THIS WEEK

Double vision

Scientists must ensure that they take the lead in the ethical debate surrounding the therapeutic use of stem cells derived from human clones.

Last week's announcement in *Cell* that a team in Oregon has successfully derived stem cells from a cloned human embryo raises many questions. Researchers must do their best to answer them. If they do not, then they risk a repeat performance of the misinformation, confusion and distraction that followed the birth in 1996 of the first mammalian clone, Dolly the sheep. Nonsense such as rumours that Saddam Hussein was going to copy himself, claims by the Raëlian sect that it had produced a cloned girl named Eve and declarations by physicist and entrepreneur Richard Seed that he would produce hundreds of thousands of human clones generated an environment of fear that distracted from the real debate and set back research.

The reaction to the Oregon team's success has been predictably mixed. Some have lauded the achievement, others have been less enthusiastic. "Barbaric," wrote one critic. Another called it a "terrible injustice".

It is true that the research faces ethical controversy on three fronts: egg donation, embryo destruction and cloning. The work needs eggs to be donated by women, who must undergo a procedure that carries risks of complications. If the research ultimately produces promising medical therapies, will we see a group of women marshalled to feed this therapeutic intervention? The procedure also requires the destruction of early-stage embryos, which many oppose, usually on religious grounds. And many criticize the fact that the procedure starts down the same path as reproductive cloning, suggesting that to advance the technology will inevitably result in reproductive cloning of humans.

These are serious points and they need to be considered. But the reality is that, if anything, the latest research suggests that the chance of creating a cloned human is more remote than it seemed two weeks ago. No group has put more effort into trying to clone monkeys reproductively than the Oregon scientists. Yet after nearly two decades of trying, they have yet to succeed. They conclude that their method, optimized for producing embryonic stem-cell lines from cloned embryos, probably cannot be used for reproductive cloning.

Still, given the controversy that surrounds this work, some argue that researchers should stick to alternatives such as induced pluripotent stem (iPS) cells. Like cloned stem cells, these offer patient-specific cells with embryo-like developmental potential, but crucially they do not require egg recruitment, embryo destruction or cloning. Leading iPS-cell researchers are now keen to compare these cells with the cloned stem cells to help them to evaluate and improve iPS cells. Until that work is done, it will be hard to say whether iPS cells are, or can be, reprogrammed to function as effectively cloned cells.

But this research will move slowly, at least in the United States. Because production of the cloned cells requires the destruction of embryos specifically created for research, US laboratories funded by the taxpayer through the National Institutes of Health cannot use the technique, nor can they work with the stem cells created by it. Progress will depend on the few laboratories with sufficient private funds.

Whatever the outcome of those investigations, there are some clinical applications for which cloned stem cells could be the best option. Cloned cells contain the nuclear DNA of whoever is being cloned but they also hold the mitochondrial DNA from the egg donor, which means that they could be used to correct defective mitochondrial DNA. That may pave the way to treatments for genetic diseases such as Leber's hereditary optic neuropathy. Such applications should not be hyped. Like any stem-cell therapy, the hurdles are high. Is it possible to get functioning cells to the right place and to make sure that they continue to work? Can enough of the cells with faulty mitochondrial DNA be replaced by the cloned ones?

> *"Armies of cloned despots are the last thing we need to worry about."*

Experiments that follow up on the Oregon team's advance will be especially valuable. If they don't prove that the technique is worth the controversy, it will no doubt fall into disuse. If it does prove to have therapeutic promise, questions of egg donation will need to be re-evaluated. Either way, armies of clones or cloned despots are the last thing we need to worry about. ∎

Uncertain times

Budgetary delays exacerbate dire outlook for US research.

Nearly three months into the across-the-board budgetary cuts known as sequestration, the US Congress seems as petulant and ineffectual as ever. There is no end in sight to the national economic stalemate driven by Senate Democrats, House Republicans and President Barack Obama. Congress has made only a few half-hearted and piecemeal attempts to fix some of the sequestration cuts, such as loosening restrictions on Federal Aviation Administration funds so that air-traffic controllers can get back to work. No such deals have been forthcoming for science agencies or research.

Before sequestration took effect on 1 March, science advocates had warned that the cuts would have dire consequences (see *Nature* **494,** 401–402; 2013). Major agencies such as the National Science Foundation would issue at least 1,000 fewer grants. Workers for federal agencies and at national laboratories would be forced to take unpaid leave, or even be laid off altogether. An entire generation of young scientists would be driven out of the field and into other careers.

Those grim predictions — or at least some of them — are now

starting to become reality. It is still too early to tell what the final effects will be, but major agencies are collecting data on how sequestration is affecting their grant recipients. Some are also taking the informal route: National Institutes of Health head Francis Collins, for instance, put out a Twitter request for stories at #NIHSequesterImpact. Tales flooded in of grants cut, delayed or denied altogether, to the point that lab heads are hiring fewer staff and delaying purchasing major equipment or other supplies.

Of course, it is hard to know which effects result from the sequestration and which are caused by general financial belt-tightening. But Washington's inability to reach a budget deal underlies nearly all the economic uncertainty plaguing US science agencies today. The last time Congress passed a proper budget was for fiscal year 2012, which ended last September. Since then, agencies have been operating on temporary, unsettled numbers, topped by a roughly 5% slash from sequestration. Then Obama was two months late in releasing his proposed budget for fiscal year 2014 (see *Nature* **496,** 277–279; 2013). Even seasoned budget wonks call the current situation the worst in recent memory.

As uncertainty swirls around the future, budget sequestration is hitting science agencies now. Some of the earliest and most drastic cutbacks have come in the field of Earth monitoring (see page 419). A small fraction of the 8,000 national streamflow gauges are being shuttered, because the US Geological Survey (USGS) can no longer pay its share of the bills. Surveys of western America's snowpack are also being slashed, eliminating crucial information about the water supply for many mountain states. All this is happening as wildfire season kicks off and as sequestration thins the ranks of federal firefighting teams. Meanwhile, the

Pavlof volcano in Alaska began erupting on 13 May, even as cutbacks at the Alaska Volcano Observatory mean that scheduled maintenance of seismic stations at remote volcanoes in the state is not carried out. Without seismic monitoring at many of these mountains, the USGS is, in essence, blind to some future eruptions.

Federal managers insist that crucial measurements will continue to be made. The National Oceanic and Atmospheric Administration (NOAA) plans to impose at least four days of unpaid leave through the summer. But officials say that mission-crucial operations will be unaffected. That includes forecasting severe storms, such as the tornadoes that swept across much of the nation's mid-section this week, as well as Atlantic hurricanes for which the season begins on 1 June.

"Even seasoned budget wonks call the current situation the worst in recent memory."

One can only hope that NOAA means it. Weather forecasting is expensive because it demands trained personnel. These are the same people who are being told to take unpaid leave, then do double duty when they return to work. Such incessant cutting can go on for only so long before morale suffers.

Already, programme managers at agencies such as the USGS and NASA are taking near-heroic steps to keep funding flowing to monitoring projects, shuffling money between programmes. But in the process they are mortgaging much of their future, by cutting back on savings for longer-term expenses such as upgraded computer servers or more personnel. Let us hope that Congress relieves the agencies, and soon, by getting its budgetary house in order. ∎

Shades of grey

It is risky to oversimplify science for the sake of a clear public–health message.

Who knows how many New Year's resolutions to diet were strangled at birth by Katherine Flegal? A group led by Flegal, an epidemiologist at the US National Center for Health Statistics in Hyattsville, Maryland, published a paper on 2 January that suggested that a bit of extra weight may be no bad thing (K. M. Flegal *et al. J. Am. Med. Assoc.* **309,** 71–82; 2013). In fact, the study — and the media coverage it generated — indicated that people deemed 'overweight' live longer than those who could fit into a pair of jeans a size or two smaller.

As we explore in a News Feature on page 428, the study was not the final word on the matter. Opponents have lined up to criticize the research and its findings. Walter Willett, chair of the nutrition department at the Harvard School of Public Health in Boston, Massachusetts, told US National Public Radio that "this study is really a pile of rubbish, and no one should waste their time reading it".

Critics of Flegal and of others who have reported similar findings take issue not just with the data used to make the claims, but the damage they feel that the claims will inflict on public-health efforts. It is much easier to gain weight than to lose it, and nobody disputes that to gain too much weight — to be obese — is bad for health. To discuss publicly results that threaten to undermine the simple message that 'fat is bad' will confuse doctors and the public, the critics say.

The debate goes beyond the health effects of weight. Just last week, the US Institute of Medicine issued a report that threatens to confuse doctors and the public on the issue of how much salt is bad for one's health. Convention holds that too much salt is harmful, and that the best approach is to reduce and limit intake as much as possible. Yet some studies show that some people can eat too little salt. Low sodium intake is also associated with poorer health in groups such as the over 50s and those with diabetes. However, too much salt remains bad for

health, the report said, as it increases the risk of heart disease. The simple message on salt, unfortunately, is that it can be good or bad for people, depending on who they are.

The political mantra on public-health advice is clear: don't send mixed messages. The media and those who get their information from the media prefer things in black and white: red wine is good for you; chocolate is bad for you. But, of course, science does not deal in black and white, hence the common criticism that scientists cannot make up their minds. One week, one group argues that extreme exercise is positive for health; the next week, a different set of researchers says the opposite.

Scientists like to believe that they can operate in shades of grey. But simple messages and themes are seductive. In the public-health and nutrition fields, the idea that excess weight might benefit health is called the 'obesity paradox' despite the fact that it does not focus on obesity, which everyone agrees is bad, and so is not a paradox. The 'mildly overweight paradox', presumably, sends mixed messages. When Willett dismissed the Flegal study as a "pile of rubbish" there were no shades of grey evident.

The problem with simple messages and black-and-white statements is that they tend to be absolutes and so the easiest to falsify. The line that the science of global warming is 'settled' must have seemed like a good idea at the time, and when taken to refer to the narrowest of scientific questions it is correct, but it was (fairly) interpreted as insistence that no queries remained. Even legitimate debates on outstanding issues — climate sensitivity, say — can now be painted as unsettling not just to the scientific position, but also to the policy response it demands.

It is easy to see why those who spend their lives trying to promote the health of others gnash their teeth when they see complex findings whittled down to a sharp point and used to puncture their message. It is more difficult, from a scientific perspective, to agree that these findings should not be published and discussed openly, warts and all, purely because they blend uncertainty into a simple mantra. Make things as simple as possible, Einstein said, but no simpler. And simple, black-and-white messages can cause confusion of their own. All things in moderation — and that should include the language we use. ∎

↻ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv
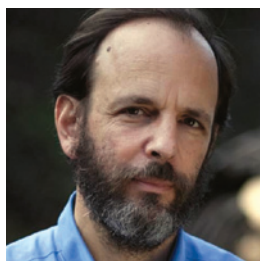
Pure hype of pure research helps no one

Congressman Lamar Smith hopes to 'improve' peer review by adding a layer of accountability, but his bill aims at imaginary ideals, argues **Daniel Sarewitz**.

What do you get when you cross science hype with conservative politics? The answer is the High Quality Research Act, a draft bill that would require the director of the US National Science Foundation (NSF) to certify the quality of every science project that the agency funds with its roughly US$5.6-billion research budget.

The proposed legislation was drawn up last month by Lamar Smith (Republican, Texas), the new chairman of the House of Representatives Committee on Science, Space, and Technology. He says it will help to "ensure that taxpayer-funded research projects are of high quality and benefit the American people".

Leaders of the US science community have a different view: they believe that the bill is an inappropriate and partisan intrusion of politics into science. The American Association for the Advancement of Science's ScienceInsider website portrayed the draft legislation as "the latest—and bluntest—attack on NSF by congressional Republicans", while explaining that the bill "would replace peer review at the [NSF] with a set of funding criteria chosen by Congress". Similar objections were voiced in letters sent to Smith by various former NSF officials and in statements by President Barack Obama's science adviser, John Holdren.

Actually, the bill doesn't say or imply anything at all about replacing peer review. It doesn't give Congress new powers over the NSF, nor does it impose on the NSF any new responsibilities. Yes, it requires that the NSF director "certifies" that projects funded by the agency are "in the interests of the United States to advance the national health, prosperity, or welfare", that the research "is of the finest quality, is ground breaking" and so on. But these vague requirements merely rearticulate the same promises that scientists and government agencies use all the time to justify their existence. Indeed, the NSF's own proposal review guidelines say that all NSF-funded science should have the potential to "benefit society or advance desired societal outcomes" and "have the potential to advance, if not transform, the frontiers of knowledge". So the bill seems mostly to be asking the NSF director to certify that the NSF is doing what the NSF says it already does.

In other words, it's not a very good bill, but neither is it much of a threat. In fact, it's just the latest skirmish in a long-running battle for political control over publicly funded science — one fought since at least 1947, when President Truman vetoed the first bill to create the NSF because it didn't include strong enough lines of political accountability.

The core argument of the scientific community and its leaders has always been that they are perfectly capable of ensuring accountability themselves, thank you. After all, the outcomes of basic research are unpredictable and

therefore politicians need only pour in the money and stand aside as the scientists make the world a better place.

This simplistic and self-serving myth was trotted out yet again in an 8 May letter from former NSF officials to Smith, which explained that the "history of science and technology has shown that truly basic research often yields breakthroughs … but that it is impossible to predict which projects (and which fields) will do that … Over the years, federal funding of basic research, using peer review evaluation, has led to vast improvements in health care, national security, and economic development."

Smith seems to believe all of this. Conservative politicians are typically loyal supporters of basic science because they recognize that it is one domain that does not provide sufficient incentives for private sector investment, so government must play a part. What Smith is doing is reminding the scientific community about Congress's authority to establish broad research spending priorities in the context of the ongoing budget gridlock, and he is reminding the NSF about its accountability to his committee: "The draft bill maintains the current peer review process and improves on it by adding a layer of accountability."

So the problem here isn't that Smith doesn't understand what the scientific community is saying, it's that after more than 60 years of hype about unpredictability and the inevitable benefits of pure science, he and other conservatives seem to understand and believe it all too deeply. Thus it's no surprise that when budgets are tight and progress towards achieving many goals — from curing cancer to revitalizing the nation's manufacturing base — is a lot slower than promised, a new conservative chairman would seek to make his mark by trying to make things run better. The grave danger here is not that he is going to interfere with peer review but that he will discover that the real world of science — in which progress is often halting and incremental, a lot of research isn't particularly innovative or valuable, and institutional arrangements are often more important than peer review or serendipity for determining the social value of science — doesn't match very well to the world on which he has been sold.

As of now, Smith's bill has not been formally introduced for congressional consideration, and perhaps it is best understood as a shot across the science community's bow. But with years of budgetary stress ahead, the science community needs to be much more assertive in articulating a vision for science that doesn't depend on continually rising budgets and isn't defended by resort to some mythical ideal of pure research. Hype is fine until people start to believe in it. ∎

> **HYPE** IS FINE UNTIL PEOPLE START TO **BELIEVE** IN IT.

Daniel Sarewitz *is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University and is based in Washington DC.*
e-mail: daniel.sarewitz@asu.edu

OPTICS

A poor man's hologram

Three-dimensional films have been a hit at the box office, but making them requires pricey equipment. Baoqing Sun at the University of Glasgow, UK, and his colleagues propose a simpler alternative. Instead of using cameras or lasers to achieve a three-dimensional (3D) effect, they illuminate an object with a rapidly changing pattern of black and white squares. Single-pixel detectors placed around the object capture light reflected in various directions, and an algorithm reconstructs a 3D image from these simple inputs.

Although the scheme takes several minutes to image a stationary object, it is cheaper than current 3D systems. In addition, it should work for wavelengths such as the far infrared that are beyond the reach of current technology.
Science 340, **844–847 (2013)**

ECOLOGY

Invasive insect's inner weapon

Ladybirds from central Asia act as a Trojan horse for a microscopic killer.

Farmers brought harlequin ladybird beetles (*Harmonia axyridis*, **pictured**) to North America and Europe to eat troublesome aphids, but the insects quickly — and mysteriously — began to supplant native ladybirds. Andreas Vilcinskas at Justus Liebig University Giessen in Germany and his colleagues discovered that a parasite lurks within harlequin ladybirds' circulatory system and eggs. And although the parasite does not harm the harlequins, the researchers found that it is deadly to a native ladybird species (*Coccinella septempunctata*), which, like many beetle species, eats its competitors' eggs and larvae.

When the harlequin ladybird was introduced to new ecosystems, the authors say, its hidden parasites probably became biological weapons that helped it to take over new territory.
Science 340, **862–863 (2013)**
For a longer story on this research, see: http://go.nature.com/cajaty

ECOLOGY

Fish mismatch makes bears eat elk

In bringing lake trout to Yellowstone National Park humans may have inadvertently triggered a cascade of changes with consequences for migratory elk.

A team led by Arthur Middleton, then at the University of Wyoming in Laramie, looked at the long-distance effects of introducing lake trout to the park in the western United States. The lake trout have caused a marked decline in native trout and, unlike the natives, the interlopers spawn on lake bottoms — out of reach of grizzly bears.

Earlier research suggested that, when fish are in short supply, grizzlies prey on elk calves. The authors used a demographic model to calculate that this dietary shift has cut the growth of elk populations by as much as 11% — even in groups that overwinter well outside the park. The decline of these elk is often blamed, perhaps erroneously, on the reintroduction of wolves.
Proc. R. Soc. B http://dx.doi.org/10.1098/rspb.2013.0870 (2013)

HUMAN MIGRATIONS

Minoans came from Europe

Ancient DNA from remains found in caves on the Greek island of Crete suggests that the Minoan civilization emerged from farmers who settled on Crete thousands of years beforehand. This challenges an early theory, which held that the Minoans — recognized as being one of Europe's first 'high cultures' from their pottery and colourful frescoes — originated from Egyptian refugees.

A team led by George Stamatoyannopoulos at the University of Washington in Seattle analysed mitochondrial DNA (mtDNA) from teeth and bone samples of 37 individuals who lived on Crete between 4,400 and 3,700 years ago. The authors found 6 mtDNA haplotypes unique to Minoans and 15 that are common in modern and ancient European populations, but none characteristic of present-day

African populations. The North African influence on Minoans probably occurred through cultural exchange, the authors say.

Nature Commun. **4**, **1871** (2013)

For a longer story on this research, see: go.nature.com/cscvvy

PALAEONTOLOGY

'Ghost' reptile lived late

The identification of a fossilized reptile indicates that extinct 'fish lizards' called ichthyosaurs were more diverse late in their history than is often thought.

Previously, it was thought that only a subset of ichthyosaurs adapted for fast swimming in the open ocean had survived into the Cretaceous period, from 145 million to 66 million years ago. Valentin Fischer at the University of Liège in Belgium and his colleagues describe a new species with archaic features, *Malawania anachronus*, from a 1.5-metre-long fossil.

The recognition of this 'ghost lineage' reveals that two distantly related groups of ichthyosaurs lived in the Cretaceous. This challenges assumptions that low numbers and diversity in ichthyosaurs during this period contributed to their extinction.

Biol. Lett. http://dx.doi.org/10.1098/rsbl.2013.0021 (2013)

MOLECULAR BIOLOGY

Small RNA tunes protein

A small RNA molecule that controls whether bacteria coalesce into biofilms has a surprising mode of action.

Small RNAs are known to regulate gene expression by binding to target messenger RNAs, typically blocking their translation into proteins. A team led by Gisela Storz at the US National Institutes of Health in Bethesda, Maryland, and Poul Valentin-Hansen at the University of Southern Denmark in Odense found that a small RNA produced by *Escherichia coli* also binds to a target protein. The binding activates synthesis of a compound needed to form a biofilm.

Mutant bacteria in which this RNA–protein binding was abolished produced less of the compound and were less able to form biofilms. The results suggest that this small RNA tunes the protein's activity by interacting with it directly.

Genes Dev. http://dx.doi.org/10.1101/gad.214734.113 (2013)

APPLIED PHYSICS

Cheap, flexible, white light

Silver nanowires offer a way to solve two problems facing those making light-emitting diodes (LEDs).

LEDs tend to be made from inorganic materials. Those based on organic materials are thinner, lighter weight, cheaper to make and more easily spread over large, flexible surfaces. But the white light that they produce is not the same at all angles, and current organic LEDs are hampered by stiff electrodes. These are not compatible with flexible manufacturing and are made of indium tin oxide, an increasingly scarce material.

Whitney Gaynor, then at Stanford University in California, and her colleagues made alternative electrodes using silver nanowires encased in transparent plastic. These produce stable white light at all angles and are better suited to manufacturing the LEDs on flexible, plastic surfaces.

Adv. Mater. http://dx.doi.org/10.1002/adma.201300923 (2013)

DEVELOPMENT

Images of a metamorphosis

A three-dimensional imaging technique makes it possible to watch butterfly pupae mature within the chrysalis.

In most studies of developing insects, seeing means destroying. Researchers must dissect a number of specimens at different time points rather than following the maturation of a single individual. Russell Garwood and Tristan Lowe at the University of Manchester, UK, and their colleagues used high-resolution X-ray computed tomography to peer into living pupae of painted lady butterflies (*Vanessa cardui*). The authors took several scans over the insects' approximately 16-day development and produced detailed models (**pictured**) of external and internal anatomy. The technique showed that butterfly breathing apparatus forms surprisingly early, and promises to allow quantitative developmental studies of a range of insects.

J. R. Soc. Interface **10**, **20130304** (2013)

ANIMAL MODELS

Multiplexed mouse mutants

Animals carrying mutations in multiple genes can be engineered in a single step with the help of a bacterial protein and targeted RNA.

Introducing a number of mutations to mice is usually tedious and time-consuming. Researchers introduce a genetic change into embryonic stem cells, mix them into embryos and transplant these into surrogate mothers. Resulting pups are then interbred for generations.

A group led by Rudolf Jaenisch at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, has succeeded in making several genetic changes at once by using a bacterial system called CRISPR/Cas that normally serves to destroy viruses. The technique allowed the authors to target both copies of a number of genes in embryonic stem cells and to manipulate genes in fertilized eggs. The method could allow researchers to create mice with a host of mutations in months rather than years, the authors say.

Cell **153**, **910–918** (2013)

EARTH SCIENCE

Cosmic rays show how boulders move

HIGHLY READ on wiley.com in March 2013

Rare isotopes generated by cosmic rays can reveal the timing of events such as megafloods and rock debris flows.

Analysis of the isotopes' concentrations can help researchers to estimate a rock's age, a method known as cosmogenic dating. Benjamin Mackey and Michael Lamb at the California Institute of Technology in Pasadena extended this technique to show how boulders move and erode. They modelled the accumulation of these isotopes in boulders and found that the isotopes' distribution near the surface indicates how often a metre-plus sized rock has been transported.

Because boulders are often used to date landforms, such information provides clues about geological events and the creation of features such as glaciers and shorelines.

J. Geophys. Res. Earth Surf. **118**, **184–197** (2013)

↻ NATURE.COM

For the latest research published by *Nature* visit:

www.nature.com/latestresearch
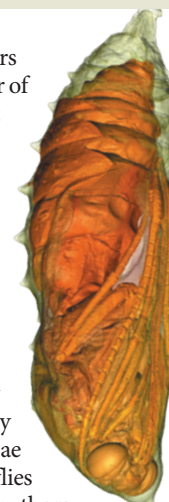
SEVEN DAYS
The news in brief

BUSINESS

Speed machine
The world's only quantum-computer seller has secured another customer — a collaboration between Google, NASA and the non-profit Universities Space Research Association. D-Wave, of Burnaby, Canada, made the announcement on 16 May. The deal was sealed after computer scientist Catherine McGeoch of Amherst College, Massachusetts, showed in a speed test that the machine can perform some tasks up to 3,600 times faster than conventional computing methods. Google, based in Mountain View, California, said that the computer would be used for machine-learning research. See go.nature.com/ufvacw for more.

POLICY

Fish progress
Fisheries ministers on 15 May reached a potential turning point in the European Union's efforts to reform its commercial fishing policy. The rules have been criticized for allowing large catches that exceed scientific recommendations, but a lack of agreement between member-state ministers, the European Parliament and the European Commission has stalled changes. Ministers have now inched towards the goals favoured by Europe's Parliament, which include restoring fish stocks to certain biomass levels and setting tougher limits on waste of unwanted fish. See go.nature.com/l6hx5y for more.

Chilean science
Chile should have a science ministry, a panel of scientists and economists has recommended to Chilean President Sebastián Piñera in

Africa's eyes on the radio sky

South Africa's astronomical ambitions received a boost on 15 May, when the first results from a seven-dish radio-telescope array were released. The KAT-7 array, based in the Northern Cape, observed two strong radio flares from the binary-star system Circinus X-1, and monitored their progression. An international team of astronomers reports the findings in a paper accepted by *Monthly Notices of the Royal Astronomical Society* (R. P. Armstrong *et al.* Preprint available at http://arxiv.org/abs/1305.3399). KAT-7 (two dishes pictured) serves as a test bed for the MeerKAT 64-telescope array, which will form part of the larger and more powerful Square Kilometre Array (SKA). The SKA is a planned €1.5-billion (US$1.93-billion) radio-telescope array, to be built across sites in South Africa and Australia.

a report delivered on 15 May. Piñera had created the advisory commission in January. The panel's report notes that science governance in the country is fragmented in uncoordinated institutions, and would benefit from being placed, along with higher education, in a dedicated ministry. Chilean scientists have repeatedly requested the same reform, most recently in a letter to the president in April.

Indonesia's forests
Indonesian president Susilo Bambang Yudhoyono on 15 May extended a moratorium on new licences to log and clear virgin rainforest and peatlands in the country for another two years. Indonesia initially established the moratorium in 2011 after committing to reduce its greenhouse-gas emissions, which arise primarily from deforestation and the destruction of peatlands. Scientists and environmentalists have criticized loopholes in the moratorium but say that the decision will give the country more time to develop a broader forest-protection plan.

Arctic group grows
At its biennial meeting last week in Kiruna, Sweden, the Arctic Council granted observer status to China, India, Italy, Japan, Singapore and South Korea. The council, established in 1996 and headed by eight Arctic nations, promotes sustainable development in the warming polar region. Its expansion, which doubled the number of observer nations, reflects growing economic interest in and environmental concerns about oil and gas resources in the Arctic. Council members also signed a cooperative agreement addressing preparedness for and responses to marine oil spills.

Stem-cell safety
Italian scientists have persuaded a reluctant parliament to amend a controversial decree that would have allowed severely ill patients to continue treatment with an unproven, and possibly unsafe, stem-cell therapy (see *Nature* **495**, 418–419; 2013). This week, the

Senate is expected to accept amendments voted for by the parliament's other house, the Chamber of Deputies, on 20 May. These would subject the therapy, developed by the Stamina Foundation in Brescia, to regulatory oversight. The government will make €3 million (US$3.9 million) available to test the therapy in clinical trials. See go.nature.com/lqqqum for more.

FACILITIES

Brazil investment

Brazil's São Paulo Research Foundation (FAPESP), one of the country's largest science-funding agencies, announced on 15 May the establishment of 17 Research, Innovation and Dissemination Centers. The research centres will focus on innovation and technology transfer in a variety of areas, from nutrition and neuroscience to functional materials and photonics. FAPESP plans to invest US$370 million over 11 years, to be matched by $310 million in salaries from host institutions.

PEOPLE

One down, one to go

The US Senate voted unanimously on 16 May to confirm physicist Ernest Moniz (**pictured**) as head of the Department of

Energy. Moniz, director of the Energy Initiative at the Massachusetts Institute of Technology in Cambridge, is known for his strong support of natural gas and nuclear power. The Senate has yet to vote on the nomination of Gina McCarthy to lead the Environmental Protection Agency. See page 418 for more.

Nobel laureate dies

Heinrich Rohrer, a Nobel-prizewinning Swiss physicist, died on 16 May at the age of 79. Together with Gerd Binnig, Rohrer shared half of the 1986 Nobel Prize in Physics for the invention of the scanning tunnelling microscope while at the IBM Research Laboratories in Zurich, Switzerland. The instrument allows researchers to image and manipulate individual atoms on electrically conductive surfaces.

EVENTS

Impact-factor abuse

More than 480 researchers and scientific organizations have signed a declaration condemning the use of the journal impact factor (JIF) to gauge scholarly success. According to the San Francisco Declaration on Research Assessment, released on 16 May, the JIF — an average measure of how often a journal's articles are cited — is misused to assess the significance of work by scientists who publish in those journals. The declaration recommends alternative metrics for evaluating researchers, including citation counts of individual articles. See go.nature.com/nvqgwj for more.

RESEARCH

Mission stalls

Kepler, NASA's planet-hunting spacecraft, may be on its last legs. On 14 May, the spacecraft entered a protective mode and stopped collecting data. See page 417 for more.

Distant particles

A handful of neutrinos detected at the South Pole probably came from beyond Earth's atmosphere, according to an international group of
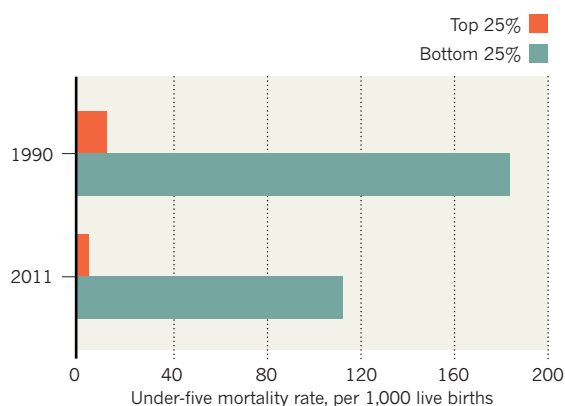
researchers. At a conference in Madison, Wisconsin, the IceCube collaboration reported on 15 May that its sensors had collected 28 neutrinos with energies far exceeding those from cosmic rays hitting Earth's atmosphere. The finding raises hopes that the particles could be traced back to some of the most energetic phenomena in the Universe, such as active black holes or γ-ray bursts from collapsing stars. See go.nature.com/pon2qj for more.

MENTORING AWARDS
Nominations for *Nature*'s annual awards for science mentoring are open until 20 August. This year, *Nature* seeks to honour two outstanding mentors in Italy, who will each receive €10,000 (US$12,800). The panel of judges is led by physicist Luciano Maiani of CERN, Europe's particle-physics lab near Geneva in Switzerland. See go.nature.com/o7bltb for more.
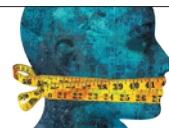
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Health gaps between the world's most- and least-advantaged countries have narrowed over the past two decades in areas such as child survival (see chart), maternal deaths in childbirth and deaths from tuberculosis. In its latest statistics, released on 15 May, the World Health Organization found improvements in many countries, but projected that progress was not fast enough to meet 2015 Millennium Development Goals, such as cutting child mortality by two-thirds from 1990 levels.

HEALTH GAP NARROWS
The difference in child mortality between the world's most- and least-advantaged countries has fallen since 1990.

Top 25%
Bottom 25%

1990
2011

Under-five mortality rate, per 1,000 live births
0   40   80   120   160   200

NEWS IN FOCUS

JPL-CALTECH/AMES/NASA
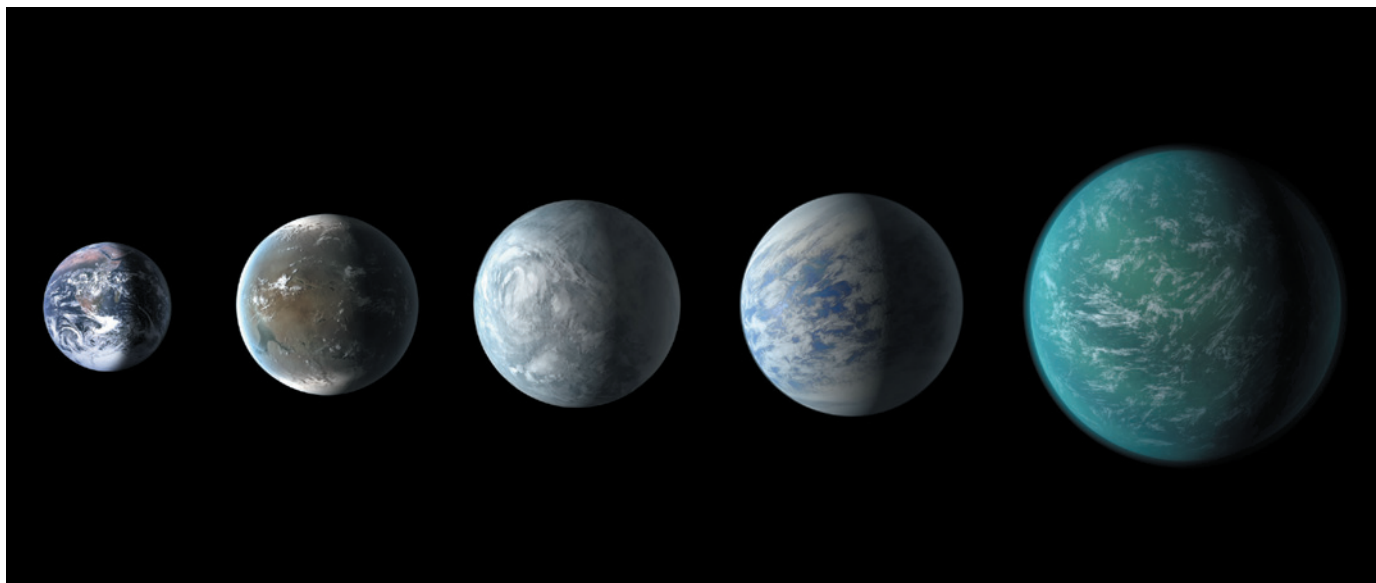
Four potentially habitable planets (artist's impression) have turned up in Kepler's data so far, but the crippled spacecraft may not find a true twin for Earth (far left).

EXTRASOLAR PLANETS

The wheels come off Kepler

Space telescope's mission to find planets outside the Solar System is probably over.

BY RON COWEN

NASA's Kepler spacecraft is not only the most prolific exoplanet detector ever; it is — or was — a marvel of engineering. Its 1.4-metre mirror funnels starlight to a 95-megapixel camera, capable of discerning dips in brightness as small as 10 parts per million — clues to the mini-eclipses caused by an exoplanet crossing the star's face. Yet on 14 May, the US$600-million craft was derailed by the failure of one of its only moving parts — a roughly $200,000 device akin to a child's gyroscope.

Mission scientists knew that the craft was vulnerable. Kepler is equipped with four metal reaction wheels — motors that, when spun up in one direction, cause the spacecraft to turn in the other. At least three are needed for the craft to remain stable, and one had failed last July. "We recognized that the wheels had a rather chequered history," says William Borucki, a space scientist at NASA's Ames Research Center in Moffett Field, California, and principal investigator of the Kepler mission. When the second wheel stopped turning — probably because of a problem with its ball bearings or housing, say NASA scientists — Kepler put itself into safe mode, pending a repair that many say is unlikely to happen.

Spinning at 1,000–4,000 revolutions per minute, reaction wheels fine-tune a spacecraft's orientation, exerting precise torques through careful control of their speed. Even before Kepler launched in 2009, the mission team was painfully aware of problems from a host of virtually identical reaction wheels made by the same manufacturer and installed on several other spacecraft.

Reactions wheels of this type failed, or were deemed too unreliable to be used, on NASA's Far Ultraviolet Spectroscopic Explorer in 2001 and Japan's Hayabusa mission in 2004 and 2005. NASA's Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics (TIMED) satellite experienced a single reaction wheel failure in 2007, and the agency's Dawn mission suffered two failures, in 2010 and 2012.

Most of these malfunctions occurred well before Kepler's launch, but the failure on TIMED created a critical mass of concern, recalls John Troeltzsch, Kepler programme manager for Ball Aerospace in Boulder, Colorado, which built the craft for NASA. "We first really realized how serious it was at the end of 2007," he says.

By then, Kepler was ready for launch. Options such as overhauling the entire reaction-wheel system or adding another set of back-ups were deemed too costly and would have meant years of further delays for a mission that had already been postponed twice.

But the team wanted to take one last look. In early 2008, Troeltzsch and his colleagues took all four wheels out of the spacecraft and sent them back to the manufacturer, Ithaco ▶

Space Systems in Ithaca, New York, for re-inspection. Changes were made, including replacing the ball bearings, which had already shown signs of pitting, says Charles Sobeck, Kepler's deputy project manager at Ames. Troeltzsch says, "The assessment was that the changes would prevent the recurrence of the types of problems that had happened before." Through a spokesman, Ithaco owner UTC Aerospace Systems declined to respond to questions about how the company had tried to fix the wheels, referring the questions back to NASA.

Borucki notes that, at the time, potential problems with the reaction wheels seemed less of an issue because the mission was supposed to last for only 3.5 years. That was how long it was expected to take for Kepler to achieve its main goal: surveying a group of about 150,000 Sun-like stars up to 920 parsecs (3,000 light years) away to determine how common Earth-sized planets in Earth-like orbits are in the Galaxy.

After launch, however, scientists discovered that most stars are more variable than the Sun (see *Nature* **477,** 142–143; 2011). The mission would need more time to distinguish a random

KEPLER'S BOUNTY

NASA's Kepler spacecraft has discovered more than 2,700 candidate extrasolar planets. But the mission may have ended just as it was becoming sensitive to Earth-sized planets in Earth-like orbits around their stars.
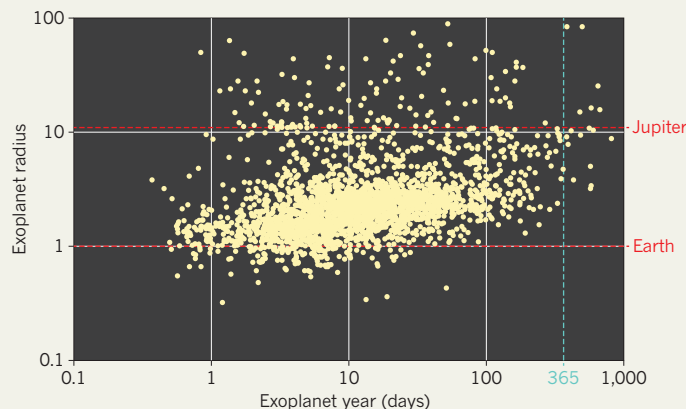
fluctuation in a star's brightness from a dip in starlight due to a passing planet, so last year, NASA extended the mission until 2016. But around the time of that decision, the first reaction wheel went bad. With the loss of the second wheel, says Troeltzsch, "the science we were doing with Kepler, as we were doing it, is over".

Although engineers will try to reactivate the failed wheels in the next couple months, Troeltzsch is not hopeful. "It's very unlikely that it can be restored to any kind of usefulness," he

says. Thrusters on board the craft may be able to serve as a crude stand-in, but they were not designed to provide the precision orientation needed. Efforts to bring the mission back to life will probably continue until the end of autumn, says Sobeck.

In the meantime, Borucki's team is analysing the two most recent years of Kepler data, which seem likely to add hundreds of candidate planets to the mission's current tally of more than 2,700 (see 'Kepler's bounty'). Because the most recent data contain information on planets with orbits of similar length to Earth's, Borucki is optimistic that they will reveal a reasonable estimate of the frequency of Earth analogues in the Galaxy.

Sara Seager, an exoplanet theorist at the Massachusetts Institute of Technology in Cambridge, is not so sanguine. She says that the data may yield decent statistics about Earth-sized planets in orbits up to about 200 days, but not all the way up to orbits in the 'habitable zone' of a Sun-like star, which are around 365 days. She is hopeful, however, that a few such planets are buried in the remaining data. "It's possible, and even likely." ■

Log–jam in agency confirmations

Political squabbles leave US science agencies without heads.

BY LAUREN MORELLO

US Senate Republicans have lots of questions for Gina McCarthy, President Barack Obama's nominee to head the Environmental Protection Agency (EPA) — 1,038 of them, to be exact. Many target alleged lapses in the EPA's policies regarding transparency and information access, but they also reflect a more general distrust of the agency that regulates greenhouse-gas emissions, sets water-pollution standards and monitors air quality.

