier, Nat. Immunol. 4, 815 (2003).
32. N. Martínez-Martín et al., Immunity 35, 208–222 (2011).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm0594
Materials and Methods
Figs. S1 to S6
Movies S1 to S5
Table S1
References (33–36)
MDAR Reproducibility Checklist

Oxygen and nitrogen production by an ammonia-oxidizing archaeon

Beate Kraft[1]*, Nico Jehmlich[2], Morten Larsen[1], Laura A. Bristow[1], Martin Könneke[3,4], Bo Thamdrup[1], Donald E. Canfield[1,5,6]

Ammonia-oxidizing archaea (AOA) are one of the most abundant groups of microbes in the world's oceans and are key players in the nitrogen cycle. Their energy metabolism—the oxidation of ammonia to nitrite—requires oxygen. Nevertheless, AOA are abundant in environments where oxygen is undetectable. By carrying out incubations for which oxygen concentrations were resolved to the nanomolar range, we show that after oxygen depletion, *Nitrosopumilus maritimus* produces dinitrogen and oxygen, which is used for ammonia oxidation. The pathway is not completely resolved but likely has nitric oxide and nitrous oxide as key intermediates. *N. maritimus* joins a handful of organisms known to produce oxygen in the dark. On the basis of this ability, we reevaluate the role of *N. maritimus* in oxygen-depleted marine environments.

Ammonia-oxidizing archaea (AOA) are only known to oxidize ammonia (NH_3) to nitrite (NO_2^-) using oxygen: $NH_3 + 1.5O_2 \rightarrow NO_2^- + H_2O + H^+$ (*1*, *2*). Yet, AOA are highly abundant in environments with very low, or even undetectable, oxygen concentrations such as marine oxygen-minimum zones (OMZs) and marine sediments (*3–7*). Certain populations of AOA become most prevalent in oxygen-depleted zones of the water column, indicating that they may be adapted to these conditions (*5*). The role of AOA in such environments is enigmatic, as they have no known anaerobic metabolism. We used trace luminescence oxygen sensors (hereafter, optodes) (*8*) to explore the physiology of AOA at low nanomolar oxygen concentrations and functional anoxia (oxygen levels below detection) as typically found in OMZs (*9*, *10*), thereby discovering oxygen production by the marine AOA *Nitrosopumilus maritimus* SCM1 (*11*). Dark, nonphotosynthetic oxygen production is rare in nature, with only three known pathways: chlorite dismutation during perchlorate and chlorate respiration ($ClO_2^- \rightarrow Cl^- + O_2$), detoxification of reactive oxygen species (e.g., H_2O_2 dismutation), and nitric oxide (NO) dismutation ($2NO_2^- \rightarrow 2NO \rightarrow N_2 + O_2$) (*12*). Although the pathway of oxygen production by *N. maritimus* is not fully resolved in this study, we show that it is not

any of the three aforementioned pathways. The oxygen-producing pathway in *N. maritimus* also involves NO dismutation but has oxygen and nitrous oxide (N_2O) as products and is thus distinct from the NO dismutation described above. Given the abundance of *N. maritimus* in oxygen-sparse environments, dark oxygen production may be common in nature. We also show that oxygen production is accompanied by N_2 production and thereby identify a previously unknown and potentially environmentally impactful N_2 production pathway.

We first grew axenic cultures of *N. maritimus* aerobically as an ammonia oxidizer. The cultures were then sparged with argon to oxygen levels below 5 μM, where the remaining oxygen was consumed by *N. maritimus* through continued ammonia oxidation, indicating physiologically active cells. Unexpectedly, after oxygen was completely consumed down to the limit of detection of 1 nM, it immediately began

to slowly increase (Fig. 1A). A series of additions of oxygen-saturated water showed the same recurring pattern: oxygen was consumed on oxygen addition and increased directly thereafter (Fig. 1A). In undisturbed incubations during which no oxygen additions were made, oxygen accumulated over hours and reached levels of between 100 and 200 nM (Fig. 1A). This pattern was observed reproducibly in multiple incubations carried out over 2 years.

In comparison, no oxygen buildup was detected in filtered abiotic controls or when cells were killed by the addition of mercuric chloride (fig. S1), ruling out the possibility of abiotic oxygen production or intrusion of oxygen into the incubation bottle. Contamination by oxygen intrusion was further ruled out by incubations in an anaerobic chamber, which showed the same trend of oxygen production (fig. S2). Involvement of medium components (e.g., HEPES, EDTA) in oxygen production was also excluded (fig. S3), and furthermore, oxygen microelectrodes, which make use of a different oxygen measurement principle, showed the same patterns of oxygen increase as did the optode measurements (fig. S4). Incubations with *N. maritimus* in medium containing pyruvate showed no difference when compared with incubations without pyruvate (fig. S5), thus ruling out oxygen production by H_2O_2 dismutation ($H_2O_2 \rightarrow H_2 + O_2$) because pyruvate reacts with and removes H_2O_2 via an abiotic decarboxylation reaction (*13*).

As described in more detail below, we measured NO accumulation in our experiments with a microelectrode and found a small NO interference with the optode measurements of O_2 (fig. S6) that we could calibrate. NO, however, did not interfere with O_2 micro-electrode measurements. When optode O_2
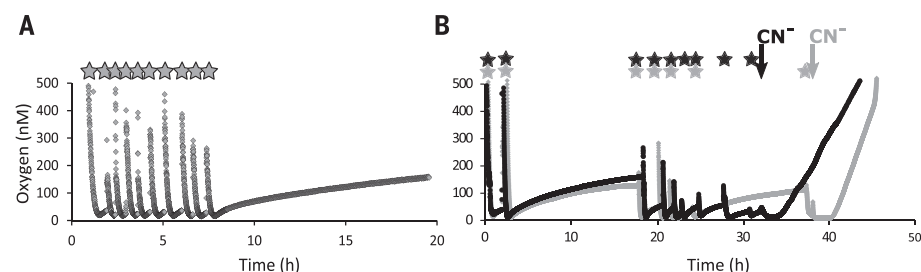

Fig. 1. Oxygen production by *N. maritimus*. (**A**) After supplied oxygen is consumed, oxygen concentrations immediately start to increase again (0 to 8 hours). When the incubation is left undisturbed and no oxygenated water is added, the oxygen concentration increases over time (8 to 20 hours). (**B**) Cyanide additions lead to a strong increase in oxygen production. Two parallel incubations showed the same pattern as observed in (A): Oxygen increased immediately after added oxygen had been consumed and accumulated over time when no oxygen additions were performed. After the additions of cyanide (CN^-, 0.5 mM final concentration; indicated with arrows), oxygen accumulations strongly increased. Star symbols indicate additions of oxygen-saturated water. Black and gray lines represent the two parallel incubations.

[1]Nordcee, Department of Biology, University of Southern Denmark, Odense, Denmark. [2]Department of Molecular Systems Biology, Helmholtz Centre for Environmental Research UFZ GmbH, Leipzig, Germany. [3]Marine Archaea Group, Center for Marine Environmental Sciences (MARUM), and Department of Geosciences, University of Bremen, Bremen, Germany. [4]Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Oldenburg, Germany. [5]Key Laboratory of Petroleum Geochemistry, Research Institute of Petroleum Exploration and Development, China National Petroleum Corporation, Beijing 100083, China. [6]Danish Institute of Advanced Study, University of Southern Denmark, Odense, Denmark.
*Corresponding author. Email: bkraft@biology.sdu.dk

concentrations and NO were simultaneously measured, a NO correction was applied to the O_2 measurements. The correction, however, was relatively small (0 to 17%) and fully predictable from the NO concentrations (fig. S7). This correction was only applied when NO and O_2 were simultaneously measured, recognizing that other optode O_2 measurements may be slight overestimates (depending on the NO concentration) of the actual O_2 concentration (fig. S7). Other potential intermediates or by-products of ammonia oxidation or nitrite conversion (N_2O, HNO, NO^-, $N_2O_3^{2+}$, and hydroxylamine) did not interfere with optode oxygen measurements (figs. S8 and S9).

Despite the small interference of NO on our oxygen measurements, *N. maritimus* clearly produces oxygen when the culture reaches anoxia. We hypothesize that oxygen accumulates as a net balance of simultaneous oxygen production and oxygen consumption by ammonia oxidation, where plateauing oxygen concentrations over time represent a balance between these processes. To test this hypothesis, cyanide (0.5 mM) was added to the oxygen-producing culture. Cyanide inhibits oxygen respiration by the heme–copper oxygen reductase and thus inhibits ammonia oxidation (14). Upon cyanide addition, and after an initial lag phase, oxygen concentrations steadily increased at rates ~5 times higher (65 ± 12 nmol liter^{-1} hour^{-1}) than before cyanide addition (14 ± 2 nmol liter^{-1} hour^{-1}) (Fig. 1B). In similar experiments where NO was also measured, O_2 increase was uncoupled from NO concentration after cyanide addition (fig. S10). These high rates of oxygen production with cyanide addition are consistent with the hypothesis that, in the absence of cyanide, some portion of the oxygen produced by *N. maritimus* is used within the cells and does not accumulate into the surroundings.

We tracked the conversion of ^{15}N-labeled ammonium (hereafter, ^{15}N-ammonium) to ni-trite to directly explore whether *N. maritimus* continues to oxidize ammonia while producing oxygen. In this experiment, cell cultures were washed to reduce the high nitrite background that accumulated (~1 mM) during normal aerobic growth. After this, ^{15}N-ammonium 50 μM was added as well as a small amount of ^{14}N-nitrite (experiment I, 5 μM; experiment II, 25 μM) to "capture" any produced ^{15}N-nitrite from further transformations. The labeling experiments showed continued ammonia oxidation to nitrite during oxygen production (Fig. 2). These results, consistent with the cyanide addition experiments, confirmed that ammonia oxidation occurs together with oxygen production and without any external addition of oxygen (Fig. 2C). Furthermore, the rates of ammonia oxidation in these duplicate experiments were 46 nM/hour (incubation I) and 39 nM/hour (incubation II) (Table 1), requiring oxygen production rates of 69 and 60 nM/hour, respectively, given the stoichiometry of ammonia oxidation ($NH_3 + 1.5O_2 \rightarrow NO_2^- + H_2O + 1H^+$). Oxygen accumulated at an average rate of only 1.2 nM/hour in both experiments. Therefore, most of the oxygen produced by *N. maritimus* in the incubations presented in Fig. 2 was immediately consumed through ammonia oxidation. The average cell density in these incubations was $1.3 \pm 0.53 \times 10^7$ cells ml^{-1}, and, therefore, the average ammonia oxidation rate per cell was between 3 and 3.5 attomoles cell^{-1} hour^{-1}. In marine OMZs, typical AOA cell densities are between 1×10^4 and 10×10^4 cells ml^{-1} (15–17). If most AOA cells are active at the rates we observed, AOA could sustain ammonia oxidation rates of 1 to 10 nM day^{-1}, which are on the same order of magnitude as anammox rates (0 to 40 nM day^{-1}) in open-ocean OMZs (10) or the Black Sea (18).

We next explored possible metabolic pathways for dark oxygen production in *N. maritimus*. Of the three known pathways of dark oxygen production, we ruled out perchlorate and chlorate respiration, as our culture medium did not include perchlorate, chlorate, or chlorite. From the pyruvate experiment discussed above (fig. S5), we also ruled out hydrogen peroxide dismutation as a source of oxygen. In addition, in the absence of photochemistry, the production of reactive oxygen species such as hydrogen peroxide, hydroxyl, or superoxide requires a reservoir of oxygen, which is incompatible with oxygen accumulation after anoxia was reached in our experiments. As *N. maritimus* metabolizes nitrogen and accumulates NO under normal aerobic ammonia oxidation (19), NO dismutation becomes a potential source of oxygen in our experiments. So far, NO dismutation is only known among the NC10 bacteria (12). These organisms are methane oxidizers and generate NO for dismutation to oxygen and dinitrogen (20), where the oxygen is used to oxidize methane. Because oxygen production and consumption are tightly coupled, methane-oxidizing NC10 bacteria are not known to liberate free oxygen into the environment (20).

Using NO microelectrodes, we found that NO and oxygen production were coupled (fig. S7). The coupling was not as strong in some cases (fig. S10), and cyanide additions caused a complete decoupling (fig. S10). Furthermore, when the NO scavenger PTIO (2-phenyl-4,4,5,5-tetramethylimidazoline-1-oxyl 3-oxide) was added, oxygen production ceased (fig. S12). Together, these results suggest that NO is a crucial intermediate in oxygen production.

We used ^{15}N-nitrite to further unravel the pathways of nitrogen and oxygen cycling during oxygen production by *N. maritimus*. In incubations with added ^{15}N-nitrite, mainly $^{30}N_2$ was produced during oxygen production (Fig. 3A and figs. S13 and S14), although some formation of $^{29}N_2$ was also detected (Fig. 3B). The production of $^{29}N_2$ most likely stems from

Table 1. Summary of rates extracted from incubations with $^{15}NH_4^+$ or $^{15}NO_2^-$ additions. Oxygen accumulation rates were taken when the accumulation rate was at its maximum at the start of oxygen production. In incubations I and II, oxygen supplied to the culture at the beginning of the incubation was consumed after 10 hours (Fig. 2C). Therefore, only time points after 10 hours were used to calculate ammonia oxidation rates during oxygen production. The means of the rates from replicate incubations and their standard deviation are presented. n.d., not determined; –, no N_2O accumulation was detected. N_2 production rates refer to the total N_2 production, which in case of incubations 1 and 2 equal $^{30}N_2$ production rates.

Incubation	O_2 accumulation (nM/hour)	NH_3 oxidation to NO_2^- (nM/hour)	N_2 production (first 20 hours; nM/hour)	N_2 production (20 to 40 hours; nM/hour)	N_2O accumulation (first 20 hours; nM/hour)
$^{15}NH_4^+$, 5 μM $^{14}NO_2^-$ (incubation I, Fig. 2)	1.2 (±0.2)	46 (±12)	49 (±12)		–
$^{15}NH_4^+$, 25 μM $^{14}NO_2^-$ (incubation II, Fig. 2)	1.2 (±0.2)	39 (±9)	51 (±6)		3 (±1)
$^{14}NH_4^+$, 1 mM $^{15}NO_2^-$ (incubation 1, Fig. 3)	21 (±8)	n.d.	9 (±3)	51 (±13)	5 (±2)
$^{14}NH_4^+$, 1 mM $^{15}NO_2^-$ (incubation 2, Fig. 3)	24 (±8)	n.d.	11 (±5)	37 (±20)	22 (±6)

the paring of ^{15}N-nitrite with residual unlabeled nitrite transferred with the inoculum when inoculating a new culture. Dinitrogen production by *N. maritimus*, or other AOA isolates, has not previously been reported. Furthermore, our results show that both nitrogen atoms in the N_2 originated from nitrite, with none coming from ammonium. This result was confirmed by incubations with ^{15}N-ammonium, where no immediate conversion of ^{15}N-ammonium to $^{29}N_2$ or $^{30}N_2$ was detected (Fig. 3, A and B). Instead, ^{15}N-ammonium was most likely converted to nitrite and diluted into the large existing nitrite pool in this experiment. In contrast, when ^{15}N-ammonium was added to washed cultures with a small nitrite pool (5 and 25 μM), the ^{15}N-nitrite produced from ammonia oxidation was further converted to N_2 (Fig. 2B and fig. S12), which is consistent with our experiments with ^{15}N-nitrite.

Thus far we have shown that NO is a likely intermediate in oxygen production and that N_2 and O_2 are produced by *N. maritimus*. Rates of O_2 accumulation and N_2 production from the different incubations shown in Figs. 2 and 3 are summarized in Table 1. These results are generally consistent with NO dismutation as a source of both N_2 and O_2, where NO is produced from the reduction of nitrite. In incubations with added ^{15}N-nitrite, however, oxygen accumulation exceeded N_2 production in the first 20 hours (incubations 1 and 2 in Table 1; Fig. 3C and figs. S13 and S14), demonstrating a decoupling of O_2 and N_2 production in this phase of the experiment. As net rates of O_2 accumulation may underestimate gross rates of O_2 production, as explored above, and as N_2 production in our experiments is a gross production rate, there is a definite imbalance between the production rates of O_2 and N_2. Such an imbalance would be inconsistent with O_2 and N_2 production directly from NO dismutation. If this was the case, the O_2 accumulation rates should not exceed N_2 production rates, but the imbalance we measure suggests that further intermediate(s) must exist between NO and N_2 and O_2 production. We suggest that N_2O may be such an intermediate, where 2NO → $N_2O + 0.5O_2$.

Indeed, in our incubations supplied with ^{15}N-nitrite, $^{46}N_2O$ accumulated before $^{30}N_2$ production accelerated (Fig. 3D), and the rates of N_2 production and N_2O accumulation taken together in the first 20 hours match the O_2 accumulation rates within the uncertainties (Table 1). Furthermore, the dismutation of NO [aqueous (aq)] to O_2 (aq) and N_2O (aq) is thermodynamically favorable (standard free energy change, $\Delta G^{0'}$ = −165 kJ/mol O_2,), and there are no abiotic mechanisms known for this dismutation re-

action. The accumulation of N_2O distinguishes the metabolism of *N. maritimus* from that in the NC10 bacterium *Methylomirabilis oxyfera*, whose O_2 production does not involve N_2O, and for which traces of N_2O have been attributed to other community members in the enrichment (*20*). However, to explain the subsequent formation of N_2 from N_2O, an unknown N_2O reductase would need to be present in the *N. maritimus* genome (*21*). Although no nitrous oxide reductases outside the NosZ family have been conclusively

identified, their existence has been proposed multiple times on the basis of the occurrence of complete denitrification in organisms that lack a NosZ enzyme (*22, 23*).

N-nitrosating hybrid N_2 formation, in which one N atom from NO_2^- and one from NH_4^+ (or an intermediate of ammonia oxidation) combine to form N_2O, has been proposed as a possible source for N_2O production in AOA (*24*), but the isotopic signature of the N_2O produced in our experiments ($^{46}N_2O$) does not support this pathway (the production of N_2O from ^{15}N-ammonia and ^{14}N-nitrite would yield $^{45}N_2O$). Thus, the N_2O formation pathway in AOA when environmental oxygen is present seems to differ from the one in our experiments (*15, 25*). However, the $\delta^{18}O$ values of N_2O produced in incubation experiments with AOA enrichment cultures support an enzymatic N_2O production pathway from nitrite in AOA (*26*). Furthermore, abiotic hybrid formation of N_2O from NO and hydroxylamine is insignificant at the NO concentrations measured in our incubations (see supplementary text in the supplementary materials).

In incubations supplied with ^{15}N-ammonium and a small nitrite pool (Fig. 2), N_2O accumulated transiently as well (fig. S15). In these incubations, N_2 production far exceeded O_2 accumulation, and no N_2O accumulation would be required for a mass balance. This is not surprising, as high rates of ammonia oxidation (Table 1 and Fig. 2) indicate that O_2 accumulation rates in these experiments far underestimate gross rates of O_2 production, as explored above. This does not mean that N_2O was not an intermediate in these experiments, only that these incubations did not demonstrate an initial imbalance between O_2 and N_2 production.

As for NO dismutation in NC10 bacteria, O_2 production in *N. maritimus* most likely has NO as an intermediate and produces N_2. However, unlike for NC10 bacteria, our results suggest that the pathway of O_2 production used by *N. maritimus* has an extra intermediate that may be N_2O and that it leads to the transient accumulation of oxygen. A proposal for the metabolic pathway associated with oxygen production in *N. maritimus* is shown in Fig. 3E. Although ammonia oxidation to nitrite is accomplished by the O_2 produced by *N. maritimus*, the conversion of nitrite to N_2 requires reducing equivalents regardless of the O_2 production pathway. The required electrons can partly, but not fully, be obtained from the ongoing oxidation of ammonia. Another source of electrons could be intra- or extracellular organic matter produced during normal aerobic ammonia oxidation (*27*) or, in situ, by dissolved organics available in the water column. Furthermore, to sustain the coupling of oxygen production and
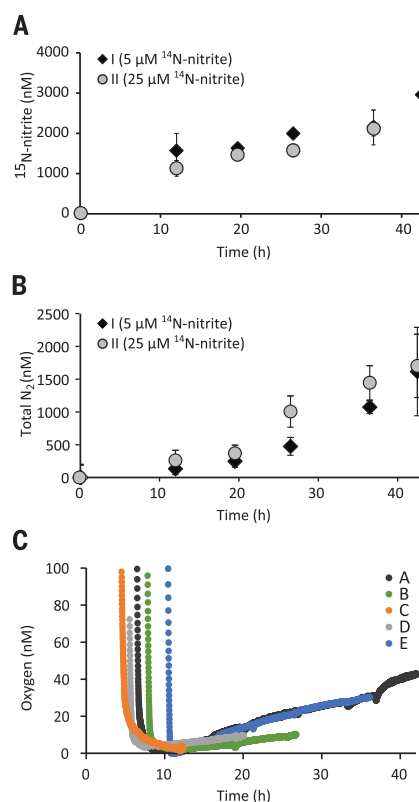

Fig. 2. Ammonia oxidation to nitrite and N_2 during oxygen production by *N. maritimus*.
(**A**) ^{15}N- nitrite production from ^{15}N-ammonium (50 μM) for two sets of incubations of washed *N. maritimus* culture. Incubation I contained a ^{14}N-nitrite pool of 5 μM, and incubation II had a ^{14}N-nitrite pool of 25 μM. ^{15}N-nitrite production continued after supplied oxygen was consumed (10 hours). (**B**) Total N_2 production in incubations I and II. Results include $^{28}N_2$ from ^{14}N-nitrite as well as $^{30}N_2$ and $^{29}N_2$ from added ^{15}N-ammonium that was converted to ^{15}N-nitrite and partly captured in the small ^{14}N-nitrite pool before further conversion to $^{30}N_2$ and $^{29}N_2$ (results in fig. S12). (**C**) Oxygen accumulation in a subset of exetainers. Exetainers A and B belong to incubation I, and exetainers C, D, and E belong to incubation II. Error bars represent the standard deviation of three replicates.
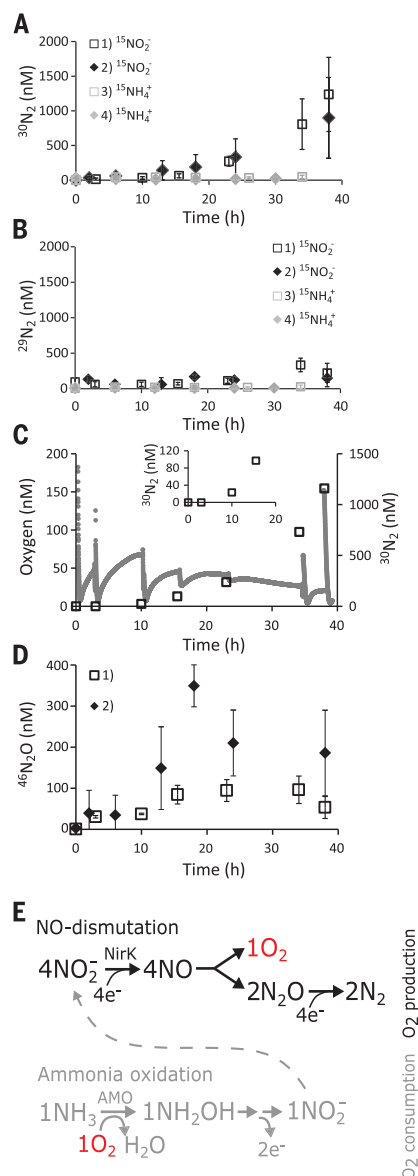
Fig. 3. N_2 production and N_2O accumulation by *N. maritimus* during simultaneous oxygen production. After sparging with argon, *N. maritimus* culture was incubated with either ^{15}N-nitrite or ^{15}N-ammonium and the production of (**A**) $^{30}N_2$ and (**B**) $^{29}N_2$ was tracked in two independent sets of incubations for each tracer addition (3 or 4 replicates each). N_2 production was only detected in incubations with ^{15}N-nitrite. For the corresponding oxygen measurements, see figs. S12 and 13. (**C**) Oxygen accumulation and $^{30}N_2$ production for a single replicate of the incubation series 1). The inset shows $^{30}N_2$ production in the first 20 hours. Disturbances in the oxygen time series at 0, 3, 10, 15.5, 23, 34, and 38 hours correspond to the time points when samples for N_2 analysis were taken, which led to slight oxygen intrusion. Gray dots, oxygen; black open squares, $^{30}N_2$. (**D**) $^{46}N_2O$ accumulation in incubations 1 and 2 supplied with ^{14}N-ammonium and 1 mM ^{15}N-nitrite. Only $^{46}N_2O$ accumulated in these incubations with a large ^{15}N-nitrite pool, indicating that all produced N_2O originated from nitrite. Error bars represent the standard deviation of four (incubations 1 and 3) or three (incubations 2 and 4) replicates. (**E**) Proposed pathway of oxygen and dinitrogen production in *N. maritimus*. Ammonia oxidation pathway (gray): For the aerobic oxidation of ammonia to nitrite, oxygen is needed to activate ammonia oxidation by the ammonia monooxygenase (AMO); two electrons per oxidized NH_4^+ enter the electron transfer chain and, under aerobic ammonia oxidation, are transferred to the terminal oxygen reductase, which reduces $0.5O_2$. Proposed oxygen production pathway (black): Nitrite is reduced to nitric oxide by the NirK nitrite reductase. Nitric oxide is dismutated to oxygen and nitrous oxide. The accumulating oxygen is consumed during ammonia oxidation. Nitrous oxide is reduced to dinitrogen. This pathway requires eight electrons per produced O_2. These electrons may partly be supplied by the ammonia oxidation reaction, which in return would reduce the oxygen demand by the ammonia oxidation pathway.

REFERENCES AND NOTES

1. C. Schleper, G. W. Nicol, *Adv. Microb. Physiol.* **57**, 1–41 (2010).
2. D. A. Stahl, J. R. de la Torre, *Annu. Rev. Microbiol.* **66**, 83–101 (2012).
3. F. J. Stewart, O. Ulloa, E. F. DeLong, *Environ. Microbiol.* **14**, 23–40 (2012).
4. V. Molina, L. Belmar, O. Ulloa, *Environ. Microbiol.* **12**, 2450–2465 (2010).
5. M. Sollai, L. Villanueva, E. C. Hopmans, G. J. Reichart, J. S. Sinninghe Damsté, *Geobiology* **17**, 91–109 (2019).
6. C. Berg, V. Vandieken, B. Thamdrup, K. Jürgens, *ISME J.* **9**, 1319–1332 (2015).
7. J. M. Beman, V. J. Bertics, T. Braunschweiler, J. M. Wilson, *Front. Microbiol.* **3**, 263 (2012).
8. P. Lehner *et al.*, *PLOS ONE* **10**, e0128125 (2015).
9. L. Tiano, E. Garcia-Robledo, N. P. Revsbech, *PLOS ONE* **9**, e105399 (2014).
10. P. Lam, M. M. M. Kuypers, *Annu. Rev. Mar. Sci.* **3**, 317–345 (2011).
11. M. Könneke *et al.*, *Nature* **437**, 543–546 (2005).
12. K. F. Ettwig *et al.*, *Front. Microbiol.* **3**, 273 (2012).
13. J.-G. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7888–7893 (2016).
14. M. T. Wilson, G. Antonini, F. Malatesta, P. Sarti, M. Brunori, *J. Biol. Chem.* **269**, 24114–24119 (1994).
15. A. E. Santoro *et al.*, *Global Biogeochem. Cycles* **35**, 1–21 (2021).
16. C. R. Löscher *et al.*, *Biogeosciences* **9**, 2419–2429 (2012).
17. L. A. Bristow *et al.*, *Nat. Geosci.* **10**, 24–29 (2017).
18. M. M. Jensen, M. M. M. Kuypers, G. Lavik, B. Thamdrup, *Limnol. Oceanogr.* **53**, 23–36 (2008).
19. W. Martens-Habbena *et al.*, *Environ. Microbiol.* **17**, 2261–2274 (2015).
20. K. F. Ettwig *et al.*, *Nature* **464**, 543–548 (2010).
21. C. B. Walker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8818–8823 (2010).
22. A. T. Fernandes *et al.*, *FEBS J.* **277**, 3176–3189 (2010).
23. D. M. Arciero, B. S. Pierce, M. P. Hendrich, A. B. Hooper, *Biochemistry* **41**, 1703–1709 (2002).
24. M. Stieglmeier *et al.*, *ISME J.* **8**, 1135–1146 (2014).
25. Q. Ji, E. Buitenhuis, P. Suntharalingam, J. L. Sarmiento, B. B. Ward, *Global Biogeochem. Cycles* **32**, 1790–1802 (2018).
26. A. E. Santoro, C. Buchwald, M. R. McIlvin, K. L. Casciotti, *Science* **333**, 1282–1285 (2011).
27. B. Bayer *et al.*, *Environ. Microbiol.* **21**, 4062–4075 (2019).

ammonia oxidation, external nitrite must be present (Fig. 3E).