McCarthy, currently the EPA's assistant administrator in charge of air and radiation, has had plenty of time to ponder the lawmakers' queries since her nomination on 7 March. She has already waited weeks longer for confirmation by the Senate than any would-be chief in the EPA's 43-year history (see 'Stalling tactics'). Obama's new energy department secretary, physicist Ernest Moniz, didn't fare much better. He was confirmed last week, but only after a key senator withdrew an objection

that had held up the confirmation for more than two months.

In his first term, Obama was able to install a

STALLING TACTICS

Gina McCarthy has waited longer than any other nominee for US Senate confirmation as head of the US Environmental Protection Agency (EPA).

EPA nominee	Presidency	Days to confirmation
Gina McCarthy, 2013	Barack Obama	77 and counting
Lisa Jackson, 2009	Barack Obama	2
Stephen Johnson, 2005	George W. Bush	46
Michael Leavitt, 2003	George W. Bush	55
Christine Todd, 2001	George W. Bush	11
Carol Browner, 1993	Bill Clinton	1
William Reilly, 1989	George H. W. Bush	13

science 'dream team', with much fanfare and little opposition (see *Nature* **489,** 488–492; 2012). But at the start of his second term, Obama is facing a number of key science-policy decisions without agency heads. The delays are not all due to Congress. Partly because of the increasingly complex background checks needed to prepare for congressional scrutiny, the Obama administration has not yet even nominated new heads for the National Oceanic and Atmospheric Administration, the US Geological Survey, and the National Science Foundation (NSF), whose previous directors have all stepped down.

Ross Baker, a political scientist at Rutgers University in New Brunswick, New Jersey, says the McCarthy delays are a prime example of 'slow-walking' — stalling a nomination in the hopes that a nominee will withdraw his or her name. Simple tactics can be enough: Republicans boycotted a 9 May committee meeting that would have allowed McCarthy's nomination to advance to the full Senate. Even answering senators' questions can end up adding days to the process. "It's basically death

by a thousand cuts," Baker says. McCarthy's nomination finally seems to be on course to a vote by the full Senate. But as *Nature* went to press, there is no word on when that vote might come.

To Calvin Mackenzie, a political scientist at Colby College in Waterville, Maine, who studies the presidential appointment process, the current stalemate is the culmination of a trend that began decades ago. Over the years, both major US political parties have taken advantage of Senate rules and customs to hinder nominations. During the presidency of George W. Bush, for example, Senate Democrats stalled the nomination of EPA chief Michael Leavitt for several weeks. "There is blood on the hands of both parties here," Mackenzie says. "The trouble with a scorched-earth policy is it keeps getting worse."

Some observers fear that the political manoeuvring will start to have an impact on US research. Like the rest of the federal government, US science agencies are grappling with the impact of automatic budget cuts known as sequestration, which has chopped roughly 5% from their funding to 30 September and is set to continue until 2021 (see page 419). Neal Lane, a former NSF director and science adviser to President Bill Clinton, says that having only an acting director can make budget negotiations problematic for agencies.

The EPA is finalizing a new rule to limit carbon dioxide emissions from power plants, and it is also advising the Department of State over whether to permit the construction of the Keystone XL pipeline to carry oil from the Canadian tar sands to the Gulf of Mexico. Those efforts are likely to be delayed until the EPA and other science agencies are again run by permanent leaders, says Lane, now at Rice University in Houston, Texas.

That's not to say interim heads aren't experienced. "They know how the agency works and they can keep the trains moving," says Lane. "But it's a serious problem if that acting arrangement lasts very long, because large policy decisions are generally on hold during that period."

Jeffrey Holmstead, who headed the EPA's Office of Air and Radiation during the George W. Bush administration, disagrees. "A lot depends on the reputation and stature" of an agency's interim chief, says Holmstead, head of the environmental-strategies group at the law firm Bracewell & Giuliani in Washington DC. He notes that the EPA eased clean-air regulations for older coal-fired power plants during a four-month period in 2003 when the agency was run by an acting administrator.

In the meantime, those watching the McCarthy nomination say they are cautiously optimistic that her long wait will end in confirmation. "I wouldn't be surprised if some day she will be confirmed," Mackenzie says. But he adds, "With these things, who knows?" ■

The flooded Red River isolated houses in North Dakota in 2011. Budget cuts now endanger monitoring.

POLICY

US budget cuts hit Earth monitoring

Sequestration threatens records of snow and stream levels in western United States.

BY ALEXANDRA WITZE

Two kilometres south of the US–Canadian border, in Pembina, North Dakota, a stream gauge measures the height of the water surging down the Red River. The instrument, one of about 8,000 maintained by the US Geological Survey (USGS), is a sentinel for communities along the river that experienced devastating floods in 2009, 2010 and 2011. Yet this spring, the USGS announced plans to shut down the Pembina stream gauge — a casualty of the sweeping federal budget cuts known as sequestration.

Implemented on 1 March, sequestration slashes about 5% from the budget of every federal agency and programme until the end of the fiscal year on 30 September, with further cuts expected until the end of 2021 unless Congress intervenes. Scientists in fields from biology to astronomy are bracing themselves for an era of smaller and fewer research grants, which will begin within months (see *Nature* **494,** 158–159; 2013). But the cuts are already hampering Earth-monitoring projects, including stream gauges and snowpack measurements, which require a constant influx of funds to keep data flowing.

Monitoring equipment frequently breaks and must be repaired or replaced, usually during the short period of summer fieldwork. That often requires expensive journeys to remote sites by helicopter or other means.

Such is the case for surveys of the United States' western snowpack, a crucial source of water in summer for many states. Continuing a tradition that began in 1906, when a University of Nevada researcher measured snow depth along a transect in the Sierra Nevada mountains, the US Department of Agriculture's Natural Resources Conservation Service conducts more than 1,100 manual 'snow courses' once a month throughout winter and spring. In 1980 it also began operating automated snow telemetry (SNOTEL) sites, and it now has around 860 spread over 13 western states. Survey data are used to produce water-supply forecasts and to analyse changes in the snowpack over time.

But in January, the snow survey announced that it would eliminate 39 snow courses in Montana. The programme was already suffering from reduced funding: it received US$9.3 million in fiscal year 2012, about 15% less than the year before. Congress has not yet finalized the 2013 budget, but the survey is probably facing another 7.5% cut this year when sequestration is taken into account, says Michael Strobel, director of the National Water and Climate Center in Portland, Oregon.

More snow courses may be at risk. "We're trying to prioritize sites that have ▶

▶ scientifically critical information", says Strobel, as well as to decide "which ones have enough SNOTEL and other sites nearby to have a lower priority".

SNOTEL sites typically need to be visited each year to have melted snow drained away, antifreeze added and damage from grizzly bears and other problems repaired. But tight budgets will probably prevent workers from getting to every location this summer, so some of the measurements are likely to fail.

Once snow melts and starts running down from the mountains, USGS stream gauges measure it — unless they, too, are shut down. The national stream-gauge network costs about $165 million to operate each year. The federal government supplies a little more than half of that, and state, local and tribal agencies make up the rest. Information from the 8,000 gauges is posted online in real time and used in everything from weather forecasting to designing bridges and planning kayaking trips.

Federal cutbacks mean that about 50 stream gauges are being shut down and some 100 more could be at risk, says Michael Norris, coordinator for the National Streamflow Information Program, based in Reston, Virginia. The USGS prioritizes gauges mandated by law, such as those that support water treaties, as well as gauges in crucial flood-forecast areas or those that have been in operation for a long time. The agency consults with local officials about which gauges can be cut.

The Pembina gauge was targeted because another station just across the Canadian border in Emerson provides much the same information, says Gregg Wiche, head of the North Dakota Water Science Center in Bismarck. But when Wiche heard how much the community used the information — a local railway relies on its forecasts, for example — he and others worked out a deal to keep the Pembina station and drop one elsewhere in the state instead.

Such small adjustments do little to reassure Johnnie Moore, a hydrologist at the University of Montana in Missoula who has used USGS streamflow data to study long-term trends in melt run-off from the northern Rocky Mountains. He worries that shutting off stream gauges that have been operating for decades could hurt climate-change studies. "Over the long run there's been a big decline in the number of gauges", says Moore, and to cut even more now "is pretty disconcerting".

The outlook is not entirely grim: some of the 50 or so gauges at immediate risk have received extra funding from state or local governments to carry them through to 30 September.

And US President Barack Obama's budget request for fiscal year 2014 asks for a 25% increase over 2012 levels for the streamflow information programme. That money, if Congress provides it, could let the USGS go back to watching the west's water. ■

EUROPEAN SCIENCE

Russian academy awaits new head

Reform is in the air at the nation's oldest research body.

BY QUIRIN SCHIERMEIER

From czarist times to the days of perestroika, the Russian Academy of Sciences (RAS) was the pillar of the nation's scholarship, boasting the country's best scientists as members. The dramatic decline in science spending after the break-up of the Soviet Union ended the academy's days of plenty, and although funding has rallied in recent years, those roubles have increasingly gone to other research centres.

Now awaiting the first new academy president in more than two decades, Russian scientists hope that the leadership will revive the struggling institution and bring about reform they feel is long overdue.
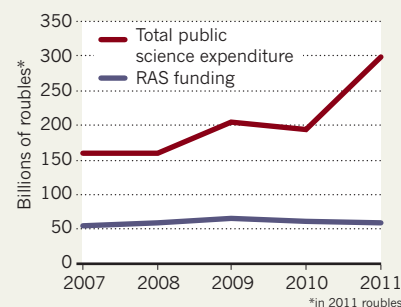
Critics say that the RAS, which employs some 45,000 scientists at 436 institutes across Russia, is burdened by a host of unproductive ageing scientists awaiting retirement and by many pursuing research of dubious value (see *Nature* **449,** 524–527; 2007). They say that current president Yuri Osipov, a mathematician who has presided over the RAS since 1991, has failed to clear out dead wood and take other steps to prevent the academy, which was founded in 1724 by Peter the Great, from declining into insignificance.

On 29 May, the academy's general assembly of more than 1,000 full and corresponding members will vote in a secret ballot on its future head. Osipov, 76, said earlier this month that he will not run for a fifth term. "I'm tired," he told the Russian news service Pravda.ru . It is time for a new person with "new views and fresh energy" to take over.

How new the views and how fresh the energy will depend on who wins on 29 May. The favoured contender, according to sources close to the academy, is 67-year-old plasma physicist Vladimir Fortov, former deputy prime minister and former science minister in the Russian government. His election manifesto includes a raft of proposals for the RAS — to cut red tape, to improve efficiency, to regularly review the performance of institutes and scientists, and to base funding on merit. "Fortov is no doubt determined to start some reform, however timidly," says Konstantin Severinov, a molecular biologist at the recently inaugurated Skolkovo Institute of Science and Technology near Moscow, and a professor at Rutgers University in Piscataway, New Jersey.

Fortov's chief opponent, Zhores Alferov, is considered less likely to modernize the academy. The 83-year-old physicist and long-time director of the respected Ioffe Physico-Technical Institute in St Petersburg shared the Nobel Prize in Physics in 2000 for his ground-breaking work on semiconductor heterostructures. He is also a member of the

FLATLINING
Funding at the Russian Academy of Sciences (RAS) has stayed flat in real terms as overall science funding has risen.

SOURCE: RAS

*in 2011 roubles

MORE ONLINE

TOP STORY

Scientists hope to get more coronavirus clues from Saudi Arabia go.nature.com/ knixeb

MORE NEWS

● South Africa slows research spending go.nature.com/kqalnt
● Drug saves fertility in mice receiving chemotherapy go.nature.com/lzuzkr
● Pregnancy test helped spread frog-killing fungus go.nature.com/ qulewc

AFP/GETTY

communist faction in the Duma, the Russian parliament, and a stalwart champion of Soviet ideals. "Alferov is a scientist of the highest repute — but his views are not always forward-looking," says Severinov.

At a meeting in April of the Council for Science and Education, a presidential advisory body chaired by Vladimir Putin, Alferov agreed with academicians who reject the use of bibliometric indicators — such as numbers of research papers and the impact factors of the journals in which they are published — for measuring scientific performance. He agreed that such assessments would encourage Russian scientists to publish their best work in foreign high-impact journals rather than in Russian ones.

Such nationalistic views threaten to cement the isolation of Russian science further, says Mikhail Feigel'man, deputy director of research at the academy's Landau Institute for Theoretical Physics near Moscow. (Feigel'man himself is not a member of the academy and will not be voting on 29 May.) "Our ties with international science are all but eliminated," he says. "We're caught in a system that was constructed in Soviet times when academic exchange was not considered important. But now it is."

Travel grants are almost unavailable to RAS researchers, and foreign scientists or students are absent at many academy institutes. "Anyone interested in working with us, or giving talks at our institute, I have to tell that it is not possible unless they come with their own money," says Feigel'man. Meanwhile, low salaries, lack of grant money and poor career prospects are deterring Russian talent from pursuing science at home, he says. "Our own people leave Russia as fast as they can."

The Russian government itself seems to view the RAS as increasingly irrelevant. The academy's budget, around 60 billion roubles (US$1.9 billion) per year, has stagnated since 2000, even as overall science spending in Russia has doubled (see 'Flatlining'). The government is funnelling the new money into university research and ventures such as the Skolkovo, a new science complex that will focus on biotechnology and information technology, among other disciplines, in partnership with international companies. German engineering and electronics giant Siemens, based in Munich, and Finnish mobile-phone company Nokia, in Helsinki, have already signed up to participate.

Science minister Dmitri Livanov has said that he expects a thorough overhaul of the academy and that productivity should become a funding prerequisite for RAS institutes and scientists. "There are some researchers who fully meet this standard," Livanov said in a 2012 interview. "Those who do not work we will have to call something different." ∎

A RISKY REGION

Geoscientists are installing 69 Global Position System stations across the Caribbean in an effort to monitor earthquakes and other hazards.

GEOPHYSICS

A network to track Caribbean hazards

Multi-nation effort is a test of scientific diplomacy.

BY ALEXANDRA WITZE IN CANCÚN, MEXICO

Some US geophysicists are going where few of their compatriots have gone before: to Cuba, Venezuela and other countries that are notoriously anti-American. But these countries are also notoriously plagued by natural disasters such as earthquakes, and so engineers are blanketing the Caribbean with a network of sensors to detect the crustal strains and brewing storms that threaten the region.

Early results from the US$6-million system, called the Continuously Operating Caribbean GPS Observational Network (COCONet) and funded by the US National Science Foundation (NSF), were presented in Cancún, Mexico, on 17 May at the American Geophysical Union Meeting of the Americas. COCONet is an offshoot of another NSF effort that, over the past decade, has sprinkled Global Positioning System (GPS) sensors across the western United States. That was child's play compared with working across 31 separate national governments, many of which distrust the United States, says Glen Mattioli, a programme director at UNAVCO, a geodetic-survey organization based in Boulder, Colorado, that is managing the project.

Under COCONet, geoscientists across the Caribbean are gaining free access to data that will help them to understand and prepare for natural hazards such as earthquakes and hurricanes. But unless they adopt the network as the long-term backbone for observations in the region, the project could go down as just another well-meaning but failed effort to bridge national differences and develop scientific capacity. To most, it is worth the risk. "If you don't bet, you don't win," says Franck Audemard, a geologist at the Venezuelan Foundation for Seismological Research in Caracas.

COCONet stations are being set up at sites as diverse as steep rainforest slopes and isolated coral-reef islands (see 'A risky region'). Each station contains meteorological instruments, along with a GPS unit that constantly monitors ground movements and, in many cases, transmits those data in near real time. So far, UNAVCO has built or upgraded 38 of its planned 69 sites. At least 61 existing stations are also being incorporated into the network.

The project has made more progress than expected in acquiring the permits needed to set up stations, says Mattioli. That holds true even in Cuba, where officials have already approved the COCONet stations slated to arrive in June, even as US officials are still working ▶

through the permissions, Mattioli says.

Spurred on by the devastating 2010 Haiti quake, the network aims to fill out the currently patchy picture of the region's geological turmoil. Off the Pacific coast of Costa Rica, for instance, a COCONet station on the isolated Isla del Coco is the only GPS station continuously tracking the Cocos plate as it dives beneath the Caribbean plate. The data show that the two plates are converging at a rate of 78 millimetres per year. Such information will help geophysicists to reconcile their ideas about plate motion in the region, says Marino Protti, a geologist at the National University of Costa Rica in Heredia, who presented his findings in Cancún.

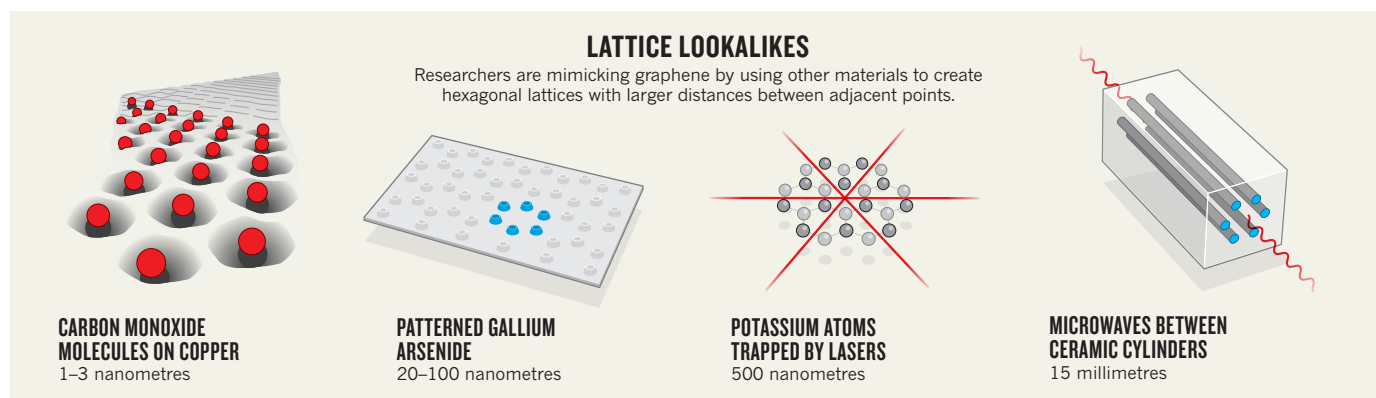Other COCONet stations dot Costa Rica's Nicoya Peninsula, one of the few places on land that lie directly above a subduction zone. In September 2012, Nicoya shuddered from a magnitude-7.6 earthquake. "Having the network helped us see that it was not a whole-plate event," Protti says. He thinks that remaining stress in the plate could trigger another large tremor.

Besides tracking movements in Earth's crust, the GPS data allow atmospheric scientists to measure how much water vapour sits in the air between each station and the GPS satellites, based on the delay in the signal's travel time. That information is a rare, real-time glimpse of how much water is available to power hurricanes, says meteorologist John Braun of the University Corporation for Atmospheric Research in Boulder.

COCONet is even yielding some unexpected societal benefits. In the Dominican Republic, land surveyors are eagerly using the nation's nine stations to comply with a real-estate law that requires precise surveying of parcels of land for sale, says Alexander Holsteinson of the National University of Pedro Henríquez Ureña in Santo Domingo.

Yet funding for COCONet runs out in 2015, and no one knows what will happen then. Each country will have to find the money to keep its stations operating, says Guoquan Wang, a geophysicist at the University of Houston, Texas, and a COCONet member who developed a GPS network for Puerto Rico and the Virgin Islands. That means that host nations will have to find value in supporting the project, says Héctor Mora Páez, who heads the GPS network for the Colombian Geological Survey in Bogota. "You have to make the local people feel involved." ∎

LATTICE LOOKALIKES
Researchers are mimicking graphene by using other materials to create hexagonal lattices with larger distances between adjacent points.

CARBON MONOXIDE MOLECULES ON COPPER
1–3 nanometres

PATTERNED GALLIUM ARSENIDE
20–100 nanometres

POTASSIUM ATOMS TRAPPED BY LASERS
500 nanometres

MICROWAVES BETWEEN CERAMIC CYLINDERS
15 millimetres

PHYSICS

Graphene knock–offs probe ultrafast electronics

Honeycomb lattices in different materials enable experiments impossible in the real thing.

BY EUGENIE SAMUEL REICH

If imitation is the sincerest form of flattery, then graphene is basking in admiration. Several labs are recreating the six-fold geometry of the carbon-based material with a range of building blocks, hoping to match — if not surpass — graphene's fascinating properties.

The approaches span a range of length scales, from nanometres to millimetres (see 'Lattice lookalikes'), and take advantage of a variety of construction techniques, including potassium atoms trapped by laser beams, hexagons etched on a gallium arsenide surface and microwaves pumped into a honeycomb arrangement of ceramic cylinders. The research is mostly motivated by curiosity, and few think that any of the 'artificial graphenes' will be used in applications. But the new materials show that hexagonal patterns are worth investigating. "It's clear the physics in electrons in graphene is very attractive," says Eros Mariani, a theoretical physicist at the University of Exeter, UK.

Graphene has unusual electronic properties as a result of its Dirac points, positions in its single-sheet honeycomb lattice of carbon atoms at which the energy bands of electrons from adjacent atoms link up. These allow electrons in the bands to speed across the material, like cars changing roads at major intersections instead of getting stuck on one jammed road. Measurements reported[1,2] in 2005 confirmed that electrons in graphene move so fast that they seem to have no mass, and are, in effect, moving at the speed of light. That fuelled hopes that graphene could be used to make ultrafast electronics; it also set the stage for the discovery of exotic quantum-mechanical effects.

In 2010, physicist Hari Manoharan, an expert in atomic manipulation at Stanford University in California, stumbled on the first graphene mimic. Manoharan and his team were using a scanning tunnelling microscope to arrange various molecules on a copper surface when they were struck by a hexagonal pattern. It looked uncannily like graphene but was eight times larger in scale. "We were thinking, why don't we see if it has graphene's properties?" says Manoharan.

⟳ NATURE.COM
For more on the properties of graphene, visit:
go.nature.com/hm41sm

He and his colleagues soon settled on carbon monoxide as their molecule of choice, and in 2012, they published evidence[3] that carbon monoxide molecules arranged in a honeycomb pattern do have Dirac points. And in the same year, Tilman Esslinger, a physicist at the Swiss Federal Institute of Technology in Zurich, showed that a gas of ultra-cold potassium atoms trapped[4] in a hexagonal lattice by criss-crossing laser beams also sustains faster-than-light conduction — but of atoms, rather than electrons.

Both groups then performed tricks that are currently impossible in real graphene. Manoharan and his colleagues deposited their molecules in a slightly warped pattern, which caused the energy levels of the electrons in the lattice to split as if they had been exposed to a magnetic field stronger than any that could be applied in a laboratory. "You can trick electrons into thinking they're in a 60-tesla magnetic field," says Manoharan.

Esslinger and his team also distorted their lattice, fading out one of the laser beams to shift the structure from a hexagonal to a dimeric one, built of pairs of atoms[4]. That prompted the precious Dirac points to merge and annihilate. With the intersections gone, the atoms got trapped in lower energy bands and were less free to move, confirming that the hexagonal laser lattice really had been mimicking the electronic structure of graphene.

In January, Fabrice Mortessagne, a physicist at the University of Nice Sophia Antipolis in France, and his colleagues demonstrated another feat of distortion[5]. His group assembled hexagonal arrangements of millimetre-scale ceramic cylinders between two metal plates. They then pumped microwaves into the lattice. The waves became trapped there, and their movement mimicked the path of electrons in graphene. When the researchers squeezed the lattice in one direction, they found that the microwaves were trapped unexpectedly along the edges of the structure, which challenges current theoretical descriptions of Dirac materials. The group suggests that similarly unexpected energy states will be observed for electrons in graphene once lab techniques have been developed to compress the material in one direction without breaking it.

Philip Kim, a physicist at Columbia University in New York who was involved in the 2005 measurements that uncovered graphene's unusual electronic properties, finds artificial graphenes exciting because they can be engineered to show effects expected in real graphene. But simply replicating graphene's properties doesn't grip everyone. "To mimic graphene sounds pretty boring scientifically and a waste of time and resources," says Antonio Castro Neto, director of the Graphene Research Centre at the National University of Singapore. "The objective of creating these artificial graphene-like lattices is to produce new systems that have properties that graphene does not have."

For example, he has predicted[6] that etching a honeycomb structure onto a semiconductor such as gallium arsenide could turn it into a topological insulator — a much-sought material that insulates in its interior but conducts superbly well on its surface, where exotic particles called Majorana fermions might emerge. These particles, which have been tentatively observed in one-dimensional nanowires, have a unique ability to encode quantum information by swapping locations with each other (see *Nature* **483**, 132; 2012). Hexagonal gallium arsenide lattices have been made on large scales, but Castro Neto says that current methods of nanolithography should enable researchers to etch them at scales of 20 nanometres or less, fine enough to see the Majoranas.

Although it will be a long time before artificial graphenes can be used in quantum computing, Castro Neto doesn't rule out a pay-off for them. "I am always surprised that people find applications for things I thought would never be useful," he says. ∎

1. Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. *Nature* **438,** 201–204 (2005).
2. Novoselov, K. S. *et al. Nature* **483,** 197–200 (2005).
3. Gomes, K. K., Mar, W., Ko, W., Guinea, F. & Manoharan, H. C. *Nature* **483,** 306–310 (2012).
4. Tarruell, L., Greif, D., Uehlinger, T., Jotzu, G. & Esslinger, T. *Nature* **483,** 302–305 (2012).
5. Bellec, M., Kuhl, U., Montambaux, G. & Mortessagne, F. *Phys. Rev. Lett.* **110,** 033902 (2013).
6. Sushkov, O. P. & Castro Neto, A. H. Preprint at http://arxiv.org/abs/1210.8186 (2013).

CLARIFICATION

We would like to clarify the following points about the News story 'Disputed results a fresh blow for social psychology' (*Nature* **497,** 16; 2013).

● Contrary to the story, unconscious thought theory, which is concerned with unconscious decision making, is not the same as intelligence priming.

● The story referred to the study by David Shanks (D. R. Shanks *et al. PLoS ONE* **8,** e56515; 2013) in which he was unable to replicate the 'professor/hooligan' intelligence-priming effect reported by Ap Dijksterhuis. We note that other researchers have observed Dijksterhuis's intelligence-priming effect.

● Contrary to the story, Dijksterhuis has stated that he will provide a protocol to assist in replicating results in the field. He had not, however, provided such a protocol at the time that the story was published.

● As the story stated, there is no suggestion of fraud in Dijksterhuis's research, and we are happy to emphasize that there is no suggestion that he has been involved in any misconduct.
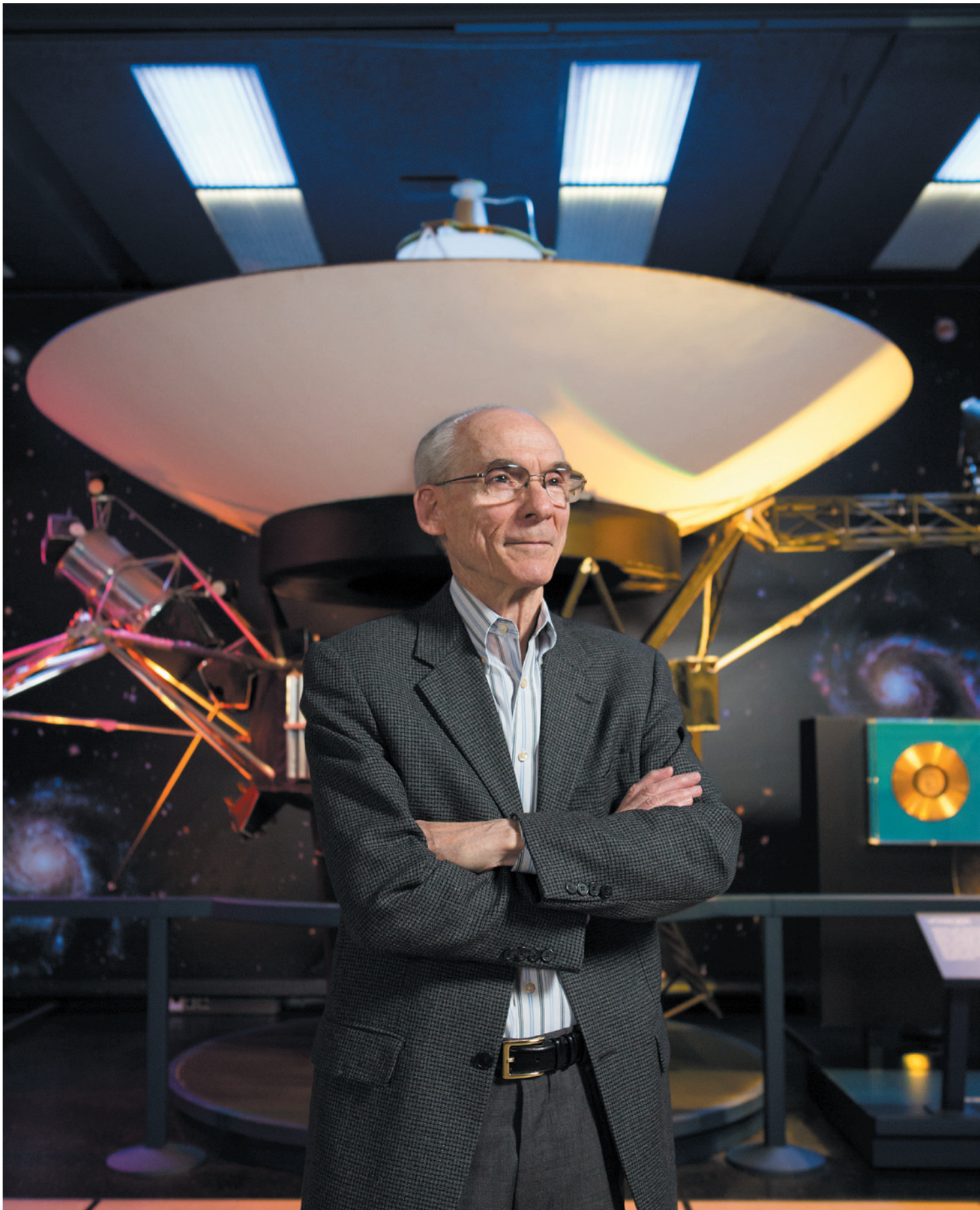
OUTWARD BOUND
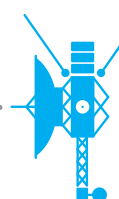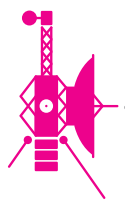
Ed Stone has spent 36 years guiding the twin Voyager spacecraft through the Solar System. Next stop, interstellar space.

BY ALEXANDRA WITZE

The 44 notebooks lined up neatly in Ed Stone's office span just half a metre of shelf space. But inside these journals, in meticulous black printing, Stone has chronicled the longest journey that humans have ever launched.

Since they left Earth in 1977, the twin Voyager spacecraft have conducted pioneering explorations of Jupiter, Saturn, Uranus and Neptune, revealing these gas giants and their moons to be far more active than scientists had expected. Now the two probes are cruising towards the edge of the Solar System — a boundary that has yet to be crossed by any emissary from Earth.

Stone has chaperoned the Voyagers from their conception. He is the mission's first and so far only project scientist, tasked with juggling the competing needs of the scientists who use the Voyagers' instruments against those of the engineers that fly the craft. By all accounts, he has succeeded. "Somehow he got that discordant orchestra to play together," says Andrew Ingersoll, a planetary scientist at the California Institute of Technology (Caltech) in Pasadena who was on the team as the spacecraft flew past Jupiter and Saturn.

To many, the Voyagers are synonymous with the unflappable Stone. The man and the mission are bound together, even as the probes enter yet another phase in their storied lifetime. Almost 19 billion kilometres from Earth, Voyager 1 is flirting with the edge of interstellar space, the medium between the stars. Last July, it saw the flood of charged particles from the Sun subside to a mere trickle — a sign that the spacecraft may soon break out of the Solar System.

Stone, who is now 77, plans to be around when it happens. He is not about to dial back his legendary work habits — not as Voyager nears such a historic milestone, with its promise of provocative science. Sitting in his Caltech office, with the Voyager logs behind him, the trim, elfin physicist looks mildly astonished at any suggestion of retirement.

Talking about his spacecraft, Stone could just as easily be describing his own drive. "What keeps Voyager still alive," he says, "is that it's still discovering."

ATOMIC AMBITIONS

Stone grew up in Burlington, Iowa, a town on the Mississippi River where his father worked in construction and was always tinkering with machines. Young Edward devoured copies of *Popular Science* magazine and *The Book of Knowledge*, and taught himself to build radio receivers.

The atomic age was well under way as Stone finished high school, and a teacher at his junior college told him about the world-renowned physics programme at the University of Chicago, Illinois, home to nuclear pioneer Enrico Fermi. Off Stone went, planning to study nuclear physics. But in October 1957, the launch of Sputnik inspired him to switch to space physics. For his graduate work, he helped to develop balloon- and satellite-borne detectors to track fast particles and cosmic rays streaming into Earth's atmosphere from space. In 1964, that led to a research position — and then a faculty job — at Caltech, which was starting a space-physics group.

His successful work on cosmic-ray detectors caught the eye of engineers at the Jet Propulsion Laboratory (JPL) in Pasadena, who were developing a mission initially called Mariner Jupiter–Saturn '77. They recruited Stone in 1972 to serve as project scientist — the person who oversees a spacecraft's scientific goals.

The mission grew into the most ambitious planetary exploration ever. Its two probes would each survey the outer Solar System with 10 instruments and a radio-science experiment — a set of investigations more sophisticated than those carried out by the Pioneer 10 and 11 craft that had reached Jupiter and Saturn before them.

The two spacecraft launched from Cape Canaveral, Florida, in August and September 1977, carrying golden records engraved with messages and recordings from Earth. En route to Jupiter, they were renamed Voyager 1 and Voyager 2.

It was not all smooth sailing. The boom carrying the scientific instruments for Voyager 2 could not deploy fully after launch. And the main radio receiver on the same probe failed completely in the spring of 1978, forcing engineers to turn to a back-up system.

Those and other glitches put added pressure on Stone, who had to coordinate between the principal investigators in charge of the instruments and the mission engineers who were troubleshooting the issues. He was the one who had to work out what science was achievable, given the constraints of the craft.

The results started to pour in during 1979, when first Voyager 1, then Voyager 2, flew past Jupiter. They spotted sulphur volcanoes belching from Jupiter's moon Io, as it was flexed by the giant planet's powerful gravity. Passing by the moon Europa, the probes photographed long, dirty-looking fractures in its icy surface, a possible clue to a subsurface ocean that could harbour extraterrestrial life. And they discovered a plasma, with temperatures of hundreds of millions of degrees Celsius, enshrouding Jupiter's magnetosphere. These and dozens of other finds now fill planetary-science textbooks.

When the probes arrived at Saturn in 1980–81, they uncovered 'shepherd' moons that herd the ice and dust in Saturn's outermost ring, which turned out to have a strange kinked appearance. The Voyagers also studied gigantic aurorae that swallowed much of the planet's northern and southern poles.

Next, mission controllers sent Voyager 1 on a trajectory out of the plane of the planets and towards the boundary with interstellar space. Voyager 2 sailed on to Uranus in 1986, where it found two new rings, ten new moons and an odd magnetic field oriented far away from the planet's rotation axis. Images taken of Miranda, the smallest and innermost of Uranus's major moons, showed a strikingly complex face with a deep, chevron-shaped groove that hints at a bizarre geological past.
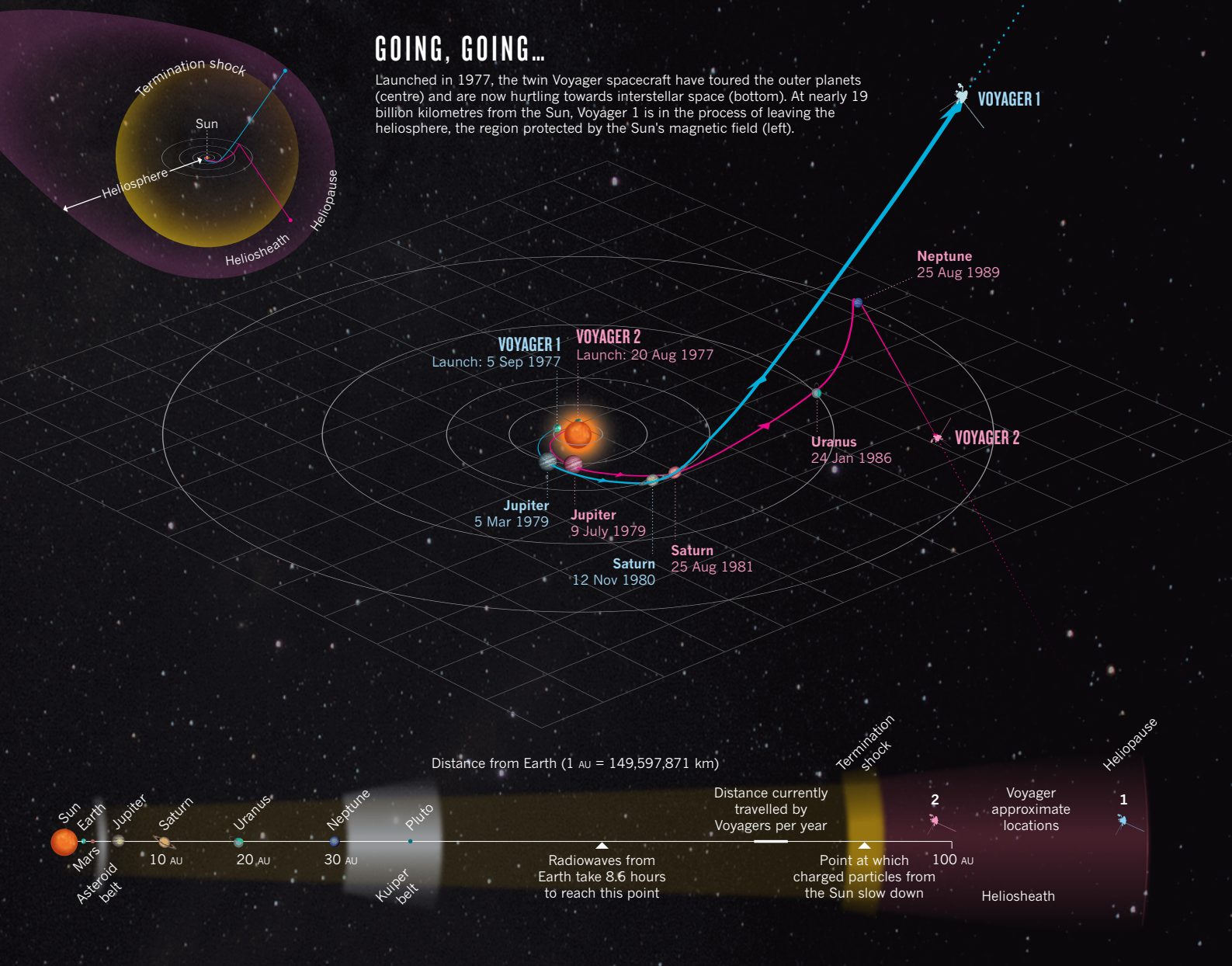
When Voyager 2 reached the seemingly placid

BRAD SWONETZ/REDUX/EYEVINE

GOING, GOING…

Launched in 1977, the twin Voyager spacecraft have toured the outer planets (centre) and are now hurtling towards interstellar space (bottom). At nearly 19 billion kilometres from the Sun, Voyager 1 is in the process of leaving the heliosphere, the region protected by the Sun's magnetic field (left).