The hypothesized pathway extends the range of AOA activity into functionally anoxic environments. By showing that *N. maritimus* produces both oxygen and N_2 under oxygen depletion, we have also uncovered a potentially environmentally significant pathway of N_2 production. By converting ammonium through nitrite to N_2, AOA perform a thus far unrecognized pathway of ammonia conversion to N_2 when oxygen is depleted. ^{15}N-tracer experiments currently performed to measure N-cycling rates in marine oxygen-depleted environments would overlook this pathway and account for its N_2 and N_2O production as canonical denitrification and/or anammox. Depending on the environmental controls and dynamics of this pathway, which are yet to be elucidated, our current understanding of how oxygen concentrations influence nitrogen and nitrous oxide production may require revision.

Considering the abundance of AOA in the global oceans and their key role in the nitrogen cycle, a widely distributed oxygen-producing pathway by AOA could have far-reaching implications for the microbial ecology and biogeochemical cycling in oxygen-depleted environments. The discovery of an oxygen-producing pathway in AOA can explain the presence and role of AOA in such environments, solving a longstanding enigma.

PALEOECOLOGY

Smaller fish species in a warm and oxygen-poor Humboldt Current system

Renato Salvatteci[1]*, Ralph R. Schneider[2], Eric Galbraith[3,4], David Field[5], Thomas Blanz[2], Thorsten Bauersachs[2], Xavier Crosta[6], Philippe Martinez[6], Vincent Echevin[7], Florian Scholz[8], Arnaud Bertrand[9]

Climate change is expected to result in smaller fish size, but the influence of fishing has made it difficult to substantiate the theorized link between size and ocean warming and deoxygenation. We reconstructed the fish community and oceanographic conditions of the most recent global warm period (last interglacial; 130 to 116 thousand years before present) by using sediments from the northern Humboldt Current system off the coast of Peru, a hotspot of small pelagic fish productivity. In contrast to the present-day anchovy-dominated state, the last interglacial was characterized by considerably smaller (mesopelagic and goby-like) fishes and very low anchovy abundance. These small fish species are more difficult to harvest and are less palatable than anchovies, indicating that our rapidly warming world poses a threat to the global fish supply.

Reduced body size has been postulated as a universal ecological response to warming in aquatic systems (*1, 2*). Shrinking body size among marine ectotherms has been hypothetically linked to temperature as predicted by Bergmann's rule, with smaller body sizes associated with warmer temperatures (*3*) and reduced oxygen availability (*4*). The energetic demand of fish and their oxygen consumption increase with water temperature, yet oxygen solubility is reduced in warmer water (*5*). Smaller fish maintain high activity in warm, oxygen-poor water (*6*) and are therefore expected to increase in relative abundance with future climate change (*7*). However, the degree to which an increase in temperature and decrease in oxygen will increase the proportion of small species (i.e., species shift hypothesis) relative to a decrease in mean body size at the population scale (i.e., population body size hypothesis) remains unclear (*2*).

We tested the relationships between water temperature, subsurface oxygen, and fish species abundance and size using paleoceanographic records from the Humboldt Current system. In this upwelling system, an extremely high biomass of anchovies (*Engraulis ringens*) yields up to 15% of the global annual fish catch (*8*). Model projections suggest that by the end of the 21st century, the Humboldt Current will have higher temperatures and lower oxygen levels (*9*) than in modern times. Concerningly, pre-fishery data have revealed biological tipping points that led to repeated multifarious regime shifts in the Humboldt Current in past centuries (*10, 11*); however, these tipping points have not been reproduced by ecological modelling experiments (*12*). Studies of sediment cores have the potential to reveal the state and dynamics of past climates and ecosystems on multiple time scales (*13*). In particular, data from sediment records independent of fisheries offer the opportunity to better understand the response of the fish community to a warmer world.

We focus on the last interglacial period or MIS5e (Marine Isotope Stage 5e), a globally warmer-than-present period to characterize the Humboldt Current at that time in comparison to the Holocene (last 11.7 thousand years before present; figs. S1 to S4). Temperature proxies include the alkenone unsaturation index (Uk'37) and the tetraether index of tetraethers consisting of 86 carbon atoms (TEXH86). Biogeochemical and ecological proxies include $\delta^{15}N$, total organic carbon, biogenic silica measurements and diatom assemblages, as well as vertebrae and other bones from fish.

Our paleoreconstruction shows that environmental changes during the last interglacial (Fig. 1, E to H) were similar to those of the RCP8.5 (Representative Concentration Pathway 8.5) projections to the year 2100 (Fig. 1, A to D, and fig. S5), including a similar magnitude of warming, intensified subsurface oxygen depletion, strengthened vertical density gradients, and comparable primary production. During the last interglacial, water temperatures adjacent to central Peru were ~2°C warmer on average than during the Holocene (Fig. 1, A and E, and fig. S1), plausibly as a result of a persistent El Niño–like state (*14*). Elevated $\delta^{15}N$ is consistent with intensified water column denitrification during the last interglacial, as would be expected with a more intense oxygen minimum zone (Fig. 1F and fig. S2), which is also supported by higher concentrations of redox-sensitive metals. Additionally, more intense stratification and deeper thermocline occurred compared with the Holocene (Fig. 1G), as indicated by the difference between the alkenone (temperatures at and above the thermocline) and the TEX^H_{86} (temperatures below the thermocline). Finally, productivity indicators are notably similar between the two interglacials (Fig. 1H and fig. S2), suggesting comparable rates of nutrient supply through upwelling activity.

The fish community inhabiting the Humboldt Current during the last interglacial differed substantially from the communities found during the Holocene (Fig. 2 and figs. S3 and S4) and the observational period. Although anchovies have clearly dominated the fish community throughout the Holocene, they occurred in only minor proportions during the last interglacial. By contrast, goby-like species accounted for almost 60% of the fish vertebrae assemblage during the last interglacial (Fig. 2). Gobies are small fishes specially adapted to survive in hypoxic or even sulfidic environments, conditions avoided by most organisms (*15*). Our inference of goby-like fishes dominating the Humboldt Current during the last interglacial is consistent with genetic analysis suggesting that a population expansion of the goby *Elacatinus puncticulatus* began at 170 to 130 thousand years ago in the Tropical Eastern Pacific (*16*).

Similar to the goby-like fishes, mesopelagic fishes were highly abundant at our coastal site during the last interglacial (Fig. 2). The vertebrae assemblage during the last interglacial included vertebrae from *Vinciguerria lucetia* and blue lanternfish (*Tarletonbeania crenularis*), in addition to many other small vertebrae likely belonging to other mesopelagic species (Fig. 2). This high coastal abundance is in marked contrast to modern observations showing that mesopelagic fishes are found at offshore locations associated with upwelling fronts (*17*). Mesopelagic fishes have evolved physiological and morphological adaptations to live in oxygen-deficient waters; for example, diel vertical migration into the oxygen minimum zone and tolerance of severe hypoxia episodes (*17*). This leads us to speculate that their incursion into coastal waters was linked to the intensification of mid-depth oxygen depletion. It is important to consider the effect of degradation of fish debris (fig. S6), which occurs in the sedimentary column (*18*) and to which small and thin vertebrae are especially susceptible. However, on its own this would be

[1]Center for Ocean and Society, Christian-Albrechts-University, Kiel, Germany. [2]Institute of Geosciences, Christian-Albrechts-University, Kiel, Germany. [3]Earth and Planetary Sciences, McGill University, Montreal, QC, Canada. [4]Institut de Ciència i Tecnologia Ambientals (ICTA-UAB), Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain. [5]College of Natural Sciences, Hawaii Pacific University, Kaneohe, HI, USA. [6]Université de Bordeaux, CNRS, EPHE, UMR 5805 EPOC, Pessac, France. [7]Sorbonne Université, LOCEAN-IPSL, CNRS/IRD/MNHN, 4 place Jussieu, 75252 Paris, France. [8]GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany. [9]Institut de Recherche pour le Développement (IRD), MARBEC, University Montpellier, CNRS, Ifremer, IRD, Sète, France.
*Corresponding author. Email: rsalvatteci@kms.uni-kiel.de

Fig. 1. Comparison of model projections for the RCP 8.5 scenario, on the basis of the Geophysical Fluid Dynamic Laboratory (GFDL) Earth system model (2086 to 2100 average minus 2006 to 2020), with paleoceanographic data for the Holocene and the last interglacial (M77/2-24).
(**A**) In R-GFDL SST, the colored circles indicate the SST difference between the last interglacial and Holocene alkenone-derived SST in cores ODP-1229 (*14*), SO147-106 (*22*), and M77/2-24. (**B**) Oxygen concentration (micromoles per liter) between 50 and 100 m in depth. (**C**) Depth of the 20°C isotherm as a proxy for thermocline depth. (**D**) R-GFDL primary productivity (millimoles of carbon per square meter per day) in the upper 500 m of the water column. The white circles in (B), (C), and (D) indicate the position of M77/2-24 and G10-B14. (**E**) Alkenone-derived SST average during the last interglacial and Holocene (M77/2-24). (**F**) $\delta^{15}N$ average during the last interglacial and Holocene as a proxy for water column denitrification (M77/2-24). (**G**) Difference between alkenone and TEX_{86}-derived temperatures, where higher values indicate a deeper thermocline (M77/2-24). (**H**) Average values (±2 SE) of total organic carbon (TOC) and biogenic silica during the last interglacial and Holocene (M77/2-24). Average biogenic silica for the Holocene is from (*23*).
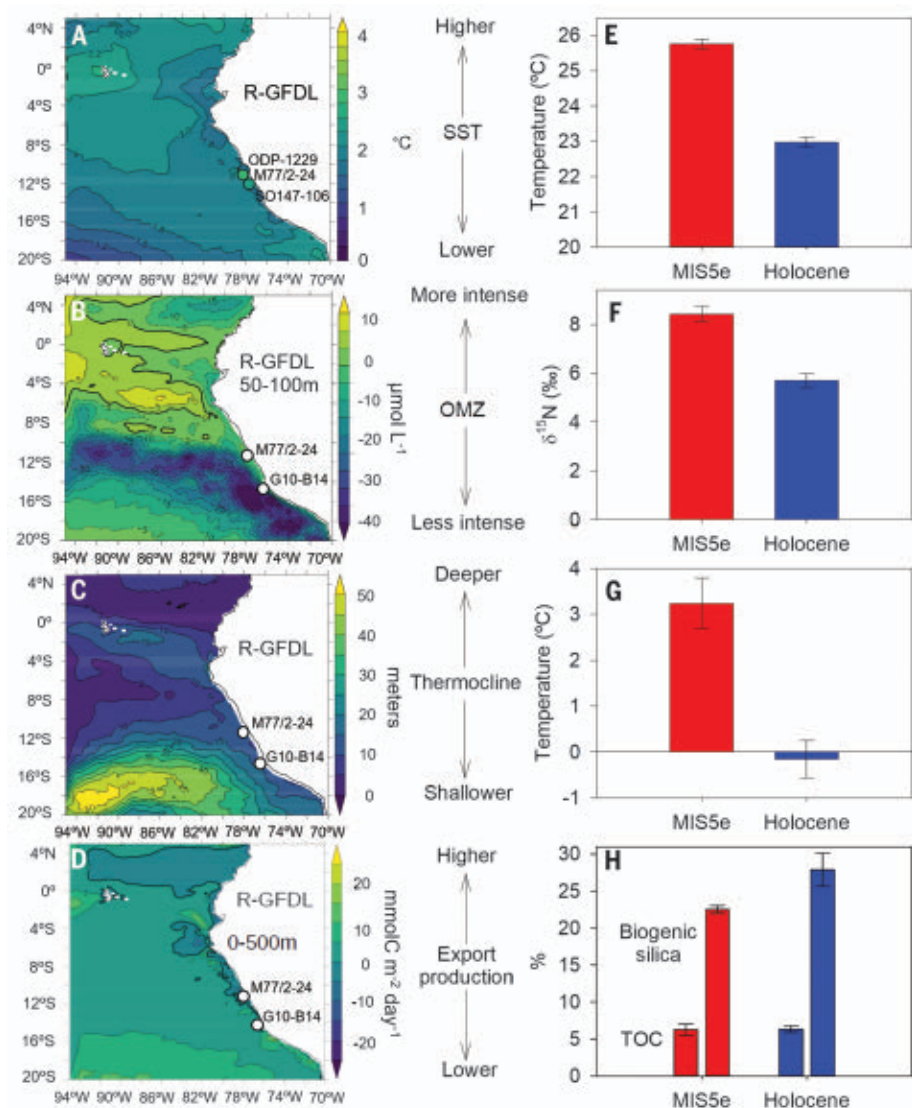
Fig. 2. No-analog fish community in a warmer world. Fish vertebrae abundances during the last interglacial and Holocene. Goby-like and mesopelagic fishes [(i.e., *V. lucetia* and blue lanternfish (*T. crenularis*)] were abundant during the last interglacial. The "other species" category includes vertebrae from unidentified adult and juvenile fish, in addition to mesopelagic fish.
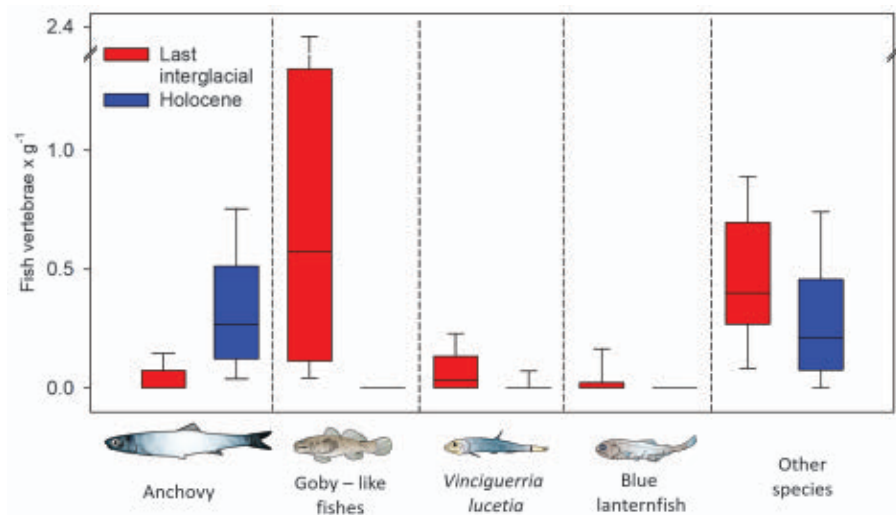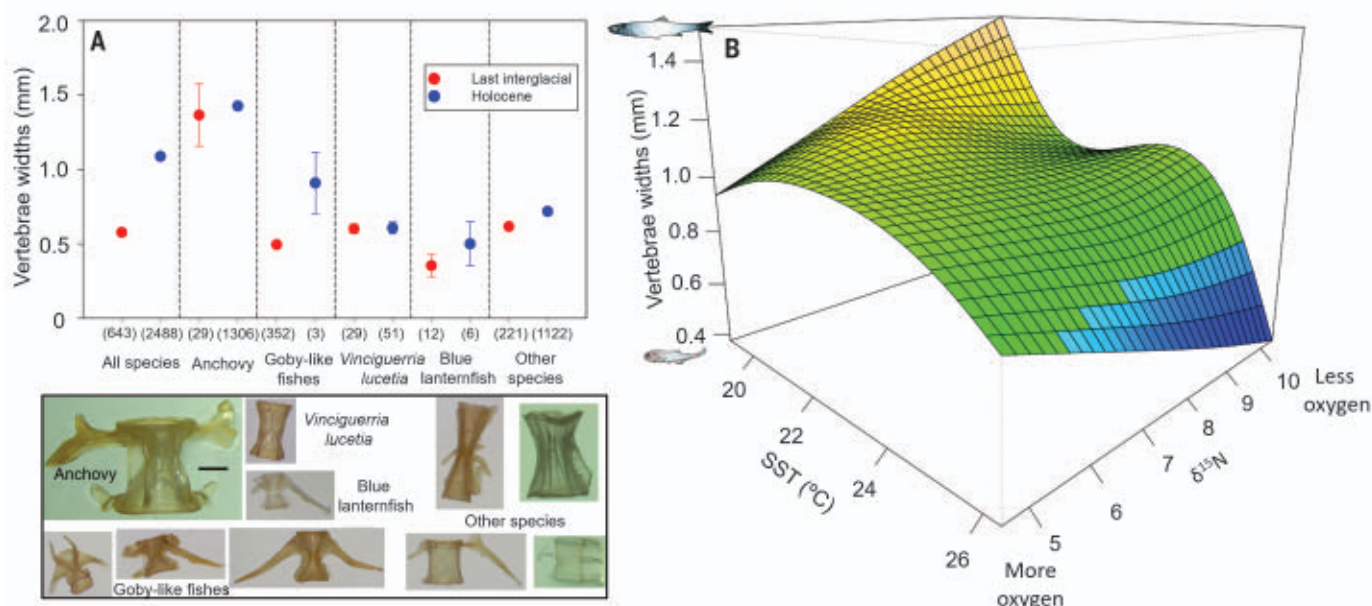
Fig. 3. Small fish in a warm ocean. (A) Average values (± 2 SE) of fish vertebrae widths as an indicator of fish size for the Holocene and the last interglacial. Values in parentheses indicate the number of vertebrae measured. In some cases, the error bars overlap with the symbol. Note that "goby-like fishes" may include multiple species. Pictures of the most abundant fish vertebrae for the Holocene and the last interglacial found in M77/2-024 are also shown (scale bar, 1 mm). **(B)** Generalized additive model (GAM) describing the significant relationship between vertebrae width and an interaction between temperature and oxygen. A linear negative correlation between vertebrae size and temperatures reiterates the relationship (fig. S7).

expected to lead to a relative underestimation of smaller fish in the older sediments of the last interglacial—the opposite of what is observed.

Our reconstruction reveals a major shift in species in response to warming but does not support a decrease in population body size, as we did not find any significant reduction in vertebrae widths of a given species during the last interglacial (Fig. 3A and table S2). A generalized additive model shows a significant ($P < 0.001$) nonlinear effect of the interaction of oxygen and temperature on the inferred average fish body size (Fig. 3B). The smallest vertebrae (mostly from goby-like fishes) are associated with the highest reconstructed water temperatures and lowest inferred oxygen concentration (i.e., last interglacial conditions). By contrast, the largest vertebrae (mostly from anchovies) are associated with periods of the lowest temperatures and lowest oxygen concentrations (i.e., some periods during the Holocene, including the past century) when strong upwelling led to low temperatures, thus promoting increased productivity and related deoxygenation (10, 11). Thus smaller fish of different species—rather than smaller fish from the same species—flourished in the Humboldt Current, suggesting the possibility of an environmentally triggered tipping point beyond which anchovies cease to be a dominant part of the ecosystem.

The overall shift to smaller body size is consistent with aspects of the gill oxygen limitation theory (GOLT), which occurred through a shift toward smaller species rather than a shrinking of the same species. The GOLT argues that the geometrical constraints of gills reduce the rate at which ectotherms can extract oxygen from water, relative to their body size, as they become larger (6). During the last interglacial, the heightened oxygen demand caused by warmer water may have required that anchovies—which reproduce in water <25°C (19)—shift poleward to satisfy oxygen demand. By contrast, smaller goby-like fishes have larger gill surfaces in relation to their body volume and can therefore withstand low-oxygen water.

Our results are concerning, as recent studies indicate that the environmental conditions and fish communities in the Humboldt Current are progressing toward those we reconstructed during the last interglacial. In the last few decades, midwater oxygen concentrations have decreased off Peru, the abundance of several warm-water mesopelagic species (like *V. lucetia*) has increased (20), and anchovy biomass and landings have steadily decreased despite a flexible and adaptive fishery management strategy (21). Conditions may therefore be approaching the tipping point in the fish community, as suggested by our analysis. Fishery managers, global markets, and other stakeholders will need to develop adaptive strategies to cope with this threat to the Humboldt Current food source while raising awareness regarding similar climate-driven shifts in fish communities elsewhere.

REFERENCES AND NOTES

1. W. W. L. Cheung *et al.*, *Nat. Clim. Chang.* **3**, 254–258 (2013).
2. M. Daufresne, K. Lengfellner, U. Sommer, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12788–12793 (2009).
3. G. Hunt, K. Roy, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1347–1352 (2006).
4. H. O. Pörtner, R. Knust, *Science* **315**, 95–97 (2007).
5. C. Deutsch, J. L. Penn, B. Seibel, *Nature* **585**, 557–562 (2020).
6. D. Pauly, The gill-oxygen limitation theory (GOLT) and its critics. *Sci. Adv.* **7**, eabc6050 (2021).
7. J. G. Rubalcaba, W. C. E. P. Verberk, A. J. Hendriks, B. Saris, H. A. Woods, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31963–31968 (2020).
8. FAO, *The State of World Fisheries and Aquaculture 2020* (FAO, 2020) www.fao.org/documents/card/en/c/ca9229en.
9. V. Echevin *et al.*, *Biogeosciences* **17**, 3317–3341 (2020).
10. R. Salvatteci *et al.*, *Glob. Change Biol.* **24**, 1055–1068 (2018).
11. R. Salvatteci *et al.*, *Prog. Oceanogr.* **176**, 102–114 (2019).
12. W. W.L. Cheung, J. Bruggeman, M. Butenschön, in "Impacts of climate change on fisheries and aquaculture: Synthesis of current knowledge, adaptation and mitigation options" (Food and Agriculture Organization of United Nations, 2018), pp. 63–85.
13. M. Yasuhara *et al.*, *Oceanography* **33**, 16–28 (2020).
14. S. Contreras *et al.*, *Geophys. Res. Lett.* **37**, n/a (2010).
15. A. Salvanes, A. Utne-Palm, B. Currie, V. Braithwaite, *Mar. Ecol. Prog. Ser.* **425**, 193–202 (2011).
16. E. R. Sandoval-Huerta *et al.*, *Mol. Phylogenet. Evol.* **130**, 269–285 (2019).
17. R. Cornejo, R. Koppelmann, *Mar. Biol.* **149**, 1519–1537 (2006).
18. R. Salvatteci, D. B. Field, T. Baumgartner, V. Ferreira, D. Gutierrez, *Paleobiology* **38**, 52–78 (2012).
19. R. Castillo *et al.*, *Fish. Oceanogr.* **28**, 389–401 (2019).
20. J. Argüelles *et al.*, *Prog. Oceanogr.* **79**, 308–312 (2008).
21. R. Oliveros-Ramos, M. Ñiquen, J. Csirke, R. Guevara-Carrasco, in "Adaptive Management of Fisheries in Response to Climate Change" (FAO Fisheries and Aquaculture Technical Paper 667, FAO, Rome, 2021), pp. 237–244.
22. B. Rein, F. Sirocko, A. Lückge, L. Reinhardt, A. Wolf, W.-C. Dullo, in *The Climate of Past Interglacials*, F. Sirocko, M. Claussen, M. Sánchez Goñi, T. Litt, Eds. (Developments in Quaternary Science series, Elsevier, 2007), pp. 305–321.
23. K. Doering *et al.*, *Paleoceanography* **31**, 1453–1473 (2016).
24. R. Salvatteci, Paleoceanographic proxies obtained from core M77/2-024-5 off Peru, Version 1, Zenodo (2021).

STRUCTURAL VIROLOGY

Structural basis of synergistic neutralization of Crimean-Congo hemorrhagic fever virus by human antibodies

Akaash K. Mishra[1]†, Jan Hellert[2]†‡, Natalia Freitas[3], Pablo Guardado-Calvo[2], Ahmed Haouz[4], J. Maximilian Fels[5]§¶, Daniel P. Maurer[6], Dafna M. Abelson[7], Zachary A. Bornholdt[7], Laura M. Walker[6], Kartik Chandran[5], François-Loïc Cosset[3], Jason S. McLellan[1]*, Felix A. Rey[2]*

Crimean-Congo hemorrhagic fever virus (CCHFV) is the most widespread tick-borne zoonotic virus, with a 30% case fatality rate in humans. Structural information is lacking in regard to the CCHFV membrane fusion glycoprotein Gc—the main target of the host neutralizing antibody response—as well as antibody–mediated neutralization mechanisms. We describe the structure of prefusion Gc bound to the antigen-binding fragments (Fabs) of two neutralizing antibodies that display synergy when combined, as well as the structure of trimeric, postfusion Gc. The structures show the two Fabs acting in concert to block membrane fusion, with one targeting the fusion loops and the other blocking Gc trimer formation. The structures also revealed the neutralization mechanism of previously reported antibodies against CCHFV, providing the molecular underpinnings essential for developing CCHFV–specific medical countermeasures for epidemic preparedness.

C rimean-Congo hemorrhagic fever virus (CCHFV) is endemic to Africa, Asia, and Europe and is transmitted by ticks and contact with bodily fluids from viremic animals or patients (*1, 2*). Although infection is asymptomatic in most vertebrates, it can cause severe disease in humans, with hemorrhage, myalgia, and high fever, eventually leading to death in ~30% of diagnosed cases (*1, 3, 4*). As a result, the World Health

Organization has shortlisted CCHFV as a priority pathogen in its research and development blueprint (*5*). The Balkan peninsula and Turkey bear the highest burden; however, global warming facilitates the spread of the tick vector into new habitats through transport by migratory birds, as exemplified by a recent outbreak in Spain and the appearance of infected ticks in Italy (*6–8*).

CCHFV is a member of the *Orthonairovirus* genus in the *Nairoviridae* family of the Bunyavirales order of viruses with a segmented, negative-strand RNA genome (*9*). New human pathogens in the *Orthonairovirus* genus (termed nairoviruses from here on) continue to be identified (*10*), highlighting the need for high–resolution structural information to guide antiviral strategies. The Bunyavirales order also includes other pathogenic arthropod-borne viruses ("arboviruses") such as the Rift Valley fever virus (RVFV, *Phlebovirus* genus, *Phenuiviridae* family) and rodent-borne viruses such as Andes virus (*Orthohantavirus* genus, *Hantaviridae* family). CCHFV infects host cells through its envelope glycoproteins Gn and Gc, which form a locally ordered lattice of heterodimers on the virus surface after they are cleaved from a poly-glycoprotein precursor by host proteases (Fig. 1A) (*11–13*). Entry into

target cells takes place by receptor-mediated endocytosis (*14*), with the acidic environment of the endosome triggering dissociation of the Gn-Gc heterodimer and the surface lattice, followed by a conformational change of Gc into a trimer of hairpin structures to drive membrane fusion (Fig. 1B). As with most bunyaviruses, CCHFV Gc is predicted to be a class II membrane fusion protein (*11, 12*) and is the only known target of CCHFV-neutralizing antibodies (*15*).