Termination shock

Sun

Heliosphere

Heliosheath

Heliopause

VOYAGER 1

Neptune
25 Aug 1989

VOYAGER 1
Launch: 5 Sep 1977

VOYAGER 2
Launch: 20 Aug 1977

Uranus
24 Jan 1986

VOYAGER 2

Jupiter
5 Mar 1979

Jupiter
9 July 1979

Saturn
12 Nov 1980

Saturn
25 Aug 1981

Distance from Earth (1 AU = 149,597,871 km)

Termination shock

Heliopause

Sun

Earth

Mars

Jupiter

Saturn
10 AU

Uranus
20 AU

Neptune
30 AU

Pluto

Asteroid belt

Kuiper belt

Distance currently travelled by Voyagers per year

2

Voyager approximate locations

1

Radiowaves from Earth take 8.6 hours to reach this point

Point at which charged particles from the Sun slow down

100 AU

Heliosheath

blue disk of Neptune in 1989, it discovered winds blowing at 2,100 kilometres per hour, the fastest in the Solar System, which fuelled storms such as an Earth-sized Great Dark Spot. Scientists had not expected such violent atmospheric activity on a planet that receives just 0.1% of the solar energy that bathes Earth and drives its weather patterns.

Stone is fond of saying that the Voyagers have returned 200% of the science he had expected: "We learned so much more than we possibly could have imagined."

TEAM PLAYERS

Voyager alumni say that Stone deserves a major share of the credit for that achievement. As project scientist, he has from the beginning chaired the science-steering group, which is composed mainly of the 11 principal investigators. When the group could not agree on which observations to prioritize, Stone stepped in and made that decision — essentially choosing who would get to make discoveries.

Hence his shelf of notebooks. By carefully recording everyone's input, Stone says, he let everyone know that he considered all perspectives. "We always knew he was fair," says Ellis Miner of the JPL, who was assistant project scientist for the Saturn, Uranus and Neptune encounters.

In a shift from usual practice at the time, Stone forced Voyager

scientists to work across the clique-like teams that had sprung up around each of the 11 investigations. Before the Jupiter fly-by, he created overarching working groups around four key themes: moons, rings, atmosphere and magnetosphere. Each group was tasked with focusing on the scientific questions that could be answered by all of Voyager's instruments. This forced members of the instrument teams to talk across boundaries, Stone says, so that "I didn't have to be a referee all the time".

Tension still cropped up. Stamatios Krimigis, the principal investigator for the instrument that measures low-energy charged particles, remembers one particular stand-off with the plasma-instrument group. Krimigis's particle detector worked by rotating sensors to scan different parts of the sky, which shook the nearby plasma instrument and irritated the scientists in charge of it. Stone negotiated a compromise schedule: sometimes the particle instrument rotated quickly, sometimes it rotated slowly to reduce the vibrations, and at other times it remained still. "We were all in the end equally unhappy," says Krimigis, of the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland, and the Academy of Athens. "That's the sign of a good negotiator."

After Voyager 2 left Neptune in 1989, the two craft earned a new name to reflect their next assignment: the Voyager Interstellar Mission (see 'Going, going…'). The title carried a fair dose of hope, given that no one

NIK SPENCER/NATURE

knew how long it would take to coast to the edge of the Solar System. In the intervening years, however, Stone had plenty to keep him busy. From 1991 to 2001 he served as director of the JPL, overseeing mission successes such as the 1997 Pathfinder landing on Mars as well as spectacular failures such as the loss of both a Mars orbiter and a Mars lander that followed Pathfinder. It was the era of NASA's 'faster, better, cheaper' approach to spacecraft, and Stone admits that the missions failed because the JPL pushed that ethos too far.

After retiring from that hectic post, Stone returned to teach and do research at Caltech. Today, he rarely travels up the 210 freeway to Voyager mission control at the JPL. There is little need, because engineers carry out the daily tasks necessary to keep the probes healthy and in touch. Their job gets tougher every day: Voyager 2 is 15.2 billion kilometres from the Sun and Voyager 1 has reached 18.6 billion kilometres — more than three times the distance between the Sun and Pluto.

One early morning in April, the team engages in an agonizingly slow conversation with Voyager 1. Engineer Roger Ludwig is testing a new command sequence that would allow Voyager to delay some tasks to better cope with the limits on Earth-bound communications. It took more than 17 hours for the commands to travel at light speed to the spacecraft, and a similar span for the response to return. Now, in the predawn hours, Ludwig is eager for the answer.

A slow stream of numbers pops up on a pair of computer monitors, and Ludwig checks to see whether the sequence has worked. The results look promising, but the Voyager team will test the code several more times before making any changes. This long into the mission, engineers don't want to make any silly mistakes. "We all feel like we're flying a national treasure," says Ludwig.

That includes Stone, who has stuck with Voyager because he is eager for more discoveries. It has been a long wait since Neptune. The team has talked so many times about the impending departure into interstellar space that office doors around mission control are decorated with a photo of a forlorn-looking Voyager with the quote: "Whenever people stop paying attention to me, I pretend to leave the Solar System."

THE FINAL FRONTIER

The exit has turned out to be more complicated than scientists had anticipated. Voyager 1 is somewhere near the edge of the heliosphere, the giant cocoon of charged particles from the Sun that surrounds the Solar System and protects the planets from the high-energy particles that streak through interstellar space (see *Nature* **489**, 20–21; 2012).

In December 2004, the low-energy-particles instrument on Voyager 1 indicated that the solar wind had slowed abruptly, a sign that the craft had entered a turbulent boundary region surrounding the heliosphere.

Then, in July and August 2012, the speed of the solar wind dropped to essentially zero even as Voyager began recording higher-energy particles. Krimigis calls those changes "totally and completely unanticipated". By themselves, they might suggest that Voyager 1 had crossed from the boundary region into interstellar space.

But the science team has remained cautious because another expected signal has not yet appeared. When Voyager 1 enters interstellar space for real, Stone and others expect the orientation of the magnetic field to change from predominantly east–west (as driven by the Sun) to randomly changing directions. So far, data from Voyager 1's magnetometer show essentially no change in the direction of the magnetic field.

The team has struggled to make sense of these signals. One day in April, Stone wrinkles his forehead while looking over some plots of particle data. "We're now seeing what's outside, even though we're not outside," he says. "The magnetic field says we haven't gone out yet." He and other mission scientists think that Voyager 1 is on some kind of 'magnetic highway' that connects the Sun's magnetic field lines to those of interstellar space, allowing charged particles to enter the border zone. The team described the latest results in December at the American Geophysical Union (AGU) meeting in San Francisco, California, and will report details in an upcoming suite of papers in *Science*.

Stone works to keep tight control over the Voyager message. In March,

for example, he moved quickly to counter a press release from the AGU announcing that Voyager 1 had left the Solar System. It will leave when he says it does.

At that point, the spacecraft will enter a realm completely new to science. The particle detector will measure galactic cosmic rays that are too weak to penetrate the heliosphere and enter the Solar System. Voyager's magnetometer will gauge the strength of the magnetic field between nearby stars for the first time. Scientists will finally get a glimpse of what truly deep space is like.

The information coming back from Voyager 1 as it leaves the heliosphere "is the only data we'll ever get from this region, and it's incredible," says Merav Opher, an astrophysicist at Boston University in Massachusetts, who is not part of the Voyager team.

With the key transition so close, Stone is getting nervous. To make sure that scientists catch the change when it happens, he made a strong pitch for more coverage by the worldwide system of giant antennas known as the Deep Space Network. The mission is now getting as many as ten precious hours of antenna time daily.

ENERGY CRISIS

But Stone and his team know that they have limited time. Voyager 1 is 124 times as far from the Sun as the Earth–Sun distance, and gaining 3.6 of these astronomical units every year. Signals are steadily fading. Both spacecraft are powered by about 315 watts from the radioactive decay of on-board plutonium-based generators, but that power drops by about 4 watts every year. Of the 10 original instruments, five are still working on Voyager 2 and four on Voyager 1.

> "WHENEVER PEOPLE STOP PAYING ATTENTION TO ME, I PRETEND TO LEAVE THE SOLAR SYSTEM."

By 2020, the power will have ebbed to the point that mission managers will have to start switching off more scientific instruments, one by one. The job of choosing which goes first will fall to the project scientist. Stone says that he has not yet thought about which instruments to let die, should he be around to decide.

By 2025, all the plutonium power will be gone, and the Voyagers will become lifeless hulks. They will probably never come as close to another star as they have been to the Sun, and their famous golden records will drift mutely through space.

Yet Stone has no time to get nostalgic; there is more discovering to be done. Among other jobs, he serves as vice-chairman for the board of the Thirty Meter Telescope, which will be the world's largest optical telescope when it is completed in the early 2020s on Mauna Kea, Hawaii. One day, its giant mirrors will gaze past the Voyagers towards distant star systems.

Stone is also helping to develop Solar Probe Plus, a mission designed to go closer to the Sun than any other spacecraft. It will follow an elaborate looping path that will take it past Venus seven times and then into a series of close solar encounters, swooping within 10 solar radii of the Sun's surface again and again. The probe's heat shield will protect it from temperatures of 2,000 °C, and its instruments will collect information about how and where the solar wind is born.

Standing outside his office, Stone gazes at a bulletin board covered with printouts of the Sun's activity over the past few solar cycles. "I've never been frustrated with exploration," he says. "I've been lucky to have the right thing to do, to be on projects that have been successful."

Solar Probe Plus is planned for a 2018 launch, which would put its first close encounter with the Sun in December 2024. By then Stone will be pushing 89. But even so, he would like to be around when the science starts streaming in — about the same time that the Voyagers will be fading beyond communication into the cold cosmos. ∎

Alexandra Witze *reports for* Nature *from Boulder, Colorado.*

The *big fat* truth

More and more studies show that being overweight does not always shorten life — but some public-health researchers would rather not talk about them.

BY VIRGINIA HUGHES

Late in the morning on 20 February, more than 200 people packed an auditorium at the Harvard School of Public Health in Boston, Massachusetts. The purpose of the event, according to its organizers, was to explain why a new study about weight and death was absolutely wrong.

The report, a meta-analysis of 97 studies including 2.88 million people, had been released on 2 January in the *Journal of the American Medical Association* (*JAMA*)[1]. A team led by Katherine Flegal, an epidemiologist at the National Center for Health Statistics in Hyattsville, Maryland, reported that people deemed 'overweight' by international standards were 6% less likely to die than were those of 'normal' weight over the same time period.

The result seemed to counter decades of advice to avoid even modest weight gain,

provoking coverage in most major news outlets — and a hostile backlash from some public-health experts. "This study is really a pile of rubbish, and no one should waste their time reading it," said Walter Willett, a leading nutrition and epidemiology researcher at the Harvard school, in a radio interview. Willett later organized the Harvard symposium — where speakers lined up to critique Flegal's study — to counteract that coverage and highlight what he and his colleagues saw as problems with the paper. "The Flegal paper was so flawed, so misleading and so confusing to so many people, we thought it really would be important to dig down more deeply," Willett says.

But many researchers accept Flegal's results and see them as just the latest report illustrating what is known as the obesity paradox. Being overweight increases a person's risk of diabetes, heart disease, cancer and many

other chronic illnesses. But these studies suggest that for some people — particularly those who are middle-aged or older, or already sick — a bit of extra weight is not particularly harmful, and may even be helpful. (Being so overweight as to be classed obese, however, is almost always associated with poor health outcomes.)

The paradox has prompted much discussion in the public-health community — including a string of letters in *JAMA* last month[2] — in part because the epidemiology involved is complex, and eliminating confounding factors is difficult. But the most contentious part of the debate is not about the science per se, but how to talk about it. Public-health experts, including Willett, have spent decades emphasizing the risks of carrying excess weight. Studies such as Flegal's are dangerous, Willett says, because they could confuse the public and doctors, and undermine public policies to curb rising obesity rates. "There is going to be some percentage of physicians who will not counsel an overweight patient because of this," he says. Worse, he says, these findings can be hijacked by powerful special-interest groups, such as the soft-drink and food lobbies, to influence policy-makers.

But many scientists say that they are uncomfortable with the idea of hiding or dismissing data — especially findings that have been replicated in many studies — for the sake of a simpler message. "One study may not necessarily tell you the truth, but a bulk of studies saying the same thing and being consistent, that really is reinforcing," says Samuel Klein, a physician and obesity expert at Washington University in St Louis, Missouri. "We need to follow the data just like the yellow brick road, to the truth."

THROWING A CURVE

The notion that excess weight hastens death can be traced back to studies from the US insurance industry. In 1960, a thick report based on data from policy-holders at 26 life-insurance companies found that mortality rates were lowest among people who weighed a few kilograms less than the US average, and that mortality climbed steadily with weight above this point. This spurred the Metropolitan Life Insurance Company (MetLife) to update its table of 'desirable weights', creating standards that were widely used by doctors for decades to come.

In the early 1980s, Reubin Andres, who was the director of the US National Institute on Aging in Bethesda, Maryland, made headlines for challenging the dogma. By reanalysing

WEIGHT WATCHING

In some studies, being overweight is associated with increased survival time, creating a U-shaped mortality curve.
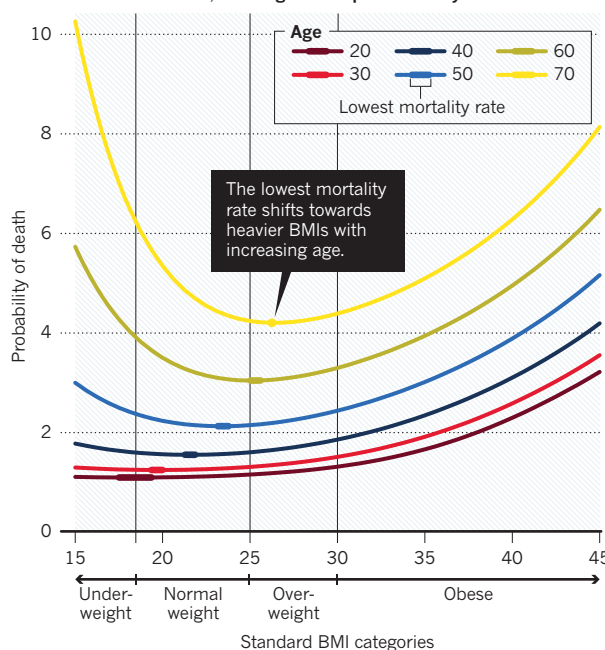
actuarial tables and research studies, Andres reported that the relationship between height-adjusted weight and mortality follows a U-shaped curve. And the nadir of that curve — the weight at which death rates are lowest — depends on age (see 'Weight watching'). The weights recommended by MetLife may be appropriate for people who are middle-aged, he calculated, but not for those in their 50s or older[3], who were better off 'overweight'. It was the first glimmer of the obesity paradox.

Andres's ideas were roundly rejected by the mainstream medical community. In an often-cited *JAMA* paper[4] published in 1987, for example, Willett and JoAnn Manson, an epidemiologist at the Harvard School of Public Health, analysed 25 studies of weight–death relationships and claimed that most were tainted by two confounders: smoking and sickness. Smokers tend to be leaner and die earlier than non-smokers, and many people who are chronically ill also lose weight. These effects could make thinness itself seem to be a risk.

Manson and Willett backed up that idea in a 1995 report that analysed body-mass index (BMI) — the 'gold-standard' measure of weight, defined as weight in kilograms divided by height in metres squared — in more than 115,000 female nurses enrolled in a long-term health study[5]. When the researchers excluded women who had ever smoked and those who died during the first four years of the study (reasoning that these women may have had disease-related weight loss), they found a direct linear relationship between BMI and death, with the lowest mortality at BMIs below 19. (That is about 50 kilograms for a woman

who is 1.63 metres tall.)

"It didn't seem to be biologically plausible that overweight and obesity could both increase the risk of life-threatening diseases and yet lower mortality rates," Manson says. The study proved, she says, that this idea "was more artefact than fact".

Around the same time, the world was waking up to obesity. Since 1980, rates of overweight and obesity had begun to rocket[6–8], and in 1997, the World Health Organization (WHO) held its first meeting on the subject, in Geneva, Switzerland. That meeting resulted in the introduction of new criteria for 'normal weight' (BMI of 18.5–24.9), 'overweight' (BMI of 25–29.9) and 'obese' (BMI of 30 or higher). In 1998, the US Centers for Disease Control and Prevention (CDC) lowered its BMI cut-offs to match the WHO's classifications. "We used to call [obesity] the Cinderella of risk factors, because nobody was paying attention to it," says Francisco Lopez-Jimenez, a cardiac physician at the Mayo Clinic in Rochester, Minnesota. They were now.

STATISTICAL SPARRING

Flegal was one of those raising the alarm. At the statistics centre, which is part of the CDC, she has at her fingertips data from the agency's National Health and Nutrition Examination Survey (NHANES). Based on interviews and physical examinations of about 5,000 people a year, the NHANES has been running since the 1960s. Flegal and her colleagues used it to show that rates of overweight and obesity in the United States were climbing[6,7].

In 2005, however, Flegal found that NHANES data confirmed Andres's U-shaped mortality curve. Her analysis showed that people who were overweight — but not obese — had a lower mortality rate than those of normal weight, and that the pattern held even in people who had never smoked[9].

Flegal's study got a lot of press, says Willett, because she works at the CDC and it seemed to be a sanction for gaining weight. "A lot of people interpreted this as being the official statement of the US government," he says. Just as they did earlier this year, Willett and his colleagues criticized the work and put together a public symposium to discuss it. The academic kerfuffle drew a lot of negative media attention to Flegal's study. "I was pretty surprised by the vociferous attacks on our work," says Flegal, who prefers to focus on the finer points of epidemiological number-crunching, rather than the policy implications of the resulting statistics. "Particularly initially, there were a lot of misunderstandings and confusion about our

findings, and trying to clear those up was time-consuming and somewhat difficult."

Over the next few years, other researchers found the same trend, and Flegal decided to carry out the meta-analysis that she published earlier this year[1]. "We felt it was time to put all of this stuff together," she says. "We might not understand what it all means, but this is what's out there." Her analysis included all prospective studies that assessed all-cause mortality using standard BMI categories — 97 studies in total. All the studies used standard statistical adjustments to account for the effects of smoking, age and sex. When the data from all adult age groups were combined, people whose BMIs were in the overweight range (between 25 and 29.9) showed the lowest mortality rates.

The Harvard group contends, however, that Flegal's approach did not fully correct for age, sickness-related weight loss and smoking. They say that the effect would have vanished in younger age groups if Flegal had separated them out. They also argue that not all smokers have the same level of exposure — people who smoke heavily tend to be leaner than occasional smokers, for example — so the best way to remove smoking as a confounder is to focus on people who have never smoked. Willett points to one of his studies[10], published in 2010, that was not included in Flegal's analysis because it did not use standard BMI categories. Analysing data from 1.46 million people, Willett and his colleagues found that among people who have never smoked, the lowest mortality occurs in the 'normal' BMI range, of 20–25.

Flegal, in turn, criticizes the Willett study for scrapping large swathes of the raw data set: nearly 900,000 people in all. "Once you delete such large numbers, and they are really large, you don't quite know how the never-smokers in the sample differ from the others," she says. Never-smokers could be richer or more educated, for example. What is more, says Flegal, Willett's study relies on participants' self-reported heights and weights, rather than objective measures. "It's a huge deal," Flegal says, because people tend to underestimate how much they weigh. This could skew death risks upwards if, for example, people who are obese and at high risk say that they are merely overweight.

HEALTHY BALANCE

Many obesity experts and health biostatisticians take issue with the harsh tone of Willett's statements about Flegal's work. They say that there is merit in both Willett's and Flegal's studies, that the two are simply looking at data in different ways and that enough studies support the obesity paradox for it to be taken seriously. "It's hard to argue with data," says Robert Eckel, an endocrinologist at University of Colorado in Denver. "We're scientists. We pay attention to data, we don't try to un-explain them."

What they are trying to explain is the reason for the paradox. One hint lies in the growing number of studies over the past decade showing that in people with serious illnesses such as heart disease, emphysema and type 2 diabetes, those who are overweight have the lowest death rates. A common explanation is that people who are overweight have more energy

> "WE NEED TO FOLLOW **THE DATA** JUST LIKE THE YELLOW BRICK ROAD, TO THE **TRUTH.**"

reserves to fight off illness. They are like contestants on the television show *Survivor*, says Gregg Fonarow, a cardiologist at the University of California, Los Angeles: "Those that started off pretty thin often don't come out successful."

Metabolic reserves could also be important as people age. "Survival is a balance of risks," says Stefan Anker, a cardiology researcher at Charité Medical University in Berlin. "If you are young and healthy, then obesity, which causes problems in 15 or 20 years, is relevant," he says. With age, though, the balance may tip in favour of extra weight.

Genetic and metabolic factors may also be at play. Last year, Mercedes Carnethon, a preventive-medicine researcher at Northwestern University in Chicago, Illinois, reported that adults who develop type 2 diabetes while they are of normal weight are twice as likely to die over a given period as those who are overweight or obese[11]. Carnethon says that the trend is probably driven by a subset of people who are thin yet 'metabolically obese': they have high levels of insulin and triglycerides in their blood, which puts them at a higher risk for developing diabetes and heart disease.

All this suggests that BMI is a crude measure for evaluating the health of individuals. Some researchers contend that what really matters is the distribution of fat tissue on the body, with excess abdominal fat being most dangerous; others say that cardiovascular fitness predicts mortality regardless of BMI or abdominal fat. "BMI is just a first step for anybody," says Steven Heymsfield, an obesity researcher and the executive director of the Pennington

Biological Research Center in Baton Rouge, Louisiana. "If you can then add waist circumference and blood tests and other risk factors, then you can get a more complete description at the individual level."

If the obesity-paradox studies are correct, the issue then becomes how to convey their nuances. A lot of excess weight, in the form of obesity, is clearly bad for health, and most young people are better off keeping trim. But that may change as they age and develop illnesses.

Some public-health experts fear, however, that people could take that message as a general endorsement of weight gain. Willett says that he is also concerned that obesity-paradox studies could undermine people's trust in science. "You hear it so often, people say: 'I read something one month and then a couple of months later I hear the opposite. Scientists just can't get it right'," he says. "We see that time and time again being exploited, by the soda industry, in the case of obesity, or by the oil industry, in the case of global warming."

Preventing weight gain in the first place should be the primary public-health goal, Willett says. "It's very challenging to lose weight once you're obese. That's the most serious consequence of saying there's no problem with being overweight. We want to have people motivated not to get there in the first place." But Kamyar Kalantar-Zadeh, a nephrologist at the University of California, Irvine, says that it is important not to hide subtleties about weight and health. "We are obliged to say what the real truth is," he says.

Flegal, meanwhile, says that the public's reaction to her results is not her primary concern. "I work for a federal statistical agency," she says. "Our job is not to make policy, it's to provide accurate information to guide policy-makers and other people who are interested in these topics." Her data, she says, are "not intended to have a message". ■ SEE EDITORIAL P.410

Virginia Hughes *is a science journalist based in New York.*

1. Flegal, K. M., Kit, B. K., Orpana, H. & Graubard, B. I. *J. Am. Med. Assoc.* **309,** 71–82 (2013).
2. Willett, W. C., Hu, F. B. & Thun, M. *J. Am. Med. Assoc.* **309,** 1681–1682 (2013).
3. Andres, R., Elahi, D., Tobin, J. D., Muller, D.C. & Brant, L. *Ann. Intern. Med.* **103,** 1030–1033 (1985).
4. Manson, J. E., Stampfer, M. J., Hennekens, C. H. & Willett, W. C. *J. Am. Med. Assoc.* **257,** 353–358 (1987).
5. Manson, J. E. *et al. N. Engl. J. Med.* **333,** 677–685 (1995).
6. Kuczmarski, R. J., Flegal, K. M., Campbell, S. M. & Johnson, C. L. *J. Am. Med. Assoc.* **272,** 205–211 (1994).
7. Flegal, K. M., Carroll, M. D., Ogden, C. L. & Johnson, C. L. *J. Am. Med. Assoc.* **288,** 1723–1727 (2002).
8. Finucane, M. M. *et al. Lancet* **377,** 557–567 (2011).
9. Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. *J. Am. Med. Assoc.* **293,** 1861–1867 (2005).
10. de Gonzalez, A. B. *et al. N. Engl. J. Med.* **363,** 2211–2219 (2010).
11. Carnethon, M. R. *et al. J. Am. Med. Assoc.* **308,** 581–590 (2012).

COMMENT
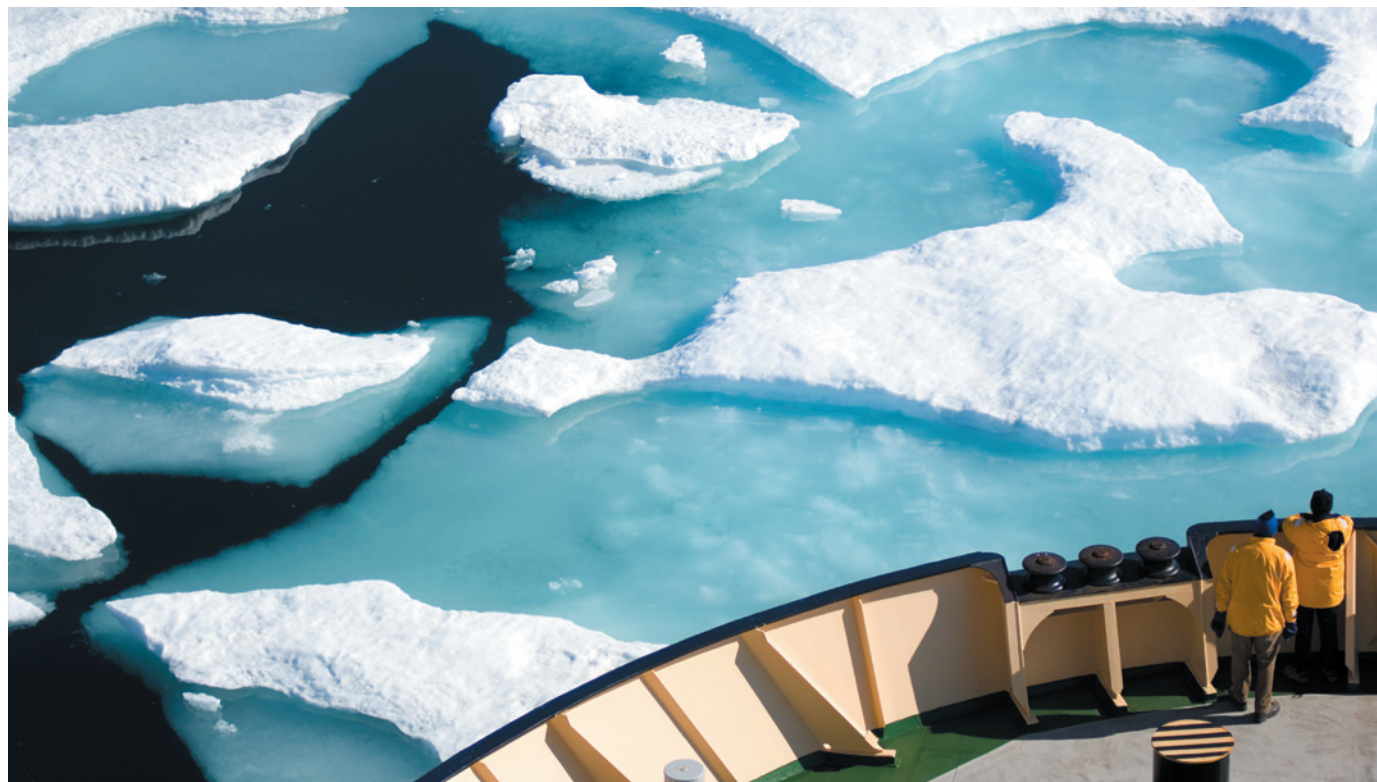
Sea ice in the Arctic is more variable and mobile than it was 30 years ago, causing problems for shipping.

Arctic sea ice needs better forecasts

Foster partnerships and share data to reduce hazards in fast-changing northern waters, says **Hajo Eicken**.

As the Arctic sea ice begins to retreat this summer, fleets of ships and drilling platforms are poised to enter the newly open waters. Each year, more of the region is accessible for longer. The seas freeze in the autumn several weeks later now than they did three decades ago[1]. The Arctic ice pack is on the verge of becoming seasonal.

Fast-changing sea ice brings hazards. Last summer, exploratory oil drilling in the Chukchi Sea off Alaska was halted on the second day when an incursion of ice threatened the safety of personnel and vessels. Such events will become more common as Arctic maritime activity grows and as the sea ice gets more variable and mobile[2]. With the area of permanent ice shrinking, social and environmental pressures will rise as people and animals compete for access[3].

Urgent action is needed to track changing conditions in the Arctic. Governments, agencies, indigenous peoples and the private sector all recognize the urgency and are engaged in a range of monitoring observations aimed at adaptation and mitigation. But, despite the partnerships that were built during the International Polar Year in 2007–08, current efforts are patchy.

What is needed are joint data-management policies and integrated observation planning. That is what was agreed at the first Arctic Observing Summit — a meeting of Arctic stakeholders held earlier this month in Vancouver, Canada, to discuss better coordination of and collaboration ▶

▶ for collecting and disseminating information. The question now is: what practical steps are needed to make this network real?

The scale of challenges in the Arctic requires more than simply moving existing approaches to the top of the world. New 'communities of practice' must be fostered alongside tools and support structures that aid information sharing and interpretation for different users.

FICKLE FLOES

Mapping sea ice is technically challenging. In summer, pools of meltwater on the ice surface confuse the microwave satellites that track sea ice so that they record the wrong ice concentrations. Newly formed ice contrasts little with open water and can be mistaken for waves or foam. My collaborator Winton Weyapuk — an ice expert and hunter who is a member of the Inupiaq people from northern Alaska — typically reports ice forming in coastal waters days before satellites detect it. This discrepancy, caused in part by the low resolution of satellite images, could be calamitous for maritime users.

Predictions of melting, freezing and movements of sea ice are difficult in a rapidly changing Arctic. Most of the data informing predictive models were derived for thick, old ice, which dominated the ice pack until a few years ago. We know little about the physical processes that govern the current mix of young (first year) and old ice.

For instance, the seasonal melt of young ice, which absorbs more solar heat and is more mobile than old ice, has proven challenging to predict. It is known to influence Earth's climate, atmosphere and ocean system, but the precise natures of atmospheric feedbacks, ocean circulation and heat retention are only now emerging and are difficult to model[4]. The Arctic ice pack also responds rapidly to weather and currents[2]. Long-term forecasting of sea-ice retreat is thus inherently limited.

Differing terminology and perspectives can be a problem. This was brought home to me last year at a workshop on sea-ice descriptions in Anchorage, Alaska, organized by the Exchange for Local Observations and Knowledge in the Arctic. Geophysicists, ice experts from indigenous communities, maritime operators and agency personnel came together for two days to share information about the seasonal ice cycle. We examined photographs, drew conceptual maps and reviewed scientific and local words for various types of sea ice.

It soon became clear that many terms are easily confused. The stages of 'new ice' that are considered hazardous by regulators, safe to walk on by the Inupiaq people, detectable through remote sensing or capable of

damaging coastal structures vary by days. Observations and predictions of sea ice therefore must be tailored to users.

Shortage of data is not always a problem. For example, around half of the 45 marine-observation programmes on the US side of the Chukchi and Beaufort Seas produce sea-ice information or relevant oceanographic data, with some programmes yielding up to 20 data sets each. But many data sets are difficult to find and access, and do not include products of immediate use to people in the Arctic. The data producers are mixed: one-third is based in academia, one-third is overseen by government agencies, one-fifth is operated by industry and the remainder are from countries other than the United States.

There is no single access point for Arctic data. Establishing one may not be necessary, as long as inter-operability and open exchange between repositories can meet many users' needs. But common priorities for Arctic monitoring still need to be developed, including a set of primary sea-ice variables that should be tracked as a minimum requirement.

"Combine sensor networks with boots on the ice and boats in the water ."

A consensus is emerging among researchers that the data should at least include ice-thickness measurements from satellites and airborne surveys; records of ice-surface and ice-bottom melt; information from drifting sensors on the evolution of ice properties; and the quantification of late-summer ocean heat using autonomous sensors and remote sensing[5–7]. Studies of interactions between the ocean, sea ice and the atmosphere at drifting ice stations will help to improve our understanding of these drivers and changes while unravelling the importance of such processes for marine ecosystems.

Observing strategies need to be aligned and streamlined. Existing efforts can help — such as the Arctic Council's initiative to catalogue various observing activities (called Sustaining Arctic Observing Networks, or SAON) and the Global Earth Observation System of Systems (GEOSS). But a broader approach to data gathering and dissemination is needed so that social and ecosystem impacts can also be addressed.

Translating the needs of Arctic users into scientifically tractable questions can be a first step toward broadening collaborations. The Arctic Sea Ice Outlook[5], a compilation of seasonal projections for the minimum extent of Arctic sea ice, for example, is moving beyond its academic origins to develop tools that are useful to governments, companies and the public. Last year, for instance, near-real-time estimates of ice thickness improved predictions of lingering

sea ice in the North American Arctic[8].

It is crucial to sustain observations and data sharing for the long term. International coordination of funding to maintain vital observing programmes is lacking, however, despite its importance for maritime safety in the Arctic.

ARCTIC CROSSROADS

New partnerships and processes are needed to track and predict sea ice. National weather and ice services and the World Meteorological Organization (WMO), based in Geneva, Switzerland, have a track record and should play an important part. But the impetus has to come from a broader base, including all Arctic stakeholders. The following steps are needed.

First, the priorities of Arctic scientists and other users must be defined for sea-ice observation, prediction variables and regions of interest. These must ensure safe maritime operations and protect ecosystem services and habitats. Observations and predictions should target ice-associated hazards, such as the remnants of old, thick ice or highly dynamic regions, as well as services that the ice provides, such as a platform for marine mammals and people.

The standard products released by ice and weather services will not fit this bill. Users must identify which aspects of the ice are most important to them. It may be the distribution of rotten, sparse ice that supports colonies of walruses and seals during the late spring, or the degrees of ice morphology, roughness and thickness that allow people safe passage across it.

Second, studies of ice processes must be carried out in a much more dynamic, meltwater-dominated Arctic Ocean environment. In the past, perennial ice served as an ideal platform for drifting sensors and field camps. The transition to seasonal ice requires amphibious drifters and autonomous airborne or submarine vehicles to do the same tasks.

Sea-ice models and remote-sensing algorithms will have to consider the decay of seasonal ice through the summer. Inter-disciplinary field campaigns that combine sensor networks with boots on the ice and boats in the water can explain how the pulse of seasonal meltwater affects ice properties, mechanics and related ecosystems as well as threats to offshore structures.

Third, to translate these priorities into viable programmes we need a forum to link experts from different stakeholder groups[3]. Governments and agencies, industry and private foundations should provide sustained funding to put efforts such as the Arctic Observing Summit on a solid foundation, administered by SAON and the International Arctic Scientific Committee.

Pressing next steps include the adoption

of a policy that promotes rapid, open access to observing data, following the protocols developed in the International Polar Year[9]. Frameworks for helping to plan and coordinate long-term observing activities across the scientific community and other sectors need to be established.

The community-based observing networks from the International Polar Year, which focus on variables related to local environmental threats or benefits, are a good start. But to be accessible to others, these data should be entered into wider networks such as those of the WMO. Similar to the practice of joint resource management[10], the scientific community, stakeholders and decision-makers all need to be included in governance from the outset to help ensure relevance and efficiency.

Opportunities remain for the private sector to contribute to such collaborative networks. Offering up commercial vessels or infrastructure as platforms for scientific observations, sharing data and engaging the research community in the planning stages of industry observing programmes would go a long way towards establishing a 'network of networks'.

Last month, I was fortunate to be out in a small boat off Toksook Bay in Alaska with ice experts and hunters from the Yup'ik people. We were surrounded by jagged, fast-moving chunks of ice that, to me, seemed hostile. To my companions, it was all in a day's work. I recalled a sentiment I had heard from a marine-mammal expert in Barrow, more than 1,000 kilometres farther north, where the ice is now unstable. He stated that the key to adapting to increasingly dynamic ice is to learn from those to the south, such as in Toksook Bay. The charge to the scientific community is to help to create a foundation for such mutual learning to occur. ∎

Hajo Eicken *is professor of geophysics at the University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA.*
e-mail: hajo.eicken@gi.alaska.edu

1. Markus, T., Stroeve, J. C. & Miller, J. *J. Geophys. Res. Oceans* **114,** C12024 (2009).
2. Rampal, P., Weiss, J. & Marsan, D. *J. Geophys. Res. Oceans* **114,** C05013 (2009).
3. Lovecraft, A. L., Meek, C. & Eicken, H. *Polar Geogr.* **36,** 105–125 (2013).
4. Screen, J. A. & Simmonds, I. *Nature* **464,** 1334–1337 (2010).
5. Calder, J., Eicken, H. & Overland, J. in *Understanding Earth's Polar Challenges: International Polar Year 2007–2008 — Summary by the IPY Joint Committee* (eds Krupnik, I. et al.) 405–410 (CCI Press, 2011).
6. National Research Council. *Seasonal to Decadal Predictions of Arctic Sea Ice: Challenges and Strategies* (National Academies Press, 2012).
7. Kauker, F. et al. *Geophys. Res. Lett.* **36,** L03707 (2009).
8. Lindsay, R. et al. *Geophys. Res. Lett.* **39,** L21502 (2012).
9. Parsons, M. A. *Nature* **458,** 830 (2009).
10. Berkes, F. *J. Environm. Mgmt.* **90,** 1692–1702 (2009).

Six red flags for suspect work

C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

A few months ago, I received a desperate e-mail from a postdoctoral scientist. Researchers — including me and my colleagues — had just reported that the majority of preclinical cancer papers in top-tier journals could not be reproduced, even by the investigators themselves[1,2]. The postdoc pleaded with me to identify those papers, saying: "I could be wasting my time working on that project." This was true, but we had signed confidentiality agreements that prevented us from revealing the specific papers. Furthermore, identifying them would not address the broader, systemic issues in research and publishing that create a plethora of papers that don't stand up to scrutiny.

There were some glaring differences ▶

between the 90% of papers that we could not reproduce and the few papers that we could. In our initial exercise[2], we contacted researchers whose work we were unable to reproduce to discuss discrepancies. Occasionally, experiments were repeated by the original authors — the most dramatic results came from investigators who could not reproduce their own work, when performed in their own laboratory, using their own reagents. The only difference the second time was that they had to perform the experiments blinded.

Many of the investigators whose work could not be reproduced were, however, prepared to honestly describe their experimental approaches to us in confidence. These non-reproducible papers shared a number of features, including inappropriate use of key reagents, lack of positive and negative controls, inappropriate use of statistics and failure to repeat experiments. If repeated, data were often heavily selected to present results that the investigators 'liked'. These, we found, are common flaws of non-reproducible papers, which apply to all basic biological research. Addressing them during the writing, editing and reviewing of research could go a long way towards creating a more robust scientific enterprise.

So, here are six questions that every author, editor, reviewer and reader should ask themselves when evaluating a research paper.

> *"Many of these flaws were identified and expunged from clinical studies decades ago."*

SIX QUESTIONS

Were experiments performed blinded? It is much easier to obtain the result that makes the best story and that best fits a hypothesis when experiments are performed by unblinded investigators. So, first, check the methods and figure legends. For instance, animal studies, *in vitro* work and reading of gels — which are used in protein or DNA separation — can and should all be done, or at the very least reviewed, by an investigator blinded to the experimental versus control groups. Even rare lone investigators can introduce some level of blinding. It is unusual to find blinded studies in basic research in top-tier journals. If experiments are performed blinded, it increases the likelihood that the work will stand the test of time.

Were basic experiments repeated? This is crucial to know in any study. Unfortunately, repetitions are seldom performed. Western blotting (a technique that uses antibodies to detect specific proteins in a mixture) and similar analyses are often performed only once, and when the desired result is obtained, that result is shown. Studies using RNA interference frequently show the results of a single experiment. Often only one or two cell lines are examined. If reports fail to state that experiments were repeated, be sceptical.