We determined the x-ray structure of the CCHFV Gc postfusion trimer using two constructs at resolutions of 2.2 and 3.0 Å (table S1), as described in the materials and methods. The trimer revealed a typical class II fold, with each protomer adopting the characteristic postfusion hairpin conformation (*16*). The inner arm of this hairpin is composed of domains I and II (red and yellow, respectively; Fig. 1C) and forms a rodlike structure with the distal tip of domain II exposing loops *bc, cd,* and *ij*, also termed "fusion loops" as they form a nonpolar host–membrane insertion surface (HMIS) required to drive membrane fusion. The domain I and II rods interact about the threefold molecular axis along their entire length to make an elongated trimeric core. The outer arm of the hairpin is formed by domain III (blue) followed by the stem (magenta) running in an extended conformation to reach the HMIS, thus completing the hairpin by bringing the downstream C-terminal transmembrane segment (not included in our structure) next to the HMIS. The turn of the hairpin at the opposite end of the rod is made of a linker region connecting domains I and III (Fig. 1C, cyan). Domain III and the stem together fill the cleft between two neighboring subunits of the core trimer, contributing to the stability of the postfusion conformation of Gc. The overall arrangement of domains I and III is similar to that of the fusion proteins of other arboviruses such as phleboviruses (*17, 18*), flaviviruses (*19, 20*), and alphaviruses (*21*). This organization is different, however, in hantaviruses (*22, 23*) and rubella virus (*24*), which do not infect arthropods. In the class II fusion proteins of these mammal-specific viruses, domain III is exchanged between neighboring protomers in the trimer (fig. S1).

Among the most potently neutralizing human monoclonal antibodies (mAbs) targeting

[1]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA. [2]Institut Pasteur, Université de Paris, CNRS UMR 3569, Structural Virology Unit, 25-28 rue du Docteur Roux, Cedex 15, Paris, 75724 France. [3]CIRI-Centre International de Recherche en Infectiologie, Univ Lyon, Université Claude Bernard Lyon 1, Inserm, U1111, CNRS, UMR5308, ENS Lyon, 46 allée d'Italie, Lyon, 69007 France. [4]Institut Pasteur, Université de Paris, CNRS UMR 3528, Crystallography Platform C2RT, 25-28 rue du Docteur Roux, Cedex 15, Paris, 75724 France. [5]Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [6]Adimab LLC, Lebanon, NH 03766, USA. [7]Mapp Biopharmaceutical Inc., San Diego, CA 92121, USA.
*Corresponding author. Email: jmclellan@austin.utexas.edu (J.S.M.); felix.rey@pasteur.fr (F.A.R)
†These authors contributed equally to this work. ‡Present address: Centre for Structural Systems Biology, Leibniz-Institut für Experimentelle Virologie (HPI), Notkestraße 85, 22607 Hamburg, Germany. §Present address: Departments of Cell Biology and Microbiology, Harvard Medical School, Boston, MA, USA. ¶Present address: Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA.
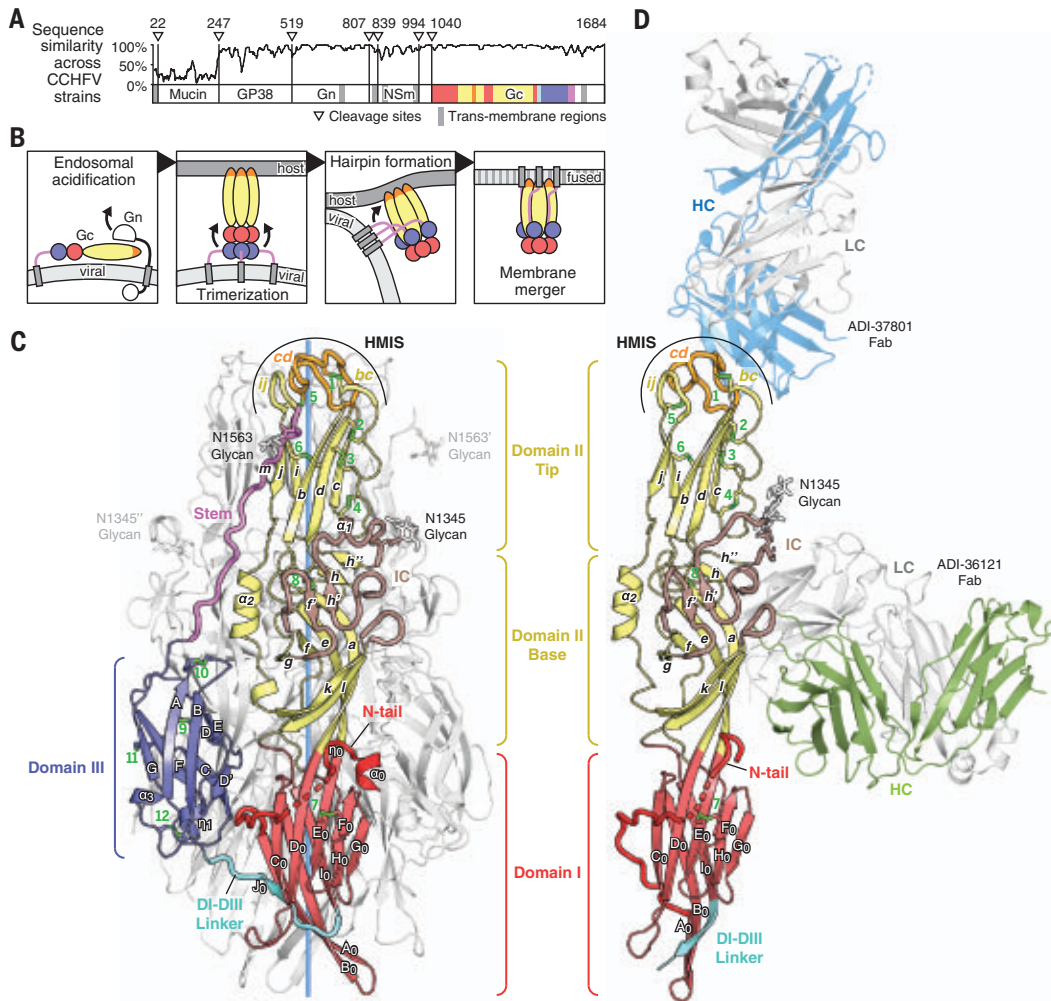
Fig. 1. Structures of CCHFV Gc. (**A**) Organization of the CCHFV glycoprotein precursor (**B**) Mechanism of bunyavirus class II membrane fusion proteins. (**C**) X-ray structure of the CCHFV Gc ectodomain in the postfusion conformation. The front protomer is colored according to domain, and the trimer axis is shown in light blue. Secondary structure elements and disulfide bonds (green numbers) are labeled. An ortho-nairovirus-specific insertions cluster (IC) is depicted in brown. (**D**) X-ray structure of the CCHFV Gc monomer in complex with the ADI-37801 and ADI-36121 Fabs. HC, heavy chain; LC, light chain.

CCHFV Gc, ADI-36121 and ADI-37801 were synergistic in coneutralization experiments (*15*). We determined the x-ray structure of both antigen-binding fragments (Fabs) in a ternary complex with monomeric Gc to 2.1-Å resolution (table S1) as described in the materials and methods. The structure showed a Gc monomer with the ADI-36121 Fab bound at the domain II base and the ADI-37801 Fab bound at the HMIS (Fig. 1D). The crystals provided interpretable electron density only for domains I and II, as well as part of the linker between domains I and III, indicating that the whole outer arm observed in the postfusion hairpin is mobile in the monomer. On the virion surface, however, the prefusion conformation of Gc is likely further stabilized by contact with Gn and neighboring Gn-Gc heterodimers. Compared with the Gc post-fusion trimer, the conformation of domain I in the monomer is different. In particular, the N- and C-terminal β strands, A_0 and J_0, display an altered topology (fig. S1A). A similar change in the conformation of domain I has been observed between the pre- and post-fusion structures of phlebovirus Gc (*17, 18, 25*)

(fig. S1B), indicating that the conformation of domain I in the CCHFV Gc monomer indeed corresponds to the prefusion form.

Unlike domain I, the conformation of the domain II tip in the Fab-bound monomer is similar to that seen in the postfusion trimer (Fig. 2A). In the flavivirus, alphavirus, and phlebovirus class II fusion proteins, the HMIS is formed exclusively by the *cd* loop (orange in our figures) (*16*). In hantaviruses, however, the HMIS is tripartite, with additional contributions from two adjacent loops, *bc* and *ij* (Fig. 2B) (*22*). CCHFV Gc has a similar tripartite configuration at its domain II tip, sharing a pattern of conserved residues with hantavirus Gc (Fig. 2C) despite an overall sequence identity of only ~20% between the two Gc orthologs. Fig. 2 compares the CCHFV Gc to that of Maporal virus (MPRLV), for which best-resolved pre- and postfusion hantavirus Gc structures are available (*22, 23, 26*). The main-chain conformation of the *bc, cd,* and *ij* loops is similar in the postfusion forms of the CCHFV and MPRLV Gc (Fig. 2, A and B, left panels), with a root-mean-square deviation (RMSD) of 0.8 Å over 29 Cα atom pairs. In

both cases, four conserved disulfide bonds (Fig. 2, A to C, green) stabilize the structure, two of which cross-link the *cd* loop with the *ij* and *bc* loops (Fig. 2C). In CCHFV, the HMIS conformation is further supported by a hydrogen bond network that involves the buried polar side chains of Asn^{1194} and Arg^{1189} of the *cd* loop in both the pre- and postfusion forms (Fig. 2, A to C). The equivalent residues in MPRLV, Asn^{769} and Asn^{764}, recapitulate the same interactions in the postfusion form (*26*) but are solvent-exposed in the prefusion Gn-Gc heterodimer, where nonpolar side chains such as MPRLV Trp^{766}, corresponding to CCHFV Trp^{1191}, are instead buried (Fig. 2B). It is likely that Gn locks the domain II tip in the conformation shown in the MPRLV prefusion Gn-Gc complex, and that release of Gn results in the HMIS conformation seen in the pre- and postfusion forms of CCHFV. This suggests that the Gc monomer observed in the ternary complex corresponds to an activated prefusion form of CCHFV capable of insertion into the host membrane.

To experimentally test the role of residues suggested by the structure to be important for

Gc function in membrane fusion, we established an assay to follow syncytia formation of cells expressing the CCHFV glycoproteins at their surface upon low-pH treatment. We used this assay to test single point mutations at the interface between domains I and III (at the turn of the postfusion hairpin) to explore their functional effect. Alanine substitutions of two conserved residues abrogated low-pH–triggered cell-cell fusion: His[1479] on domain III, which makes a salt bridge with Glu[1113] of domain I, and Trp[1068] in the N-terminal tail, which projects into a pocket at the domain I/II boundary (Fig. 2, D to E). Gc-derived linear peptides spanning the N-terminal tail (amino acids 1041 to 1060 and 1061 to 1080) around the functionally important Trp[1068] residue robustly react with CCHFV-positive human
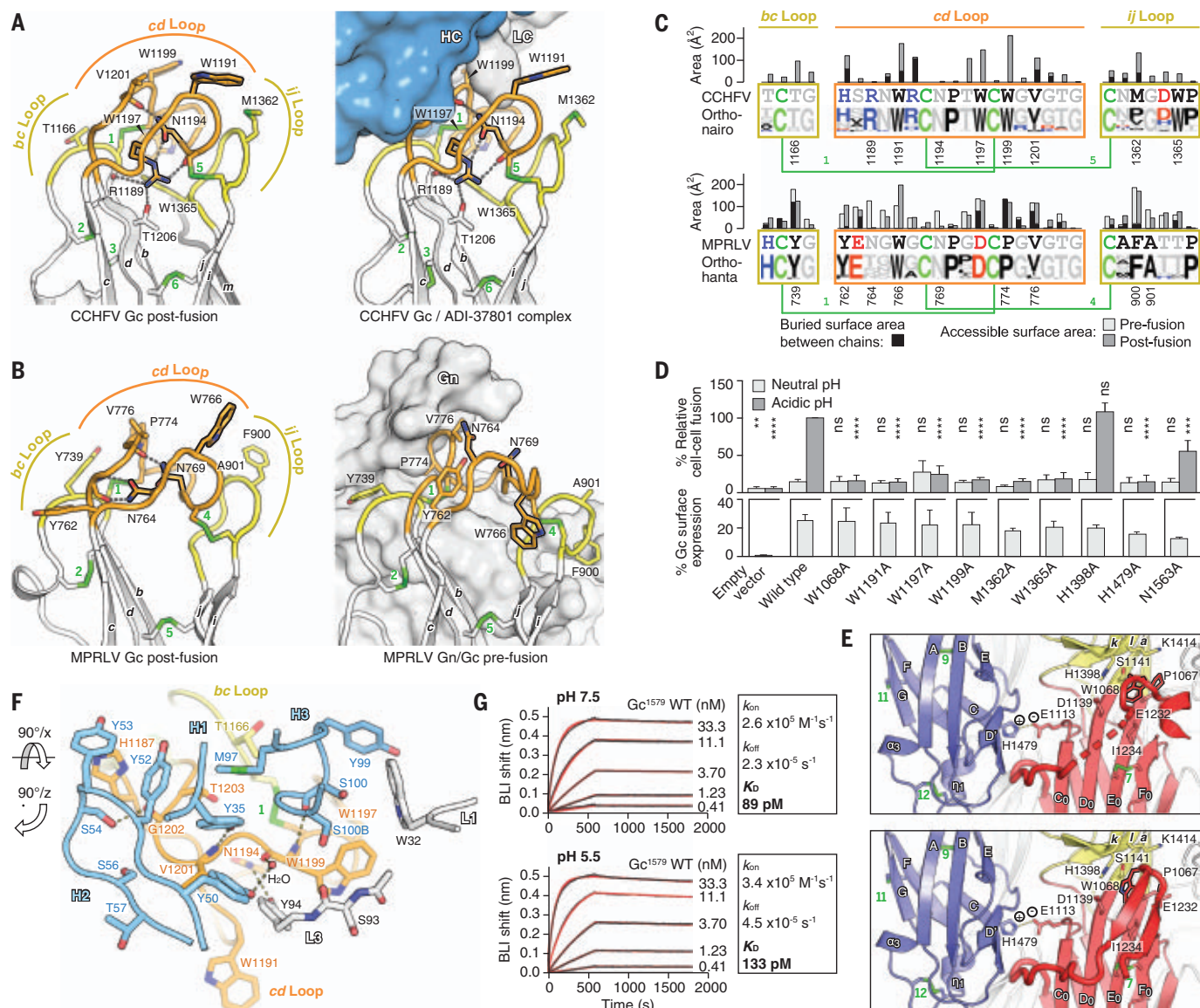
Fig. 2. ADI-37801 binds HMIS residues required for Gc-driven syncytia formation. (**A**) The CCHFV HMIS of the postfusion trimer (left) and in complex with ADI-37801 (right). In the left panel, W1191, W1197, and W1199, which were mutated to obtain the crystals, have been modeled for clarity. (**B**) Hantavirus fusion loops in the postfusion trimer forming the HMIS (left; PDB ID: 6y68, MPRLV structure) and in the prefusion Gn-Gc heterodimer, where the HMIS is not formed (right; PDB ID: 6y62) (15). (**C**) Fusion loop sequences of CCHFV Gc with consensus sequence logo for the *Orthonairovirus* (top) and *Orthohantavirus* (bottom) genera. The bar chart shows the exposed surface area per residue in pre- (hantavirus Gc) and postfusion (CCHFV and hantavirus Gc) structures. The accessible and buried surface per residue are represented in gray and black, respectively. Nonpolar residues, black; acidic, red; basic, blue; cysteines, green. (**D**) CCHFV Gc-induced syncytia formation by wild-type and indicated mutant Gc at neutral and acidic pH. The transfected cell surface expression is shown for each mutant below. (**E**) Details