Were all the results presented? Inappropriate data selection is a crucial issue. Most western blots show only a sliver of the gel with the majority of bands cropped. Although many of these cropped bands may be extraneous, their removal falsely implies that the antibody could detect only the desired protein, which is rarely the case. In addition, size standards are often not shown. Without them, the reader cannot have any confidence that the bands identified are even remotely of the correct size. It can be valuable to compare the results of other experiments in the paper that used the same antibody: the pattern of bands should be the same across experiments.

It is always beneficial to cross-check images. Since the *Journal of Cell Biology* began routinely screening images, it has had to revoke 1% of acceptances after finding digitally manipulated image files[3]. Beware the 'typical result'; ask to see all of the results. One investigator admitted to us that he selected the one atypical result that supported his hypothesis and ignored the majority of experiments that did not.

Were there positive and negative controls? Often in the non-reproducible, high-profile papers, the crucial control experiments were excluded or mentioned as 'data not shown'. Yet it is impossible to evaluate data properly without reviewing the controls. Another common practice is to show photos of gels that are over-exposed and well outside the linear range of the film. Over-exposure of the controls makes it impossible to assess the relative amounts of total material being compared. When arguing that there is a difference between samples in the intensity of a specific signal, it is crucial to know that equivalent amounts of total sample were compared. But with over-exposed controls, that difference is obscured, and an alleged difference between samples may simply be the consequence of loading more total sample. A publication that hides the controls should be viewed with caution.

Were reagents validated? Several errors are common here. Of course, it is vital to know that the selected antibodies detect only the antigen under study. Yet, typically, the crucial western blot (showing only a single band) or other analyses that validate the reagent are not shown. Instead there is often a reference to an earlier paper, which does not show the essential data either. There are also examples of investigators using an antibody even when the manufacturer declares it unfit for that particular purpose. Experiments with small-molecule inhibitors are particularly problematic. Investigators choose to attribute the desired effect to their favourite molecule, ignoring the multiple other targets affected by the inhibitor, or consign the key experiments that allegedly demonstrate their lack of relevance to 'data not shown'.

Were statistical tests appropriate? Improper statistical analysis is commonly seen in animal studies, in which results are collected over a long time. On such a time curve, two points may be highlighted and declared to be significantly different from points on the control curve, even though the totality of the two curves is essentially the same. Check that the statistical test has been applied to the whole curve, rather than just to selected points along it (the position of the asterisk marking the statistical *P* value is an important clue)[4].

Remarkably, these six flaws are common to many papers, even those that we did not include in our original analysis. As an informal exercise, I recently thumbed through the pile of high-profile journals on my desk: the vast majority included at least one paper — and often more — that contained one or more of the basic flaws outlined here. What is also remarkable is that many of these flaws were identified and expunged from clinical studies decades ago. In such studies it is now the gold standard to blind investigators, include concurrent controls, rigorously apply statistical tests and analyse all patients — we cannot exclude patients because we do not like their outcomes.

Why do we repeatedly see these poor-quality papers in basic science? In part, it is down to the fact that there is no real consequence for investigators or journals. It is also because many busy reviewers (and disappointingly, even co-authors) do not actually read the papers, and because journals are required to fill their pages with simple, complete 'stories'. And because of the apparent failure to recognize authors' competing interests — beyond direct financial interests — that may interfere with their judgement.

Every biologist wants and often needs to get a paper into *Nature* or *Science* or *Cell*, yet the scientific community fails to recognize the perverse incentive this creates. Some of these issues could be readily addressed by publishing only blinded, replicated and appropriately controlled preclinical experiments. Isn't that what my postdoc colleague expected we were doing already? ∎

C. Glenn Begley *is senior vice-president at TetraLogic Pharmaceuticals, Malvern, Pennsylvania, USA.*
e-mail: cgbegley@tetralogicpharma.com

1. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10,** 712 (2011).
2. Begley, C. G. & Ellis, L. M. *Nature* **483,** 531–533 (2012).
3. Rossner, M. 'How to Guard Against Image Fraud' *The Scientist* (1 March 2006); available at http://go.nature.com/bjjbe4
4. Vaux, D. L. *Nature* **492,** 180–181 (2012).

NOVARTIS COMPANY ARCHIVES

Albert Hofmann holds a model of an LSD molecule.

DRUG DISCOVERY

Synthesized dreams

Mike Jay contemplates the chemical legacy of Albert Hofmann, who first synthesized LSD 75 years ago.

Albert Hofmann never had a rational explanation for his discovery of the effects of lysergic acid diethylamide (LSD), the 'eureka moment' that changed his life. He first synthesized the hallucinogen in 1938, in the laboratory of pharmaceuticals company Sandoz in Basel, Switzerland, while searching for a cardiovascular stimulant based on ergot, a grain fungus. However, it had shown no particular promise. Five years later, a "peculiar presentiment" prompted him to recreate the molecule, after which he noticed a feeling of slight dizziness and a mild psychoactive effect.

To test whether this had been caused by accidental absorption, three days later Hofmann took what he imagined would be an almost imperceptible 250-microgram dose. By the time he hopped on his bicycle to go home, he found himself taking the ride of his life, entering a hallucinatory landscape of "kaleidoscopic, fantastic images". The modern psychedelic era had begun, although Hofmann had no inkling of the social revolution his discovery would foment.

The classic account of that legendary

bicycle trip appears in Hofmann's memoir, *LSD: My Problem Child*, first published in German in 1979. An English edition from 1980 was withdrawn after publisher McGraw–Hill decided that books about psychedelic drugs should not be on its science list. The memoir is now republished as a new English translation by Jonathan Ott, together with *Insights/Outlooks*, a collection of Hofmann's philosophical essays that first appeared in German in 1986.

Ott's translation joins another classic, Hofmann's *Plants of the Gods* (McGraw–Hill, 1979), the book on the botany and chemistry of plant hallucinogens he co-authored with eminent botanist Richard Schultes. Together with the abundance of valuable photographs and original documents offered in an uncritical biography of Hofmann, *Mystic Chemist*, these constitute an admirable summation of Hofmann's life and the extraordinary chemical vistas he opened up.

LSD: My Problem Child reminds us that the first generation of LSD pioneers were cultural conservatives. The tone of their investigations is captured in the evocative group photos in *Mystic Chemist* of industrial chemists in grey flannel suits and with Brylcreemed hair. The drug's biochemistry — now known to act on serotonin receptors — was a mystery to them. However, its powerful effects on perception and cognition (such as making walls seem to ripple and time to stretch out) were found to be similar to those of mescaline, which was isolated from the peyote cactus in 1897. Mescaline had been widely studied by physiologists and psychiatrists, and by writers such as Aldous Huxley, who documented his experiments in *The Doors of Perception* (Harper & Brothers, 1954) and later befriended Hofmann.

Among this coterie, psychiatrists such as Humphry Osmond saw LSD as a means of investigating the biochemical basis of schizophrenia. The Central Intelligence Agency (CIA) was also testing it covertly as a brainwashing tool, recklessly administering it to unwitting subjects for as many as 77 successive days. Unlike some of his associates, Hofmann staunchly refused to cooperate with the CIA.

The emerging California counterculture seemed shallow and degenerate to this group. Hofmann's lifelong mentor, author Ernst Jünger, wrote, "I can't agree with Huxley's idea that the masses could be given the opportunity for transcendence." The ▶

▶ flashpoint came with psychologist Timothy Leary's buccaneering drug experiments that eventually saw him expelled from Harvard University in Cambridge, Massachusetts. In January 1963, Sandoz received an order from Leary for 100 grams of LSD, enough for a million doses. Hofmann referred the request to the US Food and Drug Administration and noted that Leary was courting media publicity "that can only be injurious to the scientific clarification of questions connected with the use of these substances".

Sandoz returned Leary's cheque and in 1966 withdrew LSD from the market, where it had been in steady demand for neurological research and psychiatric use. The vacuum was swiftly filled by underground chemists, who turned out pills and LSD-soaked blotters at strengths that Hofmann considered a massive overdose. But there was no turning the tide.

Hofmann continued to work on the chemistry of ergot alkaloids, on which he became the world's leading authority, and rose to deputy managing director at Sandoz. The next 50 years of his life, as *Mystic Chemist* doggedly relates, was largely swallowed up by a procession of ethnographers, botanists, psychiatrists, shamans and politicians, seeking his wisdom and showering him with honours.

LSD has been many things to many people, but for Hofmann its core message was that objective reality "is in fact a subjective image". By enhancing idiosyncratic perceptions and ideas, it allows us to create our own reality. In Hofmann's case, it returned him to the radiant childhood vision of nature that had led him to study chemistry in the first place. Among his touchstones were the spirit-infused *Naturphilosophie* of author Johann Wolfgang von Goethe, the poetry of Johann Christoph Friedrich von Schiller and Rainer Maria Rilke and the alchemical insights of his Swiss compatriot Paracelsus, all enveloped in a mystical Christian revelation that he was "*united with the Creation*, as a part of the Universe".

After 42 years at Sandoz, Hofmann retired in 1971 to his alpine home on the Swiss–French border, writing of coming full circle to the high meadows of the Jura Mountains, "with the same flowers, and the same view into the distance". ■

Mike Jay is an author and historian who has written widely on the history of drugs, psychiatry and madness. His books include High Society: Mind-Altering Drugs in History and Culture, *which accompanied the* High Society *exhibition he co-curated at the Wellcome Collection, London.*
e-mail: mail@mikejay.net

Wolves are among the species that people are seeking to restore to their former habitats.

ECOLOGY

Into the (re)wild

Shahid Naeem examines the seductive concept of repopulating habitats with locally extinct species.

In his 1901 book *Our National Parks*, early conservationist John Muir asserted that "wildness is a necessity". Environmental journalist George Monbiot agrees, but in *Feral* we discover just how challenging it can be to fulfil that need.

Monbiot starts by recounting his adventures during the 1980s gold rush in the Brazilian state of Roraima, among murderers, desperate miners and the Yanomami tribe's faith healers. Far from showing bravado, these tales reveal how negligence, inaction and greed have led to environmental devastation in the region. After six years, Monbiot returns home to Wales to find himself "living a life in which loading the dishwasher presented an interesting challenge". Burdened with angst, he looks to wilderness as an antidote to a descent into a "small and shuffling life".

But Monbiot finds that Wales has been tamed by forces similar to those that are devastating the Brazilian wilderness. He uses his home territory as a microcosm to explore 'rewilding' — the process of reintroducing locally extinct species — and elevates this from a paradigm in conservation science to a major environmental issue. He goes even further, envisioning rewilding as a way to revitalize imprisoned human sensibilities by restoring the endless source of delights, surprises, adventures and thrills that wilderness provides.

Rewilding seems simple: let jaguars roam the American West and rhinos repopulate the Nepalese–Indian border, or allow wolves, lynxes, wildcats, wolverines, beavers, boars and even moose to return to Wales. Yet Monbiot discovers that rewilding is not a straightforward process.

For a start, which species should be prioritized? Some were hunted to extinction not long ago — for example, boars in the thirteenth century. Others, such as moose and wolverines, disappeared thousands of years ago for unknown reasons. Should we go back to the Pleistocene epoch (which ended some 12,000 years ago) and use

Feral: Searching for Enchantment on the Frontiers of Rewilding
GEORGE MONBIOT
Allen Lane: 2013. 336 pp. £20

NATURE.COM
For more from Shahid Naeem on redefining the wild, see:
go.nature.com/pmrusq

African substitutes to bring back the hippos, rhinos, hyenas, lions and elephants that once populated Europe? Who bears responsibility if released animals introduce disease, damage crops, attack pets, kill livestock or maul humans?

And how do we define the 'wilderness' on which rewilding is based? Monbiot fails to do so, yet is exasperated by policies that designate sheep pasture as 'wilderness' or promote tree removal as 'wilderness management'.

In *The Idea of Wilderness* (Yale University Press, 1991), philosopher Max Oelschlaeger notes that the concept did not arise until humanity began to imagine itself as distinct and insulated from nature. Wilderness became those vestiges of our dominion that await clearing. To combat this definition, many conservation arguments now have a utilitarian perspective, portraying wilderness as habitat that serves humanity in ways that farms, pastures and plantations cannot. Wild areas prevent soil erosion, provide homes for pollinators and act as sources of food, medicine and more.

Whatever the motivation, some feel that rewilding could re-energize conservation science. Wildlife biologist Tim Caro, for instance, sees the process as galvanizing a science "of documenting population declines, species losses and habitat destruction in excruciating detail but sadly doing little about it". Caroline Fraser's *Rewilding the World* (Metropolitan Books, 2009) and Dave Foreman's *Rewilding North America* (Island Press, 2004) echo this idea. Monbiot, however, sees the prime purpose of rewilding as a cure for ecological ennui.

At one point, Monbiot sets out to sea in his kayak to catch a longfin tuna, a fish reported to be returning to the region. The weather unexpectedly turns lethal, but Monbiot survives to reflect on what he describes as the idiocy of being lured into such danger and neglecting his duty to his daughter and partner. His repentance is short-lived, and another adventure soon tempts him back to nature.

Monbiot's prowess at wild fishing is as impressive as the fact that he survived cerebral malaria, beatings by military guards and being stung into a coma by wasps. But *Feral* made me uncomfortable. Rewilding is too often misconstrued by the public as a science by and for animal lovers and thrill-seekers. Although Monbiot covers the issues admirably, his passion for peril could confuse the central value of rewilding. The subject may be a cure for ecological boredom, but its power to invigorate the imagination needs to be tethered to the goal of attaining a more robust future for us and our fellow species. ∎

Shahid Naeem *is professor of ecology at Columbia University, New York, USA.
e-mail: sn2121@columbia.edu*

Books in brief
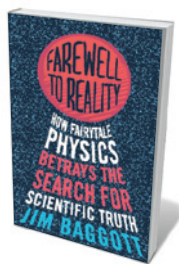
Farewell to Reality: How Fairytale Physics Betrays the Search for Scientific Truth
*Jim Baggott* CONSTABLE *352 pp. £12.99 (2013)*
"What evidence do we have for super-symmetric 'squarks'?" asks science writer Jim Baggott, bemoaning the untestable speculative theories of theoretical physics. But this is more than an energetic drubbing of them. Baggott thoroughly analyses the cracks in the authorized version of reality (such as the disjunct between quantum theory and general relativity) and equally problematic explanations, such as superstring theory. The big questions, he argues, demand context-awareness, and the courage to make doubts public.
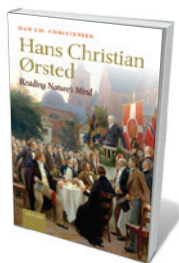
Hans Christian Ørsted: Reading Nature's Mind
Dan Charly Christensen OXFORD UNIVERSITY PRESS *768 pp. £39.99 (2013)*
Charles Darwin may have found Hans Christian Ørsted's philosophical tract *The Soul in Nature* "dreadful", but the nineteenth-century Danish physicist and chemist discovered electromagnetism, kicking open the door for Michael Faraday, James Clerk Maxwell, and a sprawling technoscape of invention. In this powerhouse of a biography-cum-science-history, historian Dan Charly Christensen reveals Ørsted as a brilliant, multifaceted figure — scientist, post-Kantian philosopher, educator and physical aesthetician.
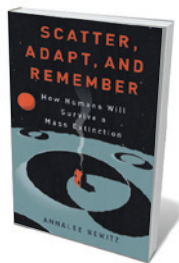
Scatter, Adapt, and Remember: How Humans Will Survive a Mass Extinction
Annalee Newitz DOUBLEDAY *320 pp. $26.95 (2013)*
Ever since the 'Snowball Earth' glaciations 2.3 billion years ago, life has squeaked through mass extinctions. Annalee Newitz examines how humans might survive the next one. Skipping through catastrophe research, she then explores how *Homo sapiens* has survived plagues, famines and more — harnessing everything from basic adaptability to pandemic surveillance and geoengineering. Together these tools and abilities are, notes Newitz, indicators of our phenomenal capacity to survive the unthinkable.

The Summits of Modern Man: Mountaineering After the Enlightenment
*Peter H. Hansen* HARVARD UNIVERSITY PRESS *392 pp. $35 (2013)*
In this treatise on peak bagging and its post-Enlightenment significance, Peter Hansen unearths stories with a crampon-like grip. We meet star summiteers such as Michel-Gabriel Paccard and Jacques Balmat, who both ascended Mont Blanc in 1786; nineteenth-century 'golden age' alpine mountaineers like physical scientist John Tyndall; and Edmund Hillary and Tenzing Norgay on Everest. But with today's retreating glaciers, Hansen reminds us that the idea of what it means to 'conquer' nature needs reconfiguring.

Consolations of the Forest: Alone in a Cabin in the Middle Taiga
Sylvain Tesson (translated by Linda Coverdale) ALLEN LANE *256 pp. £16.99 (2013)*
Six months in a hut on the shores of Siberian "cauldron of evolution" Lake Baikal spawned Sylvain Tesson's vivid memoir, part meditation on stupendous nature, part psychological experiment. This is total immersion in the seasonal minutiae of an extreme environment. Raw joys abound, from encounters with solitude-crazed meteorologists and foraging bears to the illumination of cracks in the mighty lake's ice and the interiorities of his own musing mind.
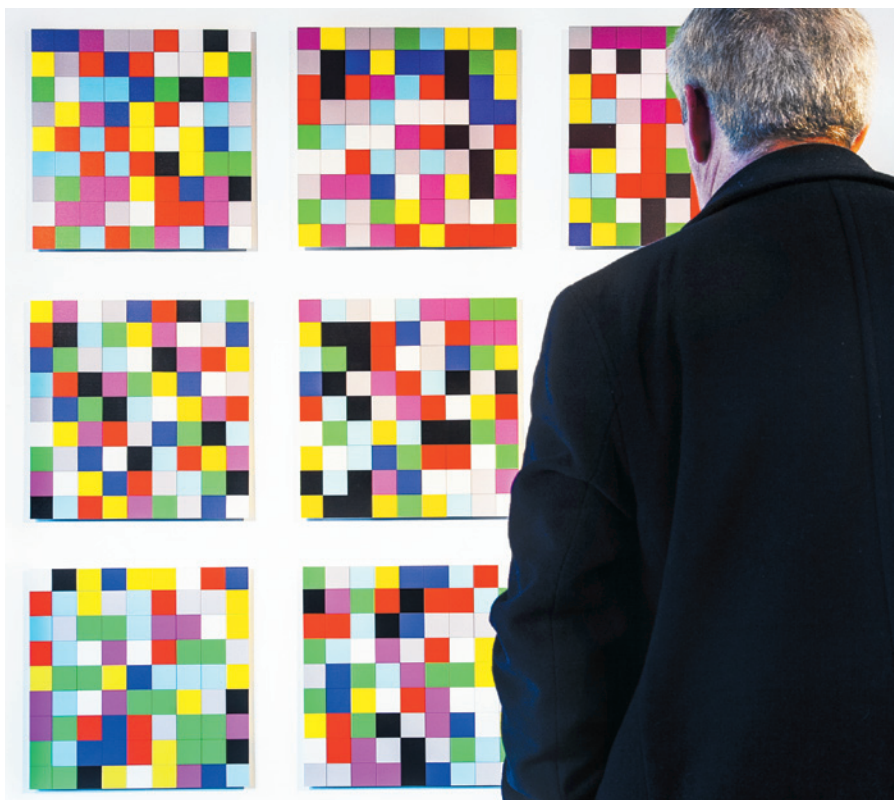
Totally Random? by David Spiegelhalter and Mike Pearson challenges the viewer to find a pattern.

STATISTICS

Charting chance

Anthony King tries his luck at a casino that lays bare the psychology of risk–taking.

Risk is a constant in our lives, permeating everything from allowing your child to walk to school alone to snowboarding late in the season. *Risk Lab*, the new exhibition at Dublin's ever-inventive Science Gallery, offers ample opportunities to explore this nexus of psychology and statistics.

Enter the exhibition's geeked-up casino and venture into the world of scientific gambling, with gallery staff on hand to help with the maths. Downstairs, you can puzzle out probabilities by playing roulette and poker or taking in themed artworks, such as one that challenges you to spot a non-random pattern among nine pictures of coloured squares. Upstairs, computers mocked up as slot machines offer psychological tests in a variety of activities that feature a large element of risk.

A central message in *Risk Lab* is that most of us use gut feelings rather than cerebral churning to do the sums when assessing risk. As Science Gallery events manager Ian Brunswick points out, intuition is generally plan A. "Humans are terrible at understanding

RISK LAB
SCIENCE GALLERY,
DUBLIN
Dublin, Ireland. Until 23 June

statistics and evaluating risk and chance," he says. Superstar statisticians can evaluate risks in elections, baseball, financial crashes and poker, but their analyses involve wading through banks of data. Most people paddle at the margins, says one of the exhibition's curators, David Spiegelhalter, who studies the public's understanding of risk at the University of Cambridge, UK. Emotions do the rest: a psychological aversion to flying often leads people to rate air travel as high-risk, even when they know that it is statistically much safer than driving.

The prime role of emotion in assessing and taking risks is shown by a strap-on biosensor device that you can wear while playing poker and which measures stress indicators such as sweat and heart rate. A poker face won't work here: physiological measurements are displayed for all to see in the form of a chair tottering off a diving

board on a TV screen. "We are trying to explore both sides of risk: the cool, rational side, and how we feel about it," says Spiegelhalter.

Daniel Kahneman's book *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011) was a major inspiration for *Risk Lab*, says gallery director Michael John Gorman. In this work, the Nobel-prizewinning economist demonstrates that powerful emotions make us overestimate the probabilities of events that have minuscule odds, such as terrorist attacks. "There's the basic question of why people are more frightened of sharks than diabetes," says Gorman. "We are trying to get to people's instinctive fears."

On the one-armed bandits, visitors can participate in five psychology experiments that pose ethical, practical and monetary conundrums. Particularly resonant in our era of potential pandemics is the exhibition's *Risky Rights and Wrongs*, a classic psychology experiment set up by Mary Parkinson and Ruth Byrne of Trinity College Dublin, which probes the moral implications of public-health choices. Visitors observe 'John', a protagonist confronted with two different options for combating an influenza-like disease — one that will definitely save some lives, and a second that has a chance of saving more lives but runs the risk of saving no lives and for which probabilities are provided. The experiment has two versions, which differ in how the options are worded, but give the same probabilities for saving or losing lives.

When participants are asked to try to save lives, they tend to be risk-averse and choose the definite option; when the options are rephrased and they are asked to prevent lives being lost, they choose the probabilities. This experiment investigates how people assess the morality of the risky choices of others. Subtle changes in word choice that alter the way a situation is framed can catch us out.

Another slot-machine study, *Price Your Vice*, explores whether people are more impulsive with their money or their vices, such as drinking and smoking, and ignore the future to live in the present. Visitors are offered various options to accept or delay rewards, but unbeknown to them, their mouse movements are tracked to gauge their certainty as they choose. "Distorted decision-making kills more people than anything else in the world," says Denis O'Hora, a psychologist involved in this study at the National University of Ireland, Galway. Heart disease, respiratory disease, liver disease, strokes and cancer all have behavioural components.

Risk Lab takes nothing from you but time, and freely provides a stack of everyday statistics and psychology lessons. You can't lose. ∎

Anthony King *is a writer based in Dublin.*
e-mail: anthonyjking@gmail.com

FREDDIE STEVENS

Correspondence

We need negative metrics too

Research metrics are ambiguous — a paper may be cited for positive or negative reasons. Funding agencies and universities focus on positive impact in evaluating research, which increasingly includes alternative metrics ('altmetrics'; see *Nature* **493**, 159 and *Nature* **495**, 437–440; 2013). We think that researchers can generate a more complete account of their impact by including seemingly negative indicators — such as confrontations with important people or legal action — as well as those that seem positive.

To explore this idea, we at the Center for the Study of Interdisciplinarity discussed ways to evaluate the impact of our research activities. We began with common quantitative indicators of scholarly impact (number and place of publications, citation indexing measures, number and size of grant awards, and so on). Warming to our theme, we came up with several other possible indicators, including negative ones (see table for examples and go.nature.com/miytf3 for a complete list).

In this age of increasing demand for accountability, we believe that academics ought to own the impact of their research, rather than have it determined by someone else.

J. Britt Holbrook, Kelli R. Barr, **Keith Wayne Brown** *University of North Texas, Denton, Texas, USA.* *britt.holbrook@unt.edu*

Campaign tactics and grants don't mix

Using political-campaign tactics to secure grant awards threatens to oversimplify the science, overwhelm the independent peer-review process, and disregard intellectual-property and confidentiality issues.

Take the European Commission's Future and Emerging Technologies 'flagship' programme, which in January selected two projects to receive about €1 billion (US$1.3 billion) each over 10 years (see *Nature* **493**, 585–586; 2013). Six projects were shortlisted after a year-long competition on the basis of scientific review — but also on the success of presentations to the European Parliament and political representatives, promotional videos and television interviews.

Substantial media exposure of the US Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative is inducing speculation on research outcomes well ahead of hard experimental evidence.

Securing broad political consensus for large-scale projects is understandable. Yet applicants for some low-budget grants are now also using social media, promoting proposals by requesting support letters through mass e-mailings and Twitter communications. However tempting, this could outweigh reasoned peer review.

Victor Maojo, Juan Pazos *Polytechnic University of Madrid, Spain.* *vmaojo@fi.upm.es* **Casimir A. Kulikowski** *Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA.*

Australian academy is fair to women

It is not true that the Australian Academy of Science, of which I am president, treats women with disdain (*Nature* **497**, 7; 2013). The absence of women elected as fellows in 2013 is of great concern, but it is an anomaly.

Election of women to our academy has steadily increased, from just one in the 1970s to 27 since 2000. Three of our executive councillors are female, and women chair five out of 22 national committees. This year, half of our early-career awardees and one of four full-career awardees are women (see go.nature.com/avrqiy).

That said, our academy still has too few women. Part of the problem is historical and universal (see, for example, *Nature* **495**, 21; 2013). Past policies either discriminated against women or failed to nurture their careers. The academy has urged Australian science agencies to prioritize career flexibility and has endorsed guidelines for gender equity (see go.nature.com/zi253a).

We are striving to increase the proportion of female fellows by ensuring that high-achieving women are not overlooked for nomination and that our criteria do not disadvantage them. Further measures to improve the gender balance will be considered at the academy's annual general meeting this month.

Suzanne Cory *Australian Academy of Science, Canberra.* *cory@wehi.edu.au*

Economic modelling could aid brain map

Debate over US President Barack Obama's brain-mapping proposal (*Nature* **495**, 19; 2013) would benefit from economic modelling. This would refine thinking on goals, funding and timing amid budget cuts and discussions of government-sponsored big science.

An example of this approach comes from modelling the time and money required for genomic research to cut adverse drug-related patient outcomes, using actual data (R. Arnaout *et al.* *Clin. Chem.* **59**, 649–657; 2013). This reveals how understanding drug-response genomics could lead to cheaper, faster progress, delivering specific, fact-based, actionable insights.

The brain-mapping proposal is broader, perhaps calling for technologies not yet invented, but the intention is still to improve health. Economic modelling could aid comparisons between the proposal and competing investments, engage stakeholders and foster accountability. It would serve the ultimate funding source and beneficiary: the taxpayer.

Ramy Arnaout* *Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.* *rarnaout@bidmc.harvard.edu* **On behalf of 4 co-authors. See go.nature.com/brfb2u for full list.*

OTHER POSSIBLE INDICATORS OF IMPACT

Public engagement	Academic community	Media
Protests, demonstrations or arrests	Invitations to present, consult or review	Article downloads
Provoking lawsuits	Interdisciplinary achievements	Website hits
Angry letters from important people	Adviser appointments	Media mentions
Meetings with important people	Reputation of close collaborators	Quotes in media
Participation in public education	Reputation as a team member	Coining of a phrase
Mention by policy-makers	Textbooks authored	Trending in social media
Public research discussions	Citation in testimonials and surveys	Blog mentions
Muckraking	Audience size at talks and meetings	Book sales
Quotes in policy documents	Developing a useful metric	Buzzword invention
Rabble rousing	Curriculum input	Social-network contacts
Engagement with citizens abroad	Faculty recommendations, prizes	Television and radio interviews

François Jacob
(1920–2013)

French freedom fighter who helped to uncover how genes are regulated.

François Jacob helped to answer a question that had been troubling geneticists for more than 30 years: how do the various tissues of multicellular organisms express different sets of genes when all cells contain the same genetic material? Along with two colleagues, Jacob devised the first conceptualization of gene regulation, opening the door to the study of development in animals and humans.

Jacob was born in 1920 in Nancy, northeastern France, to a middle-class Jewish family. He studied at the Lycée Carnot in Paris and went to medical school with the intention of becoming a surgeon. His studies were interrupted in June 1940 when, refusing to accept France's surrender to Germany, he escaped to London by boat to join the Free French Forces under General Charles de Gaulle. Jacob fought for four years as a medical auxiliary in Africa and France. When the war ended, he was named a Companion of the Liberation, the highest French military distinction of the Second World War. More than six decades later, he became Chancellor of the Order of the Liberation, the order's most important figure.

After the war, the injuries that Jacob had sustained to his arms and legs in 1944 prevented him from becoming a surgeon and he worked on producing antibiotics for the French army. He heard about a scientific revolution taking place as a result of physics coming together with genetics and microbiology. The war and its aftermath had led to a redistribution of scientists throughout Europe and the United States, which, combined with the availability of technologies such as X-ray diffraction, electrophoresis and electron microscopy, was spurring a wealth of new ideas and approaches in biology. Jacob was keen to contribute to these exciting new developments, and the microbiologist André Lwoff — one of the few French researchers involved in this new approach — invited Jacob to join his laboratory at the Pasteur Institute in Paris in 1950.

Fifteen years later, Jacob was awarded the Nobel Prize in Physiology or Medicine along with Lwoff and Jacques Monod for describing the 'operon model'. In this model, a repressor protein, encoded by a regulatory gene, binds to a cell's DNA upstream of a battery of structural genes that the trio named the operon. This binding prevents the structural genes from being transcribed into messenger RNA. The model emerged from elegant experiments investigating the factors that repress and trigger the

multiplication of viruses in bacteria, and those that activate and suppress the synthesis of the bacterial β-galactosidase enzyme needed to break down lactose into its glucose and galactose components.

During his early years at the Pasteur Institute, working with microbiologist Elie Wollman, Jacob also proposed a mechanism for bacterial conjugation, or mating. In this process, part of a bacterial chromosome is transferred from a donor bacterium to a recipient one. In all of these major research projects, Jacob's particular skill was being able to conceptualize the molecular mechanisms underlying complex observations.

In the late 1960s, Jacob switched to working with mice, to align his research with the Pasteur Institute's focus on human disease. It took time to build the genetic-engineering tools needed to analyse complex organisms and their embryonic development, but Jacob had made the right decision. Owing to their genetic similarities with humans, mice are now the models of choice for scientists studying human diseases. Jacob's choice of a type of germ-cell cancer to aid research into embryonic development in mice proved similarly prescient: the cells studied in his laboratory are the ancestors of the stem cells that are now grown in hundreds of labs across the world.

Jacob enjoyed interacting with people who were different from him, but not those who tried to imitate him, with whom he could be very harsh. When I joined his lab in 1981 as a postdoc, he gave me total freedom to pursue the questions that I deemed interesting, and this was the same for everyone who worked with him. If ever we hit on some potentially important result, he would jump up in excitement and ask to see the data. For him, only the science that was moving forward was important.

His immense scientific knowledge, combined with his wartime experiences, fuelled a lifelong commitment to combating racism and preventing the misuse of genetics. He presented his views on television, radio and in newspapers. He was also a member of the French national ethics committee for the life sciences.

Jacob was fascinated with the history of science and philosophy. In his 1973 book *The Logic of Life* (Pantheon), he explored the idea that biological knowledge has evolved in successive steps, from the study of morphology to that of the cell, and then to the study of genes and macromolecules. In several other works, such as his 1982 *The Possible and the Actual* (Pantheon), he sought to identify the features that characterize scientific knowledge as opposed to other types of knowledge. His death marks the end of a golden age of biology, in which members of a relatively small international community were free to pursue whatever question they wanted, with the possibility that they would make huge strides in discovery. ∎

Michel Morange *is professor of biology and director of the Cavaillès Centre (CIRPHLES, USR 3308) for the History and Philosophy of Science at the Ecole Normale Supérieure, Paris, France.*
e-mail: morange@biologie.ens.fr

AGEING

Beneficial miscommunication

Natural variations in the rate of protein translation in cellular organelles called mitochondria have been found to correlate with lifespan, suggesting a unified mechanism for the effects of metabolic alterations on longevity. SEE ARTICLE P.451

SUZANNE WOLFF & ANDREW DILLIN

Our existence depends on a small interloper that resides within our cells — the persistent and phantom-like presence of a once autonomous organism. More than 2 billion years ago, as one bacterium tried, but failed, to consume another, two cells forged a relationship that eventually resulted in the evolution of one into a subcellular organelle — the mitochondrion — of the other. As time passed, this organelle became a small metabolic factory for its host cell, allowing the host to produce enough energy to differentiate and to evolve into the intricate networks of cells and tissues that form the basis of a complex organism[1]. What happens to the organism when this endosymbiotic relationship is disrupted, and the surprising effects of this disruption on lifespan, are the focus of a study by Houtkooper et al.[2] on page 451 of this issue.

Across the ages, the mitochondrion has stubbornly tried to retain its identity. It has held on to its own DNA and replicates independently of the rest of the cell. And it defies the rules of Mendelian inheritance. Hundreds or even thousands of these organelles now exist within each cell, and live in a constant physical flux driven by fusion and fission, whereby separate mitochondria join to form one larger mitochondrion or individual ones suddenly split apart[3].

During this time, however, the mitochondrion has lost much of its autonomy; both its basal composition and the cellular distribution of its DNA have changed[4]. Today, most of the proteins that comprise a mitochondrion are encoded by the cell nucleus, and mitochondrial DNA encodes only 13 proteins — less than 1% of its total protein composition[5,6].

To build a mitochondrion, the nucleus must know which mitochondrial genes are needed, and when. It must also recognize what type of mitochondrion to build, because specific tissues — and perhaps even different subcellular locations — contain mitochondria of markedly different protein composition[7]. The nucleus must be ready to respond to fluctuations in the environment and to initiate mitochondrial biogenesis when metabolic
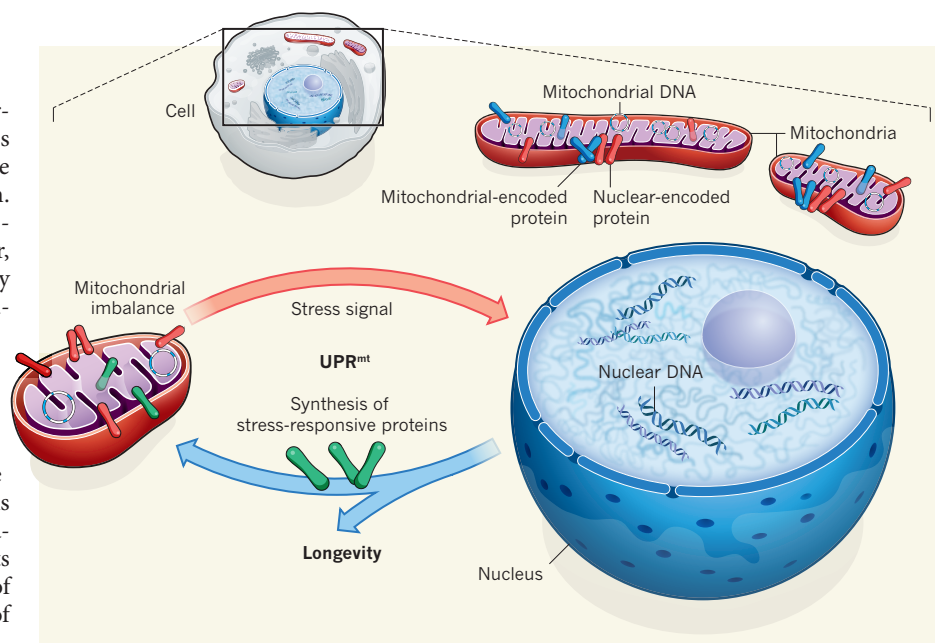
Figure 1 | Consequences of a cellular imbalance. Mitochondria carry a full complement of nuclear-encoded and mitochondrial-encoded proteins. An imbalance between mitochondrial and nuclear proteins triggers the mitochondrial unfolded protein response (UPR[mt]), whereby mitochondria send a signal to the nucleus to induce the production of stress-related proteins, which restores the mitochondrial balance. Houtkooper et al.[2] find that increased UPR[mt] is correlated with longer lifespan in mice and nematode worms.

conditions so necessitate. Finally, the cell must be poised to translate these genes into proteins in its cytoplasm, and must have sufficient chaperone proteins to help to fold and translocate the nascent proteins to the mitochondria. Synthesis and maintenance of mitochondria is thus a dazzlingly elaborate process — one that necessarily requires complex communication between the mitochondria and the nucleus to ensure synthesis of the proper ratios of proteins required for mitochondrial construction and function.

It seems impossible that a cell could keep track of all these individual fluctuations. Perhaps not surprisingly, therefore, cells have evolved intricate mechanisms specifically for detecting and responding to stress that affects their mitochondria[8–10]. An imbalance between the production of proteins encoded by the nucleus and those encoded by the mitochondria will quickly initiate defence mechanisms to restore homeostasis. During such events, mitochondria release signals

that travel to the nucleus to alter the proliferation of mitochondria by affecting the expression of nuclear-encoded mitochondrial genes. This signal also increases the translation of a network of stress-related proteins designed to protect mitochondria from further damage (Fig. 1).

It is the upregulation of one such defence mechanism, the mitochondrial unfolded protein response (UPR[mt]), that is the focus of Houtkooper and colleagues' study[2]. The authors discovered that partial loss-of-function of the mitochondrial translational machinery correlated with as much as a 2.5-fold increase in lifespan among dozens of inbred lines of mice originating from a single ancestral mating. Specifically, variation (polymorphism) in a gene encoding a single mitochondrial ribosomal protein (MRP) involved in protein translation, Mrps5, correlated with an increase in lifespan in these lines. A decrease in mitochondrial translation was also sufficient to extend lifespan and to activate

the UPR^mt in the nematode *Caenorhabditis elegans* in a dose-dependent manner.

The authors hypothesized that a deficiency in the function of MRPs might cause an imbalance in the relative levels of mitochondrial- and nuclear-encoded components of the electron transport chain, the mitochondrion's energy factory. This imbalance may secondarily activate the UPR^mt. Importantly, this effect seemed reciprocal: addition of rapamycin or resveratrol (pharmacological agents associated with attenuated cytoplasmic, rather than mitochondrial, translation, but which alter the metabolic state of the cell through the regulation of mitochondrial biogenesis) was sufficient to both upregulate the UPR^mt and extend lifespan.

This work is extremely suggestive, but it is only a start. Mitochondrial dysfunction has proved far from beneficial in most known contexts: in humans, mutations in mitochondrial genes cause a large number of extremely debilitating and life-shortening diseases[11]. And, until now, mutations in mitochondrial genes have not been associated with increased health or longevity in mammals. Therefore, the association of a natural variation in the function of MRPs with increased lifespan seems extraordinary.

The regulation of mitochondrial function and the synthesis of its proteins are necessarily complicated, however. It will be important to examine how a loss of MRPs affects the overall molar ratio of the different components of the electron transport chain. Whether other changes that affect mitochondrial proliferation affect lifespan with a dependency on the UPR^mt should also be tested. Nevertheless, the current paper illustrates the extent to which the balance of communication between the nucleus and mitochondria remains absolutely necessary for a cell to maintain its homeostasis.

After 2 billion years of partnership, then, communication between mitochondria and the nucleus may remain a core determinant of an organism's lifespan. By definition, endosymbiosis involves a balance between the needs of distinctly functioning subparts to provide a greater benefit to the whole. Our cells may be so sensitized to a loss of this equilibrium that a rapid and effective defence becomes necessary. The ageing-research community must continue to search for an understanding of the specific effects of the UPR^mt on the factors that cause ageing, and how such a response is disseminated and communicated across extremely complex organisms. We should also further our understanding of methods by which UPR^mt induction might alleviate age-onset diseases. ■

Suzanne Wolff *and* **Andrew Dillin** *are in the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720-3370, USA, and at the Howard Hughes Medical Institute.*
e-mail: dillin@berkeley.edu

1. Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. *Biol. Chem.* **382,** 1521–1539 (2001).
2. Houtkooper, R. H. *et al. Nature* **497,** 451–457 (2013).
3. Duchen, M. R. *Mol. Aspects Med.* **25,** 365–451 (2004).
4. Gray, M. W. *Cold Spring Harb. Perspect. Biol.* **4,** a011403 (2012).
5. Anderson, S. *et al. Nature* **290,** 457–465 (1981).
6. Pagliarini, D. J. *et al. Cell* **134,** 112–123 (2008).
7. Johnson, D. T. *et al. Am. J. Physiol. Cell Physiol.* **292,** C689–C697 (2007).
8. Liu, Z. & Butow, R. A. *Mol. Cell. Biol.* **19,** 6720–6728 (1999).
9. Parikh, V. S., Morgan, M. M., Scott, R., Clements, L. S. & Butow, R. A. *Science* **235,** 576–580 (1987).
10. Zhao, Q. *et al. EMBO J.* **21,** 4411–4419 (2002).
11. Wallace, D. C. *Science* **283,** 1482–1488 (1999).

DENGUE VIRUS

Two hosts, two structures

Dengue virus has a highly ordered structure when grown in mosquito cells at 28 °C. The finding that the virus expands into a less ordered form at 37 °C indicates that the human immune system does not see it as we previously thought.

FELIX A. REY

Dengue disease is caused by four viruses of the flavivirus genus that are transmitted to humans by infected mosquitoes[1]. Dengue infections impose a formidable burden: about 5.5% of the world's population is infected each year, and one infection in four is symptomatic[2]. But in spite of its impact, no effective antiviral treatment[3], nor a licensed vaccine[4], is available. Writing in *Journal of Virology* and in *Proceedings of the National Academies of Sciences*, respectively, Fibriansah *et al.*[5] and Zhang *et al.*[6] report that dengue virus particles display a different organization of surface glycoproteins when they are at temperatures above 34 °C, as in a human body, than they do at lower temperatures, such as those found in mosquitoes. These results have important implications for understanding how the virus particles are presented to the human immune system, and how to use this knowledge to develop an effective vaccine.

The difficulties in developing an anti-dengue vaccine stem from the interplay of the four related viruses (called viral serotypes) that cause dengue infections. Infection with any one of these viruses induces lifelong immunity against that serotype. Although some of the antibodies elicited during this response are cross-reactive and can neutralize the other serotypes to a degree, they do not mediate long-lasting cross-protection. Moreover, animal experiments show that the cross-reactive antibodies can actually enhance a subsequent infection with another serotype[7], which is thought to contribute to the severe forms of dengue disease seen in humans[8].

In this context, it is clear that only a vaccine that protects against all serotypes simultaneously would be successful. Although one promising candidate — which comprised four vaccines, each targeting one serotype — was shown by a large vaccine trial[7] to be safe and to confer some protection against serotypes 1, 3 and 4, it did not protect against serotype 2 despite eliciting neutralizing antibodies against all four serotypes[9]. These results highlight the importance of understanding the actual mechanisms of virus neutralization by antibodies and the correlation with protection from disease.

The main antigen targeted by neutralizing antibodies against dengue viruses is a glycoprotein called protein E, which exists as protein dimers at the virus surface. This protein is the main player during viral entry to a cell: it is responsible for receptor binding and for inducing fusion of viral and cellular membranes to release the viral RNA into the cytoplasm. Protein E contains a fusion loop that inserts into the membrane of cellular organelles called endosomes; this loop is concealed at the E-dimer interface in the mature virus particle. Receptor binding at the cell surface leads to uptake into the endosome, where the acidic environment triggers E-dimer dissociation and exposure of the fusion loop, which is accompanied by a major structural rearrangement. Antibodies against protein E can therefore block infection by interfering with receptor binding or with this conformational change.

However, the picture is complicated by another viral glycoprotein, prM, which associates with protein E during viral synthesis. PrM is cleaved during viral maturation, but a substantial amount of the protein is still found in dengue virus particles circulating in an infected host, and it elicits antibodies that are non-neutralizing and contribute to antibody-mediated enhancement of the infection[10]. This observation highlights the complexity of devising a vaccine to generate
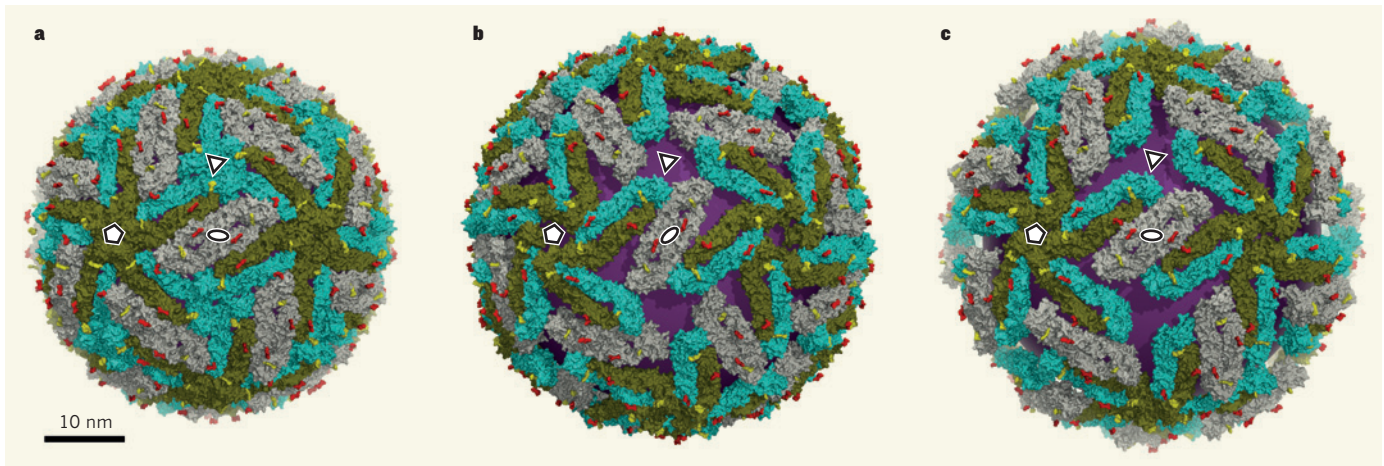
Figure 1 | Compact and expanded structures of dengue virus. a, At 28 °C, mature dengue virus particles have icosahedral symmetry and are composed of 90 dimers of the surface glycoprotein, protein E. The icosahedral axes of the particles are represented with symbols: two-fold, ellipse; three-fold, triangle; and five-fold, pentagon. Thirty dimers lie at the icosahedral two-fold axes (grey) and 60 at general positions that are not on a symmetry axis of the particle (displayed as one E protein in green and the other in cyan). Red and yellow indicate the two glycan chains that are attached to each E protein. **b, c,** Zhang *et al.*[6] (**b**) and Fibriansah *et al.*[5] (**c**) show that the particles take on an expanded form at temperatures of 34 °C and above, and no longer obey strict icosahedral symmetry. A prominent feature is a large opening at the three-fold axes, such that the viral membrane (purple) becomes exposed. The differences in dimer–dimer contacts seen in **b** and **c** may depend on the selection of particles used for three-dimensional reconstruction. The represented E dimer positions are therefore only consensus locations. (Figure prepared with the PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.)

long-lasting protective antibodies while avoiding a post-vaccination state with the potential for immune enhancement. It also emphasizes the need to understand at the structural level the relevant sites on the virus — the antigenic epitopes — that are targeted by antibodies that confer protection.

Previous electron microscopy studies[11,12] were performed with dengue virus produced in mosquito cells cultured at 28 °C, and revealed that 90 E dimers are tightly packed at the particle surface to form an icosahedral shell (Fig. 1a). These studies were recently extended[13] to near-atomic resolution (3.5 ångströms), and this work indicated that the E dimers form a rigid icosahedral surface lattice through lateral interactions. The two new reports[5,6] show that the virus particles expand when they are exposed to temperatures of 34 °C and above, and that this reveals patches of the viral membrane (Fig. 1b, c). This finding will have implications for understanding which epitopes are exposed to the human immune system. Membrane exposure is also consistent with the observation that the human proteins TIM1, 3 and 4, which bind to phosphatidylserine lipids, can also bind to dengue virus particles[14]; these data were difficult to reconcile with a structure in which tightly packed E dimers covered the membrane.

In one of the current papers, the authors report[6] that the temperature-induced structural change takes place in a narrow range between 33 °C and 34 °C, and that the change is irreversible. Both groups found that the resulting particle preparations were heterogeneous, and only a stringent particle-selection process led to an icosahedral three-dimensional reconstruction that allowed docking of the crystal structure of the E dimer[15,16]. This heterogeneity probably explains why the

packing of E dimers at the surface of the expanded particles now reported is not the same (Fig. 1b, c). Taken together, these studies show that, at the human body temperature of 37 °C, the mature dengue virus particle loses symmetry and the E dimers no longer make the specific lateral interactions seen in the 28 °C structure.

The reported particle expansion was not seen with West Nile virus, another infectious flavivirus. The structure of West Nile virus particles grown in mammalian cells at 37 °C showed[17] the same arrangement of tightly packed E dimers that is seen in dengue virus particles grown at 28 °C. In contrast to dengue virus protein E, which is in a monomer–dimer equilibrium in solution at neutral pH, the E protein of West Nile virus is monomeric[18,19] and requires the tight lateral packing at the particle surface to form a dimer concealing the fusion loop. It is therefore possible that dengue virus particles can expand while still maintaining the E dimers, whereas in West Nile virus a similar expansion would probably lead to dimer dissociation with concomitant exposure of the fusion loop, resulting in particle inactivation.

The take-home message is that the dengue virus E dimer seems to be the relevant form presented to the human immune system. De Alwis *et al.*[20] recently noted that most anti-dengue neutralizing antibodies in humans do not bind to purified E protein produced *in vitro*. Their interpretation was that these antibodies probably target epitopes shared by adjacent dimers, as in the arrangement of the 28 °C particles. However, our observations that dengue virus E protein is essentially monomeric in solution unless at high concentrations indicate that the antibodies did not bind in their experiments[20] because the E protein was monomeric under the conditions used. Together, their results and the new structural data therefore suggest that

most human neutralizing antibodies target quaternary epitopes that are confined to the E dimer, rather than ones that bridge dimers. The good news is that future attempts to identify relevant epitopes for efficient immunization in humans can be restricted to the E dimer instead of the whole particle, which will make their characterization more tractable. ∎

Felix A. Rey *is in the Unité de Virologie Structurale, Institut Pasteur and CNRS UMR 3569, 75724 Paris Cedex 15, France. e-mail: rey@pasteur.fr*

1. Morens, D. M., Folkers, G. K. & Fauci, A. S. *EcoHealth* http://dx.doi.org/10.1007/s10393-013-0825-7 (2013).
2. Bhatt, S. *et al. Nature* **496,** 504–507 (2013).
3. Simmons, C. P. *et al. PLoS Negl. Trop. Dis.* **6,** e1752 (2012).
4. World Health Organization. Dengue and severe dengue. Fact Sheet No. 117; www.who.int/mediacentre/factsheets/fs117/en (2012).
5. Fibriansah, G. *et al. J. Virol.* http://dx.doi.org/10.1128/JVI.00757-13 (2013).
6. Zhang, X. *et al. Proc. Natl Acad. Sci. USA* **110,** 6795–6799 (2013).
7. Balsitis, S. J. *et al. PLoS Pathog.* **6,** e1000790 (2010).
8. Halstead, S. B. *Adv. Virus Res.* **60,** 421–467 (2003).
9. Sabchareon, A. *et al. Lancet* **380,** 1559–1567 (2012).
10. Dejnirattisai, W. *et al. Science* **328,** 745–748 (2010).
11. Kuhn, R. J. *et al. Cell* **108,** 717–725 (2002).
12. Zhang, W. *et al. Nature Struct. Biol.* **10,** 907–912 (2003).
13. Zhang, X. *et al. Nature Struct. Mol. Biol.* **20,** 105–110 (2013).
14. Meertens, L. *et al. Cell Host Microbe* **12,** 544–557 (2012).
15. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. *Proc. Natl Acad. Sci. USA* **100,** 6986–6991 (2003).
16. Zhang, Y. *et al. Structure* **12,** 1607–1618 (2004).
17. Mukhopadhyay, S., Kim, B.-S., Chipman, P. R., Rossmann, M. G. & Kuhn, R. J. *Science* **302,** 248 (2003).
18. Kanai, R. *et al. J. Virol.* **80,** 11000–11008 (2006).
19. Nybakken, G. E., Nelson, C. A., Chen, B. R., Diamond, M. S. & Fremont, D. H. *J. Virol.* **80,** 11467–11474 (2006).
20. de Alwis, R. *et al. Proc. Natl Acad. Sci. USA* **109,** 7439–7444 (2012).

STRUCTURAL BIOLOGY

Tiny enzyme uses context to succeed

How the enzyme diacylglycerol kinase can form membrane anchors and an active site from so few amino–acid residues has long been a mystery. Crystal structures reveal that it gets by with a little help from its friends. SEE LETTER P.521

JIMIN ZHENG & ZONGCHAO JIA

The bacterial enzyme diacylglycerol kinase has always been a bit of a rebel. As an apparent evolutionary 'orphan', this integral membrane protein is functionally and structurally distinct from other kinases identified so far. Furthermore, its size (just 121 amino-acid residues) makes its modus operandi challenging to understand: how does so small an enzyme manage to span the cell membrane while still forming a functional active site that must accommodate not only a bulky, fatty substrate but also a water-soluble ATP molecule? On page 521 of this issue, Li et al.[1] report crystal structures of diacylglycerol kinase from the bacterium *Escherichia coli*. The structures reveal how the enzyme exploits the milieu in which it resides, greatly increasing our understanding of this protein and providing insight into integral membrane enzymes in general*.

Diacylglycerol kinase (DgkA) catalyses a key step in the synthesis of oligosaccharide biomolecules[2], converting diacylglycerol to phosphatidic acid by transferring a phosphate group from ATP. Most ATP-dependent kinases have the same general structure: an evolutionarily conserved, two-lobed core that has all the elements required for catalysis, including catalytic residues that are absolutely conserved; characteristic ATP-binding motifs such as P-loops and Walker motifs; and a hydrophobic spine connecting the lobes[3]. DgkA, however, lacks these features. As one of the first integral membrane enzymes to be studied, it has served as an intriguing model not only of catalysis in such enzymes, but also of the structure, stability, assembly and folding of membrane proteins[4].

DgkA is the smallest kinase known, weighing in at about half

the size of a typical kinase. It therefore has to use most of its residues to build helices that integrate it into the cell membrane, while still crafting an active site that accommodates both ATP and diacylglycerol — which contains a hydrophilic 'head' group and bulky hydrophobic 'tails'. Rather than building a complicated and energetically costly single protein molecule in the membrane, evolution has forged a trimer from three DgkA molecules. On the basis of their crystal structures of the monomers, Li and colleagues propose that each monomer 'borrows' a component from its neighbour to create a composite active site. The borrowed component is an amphiphilic (both hydrophobic and hydrophilic) amino-terminal helix. This helix is strategically located at the interface

Figure 1 | A view of diacylglycerol kinase (DgkA). Li and colleagues' crystal structure[1] of DgkA, an integral membrane enzyme from the bacterium *Escherichia coli*, is depicted here in the microbe's inner membrane; the periplasm is the space between the inner and outer membranes of the microbe. The enzyme forms a homotrimer, but for clarity only one full monomer (grey) is shown. The authors propose that the active site of DgkA, which binds ATP (yellow) and substrate (blue and red), includes both the membrane and an amino-terminal helix from a second monomer (green). Selected catalytic amino-acid residues (orange stick models) and a magnesium ion (purple sphere) are also shown. The ATP-binding pocket is in the cytoplasmic portion of the protein and is formed by all three transmembrane helices of one monomer and the N-terminal helix from the neighbouring monomer. The substrate-binding pocket is largely formed by the membrane.

between the membrane and the cytoplasm, and provides a crucial catalytic residue.

Confirmation of the putative active site will require further experimental support from an ATP-bound structure of DgkA, or, even better, from a structure of the enzyme in complex with both ATP and a substrate, to verify the locations of the ATP- and substrate-binding pockets and other fine structural features. Nevertheless, the proposed active site is fully consistent with earlier work[2] in which residues that are essential for catalysis were identified. The suggested active site also makes intuitive sense, because water-soluble ATP should bind to the cytoplasmic portion of DgkA with its transferable phosphate group (the γ-phosphate) oriented towards the membrane–cytoplasm interface, in the optimal position for attachment to the substrate (Fig. 1).

Although appropriate ATP binding is necessary for a functional active site, the substrate also needs to be stabilized and correctly oriented to allow efficient phosphate transfer from ATP. The complex visualized in the crystal structure described by Li and co-workers contains detergent molecules, which the authors used to aid crystallization of the protein and which happen to resemble the substrate. Serendipitously, two of these molecules are found in the structure at a location that primes them for phosphate transfer: the tails are embedded in and stabilized by the membrane, whereas the head groups reach into cytoplasmic space, coming close to the proposed location of ATP's γ-phosphate.

So, the membrane is not only the milieu in which DgkA resides — it also forms an integral part of the active site, dictating the orientation of the substrate to allow efficient phosphate transfer. It would be difficult for DgkA to stabilize and correctly orient the bulky, amphiphilic substrate without the membrane because, as a small kinase, it does not have the resources to create an adequate binding pocket. This suggests that DgkA has evolved together with the cell membrane[5] to create an optimal, composite active site.

Interestingly, Li and colleagues' crystal structure is at odds with a previous 'backbone-only' structure[6] of DgkA that was obtained using nuclear magnetic resonance (NMR). Although the overall homotrimeric architecture of DgkA is consistent between the two structures, the active sites differ hugely — an unprecedented observation for NMR and crystallographic structures. The dissimilarity is caused by domain swapping: in the NMR
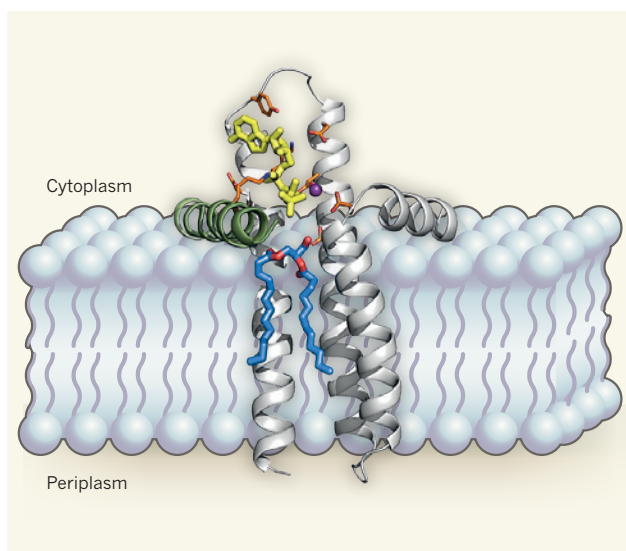
'Masters and Pupils' — The three books under review are approximately equal in price, but this is not true of their value. Two of them are hardy annuals; the third — the *Transactions of the International Ophthalmic Optical Congress, 1961* — appears ten years after the previous issue. Those who argue that opticians neglect research will find their fears both confirmed and confounded — confirmed when they read some of the uncritical work … on contact lenses, and confounded by the high-quality work emanating from British and especially American sources. The trouble with a volume such as this is two-fold: every paper presented at the congress is reproduced, and, what is worse, reproduced in full. Verbosity is a smoke-screen for self-consciousness, and by George! what a difference a whiff of editorial air would have made to this tome. This is not to convey the notion that it does not contain any paper of merit. Le Grand, reproducing data by Bonnet on the cornea, is at his usual instructive best, Ball on dark-adaptation provocative … By and large, all three volumes are well produced, adequately illustrated, and set in pleasing type. But they would be lighter, not to say cheaper, if editors kept pointing a pencil at the contributors, and — conscious of the rudeness of the gesture — kept asking them "Is this word really necessary?".
From *Nature* 25 May 1963

The Palaeolithic skull from Piltdown, Fletching, Sussex, just described in the Quart. Journ. Geol. Soc. By Dr. A. Smith Woodward as the type of a new genus and species (*Eoanthropus dawsoni*), has been placed on exhibition in a special case in the central hall of the Natural History Museum.
From *Nature* 22 May 1913

---

structure, one of the three transmembrane helices from a DgkA molecule is switched with that of another molecule.

The different structures might reflect the different conditions used in the two analyses, or the challenges involved in obtaining the NMR structure. However, given the extreme thermal stability of the membrane-embedded DgkA homotrimer (it can survive at temperatures as high as 100 °C)[4], it is difficult to imagine a seamless switch between 'swapped' and 'unswapped' homotrimers. In the light of the high-resolution crystal structure described by Li and co-workers, the NMR results might need to be revisited.

The kinases are one of the largest and most extensively studied protein families, but DgkA shows that there are still surprises to be found. Its remarkable structure and composite active site add an intriguing twist to the kinome — the set of known kinases in an organism. The enzyme's ingenious design results in an active site that is much simpler than those of other kinases. Indeed, DgkA provides a new perspective for understanding the architecture and associated catalytic mechanisms of membrane enzymes. As well as structural determination of ATP-bound and substrate-bound DgkA complexes, future work should include detailed investigation of the enzyme's catalytic mechanism and regulation, and resolution of the discrepancy between the NMR and crystal structures. ∎

Jimin Zheng *is in the College of Chemistry, Beijing Normal University, Beijing 100875, China.* **Zongchao Jia** *is in the Department of Biomedical and Molecular Science, Queen's University, Kingston, Ontario K7L 3N6, Canada.*
e-mail: jia@queensu.ca

1. Li, D. *et al.* Nature **497,** 521–524 (2013).
2. Van Horn, W. D. & Sanders, C. R. *Annu. Rev. Biophys.* **41,** 81–101 (2012).
3. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. *Phil. Trans. R. Soc. B* **367,** 2517–2528 (2012).
4. Pieringer, R. A. & Kunnes, R. S. *J. Biol. Chem.* **240,** 2833–2838 (1965).
5. Badola, P. & Sanders, C. R. *J. Biol. Chem.* **272,** 24176–24182 (1997).
6. Van Horn, W. D. *et al.* Science **324,** 1726–1729 (2009).

PHYSICAL CHEMISTRY

Handedness detected by microwaves

Traditional methods for detecting and identifying the handedness of molecules — their chirality — have been based on the same theoretical concept. A technique has been reported that departs from this paradigm. SEE LETTER P.475

LAURENCE A. NAFIE

The property of handedness known as chirality is one of the most subtle and yet profound aspects of our world. Of particular relevance to chemistry and biology is the fact that molecular structures can be chiral: as with our left and right hands, chiral molecules have otherwise equivalent mirror-image isomers of the opposite chirality[1]. Such isomers are called enantiomers, and the ability to distinguish between them is called enantiomeric detection. On page 475 of this issue, Patterson *et al.*[2] describe a method that adds to the chemist's toolbox of techniques for detecting and identifying chirality in molecules[3]. They report that, when microwave radiation associated with transitions between rotational states is emitted by molecules of opposite chirality, the radiation is exactly out of phase, thus providing a clear signal of molecular chirality.

The building blocks of life are 'homochiral': naturally occurring amino acids, proteins, sugars and nucleic acids exist as only one chiral form. If these building blocks had mixed chirality, molecular chaos would ensue and life would not be possible. Nature's lead is being followed by scientists, who have learned that single-enantiomer drugs are more efficient at binding to biological targets, and have fewer side effects, than the equivalent racemic drugs (which contain an equal mixture of opposite enantiomers). In fact, our bodies recognize drugs of the opposite chirality as different molecules, even though, apart from their chirality, they have the same structure. This is because the mirror symmetry of enantiomers is broken by the homochiral biochemistry of the human body, in the same way that only a right hand fits comfortably into a right-handed glove.

The mirror symmetry of enantiomers can also be broken by circularly polarized electromagnetic radiation. Circular polarization occurs when the electromagnetic field of radiation rotates either clockwise or anticlockwise, once per wavelength, as the beam propagates.
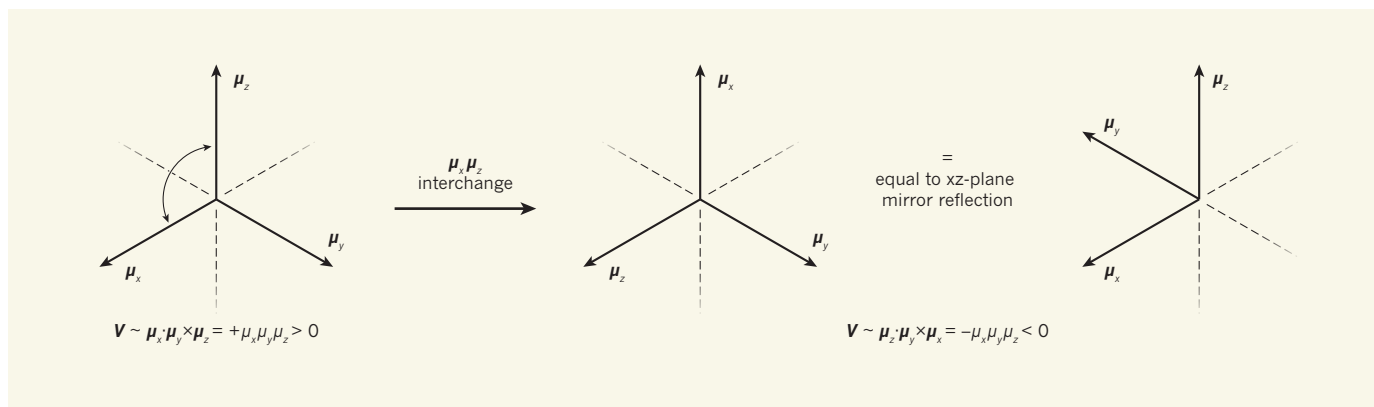
$$V \sim \boldsymbol{\mu}_x \cdot \boldsymbol{\mu}_y \times \boldsymbol{\mu}_z = +\mu_x\mu_y\mu_z > 0$$

$\boldsymbol{\mu}_x, \boldsymbol{\mu}_z$ interchange

= equal to xz-plane mirror reflection

$$V \sim \boldsymbol{\mu}_z \cdot \boldsymbol{\mu}_y \times \boldsymbol{\mu}_x = -\mu_x\mu_y\mu_z < 0$$

Figure 1 | A new mechanism for chiroptical spectroscopy. In Patterson and colleagues' method[2] for identifying molecular chirality, the measured quantity depends on the handedness of three mutually orthogonal electric-dipole rotational transition moments, $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_z$, which are associated with the three rotational degrees of freedom of a molecule. When microwave radiation interacts with these moments, energy transfer changes the rotational state of the molecule, generating a spectroscopic signal. The moments are vectors, and can be represented by a scalar triple product of $\boldsymbol{\mu}_x \cdot \boldsymbol{\mu}_y \times \boldsymbol{\mu}_z$. This product can be regarded as a dipole volume, V, that is equal to the product of the magnitudes of the three vectors. The sign of V depends on the order of the vectors. If any two vectors are interchanged (a process equivalent to a mirror reflection), the sign changes. Because V changes sign under spatial inversion (mirror reflection) and is even under time-reversal symmetry, it is a measure of true chirality[8].

Two classical forms of 'chiroptical' spectroscopy have been developed on the basis of this symmetry breaking. The older form is optical rotation[4], which measures the difference in refractive index for left-circularly polarized (LCP) and right-circularly polarized (RCP) radiation as it passes through chiral materials. The other is circular dichroism (CD), which measures the difference in the absorption of LCP and RCP radiation. Both techniques typically use ultraviolet or visible light and involve electronic transitions in molecules. Vibrational circular dichroism (VCD, which uses infrared radiation) and Raman optical activity (ROA, which mainly uses visible radiation) have also been added to the list of chiroptical spectroscopy techniques[5–7] and are based on the vibrational transitions of molecules.

Using these four techniques, two fundamental measures of chirality can be determined: enantiomeric excess, which is the excess of one enantiomer over the other in a sample; and absolute configuration, the specific handedness of an enantiomer. Every chiral drug substance approved for sale by the US Food and Drug Administration must have a known absolute configuration and a specified level of enantiomeric excess usually greater than 99% (see go.nature.com/ujg17l).

Patterson and colleagues' use of microwave radiation to detect chirality is surprising, because microwave CD would be expected to be too small for detection by modern spectrometers. All previously measured forms of chiroptical intensity are inversely proportional to the wavelength of the probing radiation. Electronic CD in the ultraviolet–visible region is typically about 100 times larger than VCD in the longer-wavelength infrared region. Microwave CD should be at least 100 times smaller than VCD. This size argument derives from the mechanism underpinning traditional chiroptical spectroscopy: optical activity arises from the interference between the electric-dipole transition moment and the weak magnetic-dipole transition moment (or, in some cases, the weak electric-quadrupole transition moment as well) that is detected when a chiral molecule is irradiated with alternating LCP and RCP light. The optical activity can be positive or negative depending on whether the electric- and magnetic-dipole transition moments point into the same or opposite halves of a sphere centred on the molecule.

The authors' method does not arise by this interference mechanism. Instead, the authors detect chirality by applying two orthogonally polarized, microwave-timescale electric fields (oriented in the x and z directions) to a sample. This causes the molecules in the sample to emit microwave radiation polarized along a third orthogonal direction (y), the intensity of which is proportional to the product of each molecule's three orthogonal rotational electric-dipole moments ($\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_z$). This product is independent of the molecule's orientation, but is sensitive to the handedness of the directions of the three rotational dipole moments. It therefore changes sign if any two of the moments are interchanged or, equivalently, if the molecule is exchanged for its opposite enantiomer, and so is a new measure of true chirality[8] (Fig. 1).

Patterson and colleagues' approach has several advantages over existing chiroptical spectroscopy techniques. Because it does not depend on a weak magnetic-dipole transition moment, the chiral signal is nearly as large as that of the applied microwaves. Furthermore, the method requires extremely cold, gaseous molecules, which exhibit sharp, narrow lines in their microwave spectra. Molecules of interest can therefore be resolved in the presence of other interfering molecules. Moreover, measurement times can be as fast as tens of seconds; measurements for traditional chiroptical spectroscopy techniques take minutes to hours.

Of course, a few drawbacks remain to be ironed out. Any molecule investigated must be sufficiently volatile to be sampled in the gas phase, potentially placing an upper limit on the size of molecules that can be analysed. And for large molecules that have high conformational freedom, it may be difficult to identify the conformation associated with the microwave line being analysed. Additionally, enantiomeric excess measured using the technique is accurate to only about 5%, and so further development is needed to reach a more desirable level of accuracy.

Finally, the determination of absolute configuration for a sample of unknown chirality has yet to be demonstrated. The scalar triple product of $\boldsymbol{\mu}_x \cdot \boldsymbol{\mu}_y \times \boldsymbol{\mu}_z$ will be positive or negative depending on the enantiomer in the sample. But how can one say which enantiomer produces a positive product and which a negative one? This information might be obtained by determining whether the chiral microwave emission of a particular enantiomer has the same phase as the driving field, or the opposite one. However, this information would still need to be connected to the absolute configuration of a chiral molecule, most probably by carrying out a quantum-mechanical calculation to determine the signs of the dipole-moment components. Most other chiroptical spectroscopy techniques require such a calculation to connect measured spectra to the molecule's absolute configuration.

If Patterson and co-workers' method can be widely applied to determine the enantiomeric excess and absolute configuration of previously unassigned chiral molecules, particularly those of pharmaceutical interest,

then their paper will be extremely important. But even if its applicability is at first limited (either for sampling reasons or by instrumentation), the unexpected demonstration of a conceptually new form of chiroptical spectroscopy makes this work a landmark in the 200-year-old history of optical activity in chemistry. ∎

Laurence A. Nafie *is in the Department of Chemistry, Syracuse University, Syracuse, New York 13244, USA.*
e-mail: lnafie@syr.edu

1. Barron, L. D. *Nature* **446,** 505–506 (2007).
2. Patterson, D., Schnell, M. & Doyle, J. M. *Nature* **497,** 475–477 (2013).
3. Wagnière, G. H. in *Comprehensive Chiroptical Spectroscopy* Vol. 1 (eds Berova, N., Polavarapu, P. L., Nakanishi, K. & Woody, R. W.) 3–34 (Wiley, 2012).
4. Arago, D. F. *Mem. de L'Inst.* **12**(1), 93 (1811).
5. Barron, L. D. *Molecular Light Scattering and Optical Activity* 2nd edn (Cambridge Univ. Press, 2004).
6. Nafie, L. A. *Vibrational Optical Activity: Principles and Applications* (Wiley, 2011).
7. Haesler, J., Schindelholz, I., Riguet, E., Bochet, C. G. & Hug, W. *Nature* **446,** 526–529 (2007).
8. Barron, L. D. *Chem. Phys. Lett.* **123,** 423–427 (1986).

MATERIALS SCIENCE

When two is better than one

Aerogels have many potential applications but usually suffer from poor elasticity. The synergistic assembly of carbon nanotubes and graphene has now allowed multifunctional, ultra–lightweight and super–elastic aerogels to be made.

WENCAI REN & HUI-MING CHENG

Solids are normally stronger but heavier than gases because their atoms are bound together more tightly. Writing in *Advanced Materials*, however, Sun *et al.*[1] describe how they have used two carbon-based nanomaterials — graphene and carbon nanotubes — to make a strong carbon aerogel that is lighter than air, is super-elastic and has a variety of useful properties.

Aerogels are highly porous solid foams that have an interconnected network of thin, solid walls. They have an extremely low density and a high specific surface area — the total surface area per unit mass. But they tend to suffer from poor strength and low elasticity: pressing them firmly can cause a catastrophic breakdown of the network. The first aerogel to be reported[2] was prepared from a silica gel by supercritical drying, in which the liquid component of the gel is dried off in a controlled manner. This silica aerogel is transparent and has the lowest thermal conductivity of any known solid, making it suitable as a thermal insulating material.

Several aerogels have since been made, including oxide and carbon aerogels. However, the synthesis of ultralight aerogels with high elasticity has remained a challenge. Aerogels with these qualities would find many uses, for example as media for energy absorption[3,4] or as reusable absorbents for liquids[1], and as pressure-responsive sensors[1].

The prospects for making ultralight, highly elastic aerogels have changed with the development of carbon nanomaterials, however, especially carbon nanotubes (CNTs) and graphene. Graphene is a single layer of carbon atoms arranged in a perfect honeycomb structure[5]. A CNT is a tubular material seamlessly rolled from a graphene sheet[6]. Both materials have a host of useful properties: low density, high strength and stiffness, high electrical and thermal conductivity, high specific surface area, excellent flexibility, and good chemical stability[5,6]. They therefore hold great promise as building blocks for aerogels with multiple functionalities.

Both graphene and CNTs have been used separately to make aerogels that have low density and high elasticity[3,4,7]. For instance, a graphene aerogel with a density of 5.1 milligrams per cubic centimetre has been shown to sustain its structural integrity under a load of more than 50,000 times its own weight and can rapidly recover from 80% compression[3]. However, the low elastic bending stiffness of graphene sheets limits improvement in the elasticity of graphene aerogels when their density is decreased[1,3]. Also, in CNT aerogels, the bundling or permanent buckling of the CNTs usually results in inefficient load transfer between the CNTs, and significant irreversible deformation of the aerogels[1,8].

Sun *et al.*[1] used giant graphene oxide (GO) sheets, tens of micrometres across, to build an ultralight structural framework, and used CNTs as 'ribs' to reinforce the framework, taking advantage of their good elasticity (Fig. 1). The authors fabricated their hybrid graphene–CNT aerogels by freeze-drying an aqueous solution of GO sheets and CNTs, and then chemically reducing the GO to graphene. This procedure resulted in a graphene–CNT aerogel that is lighter than air (it has a density of 1.0 mg cm^{-3}, compared with 1.2 mg cm^{-3} for air at ambient conditions) and that could recover its original macroscopic shape and microstructure after being repeatedly compressed by 50% in 1,000 cycles of compression
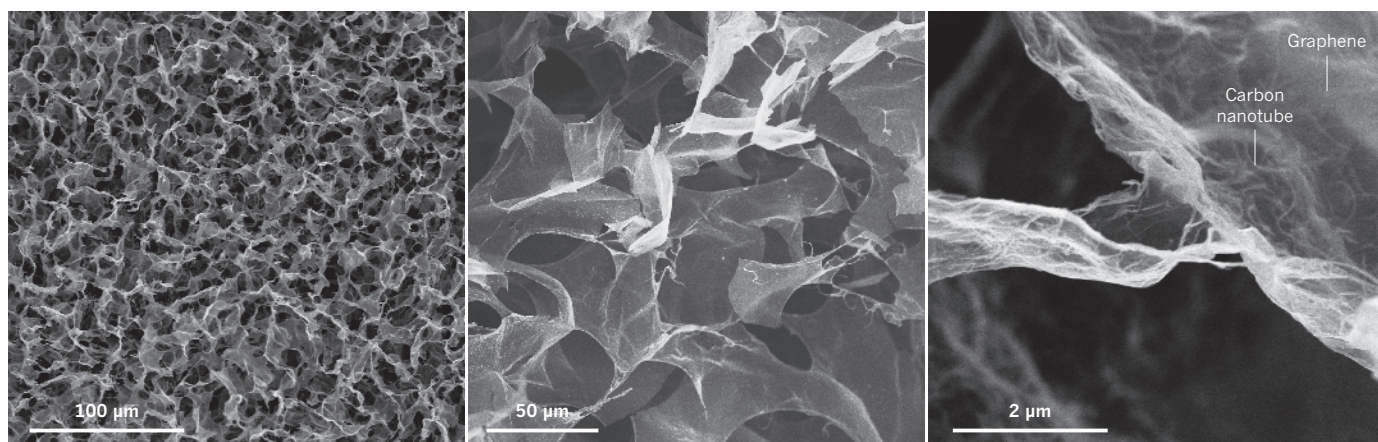

Figure 1 | Synergistically assembled carbon aerogels. Sun *et al.*[1] report the assembly of multifunctional, ultra-lightweight and super-elastic carbon aerogels from reduced graphene oxide sheets and carbon nanotubes. The microscopic structure of the aerogels, which consists of an interconnected network of randomly oriented graphene walls and carbon nanotube 'ribs', is seen here at three different scales.

and release. It could also sustain a compressive stress about ten times higher than that of an ultralight metallic microlattice with a similar density[9], and can be elongated by 16.5% before fracturing. In addition, using only reduced GO sheets, Sun *et al.* prepared the lightest known aerogel: with a density of 0.16 mg cm^{-3}, the material is more than 7 times lighter than air.

As well as producing aerogels that are ultralight and super-elastic, the synergistic effect between graphene and CNTs also yields aerogels that have excellent thermal stability, high absorption capacity for liquids, and good electrical conductivity[1]. For example, the resulting aerogels have an ultra-high oil-absorption capacity of 215–913 times their own weight, depending on the oil density, which is up to two orders of magnitude higher than those of commercial oil absorbents. What's more, the aerogel's super-elasticity means that it can be used repeatedly following mechanical extrusion to release the absorbed oil. Taken together with several other studies[3,4,7,8], Sun and colleagues' results expand the range of applications for carbon aerogels.