of two alternative conformations of the N-tail and a pH-sensitive salt bridge between domains I and III. The helical conformation (top) is dominant, whereas the β-hairpin (bottom) is well defined in only two of the six polypeptide chains in the asymmetric unit of the monoclinic crystals obtained at pH 7.5. The view is the same as in Fig. 1C. (**F**) Interface between the ADI-37801 CDRs and the Gc fusion loops. The antibody heavy and light chain CDRs are colored blue and gray, respectively. CCHFV Gc is colored orange (*cd* loop) and yellow (*bc* loop). Polar interactions are denoted by dashed lines. (**G**) Biolayer interferometry (BLI) sensorgrams showing binding kinetics of CCHFV Gc[1579] to ADI-37801 at pH 7.5 (top) or pH 5.5 (bottom). WT, wild-type; K_D, dissociation constant; k_{on}, on rate constant; k_{off}, off rate constant. ns, not significant. ($P > 0.05$); **, $P < 0.01$; ***, $P < 0.001$; and ****, $P < 0.0001$). 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.

sera (*27*), suggesting this site as a potential target for neutralizing antibodies. On the contrary, we saw no effect resulting from alanine substitution of His[1398] at the binding pocket for the N-terminal tail and only a mild effect by alanine substitution of the glycosylation site Asn[1563] on the stem (*28*) (Fig. 2, D to E).

We also tested the role of HMIS nonpolar side chains. Mutation to alanine of the highly conserved Trp[1191], Trp[1197], and Trp[1199] exposed by the *cd* loop, as well as Trp[1365] and Met[1362] exposed by the *ij* loop (see Fig. 2C), strongly impaired low-pH–triggered syncytia formation relative to wild-type Gc when substituted individually (Fig. 2D). This result is in line with the functional effect of the corresponding residues of hantavirus Gc (Fig. 2B), which have been shown to be functionally required for target membrane insertion (*22*).

The residues exposed at the HMIS make up the epitope of mAb ADI-37801, which covers 627 Å² of surface area on Gc. Two-thirds of the epitope is buried by the three complementarity-determining regions (CDRs) H1, H2, and H3 of the heavy chain, and the remainder by the light-chain CDRs L1 and L3 (Fig. 2F). There are four hydrogen bonds at the epitope-paratope interface (table S2). The core of the epitope is formed by the *cd* loop, which contributes 10 amino acids, whereas the *bc* loop contributes an additional two. The residues critical for membrane fusion—Trp[1191], Trp[1197], and Trp[1199] of the *cd* loop—are an integral part of the ADI-37801 epitope (Fig. 2, A and F). Our structure is thus consistent with yeast-display–based epitope mapping, which identified Trp[1199] as critical for ADI-37801 binding (*15*).

Our ternary complex crystals grew at pH 5.6, suggesting that the complex of Gc and ADI-37801 remains stable in the endosome during viral entry. Through biolayer interferometry (BLI) we confirmed that ADI-37801 binding is insensitive to mildly acidic conditions (Fig. 2G). Taken together, the cell-cell fusion, structural, and kinetic data suggest that ADI-37801 inhibits endosomal membrane insertion of Gc by masking its fusion loops.

The x-ray structure showed that ADI-36121 binds laterally to the domain II base adjacent to the Asn[1345] glycan and covers 943 Å² of surface area on Gc, 63 and 37% of which are buried by the heavy and light chains, respectively, involving all six CDRs (Fig. 3, A to B). The epitope is composed of 22 residues featuring 13 hydrogen bonds and one salt bridge at the interface (table S2). The structure is consistent with the yeast-display–based mutagenesis screen that identified Leu[1307] and Ile[1229] as important for ADI-36121 binding (Fig. 3B) (*15*).

Structural comparison shows that the ADI-36121 epitope becomes entirely buried at the trimer interface upon formation of the postfusion trimer of Gc (Fig. 3, C to D). To experimentally confirm that the ADI-36121 epitope
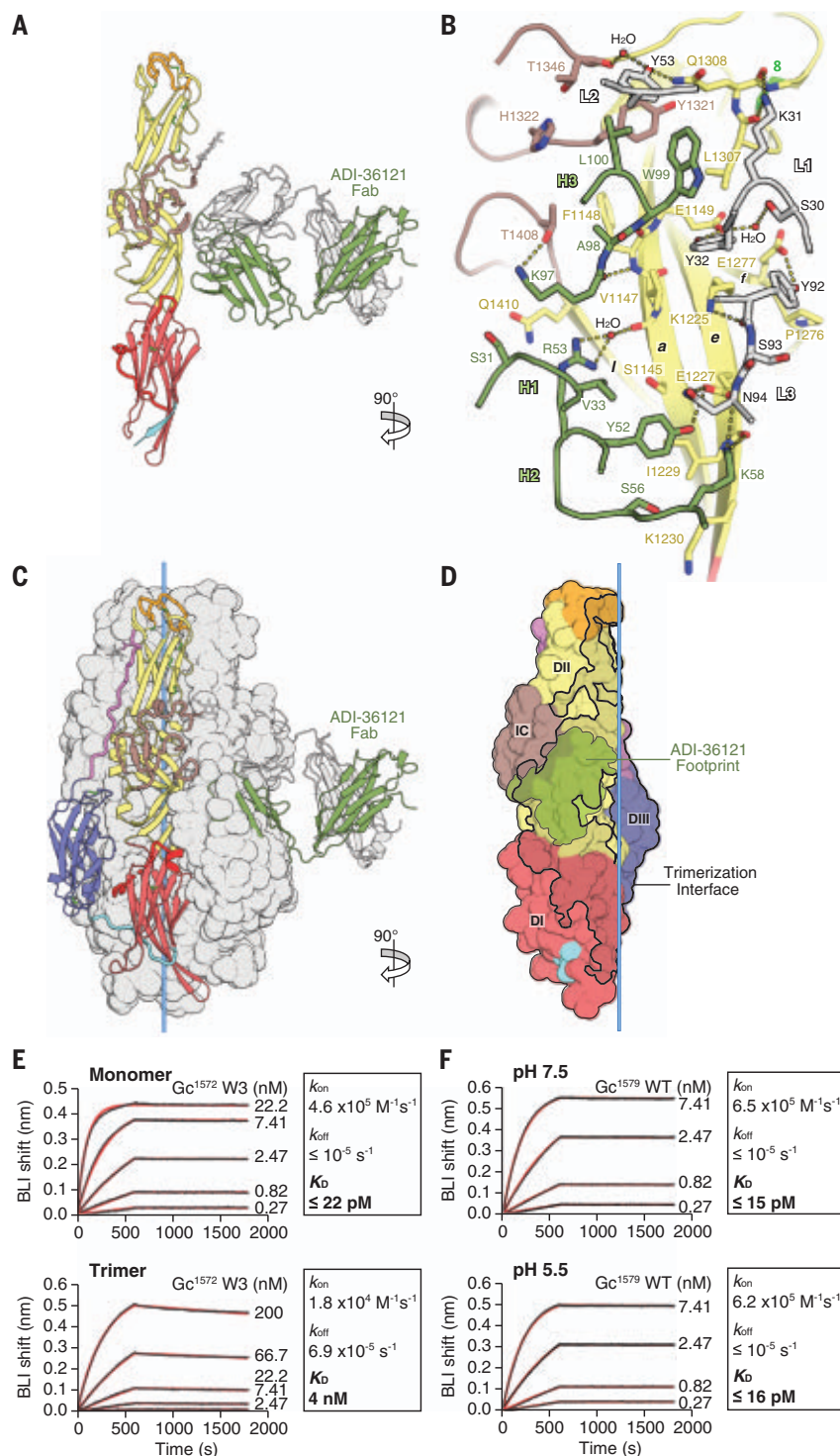

Fig. 3. The ADI-36121 epitope is buried at the trimer interface of the postfusion hairpin. (**A**) The CCHFV Gc monomer in a complex with the ADI-36121 Fab. (**B**) CDRs interacting with the Gc domain II base. Green and gray indicate heavy and light chains, respectively, and yellow indicates Gc domain II. Polar interactions are shown by dashes. (**C**) Superposition of the ADI-36121 complex with the Gc postfusion trimer. The trimer's front protomer is shown in ribbons colored according to domain, and the flanking protomers are shown as a white surface. (**D**) One protomer of the trimer shown as a surface colored according to domain, with the trimer interface outlined in black and the ADI-36121 footprint superposed in green, illustrating that the epitope is occluded in the trimer. (**E**) BLI sensorgrams showing binding kinetics of the monomeric fraction (top) or the trimeric fraction (bottom) of CCHFV Gc[1572] W3 to ADI-36121 at pH 7.5. (**F**) BLI sensorgrams showing binding kinetics of CCHFV Gc[1579] to ADI-36121 at pH 7.5 (top) or pH 5.5 (bottom). See materials and methods for details of the constructs used.
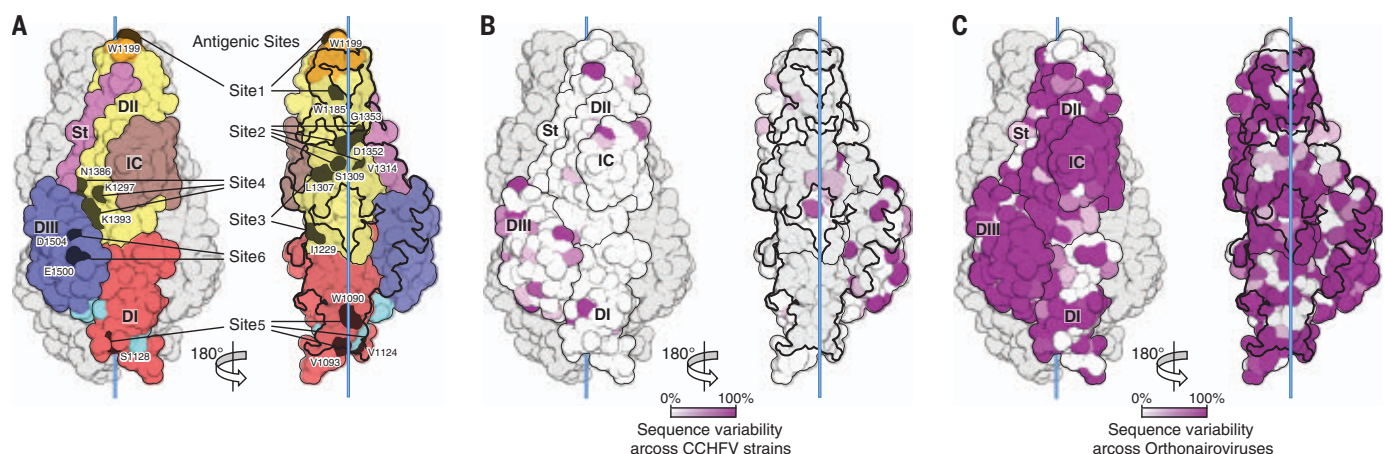
Fig. 4. The epitopes of CCHFV-neutralizing human antibodies map to Gc surfaces involved in driving membrane fusion. (**A**) Antigenic sites mapped on the surface of one CCHFV Gc protomer within the postfusion trimer. The trimer axis is shown in light blue. Only the front Gc subunit is shown in the right panel, after a 180° rotation about the trimer axis. The trimer interface is outlined in black. St, Gc stem region. (**B**) Sequence variability across 15 representative CCHFV strains (fig. S3), color plotted on the Gc surface. (**C**) Sequence variability across 14 species in the *Orthonairovirus* genus (fig. S4).

is inaccessible in the postfusion trimer, we used BLI to compare antibody binding with both monomeric and trimeric fractions of recombinant soluble Gc. The affinity of ADI-36121 for the monomeric fraction was ~200 times higher than that for the trimeric fraction (Fig. 3E). The observed residual ADI-36121 binding to the trimeric fraction suggests contamination of the sample with Gc monomers, as trimeric and monomeric fractions eluted in partially overlapping peaks in size-exclusion chromatography (fig. S2). Nevertheless, antibody binding likely outcompetes the trimerization process during viral infection, because the dissociation constant (K_D) of ADI-36121 for the Gc monomer is in the picomolar range at pH 7.5 and 5.5 (Fig. 3F). These data suggest that ADI-36121 neutralizes CCHFV by blocking Gc homotrimerization in the endosome and preventing membrane fusion.

The CCHFV-neutralizing human antibodies described previously had been tentatively assigned to six different antigenic sites by using a homology model for Gc based on the MPRLV Gc structure (*15*). Our experimental structures confirm the proposed distribution of the epitopes among the three Gc domains and also reveal the neutralization mechanisms by showing that they map to the HMIS or other surfaces buried during Gc-driven membrane fusion (Fig. 4A). The dominant antigenic site 1 maps to the *cd* loop (Fig. 4A), which is conserved across CCHFV strains as well as across members of the *Orthonairovirus* genus (Fig. 4, B to C, and figs. S3 and S4). Antigenic sites 2 to 4 map to the domain II base, with sites 2 and 3 at the trimer core interface of postfusion Gc (Fig. 4A). The most potently neutralizing antibodies—including ADI-36121—target site 3. Consistent with the high degree of conserva-

tion of its epitope across CCHFV strains (Fig. 4B), ADI-36121 displays highly potent cross-clade neutralization (*15*), which makes it a viable candidate for clinical development. It remains to be investigated whether this mAb would be effective against nairoviruses from other serogroups, such as the veterinary pathogens Dugbe virus or Nairobi sheep disease virus, which can potentially spill over to humans (*29*), as Gc from these viruses carries several point mutations in the epitope (figs. S4 and S5). Site 4 maps to the opposite face of domain II, near the interface with domain III and the stem in the postfusion structure (Fig. 4A), suggesting that antibody binding would inhibit hairpin formation. Similar to sites 2 and 3, site 5 overlaps with the Gc trimer interface but lies within domain I (Fig. 4A). Moreover, antibody binding to site 5 likely restrains the conformational change of domain I during fusion (fig. S1A). Finally, site 6 maps to domain III, where antibody binding may sterically inhibit its translocation for postfusion hairpin formation (Fig. 4A). In addition to human antibodies, this site is likely also targeted by the broadly neutralizing murine antibody 11E7, which has been mapped to a Gc fragment encompassing both domain III and the stem (amino acids 1443 to 1566) (*30*). Because the epitope was sensitive to chemical reduction, it can now be assigned to the disulfide-stabilized domain III. Because domain III contains more sequence polymorphisms across CCHFV strains than the other Gc domains (Fig. 4B), cross-clade neutralization by site 6 antibodies may be more limited compared with the other sites. Although inhibition of binding to the currently unknown entry receptor for CCHFV may also play a role in neutralization, our findings are consistent with a neutralization mechanism that inhibits membrane fusion by blocking insertion of

the HMIS into target membranes, by interfering with Gc trimerization, or by inhibiting postfusion hairpin formation.

Our structural data revealed that the HMIS of CCHFV Gc is at least transiently accessible on virus particles, as mAb ADI-37801 efficiently neutralizes the virus. However, the current paradigm is that the HMIS is protected from premature exposure by the companion protein Gn. The only available high-resolution structures of a bunyavirus Gn-Gc complex come from hantaviruses, and they indeed show that the conformation of the Gc domain II tip in interaction with Gn is such that the HMIS is not formed. Recent studies on Andes hantavirus have, however, shown a substantial degree of breathing, transiently exposing the HMIS at physiological temperatures (*31*). The strong structural similarity between their domain II tips (Fig. 2, A and C) suggests that comparable breathing dynamics can also be expected from CCHFV Gc. Because ADI-37801 neutralization was strain dependent (*15*) despite almost perfect conservation of the HMIS sequence across CCHFV strains (fig. S3), the breathing dynamics of the HMIS are likely controlled by sequences outside the fusion loops. Notably, strain-dependent breathing is also known to affect the neutralization potency of fusion loop antibodies in flaviviruses (*32, 33*).

Unlike the fusion loop antibody ADI-37801, the trimerization-inhibiting antibody ADI-36121 shows potent neutralization across CCHFV strains (*15*), indicating that accessibility of its epitope is not restricted by strain-dependent structural dynamics within the envelope. The ADI-36121 epitope on CCHFV Gc lies in the same position as the P-4G2 epitope on hantavirus Gc (fig. S6). Both antibodies bind to the same secondary structure elements on their

respective Gc targets (fig. S6C), at a surface patch that is involved in lateral interspike contacts on the hantavirus glycoprotein lattice (fig. S6, A and D). This surface patch becomes buried in the Gc postfusion trimer in both cases. Low-resolution studies of Hazara virus, a nonpathogenic nairovirus, showed tetrameric spikes arranged in a square surface lattice (*34*), similar to that of hantaviruses. The hantavirus surface lattice was visualized at a higher resolution (*26*) and was very different from the icosahedral T = 12 Gn-Gc lattice of the phlebovirus RVFV, for which relatively high-resolution structures are also available (*35*). Considering the similar square surface lattices of nairoviruses and hantaviruses and the structural similarity between the corresponding fusion proteins, it is reasonable to expect that comparable surfaces in CCHFV Gc are involved in lateral spike-spike contacts (fig. S6, B and E). It is possible, therefore, that ADI-36121 perturbs the long-range order of the CCHFV envelope in a similar way as was shown for P-4G2 (*36*). Higher-resolution cryo–electron tomography data on the nairovirus surface glycoprotein lattice are needed to identify the precise lateral spike-spike contacts and confirm the predictions illustrated in fig. S6. Our study nevertheless raises notable parallels between these two zoonotic viruses despite their different lifestyles and reservoirs—one arthropod-borne and the other transmitted by small mammals—thus highlighting the power of comparative structural studies in understanding common features of emerging viruses.

The combination of antibodies ADI-37801 and ADI-36121 displayed synergy in a neutralization assay (*15*). Moreover, a single dose of a bispecific antibody containing the variable domains of both ADI-36121 and ADI-37801 protected mice against CCHFV even when administered 24 hours after exposure, whereas the individual mAbs protected only in a prophylactic setting (*15*). To explain these findings, our structural analysis suggests that ADI-36121 binding indirectly influences the Gc fusion loop breathing dynamics by perturbation of the glycoprotein surface lattice in such a way that the HMIS becomes more exposed, allowing ADI-37801 to more easily recognize its epitope (fig. S6, E to F). Combination with ADI-36121 should therefore also broaden the reactivity of ADI-37801 with the various CCHFV strains, making these two antibodies strong candidates for therapeutic antibody cocktails. Describing CCHFV neutralization at the mechanistic level, our data guide the design of future therapeutic antibodies and will likewise support the design of protective CCHFV vaccines.

REFERENCES AND NOTES

1. O. Ergönül, *Lancet Infect. Dis.* **6**, 203–214 (2006).
2. K. Tsergouli, T. Karampatakis, A. B. Haidich, S. Metallidis, A. Papa, *J. Hosp. Infect.* **105**, 43–52 (2020).
3. E. Akinci, H. Bodur, H. Leblebicioglu, *Vector Borne Zoonotic Dis.* **13**, 429–437 (2013).
4. D. A. Bente *et al.*, *Antiviral Res.* **100**, 159–189 (2013).
5. M. S. Mehand, F. Al-Shorbaji, P. Millett, B. Murgue, *Antiviral Res.* **159**, 63–67 (2018).
6. E. Mancuso *et al.*, *Emerg. Infect. Dis.* **25**, 1418–1420 (2019).
7. A. Negredo *et al.*, *N. Engl. J. Med.* **377**, 154–161 (2017).
8. A. Negredo *et al.*, *Emerg. Infect. Dis.* **25**, 1177–1184 (2019).
9. P. Maes *et al.*, *Arch. Virol.* **164**, 927–941 (2019).
10. J. Ma *et al.*, *Nat. Med.* **27**, 434–439 (2021).
11. P. Guardado-Calvo, F. A. Rey, *Adv. Virus Res.* **98**, 83–118 (2017).
12. R. J. G. Hulswit, G. C. Paesen, T. A. Bowden, X. Shi, *Viruses* **13**, 353 (2021).
13. M. J. Vincent *et al.*, *J. Virol.* **77**, 8640–8649 (2003).
14. A. Albornoz, A. B. Hoffmann, P.-Y. Lozach, N. D. Tischler, *Viruses* **8**, 143 (2016).
15. J. M. Fels *et al.*, *Cell* **184**, 3486–3501.e21 (2021).
16. Y. Modis, *Adv. Exp. Med. Biol.* **790**, 150–166 (2013).
17. P. Guardado-Calvo *et al.*, *Science* **358**, 663–667 (2017).
18. S. Halldorsson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7154–7159 (2016).
19. S. Bressanelli *et al.*, *EMBO J.* **23**, 728–738 (2004).
20. Y. Modis, S. Ogata, D. Clements, S. C. Harrison, *Nature* **427**, 313–319 (2004).
21. D. L. Gibbons *et al.*, *Nature* **427**, 320–325 (2004).
22. P. Guardado-Calvo *et al.*, *PLOS Pathog.* **12**, e1005813 (2016).
23. S. Willensky *et al.*, *PLOS Pathog.* **12**, e1005948 (2016).
24. R. M. DuBois *et al.*, *Nature* **493**, 552–556 (2013).
25. M. Dessau, Y. Modis, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1696–1701 (2013).
26. A. Serris *et al.*, *Cell* **183**, 442–456.e16 (2020).
27. A. Fritzen *et al.*, *PLOS Negl. Trop. Dis.* **12**, e0006598 (2018).
28. B. R. Erickson, V. Deyde, A. J. Sanchez, M. J. Vincent, S. T. Nichol, *Virology* **361**, 348–355 (2007).
29. N. J. MacLachlan, E. J. Dubovi, Eds., in *Fenner's Veterinary Virology* (Academic Press, ed. 5, 2017), pp. 41–424.
30. A. A. Ahmed *et al.*, *J. Gen. Virol.* **86**, 3327–3336 (2005).
31. E. A. Bignon, A. Albornoz, P. Guardado-Calvo, F. A. Rey, N. D. Tischler, *eLife* **8**, e46028 (2019).
32. K. A. Dowd, T. C. Pierson, *Annu. Rev. Virol.* **5**, 185–207 (2018).
33. F. A. Rey, K. Stiasny, M. C. Vaney, M. Dellarole, F. X. Heinz, *EMBO Rep.* **19**, 206–224 (2018).
34. E. K. Punch *et al.*, *J. Biol. Chem.* **293**, 9937–9944 (2018).
35. S. Halldorsson *et al.*, *Nat. Commun.* **9**, 349 (2018).
36. I. Rissanen *et al.*, *eLife* **9**, e58242 (2020).

Real-Time PCR Systems

The CFX Opus 96 Dx Real-Time PCR System and the CFX Opus 384 Dx Real-Time PCR System are the latest additions to Bio-Rad Laboratories' portfolio of real-time PCR instruments. They offer accurate, precise quantification to improve assay development and workflow efficiencies and productivity for diagnostic testing and research. They are open systems that can multiplex up to five targets to deliver efficient in vitro diagnostic assay development and testing, and provide precise quantification, improved thermal cycler performance, and easy-to-use desktop management and analysis software with integrated security features.

Bio-Rad Laboratories
For info: +1-800-424-6723
bio-rad.com/cfxopusdx

Cell-Repellent Surfaces for Spheroid Cultures

Lipidure-COAT plates are one of the most effective tools for state-of-the-art 3D spheroid and embryoid body (EB) cell cultures. There are also papers reporting usage of Lipidure-COAT plates to differentiate EBs into organoids. The Lipidure coating provides a superior low-attachment cell culture condition for the formation of single spheroids in each well of multiwell plates. Spheroid and EB cell cultures are typically based on the spontaneous formation of an aggregate of cells in an environment where cell–cell interactions dominate over cell–substrate interaction. Lipidure powder is a biocompatible, hydrophilic white powder consisting of 2-(methacryloyoxy) ethyl 2-(trimethylammonio)ethyl phosphate-n-butylmethacrylate copolymer. The building unit of this copolymer is 2-(methacryloyoxy) ethyl phosphorylcholine (MPC) monomer. Lipidure mimics the cell membrane surface, and its molecular structure is the key for high hydrophilic nature and extremely low toxicity.

AMS Biotechnology
For info: +1-617-945-5033
www.amsbio.com/lipidure-coat.aspx

Exosome and EV Fluorescent Stains

ExoBrite EV Membrane Staining Kits are designed for bright and specific fluorescent staining of exosomes or extracellular vesicles (EVs) with little to no background. The stains are validated for flow cytometry, but also have potential applications in superresolution imaging or other exosome-detection platforms. In addition, ExoBrite stains are suitable for both purified and bead-bound exosomes and can be combined with antibody staining for multiparameter analysis. The kits are available with two-color options for the fluorescein isothiocyanate (FITC) or phycoerythrin (PE) channels. More color options for blue and red channels are coming soon.

Biotium
For info: +1-800-304-5357
www.biotium.com

Superresolution Confocal Microscope Module

CrestOptics announces the launch of DeepSIM, a superresolution module designed to enhance the imaging capabilities of confocal microscope systems. The DeepSIM module can be used with Crest's X-light V3 spinning disk or with alternative confocal systems, to offer high-resolution images for life science researchers. It is compatible with any existing upright or inverted microscope with a camera port. It's as easy to use as a confocal microscope, enabling scientists to access deep data from their biological samples. DeepSIM works with sample thicknesses comparable to those used in confocal microscopy, giving superresolved data over 50 µm Z in depth in nonclarified samples. This means that more meaningful data can be obtained from native heterogeneous complex samples using routine preparation protocols. DeepSIM enables the effortless study of live-cell dynamics through a temporal resolution greater than 10 frames per second (fps) [1,024×1,024 pixel field of view (FOV)], allowing biological changes to be tracked at cellular and subcellular levels.