> *"Sun and colleagues' work demonstrates how the synergistic assembly of two types of carbon nanomaterial can be used to create multifunctional materials."*

However, two major challenges remain: preparing aerogels that have a more ordered network of thin, solid walls than that achieved here; and fully translating the excellent intrinsic properties of graphene and CNTs, such as their electrical and thermal conductivity, into the aerogels. The latter aspect is plagued by the electrical and thermal resistance between CNTs and/or reduced GO sheets, and by the poor electrical and thermal conductivity of reduced GO. The technique of template-directed chemical-vapour deposition may provide a solution, because graphene foams prepared by this method[10], with nickel or other metal foams serving as a template, have an electrical conductivity two to three orders of magnitude higher than that of a graphene–CNT aerogel with a similar density.

The hybrid aerogels may also prove beneficial in composite materials, in which the pores of solid foams are filled with another material to improve their properties. Graphene-foam-based composites have already found several diverse uses, such as elastic conductors[10], flexible lithium-ion batteries[11], supercapacitors[12] and materials for lightweight electromagnetic interference shielding[13]. Basing composite structures on the hybrid aerogels may further widen the range of applications.

Sun and colleagues' work demonstrates how the synergistic assembly of two types of carbon nanomaterial can be used to create multifunctional materials. Besides graphene and CNTs, a vast number of nanomaterials have been developed since the 1980s, some of which have complementary properties to carbon-based nanomaterials. Extending the authors' synergistic approach to other nanomaterials may produce systems with various novel properties. Indeed, new electronic systems have been created by combining graphene with boron nitride or molybdenum disulphide sheets in a particular stacking order[14]. Finally, it is the nanostructured nature of these building blocks that gives them mechanical flexibility and other useful properties, and this makes them an appealing choice for building flexible devices. ∎

Wencai Ren and **Hui-Ming Cheng** are at the Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China.
e-mail: cheng@imr.ac.cn

1. Sun, H., Xu, Z. & Gao, C. *Adv. Mater.* **25,** 2554–2560 (2013).
2. Kistler, S. S. *Nature* **127,** 741 (1931).
3. Qiu, L., Liu, J. Z., Chang, S. L. Y., Wu, Y. & Li, D. *Nature Commun.* **3,** 1241 (2012).
4. Hu, H., Zhao, Z., Wan, W., Gogotsi, Y. & Qiu, J. *Adv. Mater.* **25,** 2219–2223 (2013).
5. Geim, A. K. *Science* **324,** 1530–1534 (2009).
6. Saito, R., Dresselhaus, G. & Dresselhaus, M. S. *Physical Properties of Carbon Nanotubes* (Imperial College Press, 1998).
7. Nardecchia, S., Carriazo, D., Ferrer, M. L., Gutiérrez, M. C. & del Monte, F. *Chem. Soc. Rev.* **42,** 794–830 (2013).
8. Kim, K. H., Oh, Y. & Islam, M. F. *Nature Nanotechnol.* **7,** 562–566 (2012).
9. Schaedler, T. A. *et al. Science* **334,** 962–965 (2011).
10. Chen, Z. *et al. Nature Mater.* **10,** 424–428 (2011).
11. Li, N., Chen, Z., Ren, W., Li, F. & Cheng, H.-M. *Proc. Natl Acad. Sci. USA* **109,** 17360–17365 (2012).
12. Cao, X. *et al. Small* **7,** 3163–3168 (2011).
13. Chen, Z., Xu, C., Ma, C., Ren, W. & Cheng, H.-M. *Adv. Mater.* **25,** 1296–1300 (2013).
14. Britnell, L. *et al. Science* **335,** 947–950 (2012).

DNA REPLICATION

Driving past four–stranded snags

Unusual DNA structures, such as G–quadruplexes, can stall DNA replication with drastic consequences for the cell. The Pif1 helicase family of enzymes has evolved to disentangle these structures efficiently. SEE ARTICLE P.458

SERGEI M. MIRKIN

Genomic DNA is long, shows profound sequence redundancy and is heavily tangled. These properties create formidable problems during DNA replication — the process at the heart of heredity. A particular obstacle is the presence of repetitive runs of DNA that can form unusual structures such as hairpins, cruciforms, three-stranded triplexes and four-stranded G-quadruplexes. In this issue, Paeschke *et al.*[1] (page 458) describe an evolutionarily conserved mechanism that promotes replication past G-quadruplexes*.

Unusual DNA structures are less favourable than the common DNA double helix, and form only if the duplex is stimulated to unwind, exposing single strands. The replication process provides plenty of stimuli for this. For example, during replication, replicative helicase enzymes efficiently separate the two strands to form a two-pronged replication fork (Fig. 1a). In addition, DNA polymerase enzymes — which use these DNA strands as templates for replication — function only in

*This article and the paper under discussion[1] were published online on 8 May 2013.

a particular direction along the DNA strand (5′ to 3′), whereas the two strands have antiparallel orientation. Consequently, synthesis along the leading-strand template is continuous, whereas synthesis along the lagging strand is discontinuous, proceeding in short increments known as Okazaki fragments. This leads to the presence of long, single-stranded regions at the fork, called Okazaki initiation zones, at which the next fragment will be synthesized.

Okazaki initiation zones provide convenient places for formation of unusual DNA structures, including G-quadruplexes. Earlier bacterial, yeast and mammalian studies implied that formation of such structures along the lagging-strand template slows down replication and increases the likelihood of chromosomal breakage and genomic rearrangements[2].

Formation of G-quadruplexes was also predicted[3] along the leading-strand template, in regions of single-stranded DNA that transiently arise between the replicative helicases and leading-strand DNA polymerase. These predictions were corroborated recently[4] when accumulation of G-quadruplexes was demonstrated during the S-phase of the cell

cycle — the phase at which replication occurs.

Because formation of unusual DNA structures is a recurrent problem, cells have evolved defensive mechanisms to deal with them. The first line of defence is a protein that binds to single-stranded DNA; it is known as replication protein A (RPA) in eukaryotic organisms such as animals, plants and fungi, and is an essential component of the replication fork. RPA covers the Okazaki initiation zone and prevents DNA base pairing, thus efficiently stopping it from folding into various secondary structures.

But every guardian has its Achilles heel, and for RPA that is its low affinity for DNA that is made of repetitive purine (adenine and guanine) bases[5]. Therefore, G-quadruplexes, which are mainly made of guanines, can overcome this line of defence, and fold into stable secondary structures. When this happens, DNA polymerase suddenly faces a four-stranded obstacle in the template, which it cannot disentangle under normal conditions[6] (Fig. 1b). This stalls the entire replication fork, occasionally leading to fork reversal and chromosomal fragility[7].

Enter the second line of defence — accessory DNA helicases. Paeschke and colleagues' study is devoted to one family of these enzymes, called Pif1. These helicases, which untangle DNA moving in the 5′ to 3′ direction, were initially identified in the budding yeast *Saccharomyces cerevisiae*, but have since been found in organisms as diverse as bacteria and humans[8].

S. cerevisiae has two Pif1-family helicases: Pif1 and Rrm3. Pif1 prevents replication-fork stalling at G-quadruplexes[9], whereas Rrm3 helps to resolve 'collisions' between transcription and replication processes along DNA[10]. Paeschke *et al.* demonstrate that Pif1-family helicases from various species unwind G-quadruplexes extremely quickly and efficiently. Many other DNA helicases were shown to unwind G-quadruplexes, but the Pif1 helicases seem to be the best for this task. Moreover, G-quadruplexes are by far these helicases' favourite substrates.

The authors also demonstrate that the absence of Pif1 helicases in yeast causes a drastic increase in gross chromosomal rearrangements, particularly those that originated at G-quadruplex-forming sequences. Most importantly, this high genetic instability was efficiently suppressed by expressing bacterial or human Pif1 helicases in the helicase-deficient yeast strain. Thus, unravelling
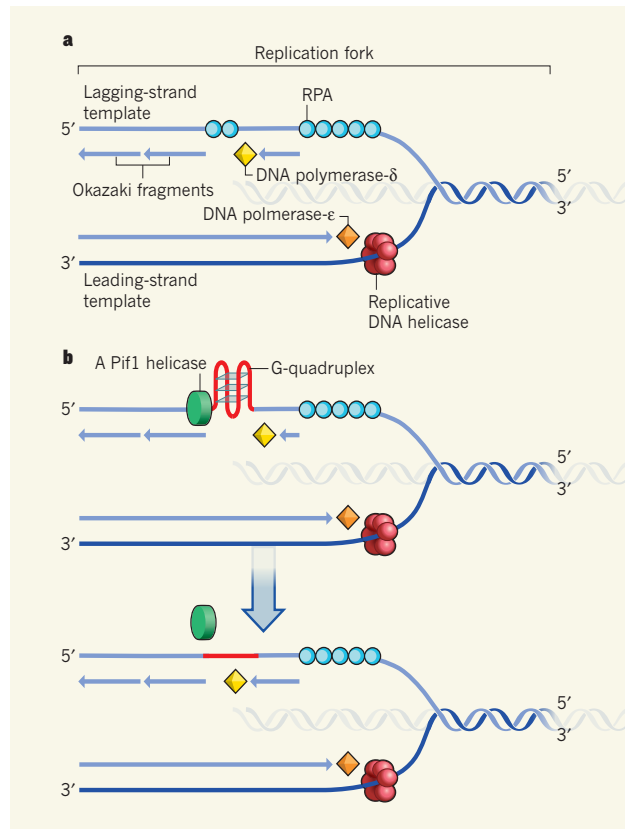
Figure 1 | The problem at G-quadruplexes. During DNA replication, replicative DNA helicases unwind the two strands into a leading- and a lagging-strand template to form a replication fork. Whereas DNA polymerase-ε promotes continuous replication along the leading-strand template, DNA polymerase-δ copies the lagging-strand template in short increments called Okazaki fragments, which are joined together later. **a**, Where the lagging strand is exposed — for instance, at Okazaki initiation zones — a single-strand DNA-binding protein, replication protein A (RPA), binds to it to ensure smooth progress of replication along the strand, and the fork as a whole. **b**, Where this strand contains DNA repeats rich in guanine bases, RPA is inefficient, allowing the strand to fold into secondary structures, including G-quadruplexes. Consequently, progress of DNA polymerase-δ is blocked, leading to stalling of the replication fork. Paeschke *et al.*[1] report that Pif1 helicases unwind G-quadruplexes, allowing replication to progress past this obstacle.

G-quadruplexes seems to be a conserved function of Pif1 helicases that arose very early in evolution.

Most unexpected are the data that came from Paeschke and co-workers' characterization of the rearrangement events triggered by G-quadruplexes in helicase-deficient yeast strains. As in their previous set of experiments, the authors introduced 'reporter' genes into a non-essential arm of the yeast chromosome to track these events. In normal strains, most events that manifested as reporter inactivation were caused by the loss of the chromosomal region carrying the reporters. By contrast, in helicase-deficient strains, the reporters' activity was not lost owing to the loss of the corresponding genes, but because the genes were epigenetically silenced — that is, their histone proteins were chemically modified, without alteration to the DNA sequence. This silencing is similar to a genetic phenomenon known as position-effect

variegation[11], which also depends on the activity of the histone-modifying deacetylase enzyme sirtuin-2. As the yeast clones continued to grow on selective media, they underwent genetic rearrangements leading to actual loss of the reporter genes.

Paeschke and colleagues' findings become curiouser and curiouser, with this provocative result prompting them to coin the term complex genetic–epigenetic events. It is difficult to explain this concept in known molecular terms. The authors speculate that the presence of a G-quadruplex motif could trigger local epigenetic silencing, which somehow can be extended to the rest of the chromosome when Pif1 helicases are absent, impairing replication. Theoretically, one can think of such potential mechanisms as altered histone deposition and/or accumulation of epigenetic marks on them owing to impaired replication. Many other scenarios should also be considered.

From this study, the Pif1 family of DNA helicases emerges as a principal player in rescuing genomes from the negative effects of G-quadruplexes — which include replication-fork impairment, epigenetic silencing of surrounding genes and gross chromosomal rearrangements. G-quadruplexes are extremely common elements of the human genome. They are required for the genome's normal activities (such as telomere maintenance and immunoglobulin-switch recombination), but are also implicated in its malfunction, resulting in genetic rearrangements and human disease, including cancer[12]. This will, no doubt, stimulate studies of Pif1 helicases in health and disease for years to come. ∎

Sergei M. Mirkin *is in the Department of Biology, Tufts University, Medford, Massachusetts 02155, USA.*
e-mail: sergei.mirkin@tufts.edu

1. Paeschke, K. *et al. Nature* **497,** 458–462 (2013).
2. Mirkin, E. V. & Mirkin, S. M. *Microbiol. Mol. Biol. Rev.* **71,** 13–35 (2007).
3. Lopes, J. *et al. EMBO J.* **30,** 4033–4046 (2011).
4. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. *Nature Chem.* **5,** 182–186 (2013).
5. Wold, M. S. *Annu. Rev. Biochem.* **66,** 61–92 (1997).
6. Woodford, K. J., Howell, R. M. & Usdin, K. *J. Biol. Chem.* **269,** 27029–27035 (1994).
7. Anand, R. P. *et al. Nucleic Acids Res.* **40,** 1091–1105 (2012).
8. Bochman, M. L., Sabouri, N. & Zakian, V. A. *DNA Repair* **9,** 237–249 (2010).
9. Ribeyre, C. *et al. PLoS Genet.* **5,** e1000475 (2009).
10. Ivessa, A. S. *et al. Mol. Cell* **12,** 1525–1536 (2003).
11. Perrod, S. & Gasser, S. M. *Cell. Mol. Life Sci.* **60,** 2303–2318 (2003).
12. Maizels, N. & Gray, L. T. *PLoS Genet.* **9,** e1003468 (2013).

ARTICLE

Mitonuclear protein imbalance as a conserved longevity mechanism

Riekelt H. Houtkooper[1,2]*, Laurent Mouchiroud[1]*, Dongryeol Ryu[1], Norman Moullan[1], Elena Katsyuba[1], Graham Knott[3], Robert W. Williams[4] & Johan Auwerx[1]*

Longevity is regulated by a network of closely linked metabolic systems. We used a combination of mouse population genetics and RNA interference in *Caenorhabditis elegans* to identify mitochondrial ribosomal protein S5 (*Mrps5*) and other mitochondrial ribosomal proteins as metabolic and longevity regulators. MRP knockdown triggers mitonuclear protein imbalance, reducing mitochondrial respiration and activating the mitochondrial unfolded protein response. Specific antibiotics targeting mitochondrial translation and ethidium bromide (which impairs mitochondrial DNA transcription) pharmacologically mimic *mrp* knockdown and extend worm lifespan by inducing mitonuclear protein imbalance, a stoichiometric imbalance between nuclear and mitochondrially encoded proteins. This mechanism was also conserved in mammalian cells. In addition, resveratrol and rapamycin, longevity compounds acting on different molecular targets, similarly induced mitonuclear protein imbalance, the mitochondrial unfolded protein response and lifespan extension in *C. elegans*. Collectively these data demonstrate that *MRPs* represent an evolutionarily conserved protein family that ties the mitochondrial ribosome and mitonuclear protein imbalance to the mitochondrial unfolded protein response, an overarching longevity pathway across many species.

Longevity is coordinated by intersecting pathways, often converging on metabolic networks[1–4]. A key player in lifespan regulation is the mitochondrion. Over a thousand proteins encoded by nuclear DNA (nDNA) translocate to and function in mitochondria[5], in synchrony with 13 proteins encoded by the mitochondrial DNA (mtDNA) that require a separate translation machinery, including mitochondrial ribosomal proteins (MRPs)[6,7]. Many molecular studies of longevity have exploited simple organisms and loss- or gain-of-function mutations, but the complex connectedness of mitochondrial and metabolic longevity networks benefits from an integrative cross-species approach[2].

Here we pioneered such a strategy and used the BXD reference population of mice[2,8–10] to identify mitochondrial ribosomal protein S5 (*Mrps5*) and other members of the *MRP* family as longevity genes. In *C. elegans*, we confirmed this role of MRPs and demonstrated that they induce a stoichiometric imbalance between nDNA- and mtDNA-encoded oxidative phosphorylation proteins, hereafter termed 'mitonuclear protein imbalance', which activates the mitochondrial unfolded protein response (UPR^mt). Our conclusions were corroborated using specific antibiotics targeting bacterial/mitochondrial translation, and ethidium bromide, which inhibits mtDNA transcription. This mechanism is shared with pathways that induce mitonuclear protein imbalance from a nuclear perspective, such as the UPR^mt and lifespan enhancing effects of rapamycin and resveratrol. Our data hence tie mitochondrial translation and metabolism to natural lifespan regulation across species.

A QTL for mouse longevity

The BXD family consists of fully inbred progeny of a cross between C57BL/6J and DBA/2J mice, with a complexity that matches many human populations[11]. Both parental strains have been sequenced, enabling analysis of sequence variants linked to phenotypes[12]. We

used new genomic and genetic resources to re-analyse longevity data for BXD lines[13] using forward and reverse genetic methods[9].

The forward strategy exploits longevity data and updated high-density single nucleotide polymorphism (SNP) genotypes[14] archived in http://www.GeneNetwork.org. As reported[13], lifespan of BXDs varies from ~365 days for the shortest lived strain to ~900 days for the longest lived strain (Fig. 1a). We remapped longevity using the new genotypes and detected one genome-wide significant locus on chromosome 2 with a peak at 124–129 Mb (Fig. 1b, log odds ratio (lod) = 4.0). Two additional loci, on chromosomes 4 and 7, were not significant, but suggestive (lod = 2.8 and 3.0, respectively). However, neither was suggestive after controlling for SNP rs6374387 on chromosome 2 using composite interval mapping.

The chromosome 2 locus contains ~70 genes (Supplementary Table 1), none of which were previously linked to longevity. To evaluate and rank candidates, we correlated lifespan with multiple gene expression data sets. Only three genes in the locus correlate strongly with lifespan (Fig. 1c, $P < 0.01$; Supplementary Fig. 1): solute carrier family 12 member 1 (*Slc12a1*), mitochondrial ribosomal protein S5 (*Mrps5*) and tubulin tyrosine ligase (*Ttl*). From the natural variation in expression of these genes, we deduced that 50% reduction of expression corresponds to a ~250 day lifespan difference.

Conservation of longevity in *C. elegans*

We identified Y37A1C.1/*nkcc-1*, E02A10.1/*mrps-5*, and F25C8.5/*ttll-9* as worm homologues of *Slc12a1*, *Mrps5* and *Ttl*, respectively. RNA interference (RNAi)-mediated knockdown of *nkcc-1* and *mrps-5*, but not of *ttll-9*, extended lifespan (Fig. 2a).

Next, we compared expression of *Mrps5* and other *Mrp* family members in a muscle microarray of ageing and caloric restriction in C57BL/6J (ref. 15). *Mrp* expression decreased with age, an effect rescued by
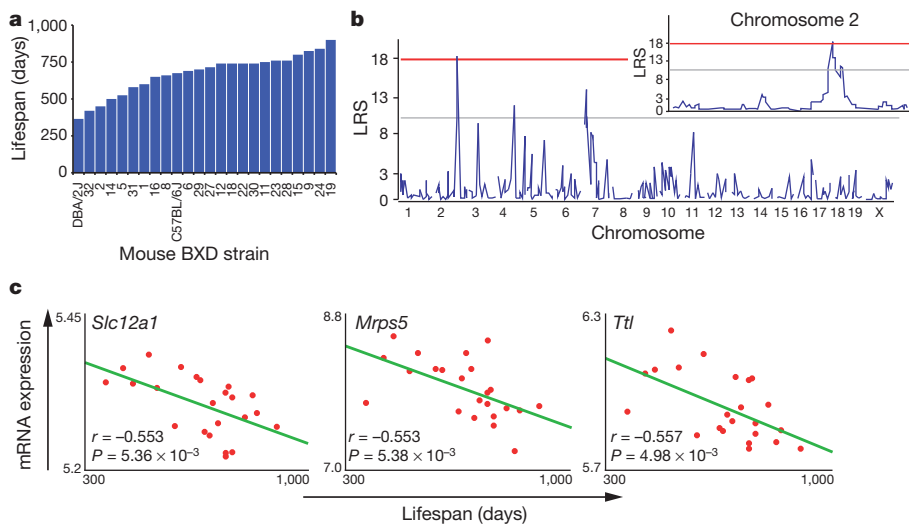
Figure 1 | Lifespan regulation in BXD recombinant inbred mice.
a, Lifespan in different BXD strains. **b**, Interval mapping using the BXD lifespan data reveals a strong QTL on chromosome 2, between 124–129 Mb. The red line depicts the cut-off for statistical significance (P genome-wide <0.05), the grey line represents the limit for suggestive QTLs. See also Supplementary Table 1. LRS, likelihood ratio statistics. **c**, Pearson's r correlation coefficient with corresponding P values for the co-variation between BXD lifespan (x axis) and mRNA expression of the indicated gene in the BXD eye microarrays (y axis). Decreased expression of *Slc12a1*, *Mrps5* and *Ttl* robustly correlates with longevity ($P < 0.01$). Correlation coefficient trend line is shown in green.

caloric restriction; in contrast, expression of *Slc12a1* and *Ttl* was un-affected (Fig. 2b). Linkage of MRPs with lifespan is strengthened as many other *Mrp* family members also correlate with longevity (Fig. 2c). We extended our analyses to the DNA level using sequence data for *Mrps5* in both parental strains and identified missense variants in exon 3 (rs29667217 and rs13471334; V60A and V67I, respectively). Other sequence variants in *Mrps5* contribute to variation in transcript abundance; *Mrps5* mRNA levels among the BXDs are associated with a strong quantitative trait loci (QTL) superimposed over the gene itself—a cis-expression QTL.

Using a reverse genetics approach, we studied the *Mrps5*-associated network. *Mrps5* expression co-varies with genes involved in oxidative phosphorylation. Considering that oxidative metabolism is involved in known longevity pathways[2], the set of transcripts that co-vary with *Mrps5* qualified as an appealing longevity network. Oxidative phos-phorylation was the most enriched network of *Mrps5* co-variates in both BXDs[16] and a conventional F2 intercross[17] ($P = 1.53 \times 10^{-21}$, $P = 5.78 \times 10^{-10}$, respectively). Finally, we generated an interaction net-work of oxidative phosphorylation genes with *Mrps5* (Fig. 2d), in which *Ndufb7* provides the hinge that links *Mrps5* to oxidative phosphorylation. Knockdown of the worm homologues for the network components *Ndufb7* and *Ndufa6* robustly extended lifespan[18–20]. *Mrps5* hence emerged as a strong longevity candidate, integrating protein synthesis and mitochondrial metabolism—both important longevity modulators.
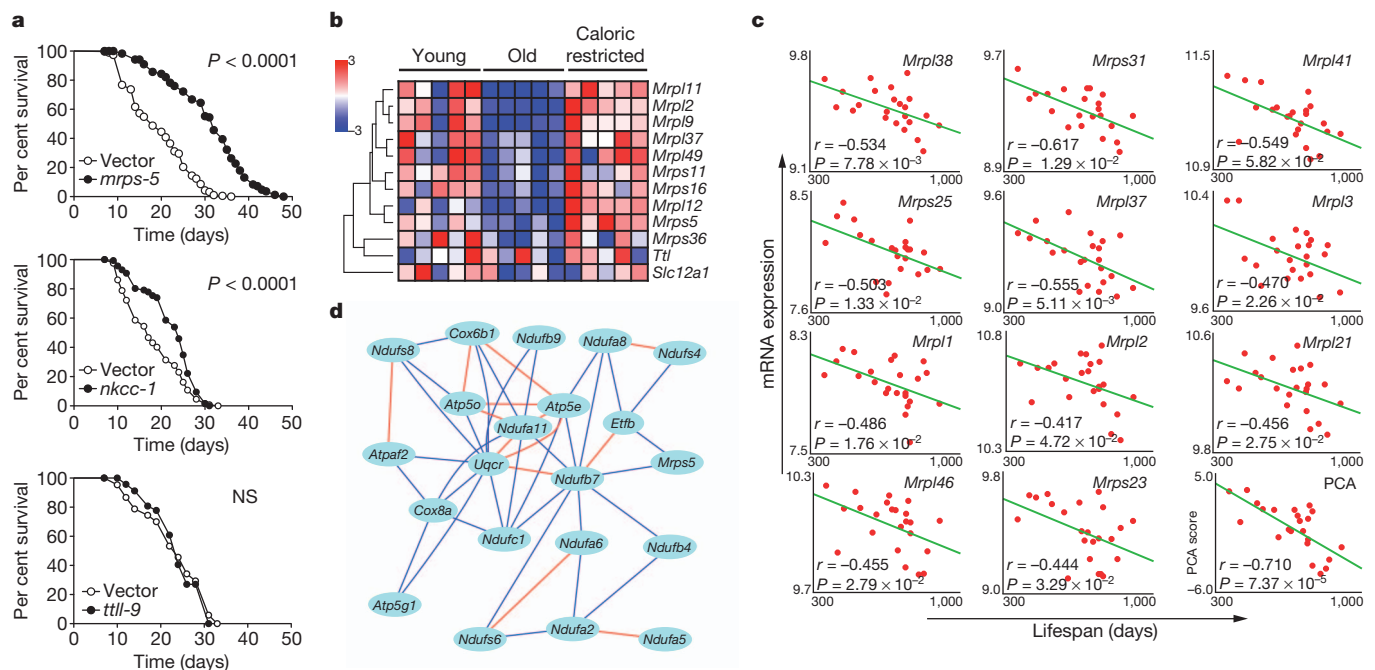
Figure 2 | Validation of *Mrps5* as a candidate longevity gene. a, Knockdown of *mrps-5*, *nkcc-1* or *ttll-9* throughout the entire life of *C. elegans* increased lifespan by 60%, 23% or 3%, respectively. See also Supplementary Table 2. NS, not significant. **b**, One-way hierarchical clustering showing gene expression differences in gastrocnemius muscle between young (5 months), old (25 months) and caloric restricted C57BL/6J mice[15]. Expression of mouse *Mrp* genes decreases upon ageing, and reverts with caloric restriction, whereas *Slc12a1* and *Ttl* do not change. The colour-coded heat map represents relative gene expression differences (red, large gene expression values; blue, small gene expression values). **c**, Pearson's r correlation coefficient with corresponding P values for co-variation between BXD lifespan (x axis) and mRNA expression of eleven other *Mrp* genes (y axis) indicates robust correlation. Principle component analysis (PCA) reveals a highly significant correlation between the *Mrp* gene family and BXD lifespan. **d**, *Mrps5* strongly correlates with genes involved in oxidative phosphorylation. Red lines indicate a positive Pearson correlation coefficient of 0.7–1.0, and blue lines indicate a correlation coefficient of 0.5–0.7.

Mitonuclear protein imbalance and ageing

To define causality of the *MRP*s in determining lifespan, we knocked down *mrp* genes during the entire life of the worm and robustly increased lifespan (Fig. 3a and Supplementary Table 2). Similar to well-characterized mitochondrial mutants that live longer, larval development was delayed (Supplementary Fig. 2a)[21]. Knockdown during development proved crucial and sufficient to extend lifespan, whereas RNAi during adulthood alone did not (Fig. 3b, Supplementary Fig. 2b and Supplementary Table 2), as reported in other long-lived mitochondrial mutants[22]. Increased lifespan was not due to effects on feeding, as pharyngeal pumping rates were normal (Supplementary Fig. 2c). *mrps-5* RNAi also delayed physiological decline with age. Even though they moved slightly less in early adulthood (day 3), *mrps-5* RNAi worms move twice as much as controls at day 13, and this effect becomes more pronounced at day 20 (Supplementary Fig. 2d, e and Supplementary Videos 1–4). This difference was accompanied by a delay in decline of pharyngeal pumping (Supplementary Fig. 2c) and in muscle fibre disorganization (Fig. 3c), hallmarks of fitness of aged *mrps-5* RNAi worms.

In line with the mitochondrial connection of *Mrps5*, basal respiration was reduced upon *mrp* knockdown and unresponsive to the mitochondrial uncoupler carbonyl cyanide *p*-trifluoromethoxyphenylhydrazone (FCCP) (Fig. 3d). As a consequence, *mrps-5* RNAi worms displayed reduced ATP levels and citrate synthase activity (Fig. 3e, f), indicative of reduced mitochondrial abundance or activity. Consistent with its role in mitochondrial translation, *mrps-5* RNAi induced a stoichiometric imbalance between nDNA- and mtDNA-encoded oxidative phosphorylation subunits, termed mitonuclear protein imbalance, visualized by selective reduction in MTCE.26

(MTCO-1 homologue; from mtDNA) relative to H28O16.1 (ATP5A homologue; from nDNA) expression (Fig. 3g). The mitonuclear protein imbalance and consequences for mitochondrial function was similar to the long-lived *cco-1* mutant—deficient for the nDNA-encoded worm homologue of complex IV, subunit Vb/COX4—but not observed in the short-lived complex II SDHC mutant *mev-1* (Fig. 3g). Furthermore, mitochondria had a more punctuate globular pattern instead of the regular reticular or tubular appearance in both muscle (Fig. 3h) and intestine, a finding confirmed by electron microscopy (Supplementary Fig. 3a, b).

To identify which 'longevity pathways'—insulin/IGF-1 signalling[23], caloric restriction[24] or mitochondrial dysfunction[22]—are required for the lifespan phenotype, we reduced *mrps-5* expression in worms carrying mutations in these pathways. *mrps-5* RNAi increases lifespan by ~40% in wild type (Fig. 3i), similar to the effect in *daf-2, daf-16, eat-2, sir-2.1* and *aak-2* mutants (Fig. 3j–l and Supplementary Fig. 4a–e) indicating that *mrps-5* regulates longevity independently of insulin/IGF-1 (*daf-16/daf-2*) and caloric restriction (*eat-2/sir-2.1*) and acts downstream of mitochondrial regulator *aak-2*.

We focused on the mitochondrial pathway, because (1) it robustly effects longevity[22]; (2) *MRP*s function in the translation of mtDNA-encoded oxidative phosphorylation subunits[6,7]; and (3) in the BXDs, *Mrps5* networked with several oxidative phosphorylation components (Fig. 2d). *mrps-5* RNAi reverts the short-lived phenotype of *mev-1* mutants, with a dramatic 112% lifespan extension (Fig. 3m and Supplementary Fig. 4f). *mrps-5* RNAi in the *cco-1* mutants did not extend lifespan compared to *mrps-5* RNAi alone, indicating that *cco-1* and *mrps-5* act in a similar fashion (Fig. 3n and Supplementary Fig. 4g). The same is true for *mrps-5* RNAi in the mitochondrial *clk-1(e2519)*
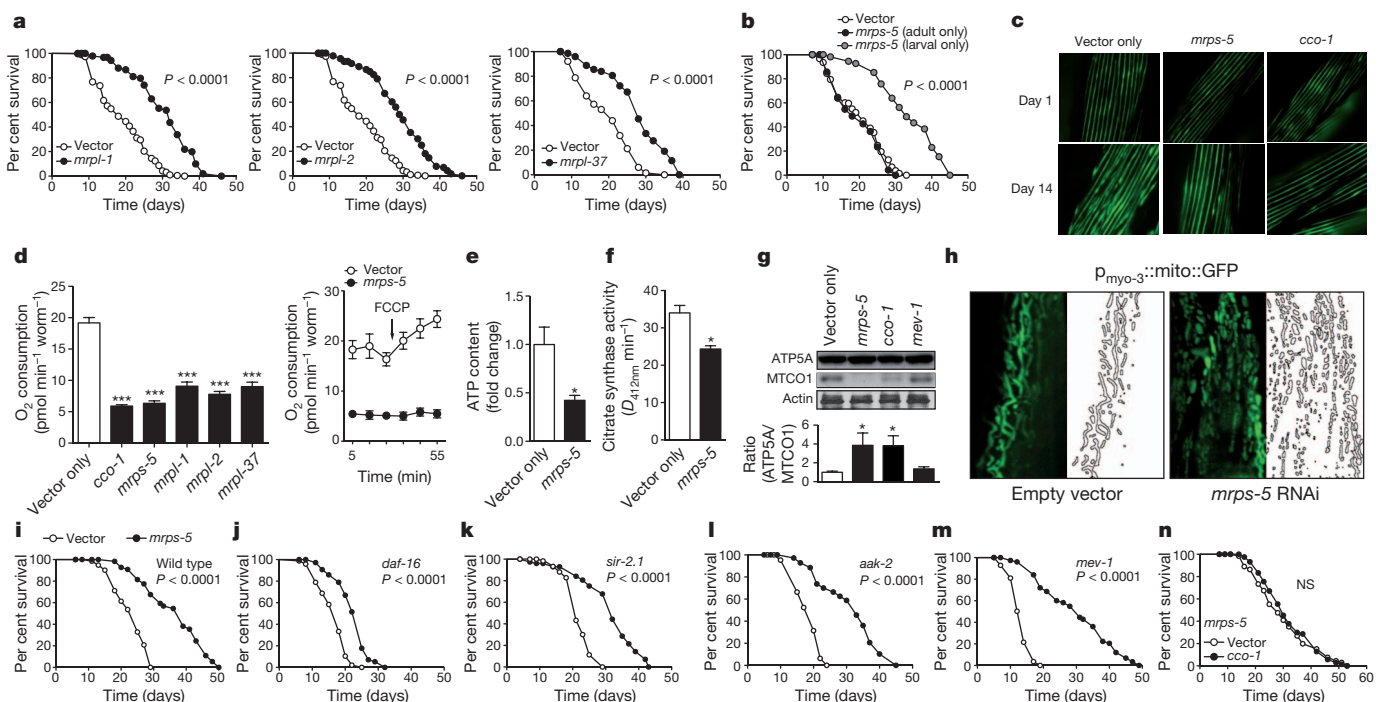
Figure 3 | *mrps-5* RNAi prevents ageing-associated functional decline and alters mitochondrial function. a, Knockdown of *mrpl-1, mrpl-2* or *mrpl-37* increased lifespan by 57%, 54%, or 41%, respectively. **b**, When RNAi of *mrps-5* was performed during the larval stages only, lifespan increased by 48%, whereas RNAi started from the L4 stage had no effect. $P \le 0.001$ is for larval-only versus either vector control or adult-only. **c**, *mrps-5* or *cco-1* RNAi prevented age-related changes in muscle morphology as evidenced by a worm p_myo-3_MYO-3::GFP reporter marking myosin heavy chain. **d**, *mrp* RNAi in *C. elegans* decreased respiration. Respiration per worm is shown, respiration is similarly decreased when corrected for protein. FCCP was added at the indicated time. Values are mean ± s.e.m. ($n = 10$), ***$P \le 0.001$. **e–g**, *mrps-5* RNAi decreased

ATP levels (**e**, $n = 3$), citrate synthase activity (**f**, $n = 3$), and altered the ratio between nDNA (ATP5A) versus mtDNA-encoded (MTCO1) oxidative phosphorylation proteins, similar to *cco-1*, but not *mev-1* (**g**, $n = 4$). *$P \le 0.05$. A, attenuance. **h**, *mrps-5* RNAi resulted in fragmented mitochondria, as visualized in body wall muscle (day 2, adult worms) using the p_myo-3_::mito::GFP reporter, which expresses mitochondrial-targeted GFP driven by the muscle-specific *myo-3* promoter. **i**, *mrps-5* RNAi increased mean lifespan by 40%. **j–m**, *mrps-5* RNAi extends lifespan of *daf-16(mu86)* (**j**), *sir-2.1(ok434)* (**k**), *aak-2(ok524)* (**l**), *mev-1(kn1)* (**m**) mutants by 37%, 40%, 69% and 112%, respectively. **n**, Knockdown of *cco-1* does not extend lifespan of *mrps-5* RNAi worms. See Supplementary Table 2 and Fig. 4.

mutant, confirming the link with mitochondrial longevity pathways (Supplementary Fig. 4h).

Mitochondrial unfolded protein response

The mitochondrial unfolded protein response accounts for longevity upon *cco-1* loss-of-function and is selective for the mitochondrial pathway and not involved in the caloric restriction or insulin/IGF-1 pathways[25]. UPR[mt] is induced by mitochondrial stress, subsequently activating a nuclear transcriptional response, inducing the chaperones HSP-6 (HSP-70 in mammals) and HSP-60 to restore mitochondrial proteostasis[26,27]. We monitored UPR[mt] using *hsp-6*::GFP (green fluorescent protein) and *hsp-60*::GFP reporter worms with reduced *mrp* expression. Similar to the *cco-1* mutant[25], *hsp-6* and *hsp-60* were induced in worms with reduced *mrp* (Fig. 4a–c and Supplementary Fig. 5a, b). This was specific for UPR[mt], as *mrps-5* RNAi did not affect UPR in the endoplasmic reticulum (UPR[ER]) and cytosolic heat shock response (Supplementary Fig. 5c). As for lifespan, UPR[mt] was not induced when *mrp* expression was only inhibited during adulthood (Supplementary Fig. 5d). We measured UPR[mt] upon combined *mrps-5* and *mev-1* inactivation. Whereas *mev-1* RNAi alone did not induce UPR[mt] (Fig. 4c, d), combined inactivation induced mitonuclear protein imbalance (Supplementary Fig. 5e) and synergistically induced UPR[mt] (Fig. 4d), accounting for the extended lifespan. Double inactivation of *mrps-5* and *cco-1* did not further enhance UPR[mt] compared to *mrps-5* alone (Fig. 4d), in line with the similar lifespan.

There are individual differences in the degree of UPR[mt] within the *mrps-5* RNAi worm population, which tightly correlate with lifespan extension (Supplementary Fig. 5f, h). GFP expression stayed similar throughout life, demonstrating that this is not an artefact of transiently reduced food intake (Supplementary Fig. 5i). To further link the level of UPR[mt] to longevity, we measured UPR[mt] in worms treated with RNAi against *mrp* genes[28]. Reduced expression of each *mrp* gene activated UPR[mt] to a different degree (Supplementary Fig. 6a). The level of UPR[mt] again correlated significantly with lifespan extension (Fig. 4e, Supplementary Fig. 6a, b and Supplementary Table 3).

Two downstream effectors of UPR[mt] are HAF1, a mitochondrial peptide transporter[29], and UBL5, a ubiquitin-like protein that regulates the transcriptional activation of mitochondrial chaperones[30]. Knockdown of *haf-1* along with *mrps-5* RNAi reduced lifespan extension, UPR[mt] and increased respiration (Fig. 4f–h). Similarly, when both *ubl-5* and *mrps-5* were knocked down, lifespan extension, the respiration phenotype and UPR[mt] were partially lost, in line with the double *cco-1* and *ubl-5* RNAi treatment[25] (Supplementary Fig. 7a–f).

This network could be traced back to mice, as *Ubl5* and the most likely mouse *haf-1* homologue—*Abcb10*—correlated tightly with *Mrp* genes, for instance in the hippocampus of the BXDs[31] and in adipose tissue of F2-intercrossed mice[17] (Fig. 4i and Supplementary Fig. 7g). Additionally, *Hspd1* (also known as *Hsp60*) correlated with several *Mrp* genes (Fig. 4i and data not shown). Gene ontology analysis showed strong connectivity between *Ubl5* and oxidative phosphorylation genes ($P = 9 \times 10^{-4}$ in eye; $P = 8.62 \times 10^{-10}$ in hippocampus), the translation process or ribosome ($P = 6 \times 10^{-4}$ eye; $P = 6.03 \times 10^{-10}$ hippocampus) and the mitochondrial inner membrane ($P = 1 \times 10^{-4}$ eye; $P = 3.31 \times 10^{-27}$ hippocampus) in the BXDs. Finally, we tied *Hspd1* in a close correlation network with various *Mrp* and oxidative phosphorylation genes (Fig. 4j).