CrestOptics
For info: +39-(0)-06-61660508
crestoptics.com/deepsim

Flow Chemistry Support Stands

Designed with the needs of process chemistry laboratories in mind, the ReactoMate ATOM support is suitable for all sizes of vessels up to 30,000 mL within a compact footprint and with all-round visibility and access. The ATOM support features an innovative design allowing you to easily adjust the height of your reaction vessel with just one hand. With a chemically resistant build and overhead stirrer support, the ATOM support allows you to reach pilot-scale reactions all within the footprint of a benchtop system. Alongside the standard configurations shown, the ATOM support is available with several extra options including castors, drip trays, and clamps for additional vessels.

Asynt
For info: +44-0-1638-781-709
www.asynt.com/product/reactomate-atom-support

Liquid Chromatography Flowmeter

TESTA Analytical Solutions e.K. has launched a liquid chromatography flowmeter that allows users to set the optimal integration time to suit different applications. It also enables users to store different calibration factors onboard for different solvents, while saving a clear text name of the solvent in use. Compact and conveniently powered via a USB connection, TESTA Analytical's Liquid Chromatography Flowmeter uses a thermal flow sensor that is not only extremely accurate, sensitive, and high-resolution, but also offers the advantage of being noninvasive. As the sensor does not interfere with the measurement, it can operate over a wide dynamic range and has been demonstrated to be the perfect real-time flow-monitoring tool for liquid chromatography applications, including calibrating the flow rate of HPLC pumps, troubleshooting defective systems, and performing continuous validation for regulated products.

TESTA Analytical Solutions
For info: +49-(0)-30-864-24076
www.testa-analytical.com/flowmeter-request.html

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Dartmouth
GEISEL SCHOOL OF MEDICINE

TENURE-TRACK FACULTY POSITION IN BIOMEDICAL DATA SCIENCE

The Department of Biomedical Data Science and the Center for Quantitative Biology at the Geisel School of Medicine at Dartmouth invite applications for a tenure-track appointment at the rank of **Assistant Professor** as part of a targeted expansion of Dartmouth's Biomedical Data Science Faculty. The Department of Biomedical Data Science offers a dynamic and interactive environment, with a commitment to research excellence. The Center for Quantitative Biology is focused on developing methods and faculty expertise in single cell genomics at Dartmouth. Dartmouth faculty members benefit from highly collaborative and collegial interactions across campus, inclusive of departments at Geisel, the Dartmouth Faculty of Arts & Sciences, the Norris Cotton Cancer Center, and the Thayer School of Engineering.

Successful applicants will develop vigorous and innovative research programs focused on bioinformatics, biostatistics, or biomedical informatics to study and understand a wide range of biological or clinical questions, including but not limited to single-cell analyses, systems biology, spatial transcriptomics, genomics, epigenomics, and visualization of complex data. Applicants with expertise in machine learning, artificial intelligence, computational biology, or biostatistics are strongly encouraged. Candidates are expected to have a strong record of scholarship, to develop an independent research program, and to participate in graduate-level teaching. Evidence of an ability to secure extramural funding is desirable. Applicants should have a PhD, MD, or equivalent degree with relevant post-doctoral training.

A generous start-up package as well as access to state-of-the-art research facilities will be provided. Individuals will be given opportunities for mentorship and teaching in relevant PhD graduate programs (e.g., Program in Quantitative Biomedical Sciences, Program in Molecular and Cellular Biology). Training grants (e.g., the Burroughs Wellcome Fund Big Data in the Life Sciences for Unifying Population and Laboratory-based Sciences Training Grant, and the Quantitative Biomedical Sciences Training Grant) are available to support trainees.

Opportunities for innovative research methods development and application are facilitated through Dartmouth's high-performance computing resources (e.g., Discovery), access to Dartmouth's Shared Resources, and interactions with existing NIH Centers for Biomedical Research Excellence (COBREs) at Dartmouth. Dartmouth is located in the picturesque Upper Connecticut River Valley on the New Hampshire and Vermont border, a vibrant, academic and professional community offering excellent schools, lively arts, and an unmatched quality of life in a beautiful, rural setting. Boston, MA; Burlington, VT; and Montreal, QC are all within a few hours' drive.

Dartmouth College is an equal opportunity/affirmative action employer with a strong commitment to diversity and inclusion. We prohibit discrimination on the basis of race, color, religion, sex, age, national origin, sexual orientation, gender identity or expression, disability, veteran status, marital status, or any other legally protected status. Applications by members of all underrepresented groups are encouraged. Dartmouth is highly committed to fostering a diverse and inclusive population of students, faculty, and staff. We are especially interested in applicants who are able to work effectively with students, faculty, and staff from all backgrounds, including but not limited to racial and ethnic minorities, women, individuals who identify with LGBTQ+ communities, veterans, individuals with disabilities, individuals from lower income backgrounds, and first-generation college graduates.

Applicants should upload a cover letter addressed to Professor Jiang Gui, search committee Chair, along with a curriculum vitae, a one-page research statement, teaching statement (500 words or less), diversity statement (500 words or less), and three letters of recommendation from referees to **apply.interfolio.com/100008.**

Screening of applications will begin **Jan 10, 2022** and continue until the position is filled.

andlinger center
for energy + the environment

DIRECTOR, ANDLINGER CENTER FOR ENERGY AND THE ENVIRONMENT

Princeton University seeks an internationally renowned researcher to become director of the Andlinger Center for Energy and the Environment.

The Andlinger Center for Energy and the Environment brings an interdisciplinary perspective to the realm of energy and the environment. The center seeks to provide leadership in research to examine and address the global energy and environmental challenges we face today to ensure a sustainable future. It aims to facilitate engagement and lower barriers to collaboration and innovation through its programs.

The university seeks a distinguished scholar with a strong record of leadership and service, and a vision in transforming how we power the world. The successful candidate will hold a doctorate degree and have the scientific breadth and vision needed to advance a multidisciplinary research center. With substantial experience working in energy technologies and/or energy policy, the director will have considerable understanding of energy challenges at the national and international scale. The director's overarching responsibilities are to build upon the tradition of excellence of the Andlinger Center and enhance its visibility in the research community and industrial and public policy sectors. The director will lead the development of the center's strategic plan and develop objectives and goals that build research teams and form new partnerships and new ways of thinking to solve some of the world's most difficult energy and environmental challenges.

Applicants who are committed to a culturally diverse environment and demonstrate a willingness to nurture an inclusive Princeton community, particularly women and members of traditionally underrepresented minority groups, are encouraged to apply.

Review of applications will begin **February 1** and will continue until the position is filled.

To learn more about the position and how to apply, go to: **https://acee.princeton.edu/opportunities.**

myIDP:
A career plan customized
for you, by you.

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.

- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.

- Articles and resources to guide you through the process.

- Options to save materials online and print them for further review and discussion.

- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.

- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers In partnership with:

AAAS

FASEB
Federation of American Societies
for Experimental Biology

University of Massachusetts
UMASS. Medical School

UCSF
University of California
San Francisco

UIC

BURROUGHS
WELLCOME
FUND

Technische Universität Braunschweig

Zertifikat seit 2007
audit familiengerechte
hochschule

The **Faculty of Electrical Engineering, Information Technology, Physics,** Institute for Theoretical Physics offers a

University professorship (m/f/d) for "Computational Astro- and Space Physics" Salary-Gr. W2

to begin on **October 1st, 2023**.
The successful candidate will join the Institute for Theoretical Physics and is expected to teach any undergraduate course in theoretical physics. They should have a general background in the field of Computational Astro- and Space Physics and should have specialized in one or more of the following areas: Space Plasma Physics, Physics of Planetary Bodies, Solar and Heliospheric Physics, or High-Energy Particles in the Solar System.

The appointee is expected to become an integral part of the existing theoretical physics community but should broaden its current foci beyond mathematical and condensed matter physics. They must have a track record of well-recognized, independent research, an extensive publication list, and should have demonstrated the ability to secure external funding. Having participated in international projects or missions is advantageous. We expect that the successful candidate will actively participate in collaborative research efforts in Space Science at the TU Braunschweig and will engage with institutions such as Deutsches Zentrum für Luft- und Raumfahrt (DLR), the Max Planck Institute for Solar System Research (MPS), or the Physikalisch-Technische Bundesanstalt (PTB).

Vacancy offer: https://www.tu-braunschweig.de/en/vacancies

Please send your application until **February 28th, 2022** using the **online form** of TU Braunschweig.

N Northeastern University College of Engineering

With over **195** tenured/tenure-track faculty, and **17** multidisciplinary research centers and institutes, and funding by eight federal agencies, Northeastern's College of Engineering is in a period of dynamic growth. Our emphasis on interdisciplinary, transformative research–tied to Northeastern's unique history of industry collaboration through the university's signature cooperative education program–enables partnerships with academic institutions, medical research centers, and companies near our centrally located Boston campus and around the globe.

Learn more and apply at coe.northeastern.edu/faculty-hiring

Consideration will be given to candidates at the assistant, associate, and full professor levels; successful applicants will be expected to lead internationally recognized research programs aligned with one or more interdisciplinary research themes. We are also seeking to recruit and support a broadly diverse community of faculty and staff, and strive to foster an inclusive culture built on respect that affirms inter-group relations and builds cohesion. Applicants will be asked to submit a diversity statement discussing how they view their contributions to sustainment and improvement of diversity in the college and community at large.

Northeastern University is an equal opportunity employer, seeking to recruit and support a broadly diverse community of faculty and staff. Northeastern values and celebrates diversity in all its forms and strives to foster an inclusive culture built on respect that affirms inter-group relations and builds cohesion. All qualified applicants are encouraged to apply and will receive consideration for employment without regard to race, religion, color, national origin, age, sex, sexual orientation, disability status, or any other characteristic protected by applicable law.

To learn more about Northeastern University's commitment and support of diversity and inclusion, please see: northeastern.edu/diversity.

Beyond Title IX

"**H**ave you ever thought about taking a leave of absence?" The dean's question surprised me. I had never imagined I might need to take a break from my Ph.D. But after enduring an abusive adviser and a grueling Title IX investigation, I had discovered some experiences aren't easy to put behind you. I had also recently failed my qualifying exams a second time, which meant I was on the brink of being kicked out of graduate school. Something had to change.

Up to that point, my doctoral experience had been volatile: I was one of many complainants in a Title IX investigation against my first Ph.D. adviser, which looked into allegations of sexual misconduct. My adviser ended up resigning. But the abuse and the aftermath were the most difficult experiences of my life. I met and communicated with the investigator dozens of times over the course of a year, which was emotionally challenging and took time away from my research. I was also tormented by a stream of anonymous texts and emails from people who thought I wasn't cooperating with the investigation.

Still, I plowed forward. I had an overwhelming urge to prove myself—to demonstrate that I belonged in my program, and that my former adviser had not affected my ability to become a scientist. I didn't want to miss out on opportunities and stall my research. I was nervous to be away from my lab mates, who had become a strong support system. A leave of absence seemed out of the question. I was afraid I'd be seen as weak, be forgotten, or never come back.

I didn't realize I was setting myself up to implode. I could barely keep up with my workload and increasingly fell behind on deadlines. My physical health was a wreck. My weight dropped dangerously low. What little sleep I got was littered with nightmares. I finally began therapy and was told I suffered from post-traumatic stress disorder. But I continued to use work as a coping mechanism.

I scheduled my qualifying exams only a couple of months into the investigation, and I promptly failed them. Ten months later, I tried again, but that didn't go well either. I felt I was up against a wall.

When I met with the dean of the graduate school, who was accompanied by the university's Title IX coordinator, they were sympathetic and offered me a deal: I'd be allowed a third chance to take my exams, but only after I took time off. They also arranged for an assistant dean—a psychology professor

"Targets of sexual assault and harassment need time to process what they've been through."

who is an expert in trauma—to meet with me via Zoom regularly during my leave and provide support.

Taking a break terrified me. But deep down, I knew it was a good idea. I was also fortunate to have a partner with a stable job and parents who could help us financially during the time I wasn't being paid.

The hiatus helped me embrace some of the quiet I had been desperately avoiding and concentrate more on my partner and our child. I underwent trauma therapy, my sleep patterns began to stabilize, and I rapidly gained back weight. It wasn't easy, though. As the demanding grad student lifestyle to which I was accustomed shrunk away, there were moments when I desperately missed my work. I am an impatient person, and taking time to deal with my grief was not easy.

A year later, I returned to my program. My department gave me a few months to reacclimate before I sat for my exams for a final time. That time, I passed. I also made progress in other ways: I began to submit first-author papers, was awarded a grant, and won an award for my research. I don't think I would've accomplished those things had I not taken a break.

The impact of an abusive adviser doesn't disappear once the abusive behavior stops. Targets of sexual assault and harassment need time to process what they've been through, and it's important for universities to give them the space to do that. Fortunately, I had a community of administrators and mentors whose efforts didn't stop once the Title IX investigation was over. They continued to support and advocate for me—and are a big reason I wasn't lost from the system. I hope those in positions of power do what they can to support others in similar situations. It can make all the difference in the world. ∎

Savannah A. Collins-Key is a Ph.D. student at the University of Tennessee, Knoxville. Send your career story to SciCareerEditor@aaas.org.

TRILLIONS OF MICROBES ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators, under the age of 35, who are working on the functional attributes of the microbiota. The research can include any organism that has potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent scientist, chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and $25,000 (USD). Submit your research essay today.

Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 1/24/22 at www.science.org/noster
Sponsored by Noster, Inc