Pharmacological mitonuclear protein imbalance

Many mitochondrial functions can be traced back to their endosymbiotic 'bacterial' origin. Consequently, antibiotics that target bacterial translation also inhibit mitochondrial translation. We therefore used doxycycline to confirm the role of mitochondrial translation in longevity, using carbenicillin—targeting the bacterial cell wall—as a control. We used heat-killed OP50 or live HT115 bacteria—the latter insensitive to low concentrations of doxycycline (data not shown)—to feed worms, to prevent antibiotic effects on bacteria. Doxycycline, given throughout life, dose-dependently extended lifespan, induced UPR[mt] not UPR[ER], and reduced oxygen consumption, without affecting ATP levels or citrate synthase activity (Fig. 5a–e and Supplementary Fig. 8a, b). Doxycycline at $60\,\mu g\,ml^{-1}$ caused developmental
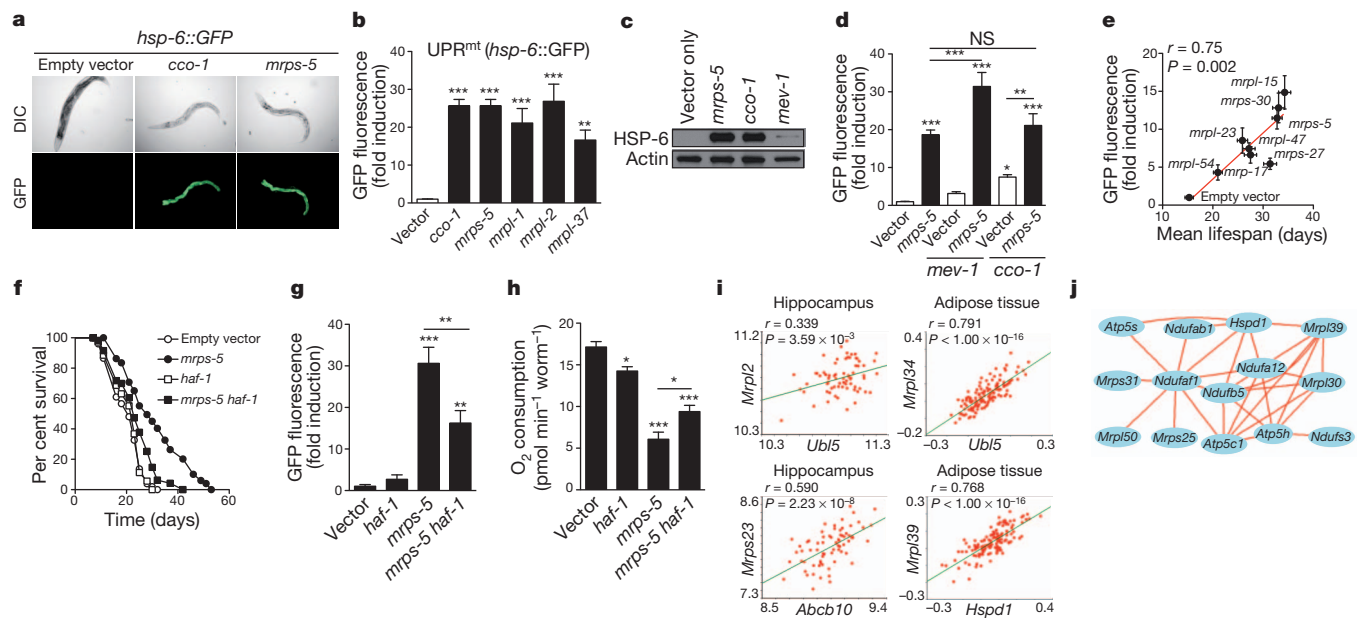
Figure 4 | *mrp* genes confer longevity effects through UPR[mt]. a, RNAi of *mrp* genes induced UPR[mt] (*hsp-6*::GFP reporter), similar to *cco-1* knockdown. Worms were synchronized at day 1 of adulthood. DIC, differential interference contrast. **b**, Quantification of UPR[mt] upon knockdown of *mrp* or *cco-1* ($n = 4$). **c**, *mrps-5* and *cco-1*, but not *mev-1*, RNAi induce UPR[mt] as reflected by the induction of HSP-6::GFP protein. **d**, Combined RNAi of *mrps-5* and *mev-1* synergistically increased UPR[mt], whereas combined *cco-1* and *mrps-5* RNAi did not further increase UPR[mt] ($n = 6$). **e**, Knockdown of different *mrp* genes results in different levels of UPR[mt], which correlates with mean lifespan

($n = 33$–61 worms for lifespan, $n = 3$ for GFP). **f–h**, Epistasis with UPR[mt] regulator *haf-1*. Double RNAi of *mrps-5* and *haf-1* partially prevented lifespan extension (**f**), UPR[mt] (**g**, $n = 5$), and reduction in respiration (**h**, $n = 10$), compared to *mrps-5* RNAi alone. **i**, In various tissues of mouse crosses, *Ubl5*, *Abcb10* and *Hspd1* expression correlated with *Mrp* expression. **j**, *Hspd1* (also known as *Hsp60*) ties in a correlation network with *Mrp* and oxidative phosphorylation genes. Connecting lines indicate a Pearson correlation coefficient of 0.75–1.0. Bar graphs show mean ± s.e.m., *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$. See also Supplementary Figs 5–7 and Supplementary Table 3.

delay, like *mrps-5* RNAi, but no abnormalities were apparent at lower concentrations (data not shown). A low concentration of doxycycline ($6 \mu g\,ml^{-1}$), given only during development, also increased lifespan and UPR^{mt} and attenuated respiration (Fig. 5f–h). Chloramphenicol—belonging to a different class of antibiotics targeting translation—also increased lifespan and UPR^{mt}, at the same time as decreasing respiration (Fig. 5i–k), when administered during development. Similar to *mrps-5* RNAi, doxycycline increased the ratio of nDNA- (ATP5A) over mtDNA-encoded (MTCO1) oxidative phosphorylation proteins (Fig. 5l).

Linking back to mammals, doxycycline decreased respiration in a mouse hepatocyte cell line (Fig. 5m). Doxycycline also induced UPR^{mt}, as evidenced by induction of *Hsp60* (Fig. 5n) and the UPR^{mt} protease *ClpP* (Supplementary Fig. 8c), and increased HSP60 protein expression in hepatocyte cell lines and primary murine hepatocytes (Fig. 5o, p). Doxycycline induced a striking mitonuclear protein imbalance in hepatocytes (Fig. 5o, p). Finally, feeding mice with doxycycline for 10 days lowered oxygen consumption *in vivo*, indicative of attenuated mitochondrial function (Fig. 5q).

Similar effects on mitonuclear protein imbalance, UPR^{mt}, respiration and lifespan, without affecting mitochondrial morphology, were also observed in worms exposed to low concentrations of ethidium bromide, which inhibits mtDNA transcription specifically[32] (Supplementary Fig. 8d–h). This indicates that mitonuclear protein imbalance is the common underlying mechanism that links basic mitochondrial function to lifespan regulation.

A conserved longevity mechanism

To define how intricately mitonuclear protein imbalance and UPR^{mt} are involved in longevity, we analysed its activation in worms exposed to rapamycin[33,34]. Rapamycin inhibits TOR signalling to alter nDNA translation, inducing mitonuclear protein imbalance[35], and increases lifespan in various species, including mice[33]. Rapamycin also increased mean worm lifespan (by 16%)[34] in a *ubl-5*-dependent manner, induced UPR^{mt}, but not UPR^{ER} or heat shock response, and increased respiration (Fig. 6a, c and Supplementary Fig. 9a). This was associated with increased ATP levels, equal citrate synthase activity and altered nDNA/mtDNA oxidative phosphorylation protein ratio (Fig. 6d, e). Additionally, rapamycin changed the balance between nDNA- and mtDNA-encoded oxidative phosphorylation subunits in mouse hepatocytes in a dose dependent manner (Fig. 6f, g). This mitonuclear protein imbalance induced HSP60 and ClpP (Fig. 6f–h). Similarly, the lifespan enhancer resveratrol induced mitonuclear protein imbalance in hepatocytes (Fig. 6i) and *ubl-5*-dependently increased worm lifespan and UPR^{mt}, but not UPR^{ER} or heat shock response, at the same time as increasing respiration and maintaining ATP levels and citrate synthase activity (Supplementary Fig. 9b–f). Mitonuclear protein imbalance and UPR^{mt} hence represent an overarching mechanism of longevity that also can be engaged by pathways that signal mainly through the nucleus.

Finally, we tested whether reactive oxygen species (ROS) and mitohormesis, a theory which posits that an initial ROS burst (after 24 h) induces adaptive long-term protection[36], could explain our worm phenotypes. However, no ROS was produced after 24 h of *mrps-5* RNAi or doxycycline, rapamycin or ethidium bromide treatment (Supplementary Fig. 10a). In addition, the mitohormesis regulators *daf-16* and *aak-2* (ref. 36, 37) were not involved in UPR^{mt} induction (Supplementary Fig. 10b) or lifespan extension (Fig. 3j, l) following *mrps-5* RNAi. Finally, the ROS scavenger *N*-acetylcysteine (NAC) did
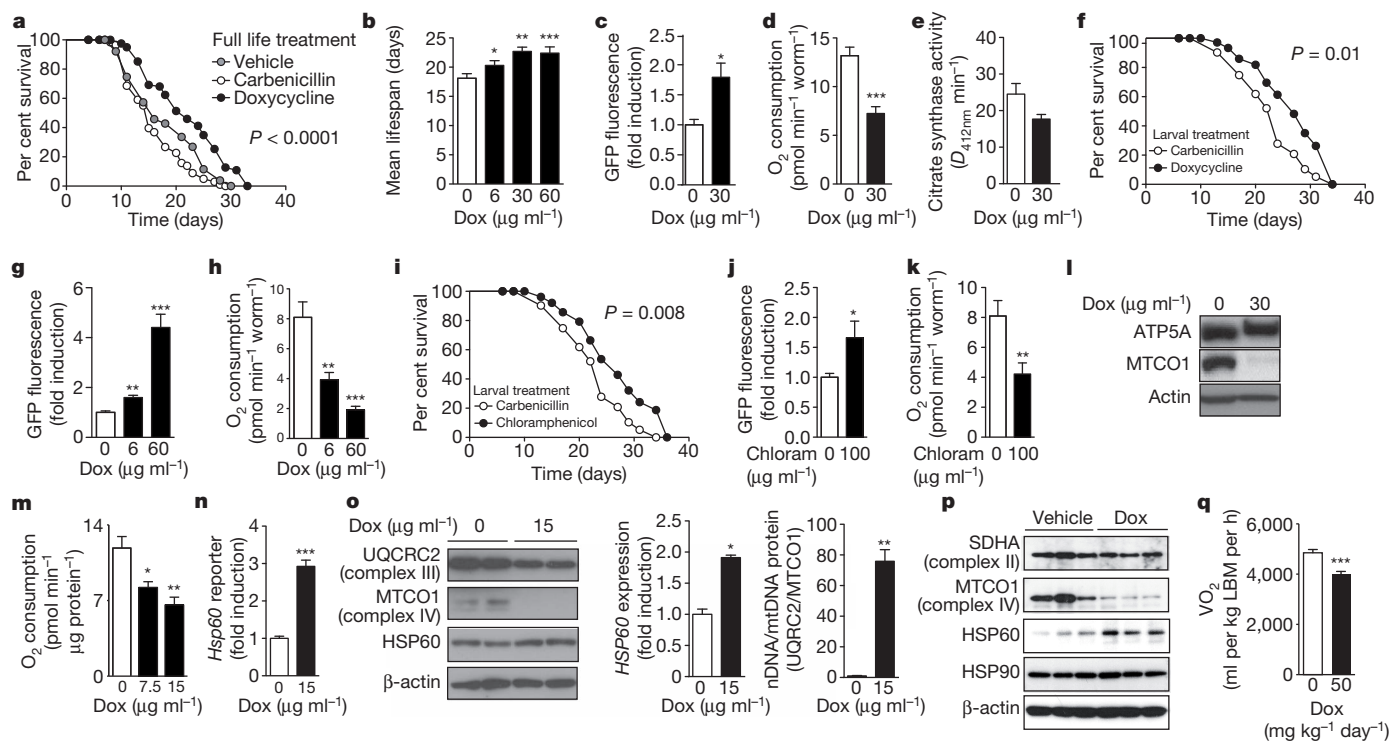
Figure 5 | Specific antibiotics extend lifespan by phenocopying *mrps-5* knockdown. a, Effects on worm lifespan of doxycycline ($30 \mu g\,ml^{-1}$), compared to carbenicillin ($30 \mu g\,ml^{-1}$) or vehicle. $P \le 0.001$ refers to statistical significance of doxycycline compared to either vehicle or the carbenicillin control. Antibiotics were given throughout life in panels **a–e**. **b**, The effects of doxycycline (Dox) on lifespan are dose-dependent. **c–e**, Doxycycline induced UPR^{mt} (**c**, $n = 5$) and reduced respiration (**d**, $n = 10$), without changing citrate synthase activity (**e**, $n = 3$). **f–k**, When treated only during larval development, doxycycline ($6 \mu g\,ml^{-1}$, **f–h**) and chloramphenicol (Chloram, $100 \mu g\,ml^{-1}$; **i–k**) extend lifespan (**f, i**), induced UPR^{mt} (**g, j**, $n = 5$) and reduced respiration (**h, k**, $n = 6$).

l, Doxycycline alters the ratio between nDNA- (ATP5A) and mtDNA-encoded (MTCO1) oxidative phosphorylation proteins in worms. Doxycycline decreased respiration in a cultured hepatocyte cell line (**m**, $n = 5$), induced *Hsp60* transcription, as measured using an *Hsp60* promoter reporter (**n**, $n = 8$) and increased HSP60 protein expression and altered the ratio of nDNA- (UQCRC2) versus mtDNA- (MTCOI) encoded proteins (**o**, $n = 2$). **p**, Doxycycline increased HSP60 protein and altered the ratio of nDNA- versus mtDNA-encoded proteins in primary murine hepatocytes. **q**, Doxycycline ($50 \text{ mg kg}^{-1}\text{ day}^{-1}$) for 10 days in C57BL/6N mice decreased oxygen consumption ($n = 10$). LBM, lean body mass. See also Supplementary Table 2. *$P \le 0.05$; **$P \le 0.01$; ***$P \le 0.001$.
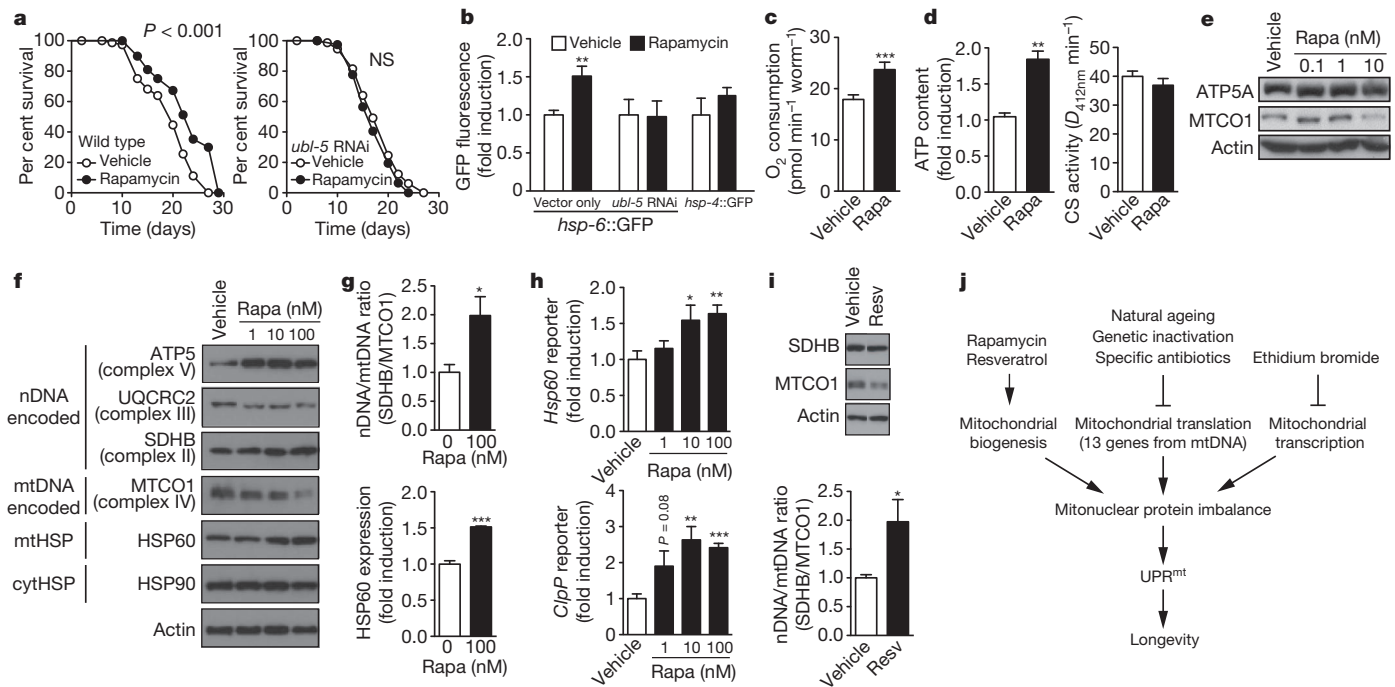
Figure 6 | Conserved function of mitonuclear protein imbalance and UPR^{mt} in longevity. a, Rapamycin (Rapa, 1 nM) extends worm lifespan in a *ubl-5*-dependent manner; **b**, *ubl-5*-dependently induced UPR^{mt} (*hsp-6*::GFP) but not UPR^{ER} (*hsp-4*::GFP) ($n = 4$). **c–e**, Rapamycin increased respiration (**c**, $n = 10$) and ATP content but not citrate synthase activity (**d**, $n = 3$) and induced mitonuclear protein imbalance (**e**). **f–h**, In mouse hepatocytes, rapamycin induces mitonuclear protein imbalance (**f, g**) and induces UPR^{mt} as shown at the protein (**f, g**, $n = 3$), and transcriptional (**h**, $n = 8$) level. **i**, Resveratrol (Resv, 25 μM) induced mitonuclear protein imbalance in mouse hepatocytes ($n = 4$). **j**, Hypothetical scheme of the mechanism by which reduced *Mrp* expression (during ageing or genetic inactivation), specific antibiotics, ethidium bromide and rapamycin and resveratrol extend lifespan by inducing UPR^{mt}. Bar graphs are shown as mean ± s.e.m., *$*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$. See also Supplementary Table 2.

not abrogate the *mrps-5* RNAi- or rapamycin-induced UPR^{mt}, nor did it suppress longevity (Supplementary Fig. 10c–f), similar to NAC-treated *cco-1* RNAi worms[25]. Together, these data demonstrate that, even though mitohormesis is important for insulin/IGF1-dependent ageing[37], UPR^{mt}-mediated longevity is independent of ROS.

Discussion

Using the BXD inbred wild type mouse panel, we identified a chromosome 2 QTL that is responsible for longevity, with *Slc12a1*, *Mrps5* and *Ttl* showing strong correlation with lifespan. A holistic approach, involving bioinformatics, genetics and pharmacological strategies in both worms and mammals, established that *Mrps5* and the *Mrp* protein family are the main actors in metabolic lifespan regulation. The MRPs constitute the mitoribosome that regulates translation of 13 mtDNA-encoded proteins, underscoring the vital importance of mitochondrial protein production[6,7].

Inhibiting mitochondrial translation reduced respiration and extended lifespan. There is an apparent dichotomy, however, as rapamycin (this study) as well as NAD$^+$ boosters—resveratrol (this study), nicotinamide riboside, nicotinamide and PARP inhibitors (L.M., R.H.H. and J.A., unpublished observations)—couple longevity to increased respiration. Abnormal mitochondrial proteostasis could reconcile these disparate observations. Knockdown of *mrps-5* or *cco-1* affect proteostasis and activate UPR^{mt} (refs 25, 28). From the *cco-1* study[25] it was, however, unclear if there was a direct connection between the level of UPR^{mt} and the lifespan extension. Our data demonstrate a positive correlation between the level of UPR^{mt} and lifespan. Moreover, UPR^{mt} seems to result from an imbalance between mtDNA- and nDNA-encoded proteins and is a common feature linking mitochondrial longevity pathways. Genetic defects in *mrp* or respiratory chain genes, antibiotics that inhibit translation, or moderate mtDNA transcription inhibition, induce such a mitonuclear protein imbalance from within mitochondria. Conversely,

resveratrol and rapamycin change the production of nDNA-encoded mitochondrial proteins and if this is not matched with the levels of mtDNA-encoded proteins, mitonuclear protein imbalance and UPR^{mt} will also ensue, which favours longevity (Fig. 6j). The reason why *mev-1* mutants do not display UPR^{mt} is consistent with the fact that complex II is entirely nDNA-encoded and therefore does not require a balanced production of proteins from the nDNA and mtDNA. Additionally, complex II can be bypassed for mitochondrial ATP generation and is not part of oxidative phosphorylation supercomplexes[38]. Although further work to validate this hypothesis is warranted, this could explain apparent contradictions such as why mutations that either decrease or increase respiration can both induce longevity.

Our data identify MRPs as a novel longevity protein family, conserved from worms to mammals. The identification of these genes was triggered by analysis of murine reference populations. Hence, it is natural variation in *Mrp* expression, not artificial loss- or gain-of-function, that translates to longevity. In worms, longevity involves enhanced fitness and UPR^{mt}, and correlates tightly with levels of *mrp* knockdown. Our data suggest that stoichiometric imbalance between nDNA- and mtDNA-encoded oxidative phosphorylation proteins, or mitonuclear protein imbalance, is at the core of UPR^{mt} activation, both in worms and mammals. The apparent conservation of mitonuclear protein imbalance and UPR^{mt} as a general longevity mechanism should invite further studies to explore whether targeting UPR^{mt} can prevent ageing-associated functional decline and treat diseases linked with ageing.

METHODS SUMMARY

For the identification of novel longevity genes we used publicly available longevity data of the BXD genetic reference population. We used interval mapping in GeneNetwork (http://www.genenetwork.org; Trait ID 10112) for QTL analysis. Pearson's *r* genetic correlation was performed to establish genetic correlations with longevity. Candidate longevity genes were knocked down in wild type *C. elegans* strain Bristol N2, which were tested for lifespan at 20 °C. Worm UPR^{mt} was measured in the *hsp-6*::GFP or *hsp-60*::GFP reporter strains, and additional

mechanistic information was obtained after knockdown in other worm strains, as described in the online Methods. Treatments with carbenicillin, doxycycline, chloramphenicol, ethidium bromide, *N*-acetylcysteine, resveratrol and rapamycin (all from Sigma) were performed with heat-killed bacteria or with HT115 bacteria. Oxygen consumption in worms was measured using the Seahorse XF24 (Seahorse Bioscience) using 50 worms per well.

Mouse hepatocyte cell lines AML-12 or Hepa1-6, or primary mouse hepatocytes were used for mammalian conservation assays. UPR^{mt} was measured using luciferase reporters containing human *Hsp60* and *ClpP* promoter fragments and cellular oxygen consumption was measured using the Seahorse XF24 equipment. Mouse indirect calorimetry was performed in wild type C57BL/6N mice treated for 10 days with $50\,mg\,kg^{-1}\,day^{-1}$ doxycycline as a food admix.

Full Methods and any associated references are available in the online version of the paper.

1. Fontana, L., Partridge, L. & Longo, V. D. Extending healthy life span–from yeast to humans. *Science* **328**, 321–326 (2010).
2. Houtkooper, R. H., Williams, R. W. & Auwerx, J. Metabolic networks of longevity. *Cell* **142**, 9–14 (2010).
3. Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
4. Mair, W. & Dillin, A. Aging and survival: the genetics of life span extension by dietary restriction. *Annu. Rev. Biochem.* **77**, 727–754 (2008).
5. Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
6. Sharma, M. R. *et al.* Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell* **115**, 97–108 (2003).
7. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
8. Liao, C. Y., Rikke, B. A., Johnson, T. E., Diaz, V. & Nelson, J. F. Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. *Aging Cell* **9**, 92–95 (2010).
9. Argmann, C. A., Chambon, P. & Auwerx, J. Mouse phenogenomics: the fast track to "systems metabolism". *Cell Metab.* **2**, 349–360 (2005).
10. Andreux, P. A. *et al.* Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* **150**, 1287–1299 (2012).
11. Peirce, J. L., Lu, L., Gu, J., Silver, L. M. & Williams, R. W. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* **5**, 7 (2004).
12. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
13. De Haan, G. & Van Zant, G. Genetic analysis of hemopoietic cell cycling in mice suggests its involvement in organismal life span. *FASEB J.* **13**, 707–713 (1999).
14. Shifman, S. *et al.* A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol.* **4**, e395 (2006).
15. Edwards, M. G. *et al.* Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program. *BMC Genomics* **8**, 80 (2007).
16. Geisert, E. E. *et al.* Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. *Mol. Vis.* **15**, 1730–1763 (2009).
17. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
18. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
19. Hamilton, B. *et al.* A systematic RNAi screen for longevity genes in *C. elegans*. *Genes Dev.* **19**, 1544–1555 (2005).
20. Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS Genet.* **1**, e17 (2005).
21. Wong, A., Boutis, P. & Hekimi, S. Mutations in the *clk-1* gene of *Caenorhabditis elegans* affect developmental and behavioral timing. *Genetics* **139**, 1247–1259 (1995).
22. Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298**, 2398–2401 (2002).
23. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A. *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
24. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
25. Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144**, 79–91 (2011).
26. Haynes, C. M. & Ron, D. The mitochondrial UPR - protecting organelle protein homeostasis. *J. Cell Sci.* **123**, 3849–3855 (2010).
27. Zhao, Q. *et al.* A mitochondrial specific stress response in mammalian cells. *EMBO J.* **21**, 4411–4419 (2002).
28. Yoneda, T. *et al.* Compartment-specific perturbation of protein handling activates genes encoding mitochondrial chaperones. *J. Cell Sci.* **117**, 4055–4066 (2004).
29. Haynes, C. M., Yang, Y., Blais, S. P., Neubert, T. A. & Ron, D. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* **37**, 529–540 (2010).
30. Benedetti, C., Haynes, C. M., Yang, Y., Harding, H. P. & Ron, D. Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response. *Genetics* **174**, 229–239 (2006).
31. Overall, R. W. *et al.* Genetics of the hippocampal transcriptome in mouse: a systematic survey and online neurogenomics resource. *Front. Neurosci.* **3**, 55 (2009).
32. Zylbee, E., Vesco, C. & Penman, S. Selective inhibition of the synthesis of mitochondria-associated RNA by ethidium bromide. *J. Mol. Biol.* **44**, 195–204 (1969).
33. Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).
34. Robida-Stubbs, S. *et al.* TOR signaling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO. *Cell Metab.* **15**, 713–724 (2012).
35. Zid, B. M. *et al.* 4E-BP extends lifespan upon dietary restriction by enhancing mitochondrial activity in *Drosophila*. *Cell* **139**, 149–160 (2009).
36. Schulz, T. J. *et al.* Glucose restriction extends *Caenorhabditis elegans* life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell Metab.* **6**, 280–293 (2007).
37. Zarse, K. *et al.* Impaired insulin/IGF1 signaling extends life span by promoting mitochondrial L-proline catabolism to induce a transient ROS signal. *Cell Metab.* **15**, 451–465 (2012).
38. Schägger, H. & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.* **19**, 1777–1783 (2000).

Author Contributions D.R., N.M. and E.K. contributed equally to this work. R.H.H., L.M. and J.A. conceived and designed the project. R.H.H. and R.W.W. performed QTL mapping and sequence analyses. R.H.H., L.M., E.K., D.R., N.M., G.K. performed experiments. R.H.H. and J.A. wrote the manuscript with contributions from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A. (admin.auwerx@epfl.ch).

METHODS

Forward genetics. For the *in silico* search for longevity genes, we have used publicly available longevity data of the BXD genetic reference population[39]. We used interval mapping in GeneNetwork (http://www.genenetwork.org), Trait ID 10112 for QTL analysis. Pearson's *r* genetic correlation was performed to establish genetic correlations with longevity (Eye M430v2 Data Set (Sept08) RMA)[40]. Principle component analysis (PCA) was performed using tools implemented in GeneNetwork.

Reverse genetics. We used BXD mice (Trait ID 1448488_at from Eye M430v2 (Sep08) RMA) and an F2 cross (Trait ID 10024407239 from UCLA BHF2 Adipose (June05) mlratio)[41] to analyse *Mrps5* correlates. We primarily used microarray data from the eye of BXD mice, as this organ contains multiple tissues and cell types—for example, neuronal and muscle—and because this microarray is very well annotated. We analysed correlates as follows: we selected the top 250 *Mrps5* correlates in both databases and performed Kyoto encyclopedia of genes and genomes (KEGG) and gene ontology analysis. For the network approach, we selected the oxidative phosphorylation genes within the top 250 traits associated with *Mrps5* and used the Pearson correlation coefficients to create a network (threshold = 0.6). For *Ubl5* and *Hspd1* correlates, we used a BXD hippocampus database[31], and adipose tissue data from an F2 intercross[41]. The *Hspd1* network was generated by selecting the oxidative phosphorylation and *Mrp* genes in the top 500 correlates in the F2 intercross adipose tissue (threshold = 0.75).

Hierarchical clustering. Unsupervised hierarchical clustering was performed using complete linkage and Pearson rank correlation distance on the *z*-score normalized gene expression data using software implemented in GenePattern[42,43].

***C. elegans* strains and RNAi experiments.** Strains used were wild-type Bristol N2, CB1370 *daf-2(e1370)* III, CB4876 *clk-1(e2519)* III, CF1038 *daf-16(mu86)* I, DA465 *eat-2(ad465)* II, RB754 *aak-2(ok524)* X, VC199 *sir-2.1(ok434)* IV, TK22 *mev-1(kn1)* III, SJ4100 (zcIs13[*hsp-6*::GFP]), SJ4058 (zcIs9[*hsp-60*::GFP]), SJ4005 (zcIs4[*hsp-4*::GFP]), SJ4103 (zcIs14[*myo-3*::GFP(mit)]), SJ4143 (zcIs17[*ges-1*::GFP(mit)]), RW1596 stEx30[*myo-3*::GFP +*rol-6*(su1006)], and CL2070 (dvIs[*hsp-16.2*::GFP]). Strains were provided by the *Caenorhabditis* Genetics Center (University of Minnesota).

Bacterial feeding RNAi experiments were carried out as described[44]. Clones used were *mrps-5* (E02A10.1), *mrpl-1* (F33D4.5), *mrpl-2* (F56B3.8), *mrpl-37* (Y48E1B.5), *cco-1* (F26E4.9), *nkcc-1* (Y37A1C.1), *ubl-5* (F46F11.4), *haf-1* (C30H6.6), *mev-1* (T07C4.7), *mrp-17* (R12E2.12), *mrpl-47* (B0261.4), *mrpl-23* (T08B2.8), *mrpl-54* (F25H5.6), *mrps-30* (B0511.8), *mrps-27* (K11B4.1) and *mrpl-15* (Y92H12BR.8). Clones were purchased from GeneService and sequenced and RNAi clone information is shown in Supplementary Table 4. Double RNAi experiments were carried out by mixing the bacterial cultures directly before seeding on NGM plates. Controls were RNAi clones 50% diluted with control empty vector RNAi bacteria.

Lifespan tests were performed at 20 °C as described[45]. Treatments with carbenicillin, doxycycline, chloramphenicol, ethidium bromide, *N*-acetylcysteine, resveratrol and rapamycin (all from Sigma) were performed with heat-killed bacteria[46] or with HT115 bacteria.

Construction of *ttll-9* RNAi clones. The RNAi clone was cloned by reverse transcription PCR amplification of the corresponding cDNA from total RNA with primers: *ttll-9*-XbaI (601–960): 5′-GGGTCTAGATCATAGCC ATACAGCTCGAAACAGTGGTTATCT-3′, *ttll-9*-KpnI (601–960): 5′-GGGG GTACCAAGTTCGATGTCAGAATCTATGTGCTGGTT-3′. PCR products were digested with KpnI/XbaI, and ligated into appropriately digested plasmid L4440.

Microscopy and GFP analysis. GFP expression and quantification were carried out as described[47]. Briefly, eighty worms (day 1 adults) were picked (20 per well of a black-walled 96-well plate) and GFP was monitored on a Victor X4 plate reader (Perkin Elmer). Each experiment was repeated at least twice. For picture acquisition of *hsp-6*::GFP expression, animals were mounted on 2% agarose pads in 10 mM tetramisole (Sigma) and examined using a Zeiss Axioplan-2 microscope (Carl Zeiss). Time lapse was performed by recording 10 pictures of mobile worms with a 10 s interval. Worm tracking was carried out using ImageJ software. Experiments were conducted with three worms from different plates.

Confocal microscopy and image processing. Worms were immobilized with tetramisole (Sigma) and mounted on 6% agarose pads on glass slides. Images were acquired using Zeiss LSM 700 upright confocal microscope (Carl Zeiss). For each condition, several worms were imaged. Image processing was performed with Fiji software. Tracing of the mitochondrial network contour was done by the use of Gaussian blur filter followed by the application of Laplacian operator.

Worm respiration assays. Oxygen consumption was measured using the Seahorse XF24 (Seahorse Bioscience). Typically, 200 animals were recovered from NGM plates with M9 medium, washed three times to eliminate residual bacteria, and resuspended in 500 µl M9 medium. Worms were transferred to 24-well Seahorse plates (50 worms per well) and oxygen consumption was measured 6 times. FCCP treatment was performed at 10 µM final concentration.

MitoSox staining. MitoSox staining was performed as previously described with slight modification[48]. Briefly, a population of 100 worms was recovered in 1 ml of M9 buffer, washed three times to remove residual bacteria, and resuspended in 200 µl of 1:200 MitoSox stock solution (initial stock solution was dissolved at 1 mM in dimethylsulphoxide (DMSO)). After 20 min of treatment, worms were washed five times in 1 ml of M9 buffer to eliminate the MitoSox reagent and then transferred in a black-walled 96-well plate for reading.

Quantification of ATP levels and citrate synthase enzymatic activity. Total ATP content was measured by the CellTiter-Glo luminescent cell viability assay (Promega). The luminescence was recorded with a Victor X4 plate reader (PerkinElmer) and values normalized by the total protein concentration determined using a Bradford assay.

Citrate synthase enzymatic activity was determined using the CS assay kit (Sigma). Absorbance at 412 nm was recorded on a Victor X4 (PerkinElmer) with 10 readings over the 1.5 min timespan. These readings were in the linear range of enzymatic activity. The difference between baseline and oxaloacetate-treated samples was obtained and used to calculate total citrate synthase activity according to the formula provided in the manual.

Worm electron microscopy. Worms were fixed using high pressure freezing (Leica Micosystems, HPM100) in low melting point agarose, freeze substituted in 0.5% osmium tetroxide, and 0.5% uranyl acetate in acetone at −90 °C, and then slowly warmed to −10 °C and transferred to pure acetone. Worms were embedded in increasing concentrations of epon resin at 20 °C, transferred to flat embedding moulds in pure resin, and cured at 65 °C for 48 h. Serial sections were cut at 50 nm, and placed onto formvar support films on single slot copper grids. These were imaged at 80 kV filament tension in a transmission electron microscope with a charge coupled device (CCD) camera (Tecnai Spirit, FEI Company, with Eagle 4k × 4k CCD camera).

Hsp60 and ClpP reporter assays. Human *HSPD1* (also known as *Hsp60*) and *CLPP* promoter fragments (−603 to +735 for *Hsp60*, and −1272 to +337 for *CLPP*) were amplified and ligated into the pGL3 basic vector (Promega). Primers used were HSP60-forward (5′-GACAACGCGTAACAAAAGAGGGGCGTCAG-3′) and HSP60-reverse (5′-GACACTCGAGCCCTGAGAAACCAAGTCAGC-3′), and ClpP-forward (5′-GACAACGCGT CTTTCCGGTCTGATCTCCAG-3′) and ClpP-reverse (5′-GACACTCGAGGTACCGTCTGCTCCACCAC-3′). The primers were tailed with MluI site (forward) and XhoI site (reverse).

The mouse hepatocyte cell line AML-12 (alpha mouse liver 12) was obtained from ATCC (Manassas). Cells were grown according to the supplier guidelines but in the absence of antibiotics unless specified.

Transfections were performed in 96-well plates using jetPEI (PolyPlus). Each well contained 30 ng of luciferase reporter and 5 ng of β-galactosidase expression plasmid. After 6 h of incubation with the DNA–jetPEI complexes, the transfection medium was exchanged for medium with or without doxycycline. Doxycycline was dissolved in DMSO and added to the cells in the medium.

Luciferase activity was measured with the luciferase assay system (Promega) in the Victor X4 (PerkinElmer) and normalized to β-galactosidase activity.

Cell culture and oxygen consumption. Hepa1-6 mouse liver cells were maintained in DMEM medium containing 4.5 g per litre glucose and 10% fetal calf serum. Hepa1-6 cells were incubated for 48 h with doxycycline or chloramphenicol (both Sigma) in medium containing 1 g per litre glucose, with 1% (v/v) oleate-BSA (Sigma) supplementation, and without any other antibiotics. Oxygen consumption was measured using the Seahorse XF24 equipment[49], and normalized for protein content.

Mouse indirect calorimetry. Wild type C57BL/6N mice were treated for 10 days with 50 mg kg^{-1} day^{-1} doxycycline (Sigma) as a food admix. Oxygen consumption (VO$_2$) was monitored by indirect calorimetry using the comprehensive laboratory animal monitoring system (CLAMS) (Columbus Instruments)[50].

Culture of primary hepatocytes. Primary hepatocytes were prepared from 8–10-week-old C57BL/6 mice by collagenase perfusion method as described[51,52]. Isolated primary hepatocytes were plated with medium 199 (Gibco) including 10% FCS, 10 units per ml penicillin and 10 µg ml^{-1} streptomycin. After 3–6 h attachment, cells were replaced with media without FCS and treated with 30 µg ml^{-1} doxycycline or DMSO every 24 h. Primary hepatocytes were harvested 48 h later.

Western blotting. Western blotting was performed with antibodies against HSP60 (N-20), HSP90 (BD Transduction Laboratories), β-actin (Santa Cruz Biotechnology), MitoProfile Total OXPHOS Rodent WB Antibody Cocktail against ATP5A (H28O16.1 in worms), MTCO1/COX1 (MTCE.26 in worms) and UQCRC2 (Abcam), green fluorescent protein (Cell Signaling) HRP-labelled anti-goat and anti-mouse secondary antibodies.

39. De Haan, G. & Van Zant, G. Genetic analysis of hemopoietic cell cycling in mice suggests its involvement in organismal life span. *FASEB J.* **13,** 707–713 (1999).

40. Geisert, E. E. *et al.* Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. *Mol. Vis.* **15,** 1730–1763 (2009).
41. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452,** 429–435 (2008).
42. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38,** 500–501 (2006).
43. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (2004).
44. Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2,** research0002—research0002.10 (2000).
45. Mouchiroud, L. *et al.* Pyruvate imbalance mediates metabolic reprogramming and mimics lifespan extension by dietary restriction in *Caenorhabditis elegans*. *Aging Cell* **10,** 39–54 (2011).
46. Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430,** 686–689 (2004).
47. Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144,** 79–91 (2011).
48. Yang, W. & Hekimi, S. A mitochondrial superoxide signal triggers increased longevity in *Caenorhabditis elegans*. *PLoS Biol.* **8,** e1000556 (2010).
49. Watanabe, M. *et al.* Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **439,** 484–489 (2006).
50. Lagouge, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1α. *Cell* **127,** 1109–1122 (2006).
51. Ryu, D. *et al.* Endoplasmic reticulum stress promotes LIPIN2-dependent hepatic insulin resistance. *Diabetes* **60,** 1072–1081 (2011).
52. Noriega, L. G. *et al.* CREB and ChREBP oppositely regulate SIRT1 expression in response to energy availability. *EMBO Rep.* **12,** 1069–1076 (2011).

ARTICLE

Pif1 family helicases suppress genome instability at G–quadruplex motifs

Katrin Paeschke[1]†*, Matthew L. Bochman[1]*, P. Daniela Garcia[1], Petr Cejka[2]†, Katherine L. Friedman[3], Stephen C. Kowalczykowski[2] & Virginia A. Zakian[1]

The *Saccharomyces cerevisiae* Pif1 helicase is the prototypical member of the Pif1 DNA helicase family, which is conserved from bacteria to humans. Here we show that exceptionally potent G-quadruplex unwinding is conserved among Pif1 helicases. Moreover, Pif1 helicases from organisms separated by more than 3 billion years of evolution suppressed DNA damage at G-quadruplex motifs in yeast. The G-quadruplex–induced damage generated in the absence of Pif1 helicases led to new genetic and epigenetic changes. Furthermore, when expressed in yeast, human PIF1 suppressed both G-quadruplex–associated DNA damage and telomere lengthening.

G-quadruplex (also known as G4) DNA is a four-stranded DNA structure held together by guanine (G) base pairing, and most genomes are replete with G4 motifs—that is, sequences that can form G4 structures *in vitro*[1]. Several DNA helicases unwind G4 structures *in vitro*, including several human helicases (WRN, BLM, FANCJ and PIF1), the mutation of which is associated with genome instability, premature ageing and/or increased cancer risk (Supplementary Table 1).

The *S. cerevisiae* 5′-to-3′ DNA helicase Pif1 has multiple nuclear functions, including inhibition of telomerase at both telomeres and double-strand breaks[2–5] and preventing replication pausing and double-stranded breaks at G4 motifs[6,7]. Unlike most eukaryotes, which encode one Pif1 helicase, *S. cerevisiae* encodes two, Pif1 and Rrm3 (ref. 8). However, Pif1 and Rrm3 have different functions. Until now, the only known nuclear functions of Rrm3 were to promote replication past stable protein complexes[9] and to separate converged replication forks[8,10]. Although the functions of human PIF1 are not known, mutation of a conserved PIF1 residue in the Pif1 family signature motif[11] is associated with increased cancer risk[12].

Pif1 is a potent G4 binder and unwinder

So far, more than 20 tested helicases, including both *S. cerevisiae* Pif1 and human PIF1, bind and/or unwind G4 structures *in vitro* (Supplementary Table 1). To determine whether *S. cerevisiae* Pif1 is particularly adept at unwinding G4 structures, we analysed its G4 binding and unwinding activities in a quantitative manner. Filter-binding assays were used to quantify Pif1 binding to different DNA substrates (Fig. 1 and Supplementary Fig. 1; oligonucleotides in Supplementary Table 2). Pif1 had a preference for poly-purine tracts (Fig. 1a), which was consistent with its preference for G-rich (dissociation constant (K_d) = 0.04 nM) over non-G-rich (K_d = 0.2 nM) single-stranded DNA (ssDNA) (Fig. 1c). Pif1 displayed similarly high binding to G4 DNA (average K_d = 0.08 nM for three G4 motifs; Fig. 1e), which was roughly 500-fold better than its binding to Y-structures (that is, Y-shaped DNA structures consisting of ssDNA arms connected to a double-stranded DNA stem) (Supplementary Fig. 1).

S. cerevisiae Pif1 efficiently unwound seven out of seven G4 substrates, six from *S. cerevisiae* chromosomal DNA and TP$_{G4}$, and a standard G4 substrate from the mouse immunoglobulin locus

(Fig. 2a, d–f and data not shown; sequences in Supplementary Table 2). The apparent Michaelis constant (K_m) of unwinding for each G4 structure occurred at equimolar concentrations of Pif1 and the G4 substrate (0.1 nM) (Fig. 2a, d). By contrast, using the same enzyme preparation, a fivefold molar excess of Pif1 was required to unwind Y-structures (Fig. 2a), even though Y-structures are considered preferred Pif1 substrates[13].

S. cerevisiae Pif1 unwinding rates of G4 structures (Fig. 2a, d) were too fast at 30 °C to quantify. Therefore, time-course analyses were performed at a suboptimal temperature (25 °C; Fig. 2e). Even at 25 °C, Pif1 unwound 100% of the G4 substrate in 2 min. Although Pif1 cannot unwind Y-structures under single-cycle conditions[14] (that is, in the presence of a 500-fold excess of unlabelled G4 DNA), it unwound G4 structures under single-cycle conditions with no change in kinetics (Fig. 2f). Thus, G4 structures are a preferred Pif1 substrate.

Bacterial Pif1s are potent G4 unwinders

We are unable to purify full-length Rrm3, Pfh1 (the *Schizosaccharomyces pombe* Pif1 family helicase), or human PIF1 (ref. 11). However, the sequences of many bacterial Pif1 proteins are available[15]. To determine whether vigorous G4 unwinding is conserved among Pif1 family helicases, we purified Pif1 proteins from four diverse bacteria and a bacteriophage. All five enzymes robustly unwound the rDNA$_{G4}$ and TP$_{G4}$ substrates with apparent K_m values in the subnanomolar to nanomolar range (Fig. 2d and Supplementary Fig. 2a–f). As with *S. cerevisiae* Pif1, each of these Pif1 family helicases unwound G4 DNA rapidly (Fig. 2e and Supplementary Fig. 2) and under single-cycle conditions (although not to completion) (Fig. 2c).

To determine whether a helicase is particularly good at G4 DNA unwinding, one can compare its unwinding of G4 DNA to the unwinding of other substrates (for example, *S. cerevisiae* Pif1 unwinding of G4 DNA versus Y-structures; Fig. 2a) or compare the unwinding activity of multiple helicases on the same G4 substrate. As several RecQ family helicases unwind G4 structures *in vitro* (Supplementary Table 1), we tested Sgs1, an *S. cerevisiae* RecQ helicase, and *Escherichia coli* RecQ. Sgs1 bound ssDNA and unwound Y-structures at reported rates[16] (Fig. 2b). However, Sgs1 did not bind preferentially to G-rich DNA (Fig. 1b, d), and the apparent Sgs1 binding affinity for four G4

[1]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [2]Departments of Microbiology and Molecular Genetics, and Molecular and Cellular Biology, University of California, Davis, California 95616, USA. [3]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37232, USA. †Present addresses: Department of Biochemistry, Theodor Boveri-Institute, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany (K.P.); Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland (P.C.). *These authors contributed equally to this work.
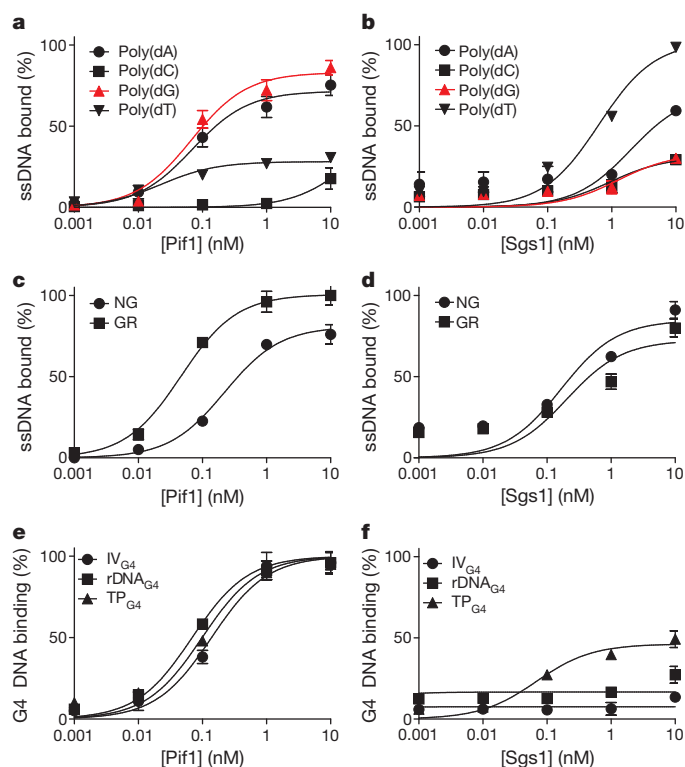
Figure 1 | Pif1 preferentially binds G4 DNA. a, b, Pif1 (**a**) and Sgs1 (**b**) binding to homopolymeric oligonucleotides. **c, d**, Pif1 (**c**) and Sgs1 (**d**) binding to 20-nucleotide substrates consisting of 25% each dNTP (non-G-rich, NG) or 75% purines (G-rich, GR; Supplementary Table 2). **e, f**, Pif1 (**e**) and Sgs1 (**f**) binding to G4 structures. Data are mean ± s.d. from ≥3 experiments.

Figure 2 | Pif1 helicases preferentially unwind G4 structures. a–f, G4 and Y-structure (both 100 pM) unwinding after 20 min (or as indicated) was assessed in standard assays at 37 °C (*E. coli* RecQ (EcRecQ)), 30 °C (Sgs1, *Bacteroides* Pif1 (BacPif1)), or 25 °C (*S. cerevisiae* Pif1 (ScPif1)). A given G4 substrate differed only in the location of the poly(dA) tail (5′ for Pif1 and 3′ for RecQ). For all G4 experiments, graphs show mean unwinding by Pif1 for three G4 structures (IV$_{G4}$, rDNA$_{G4}$ and TP$_{G4}$) or TP$_{G4}$ unwinding by BacPif1, Sgs1 and RecQ. **a–c**, G4 and Y-structure unwinding as a function of [ScPif1] (**a**), [Sgs1] (**b**) and [EcRecQ] (**c**). **d**, G4 unwinding as a function of [helicase]. **e**, G4 unwinding time course by 100 pM ScPif1, 10 nM BacPif1, 10 nM Sgs1 and 50 nM EcRecQ. Higher Sgs1 concentrations and/or the addition of *S. cerevisiae* RPA did not increase unwinding (data not shown). **f**, G4 unwinding under single-cycle conditions by 100 pM ScPif1, 10 nM BacPif1 and 50 nM EcRecQ.

structures was more than 40-fold lower than that of *S. cerevisiae* Pif1 (Fig. 1f). Similarly, Sgs1 was considerably less efficient than all tested Pif1 family helicases at unwinding G4 structures (for example, 1,000-fold molar excess of Sgs1 was needed to unwind 50% of the G4 structures; Fig. 2e). *E. coli* RecQ displayed better unwinding of the TP$_{G4}$ substrate than Sgs1 (Fig. 2c, e, f), but the apparent K_m value of this activity was 160-fold greater than that of Pif1.

Time-course experiments revealed slower unwinding of G4 structures by Sgs1 and *E. coli* RecQ (Fig. 2e) relative to Pif1, and Sgs1 was unable to unwind G4 DNA under single-cycle conditions. Although *E. coli* RecQ did unwind the TP$_{G4}$ substrate under single-cycle conditions (Fig. 2f), 500-fold more protein relative to Pif1 was necessary for activity, yielding a half-life ($t_{1/2}$) approximately tenfold greater than that of Pif1 in the same assay. The same preparations of Sgs1 and *E. coli* RecQ unwound a conventional Y-structure 100- and 10-fold better, respectively, than G4 structures (Fig. 2b, c). With WRN, a human RecQ helicase, we obtained a similar unwinding rate for TP$_{G4}$ and rDNA$_{G4}$ as reported for TP$_{G4}$ (ref. 17); both were similar to unwinding by Sgs1 (Fig. 2e and Supplementary Fig. 3). Thus, three evolutionarily diverse RecQ helicases were much less effective than any tested Pif1 family enzyme at G4 unwinding.

Pif1s suppress G4–induced instability

We developed a quantitative assay to monitor G4-induced genome instability by modifying the gross-chromosomal rearrangement (GCR) assay[18]. The GCR assay detects complex genome rearrangements by simultaneous selection against *URA3* and *CAN1* (Fig. 3a). We modified this assay by inserting four strong Pif1-binding sites[6], two G4 motifs (Chr I$_{G4}$, Chr X$_{G4}$) and two non-G4 sites (Chr VII$_{NG}$, not G-rich; Chr I$_{GR}$, G-rich, not G4-forming; Supplementary Table 3), within the *PRB1* locus, a non-essential gene that is centromere-proximal to the two counterselectable genes (Fig. 3a). As reported[19], the GCR rate in the 'no insert' wild-type control was approximately

1×10^{-10} events per generation, and none of the inserts affected this rate (Table 1). However, the already high GCR rate in the no-insert *pif1-m2* strain was increased ~3-fold in the presence of either of the G4 motifs but was unaffected by either of the other Pif1-binding sites (Table 1). The G4 inserts did not increase GCR rates in *rrm3Δ* or *sgs1Δ* cells compared to no-insert controls. Likewise, the GCR rate in *pif1-m2 sgs1Δ* cells was not increased by the G4 inserts. However, the GCR rate in *pif1-m2 rrm3Δ* cells was approximately eight times higher in the presence of the G4 motifs compared to the *pif1-m2 rrm3Δ* strains containing no insert or non-G4 Pif1-binding sites (1,700-fold over the background no-insert wild-type levels; Table 1). These data suggest that when Pif1 levels are low, Rrm3 (but not Sgs1) suppresses G4-induced genome instability. Consistent with these findings, Rrm3 bound preferentially to G4 motifs in *pif1-m2* but not wild-type cells[20] (Supplementary Fig. 4a).

To determine whether Pif1 suppression of DNA damage at G4 motifs is evolutionarily conserved, we tested diverse Pif1 proteins for their ability to suppress the high GCR rate in *pif1-m2 rrm3Δ*+G4 cells. Helicases were introduced on a single-copy plasmid and expressed from the *RRM3* promoter (see Supplementary Fig. 6 for western analysis of protein expression). A simple spot assay was used to monitor the frequency of GCR events; cells were spotted 150 times at high density on 5-fluoroorotic acid (5-FOA) plus canavanine sulphate (Can) plates, and incubated until resistant colonies formed (~20 GCR events per spot for the *pif1-m2 rrm3Δ*+G4 strain containing no or empty vector; Fig. 4c). As expected, Pif1 and Rrm3 both suppressed the *pif1-m2 rrm3Δ*+G4 GCR rate (0.06–0.09 events per spot), whereas helicase-dead *S. cerevisiae* Pif1 (Pif1-KA) did not
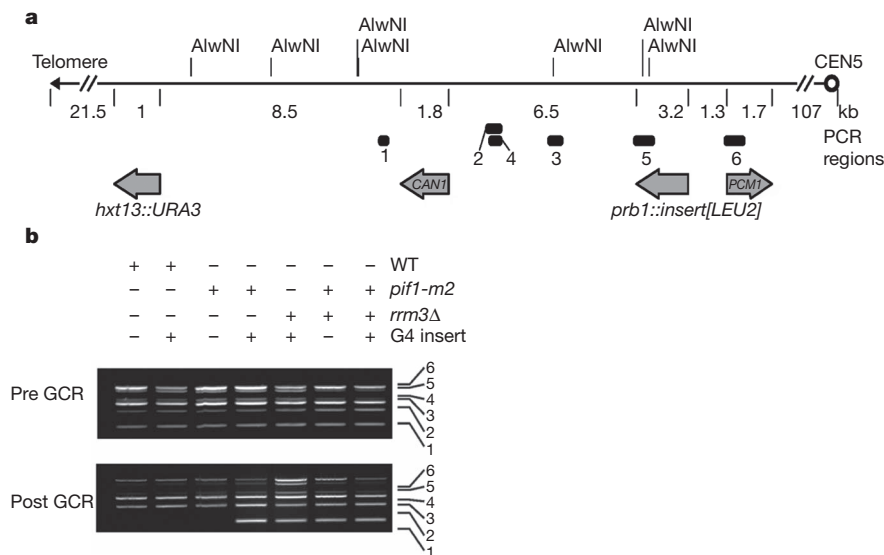
Figure 3 | Effects of G4 motifs on GCRs. a, Schematic of Chr. V-L in GCR strains. Numbered bars denote the positions of PCR products in **b**. *URA3* and *CAN1* denote counterselectable markers; *PCM1* is the most telomere-proximal essential gene. AlwNI sites are marked. **b**, Multiplex PCR analysis before (pre) and after (post) GCR events. Plus symbols denote that the strain was wild type (WT), had the indicated allele, and/or contained a G4 insert; minus symbols denote that the strain was not wild type, lacked the allele, and/or did not contain the G4 insert. The numbers indicate the region amplified as shown in **a**. The loss of band 1 in is consistent with telomere addition.

(19 events per spot; Fig. 4c). Remarkably, all seven heterologous Pif1 helicases, including human PIF1 (0.3 events per spot) and six prokaryotic/viral Pif1 helicases (0.07–1.0 events per spot), suppressed the high GCR rate.

Novel G4–induced events in *pif1* cells

We used several methods to determine whether G4 motifs affected the structure of the distal portion of chromosome V in the GCR events (Figs 3 and 4a, b and Supplementary Figs 4c–e and 5). As predicted[21], multiplex PCR (Fig. 3b) and Southern blot (Supplementary Fig. 4c–e) analyses revealed that almost all GCR events in the no-insert *pif1-m2* strain were due to telomere addition centromere-proximal to *CAN1* (52 out of 56 events). However, apparent telomere addition was rare (5 out of 27) or not detected (0 out of 28) in GCR clones from *pif1-m2*+G4 or *pif1-m2 rrm3Δ*+G4 cells, respectively (that is, *CAN1* fragment retained in multiplex PCR (Fig. 3b); new telomere bands were rare in Southern blots (Supplementary Fig. 4d, e)).

We also sequenced the 1,000-base-pair (bp) region around the G4 insert in individual GCR clones (Fig. 4a, b and Supplementary Fig. 5). There were no changes in this region in 17 out of 17 GCR clones from *sgs1Δ*+G4 cells. However, all (19 out of 19) G4 inserts were altered in *pif1-m2*+G4 GCR clones. These changes included mutations limited to the G4 motif (5% of clones); partial or complete deletion of the G4 motif and/or flanking DNA (10%); and more complex events involving deletions, mutations and insertions (85%) (Supplementary Fig. 5). A similar pattern was seen in most (82%) of the *pif1-m2 rrm3Δ* GCR events (Fig. 4 and Supplementary Fig. 5).

As expected, *URA3* and *CAN1* were lost or moved to new locations in all GCR clones from the wild-type strain containing a G4 insert (8 out of 8) and *sgs1Δ*+G4 (11 out of 11) cells. However, the positions of *URA3* and *CAN1* were unchanged in most *pif1-m2*+G4 (19 out of 27) and *pif1-m2 rrm3Δ*+G4 (27 out of 28) GCR clones. On the basis of the high mutation rate of the G4 inserts, we predicted that loss of *URA3*

and *CAN1* expression would be due to mutations in the genes. However, cloning and sequencing of *URA3* and *CAN1* from six clones each from the *pif1-m2*+G4 and *pif1-m2 rrm3Δ*+G4 strains revealed that the *URA3* and *CAN1* sequences, including ~200 bp up and downstream of the genes, were unchanged, but the left arm of chromosome V (Chr. V-L) was unstable in many of these clones. Subsequent analyses of the same clones, for example, after restreaks or growth in liquid culture, revealed that either *URA3* or both *URA3* and *CAN1* were lost
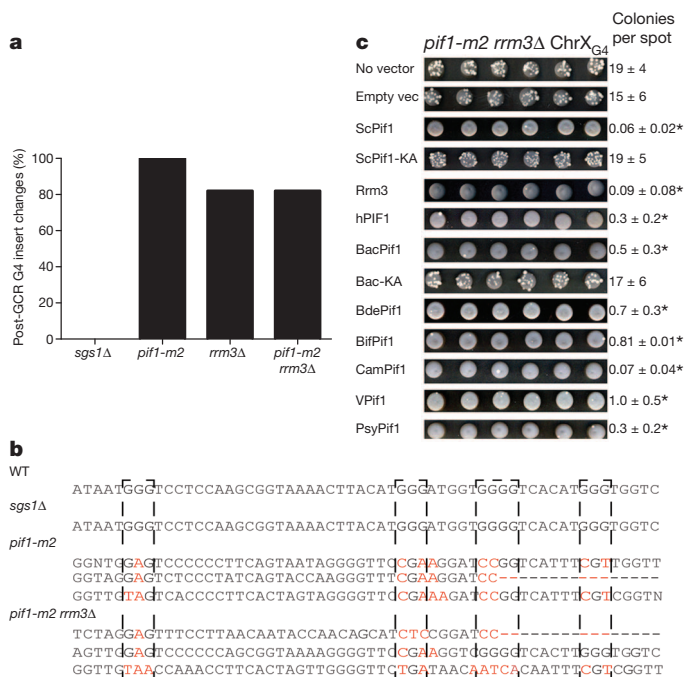
Figure 4 | Pif1 family helicases suppress G4-induced GCR events in *pif1-m2 rrm3Δ*+G4 cells. a, **b**, The G4-insert region was PCR-amplified and sequenced from 19 (*pif1-m2 rrm3Δ*) or 17 (others) GCR clones. **b**, Examples of G4 mutations. G-tracts involved in G4 formation are denoted by dashed boxes. Mutated G residues are in red; dashes denote deletions. **c**, GCR events in *pif1-m2 rrm3Δ*+G4 cells expressing the indicated helicase. Six out of one-hundred-and-fifty spots per strain are shown. GCR events are white colonies on a grey background of non-growing cells. The mean ± s.d. colonies per spot are indicated to the right. *P < 0.016 for the colonies per spot compared to the 'no vector' control as calculated by Student's *t*-test. BdePif1, *Bdellovibrio* Pif1; BifPif1, *Bifidobacterium* Pif1; CamPif1, *Campylobacter* Pif1; hPIF1, human PIF1; PsyPif1, *Psychrobacter* Pif1; VPif1, V99B1Pif1.

Table 1 | Mean GCR rates

Genotype	No insert	Chr I$_{G4}$	Chr X$_{G4}$	G-rich	Non-G-rich
		Sequence inserted at the *PRB1* locus			
Wild type	1	1.2 ± 0.5	1.4 ± 0.5	1.2 ± 0.2	1.3 ± 0.6
pif1-m2	76 ± 8	200 ± 20	210 ± 10	70 ± 30	60 ± 20
rrm3Δ	6 ± 5	12 ± 6	9 ± 4	3 ± 1	3.2 ± 0.9
sgs1Δ	16 ± 5	12 ± 8	19 ± 8	ND	ND
pif1-m2 rrm3Δ	210 ± 32	1,500 ± 500	1,900 ± 200	200 ± 10	250 ± 40
pif1-m2 sgs1Δ	20 ± 80	190 ± 35	200 ± 50	ND	ND

Data are mean ± s.d. calculated from ≥3 independent experiments and normalized to the rate ($1.5 (\pm 0.7) \times 10^{-10}$ GCR events per generation) in the wild-type strain with no insert at the *PRB1* locus. ND, not determined.

as often as 95% of the time (data not shown). However, some clones maintained wild-type *URA3* and *CAN1* genes for ≥200 generations.

Mechanism of G4–induced silencing

Given that the sequences and positions of *URA3* and *CAN1* were unchanged in the *pif1-m2*+G4, *pif1-m2 rrm3Δ*+G4 and *pif1-m2 rrm3Δ* GCR clones that retained these genes, their 5-FOA- and Can-resistant (FOA^R Can^R) phenotypes must be due to epigenetic silencing. To determine whether the silencing occurred at the transcriptional level, we used real-time quantitative PCR (qPCR) to assess the amounts of *URA3* and *CAN1* messenger RNA in four independent *pif1-m2 rrm3Δ*+G4 GCR clones and the parental pre-GCR strain. Depending on the clone, *URA3* mRNA levels ranged from 9 to 24% of the levels in the pre-GCR strain; *CAN1* mRNA ranged from 20 to 53% of the control in the same clones (Fig. 5a, b). Thus, silencing was not due to translational regulation.

In many organisms, including *S. cerevisiae* and humans[22,23], genes that are near telomeres are transcriptionally repressed (telomere position effect (TPE); reviewed in ref. 24). To determine whether the silencing observed in the *pif1-m2*+G4 GCR clones was due to TPE, we deleted *SIR2*, which encodes a protein that is required for TPE, in two independent *pif1-m2*+G4 GCR clones that retained *URA3* and *CAN1* in their original positions. Both clones lost their FOA^R Can^R phenotypes, suggesting that silencing was due to a TPE-like mechanism.

Human PIF1 inhibits telomere lengthening

Telomeres are longer in Pif1-deficient cells owing to the ability of *S. cerevisiae* Pif1 to remove telomerase from DNA ends[2,3]. To determine whether other Pif1 helicases inhibit yeast telomerase, we determined telomere length in *pif1-m2* cells expressing heterologous Pif1 helicases (Fig. 5c). Empty vector or the expression of bacterial Pif1s (*Bacteroides* Pif1 and *Campylobacter* Pif1) or *S. pombe* Pfh1 did not suppress the long telomere phenotype of *pif1-m2* cells. Indeed, telomeres were even longer in *pif1-m2* cells expressing bacterial Pif1s or Pfh1 than in *pif1-m2* cells alone. However, human PIF1 was nearly as effective as *S. cerevisiae* Pif1 in restoring telomere length to *pif1-m2* cells (Fig. 5c), even though it was expressed at much lower levels (Supplementary Fig. 6).

Discussion

S. cerevisiae Pif1 and five prokaryotic Pif1 helicases were extremely proficient at unwinding G4 structures (Fig. 2 and Supplementary Fig. 2), whereas three RecQ helicases had ~1,000-fold lower G4 unwinding activity than the Pif1 helicases (Fig. 2 and Supplementary Figs 2 and 3). Moreover, although *S. cerevisiae* Pif1 unwound G4 structures much better than Y-structures (Fig. 2a), which are themselves a preferred Pif1 substrate[13], Sgs1 and *E. coli* RecQ were more active on Y-structure than G4 substrates (Fig. 2b, c). Thus, vigorous G4 unwinding is a conserved feature of Pif1 helicases.

Suppression of G4-induced DNA instability was also conserved (Table 1). In *pif1-m2* cells, GCR rates were increased when the substrate contained a G4 motif but not when it contained other strong Pif1-binding sites; this effect is probably underestimated as *pif1-m2* is not a null allele[3]. Similarly, the human minisatellite CEB1, a tandem array of ~40 [GC]-rich repeats, increases the GCR rate in *pif1Δ* cells[25]. In contrast to no-insert *pif1-m2* cells, G4-mediated GCR events were rarely due to telomere addition (52 telomere additions per 56 GCR events in *pif1-m2* cells, versus 5 telomere additions per 27 GCR events in *pif1-m2*+G4 cells). In both *pif1-m2* and *pif1-m2 rrm3Δ* cells, the G4-induced events were usually associated with mutation of the G4 insert so that it could no longer form a G4 structure (Fig. 4a), suggesting that the process enabling cells to replicate and/or repair a G4 motif in the absence of Pif1 helicases is error-prone. Remarkably, the double drug-resistant phenotype of the G4-induced clones recovered from *pif1-m2*+G4 (75 out of 104 clones) and *pif1-m2 rrm3Δ*+G4 (50 out of 64 clones) cells was usually due to epigenetic

silencing, although the genes could be lost during further outgrowth. Although silencing of *URA3* and *CAN1* in these complex genetic-epigenetic (CGE) clones was Sir2-dependent, as is TPE[26], this silencing was considerably more effective than classical TPE. When *URA3* is immediately adjacent to the chromosome VII-L telomere, mRNA levels are ~20% of control levels, but when *URA3* is ~20 kilobases (kb) from the same telomere, FOA^R colonies are not detected (<6 × 10^{-7}). By contrast, in CGE clones, the average *URA3* mRNA level was 19% of the control, even though *URA3* was 21 kb from the telomere. The extension of silencing to more internal sites may be associated with impaired replication through a G4 structure, as changes in silencing occur in translesion polymerase-defective avian DT40 cells[27]. The unusual *URA3* and *CAN1* silencing seen here also required or was enhanced by lack of Pif1 and/or Rrm3, as it was not detected in *sgs1Δ*+G4 GCR clones (0 out of 17 clones). Furthermore, it was enhanced by a nearby G4 motif as it was not seen in GCR clones from the no-insert *pif1-m2* cells (0 out of 56 clones). The new events at both G4 motifs and structural genes in the absence of Pif1 family helicases are distinct from previously described GCR events. Thus, we term them CGE events. The epigenetic silencing of *URA3* and *CAN1* is reminiscent of the gene silencing that occurs in some human tumours that can lead to loss of heterozygosity.

Although Pif1 and Rrm3 have largely non-overlapping functions[11], they both suppressed damage at G4 motifs, as did seven out of seven heterologous Pif1 helicases (Fig. 4c). This suppression was efficient. For example, human PIF1 suppressed CGE events ~20% as effectively as *S. cerevisiae* Pif1, even though it was expressed at considerably lower
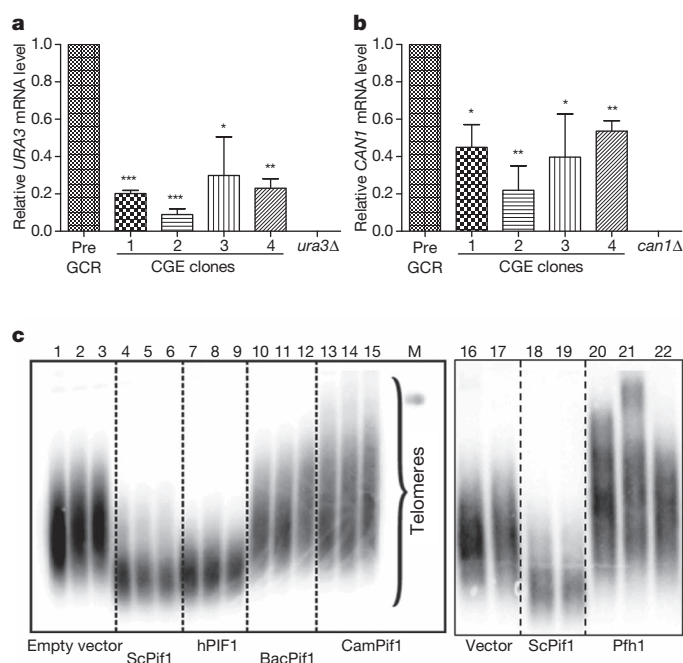

Figure 5 | Mechanism of CGE silencing and effect of Pif1 helicases on telomere length. **a**, **b**, *URA3* (**a**) and *CAN1* (**b**) mRNA levels in *pif1-m2 rrm3Δ*+G4 CGE clones and controls (pre-GCR parental strains and *ura3Δ* and *can1Δ* cells). Quantitative reverse-transcriptase PCR (qRT–PCR) was used to determine the *ACT1*, *URA3* or *CAN1* mRNA levels in the indicated strains. *URA3* and *CAN1* values were normalized to *ACT1* levels in each strain; the $2^{-\Delta\Delta Ct}$ method[38] was used to determine *URA3* and *CAN1* levels relative to parental pre-GCR cells. *P < 0.05; **P < 0.01; ***P < 0.001. **c**, Telomere blot of DNA from *pif1-m2* spore clones expressing vector only (lanes 1–3, 16, 17), ScPif1 (lanes 4–6, 18, 19), human PIF1 (7–9), two different bacterial Pif1s (10–12, 13–15), or Pfh1 (20–22). M, markers. DNA was prepared ~50, 75 and 100 generations after sporulation (first, second and third lanes in each set) or 100 generations after sporulation from two or three spore clones (lanes 16–22). See Supplementary Fig. 7 for full gel images.

levels (Supplementary Fig. 6). Thus, activity at G4 DNA by both *in vitro* and *in vivo* assays is a conserved feature of Pif1 family helicases.

S. cerevisiae Pif1 (but not Rrm3 or Pfh1) inhibits telomerase[4,28,29]. Human PIF1 (but not prokaryotic Pif1 helicases or Pfh1) restored telomere length in *pif1-m2* cells (Fig. 5c), suggesting that PIF1 is a regulator of both telomerase and G4 structures in its endogenous context. One or both of these activities might explain why mutation of human PIF1 is associated with cancer[12].

METHODS SUMMARY

Strains were YPH500 (ref. 30) derivatives (Supplementary Tables 4 and 5). Cloning oligonucleotides are listed in Supplementary Table 6. The *pif1-m2* allele was used instead of *pif1Δ* because *pif1-m2* cells are mitochondrial proficient and grow at near-wild-type rates[3]. Pif1, Sgs1 and *E. coli* RecQ were purified and assayed as described[14,16,31]. Bacterial Pif1 helicases were cloned (Supplementary Table 8), overexpressed and purified as described in the Methods. G4 motifs were from the yeast genome[32] or mouse immunoglobulin locus (TP$_{G4}$) (Supplementary Tables 2 and 3). G4 structures were formed *in vitro*[33] and 5′-end labelled with [γ-^{32}P]ATP. Protein–DNA binding was analysed by the double-filter binding method[34]. Activity of all Pif1 enzymes was measured as described previously[14]. WRN helicase assays were as described[35]. GCR assays were performed as described[36]. GCR rates were calculated using the FALCOR web server and MMS maximum likelihood method[37]. Multiplex PCR oligonucleotides are in Supplementary Table 7. G4 inserts were sequenced from genomic DNA and analysed with Biology WorkBench tools (http://workbench.sdsc.edu/). Suppression analyses of GCR phenotypes were performed in *pif1-m2 rrm3Δ*+G4 cells carrying a single-copy *TRP1*-marked plasmid with 3×Flag-tagged helicase genes under control of the *RRM3* promoter (Supplementary Table 8). Chromatin immunoprecipitation was performed as described previously[6,20]. Total RNA was isolated using a Quick-RNA MiniPrep kit (Zymo Research), reverse-translated into complementary DNA using an iScript One-Step RT–PCR kit with SYBR Green (Bio-Rad), and analysed by real-time PCR using a Bio-Rad CFX96 real-time system.

Full Methods and any associated references are available in the online version of the paper.

1. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Rev. Genet.* **13**, 770–780 (2012).
2. Boulé, J.-B., Vega, L. & Zakian, V. The Yeast Pif1p helicase removes telomerase from DNA. *Nature* **438**, 57–61 (2005).
3. Schulz, V. P. & Zakian, V. A. The *Saccharomyces PIF1* DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76**, 145–155 (1994).
4. Zhou, J.-Q., Monson, E. M., Teng, S.-C., Schulz, V. P. & Zakian, V. A. The Pif1p helicase, a catalytic inhibitor of telomerase lengthening of yeast telomeres. *Science* **289**, 771–774 (2000).
5. Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**, 1073–1076 (2001).
6. Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G–quadruplex motifs Is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678–691 (2011).
7. Lopes, J. *et al.* G–quadruplex-induced instability during leading-strand replication. *EMBO J.* **30**, 4033–4046 (2011).
8. Ivessa, A. S., Zhou, J.-Q. & Zakian, V. A. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* **100**, 479–489 (2000).
9. Ivessa, A. S. *et al.* The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol. Cell* **12**, 1525–1536 (2003).
10. Fachinetti, D. *et al.* Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol. Cell* **39**, 595–605 (2010).
11. Bochman, M. L., Sabouri, N. & Zakian, V. A. Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst.)* **9**, 237–249 (2010).
12. Chisholm, K. M. *et al.* A genomewide screen for suppressors of *Alu*-mediated rearrangements reveals a role for PIF1. *PLoS ONE* **7**, e30748 (2012).
13. Lahaye, A., Leterme, S. & Foury, F. PIF1 DNA helicase from *Saccharomyces cerevisiae*. Biochemical characterization of the enzyme. *J. Biol. Chem.* **268**, 26155–26161 (1993).
14. Boulé, J. B. & Zakian, V. A. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res.* **35**, 5809–5818 (2007).
15. Bochman, M. L., Judge, C. P. & Zakian, V. A. The Pif1 family in prokaryotes: what are our helicases doing in your bacteria? *Mol. Biol. Cell* **22**, 1955–1959 (2011).
16. Cejka, P. & Kowalczykowski, S. C. The full-length *Saccharomyces cerevisiae* Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. *J. Biol. Chem.* **285**, 8290–8301 (2010).
17. Mohaghegh, P., Karow, J. K., Brosh Jr, R. M., Bohr Jr, V. A. & Hickson, I. D. The Bloom's and Werne's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **29**, 2843–2849 (2001).
18. Schmidt, K. H., Pennaneach, V., Putnam, C. D. & Kolodner, R. D. Analysis of gross-chromosomal rearrangements in *Saccharomyces cerevisiae*. *Methods Enzymol.* **409**, 462–476 (2006).
19. Chen, C. & Kolodner, R. D. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nature Genet.* **23**, 81–85 (1999).
20. Azvolinsky, A., Giresi, P., Lieb, J. & Zakian, V. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell* **34**, 722–734 (2009).
21. Smith, S. *et al.* Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **101**, 9039–9044 (2004).
22. Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
23. Baur, J. A., Zou, Y., Shay, J. W. & Wright, W. E. Telomere position effect in human cells. *Science* **292**, 2075–2077 (2001).
24. Mondoux, M. & Zakian, V. in *Telomeres* 2nd edn (eds de Lange, T., Lundblad, V. & Blackburn, E. H.) 261–316 (CSHL, 2005).
25. Piazza, A. *et al.* Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in *Saccharomyces cerevisiae* depends on G-quadruplexes or Cdc13. *PLoS Genet.* **8**, e1003033 (2012).
26. Aparicio, O. M., Billington, B. L. & Gottschling, D. E. Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell* **66**, 1279–1287 (1991).
27. Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* **40**, 703–713 (2010).
28. Ivessa, A. S., Zhou, J.-Q., Schulz, V. P., Monson, E. M. & Zakian, V. A. *Saccharomyces* Rrm3p, a 5′ to 3′ DNA helicase that promotes replication fork progression through telomeric and sub-telomeric DNA. *Genes Dev.* **16**, 1383–1396 (2002).
29. Pinter, S. F., Aubert, S. D. & Zakian, V. A. The *Schizosaccharomyces pombe* Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Mol. Cell. Biol.* **28**, 6594–6608 (2008).
30. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
31. Harmon, F. G. & Kowalczykowski, S. C. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.* **12**, 1134–1144 (1998).
32. Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLOS Comput. Biol.* **6**, e1000861 (2010).
33. Bachrati, C. Z. & Hickson, I. D. Analysis of the DNA unwinding activity of RecQ family helicases. *Methods Enzymol.* **409**, 86–100 (2006).
34. Wong, I. & Lohman, T. M. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl Acad. Sci. USA* **90**, 5428–5432 (1993).
35. Brosh, R. M. Jr, Opresko, P. L. & Bohr, V. A. Enzymatic mechanism of the WRN helicase/nuclease. *Methods Enzymol.* **409**, 52–85 (2006).
36. Putnam, C. D. & Kolodner, R. D. Determination of gross chromosomal rearrangement rates. *Cold Spring Harbor Protoc.* **2010**, pdb.prot5492 (2010).
37. Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
38. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).

METHODS

Methods Summary. Strains were YPH500 (ref. 30) derivatives (Supplementary Tables 4 and 5). *E. coli* RecQ was purified and assayed as described[31]. G4 motifs were from the yeast genome[32] or mouse immunoglobulin locus (TP$_{G4}$) (Supplementary Tables 2 and 3). G4 structures were formed *in vitro*[33] and 5'-end labelled with [γ-^{32}P]ATP. Protein–DNA binding was analysed by the double-filter binding method[34]. WRN helicase assays were as described[35]. GCR assays were performed as described[36]. GCR rates were calculated using the FALCOR web server and MMS maximum likelihood method[37].

Yeast strains. All experiments were performed in the YPH500 background[30]. Yeast strains are listed in Supplementary Table 4, except for those used in the GCR assays (Supplementary Table 5). Gene disruptions and epitope tagging of proteins were confirmed by colony PCR, sequencing, Southern blotting and/or phenotypic analysis. The *pif1-m2* allele was introduced as previously described[3] (see Supplementary Table 6 for oligonucleotide sequences used for cloning). The carboxy terminus of Rrm3 was tagged at its endogenous locus with 13 Myc epitopes using PCR[39]. Tagged Rrm3 was expressed from its own promoter as the only version of the protein in the strain. Plasmids are listed in Supplementary Table 8. All GCR strains (Supplementary Table 5) are derivatives of YPH500 in which *HXT13* was deleted with the *Kluyveromyces lactis URA3* gene using pUG72 (ref. 40) and oligonucleotides MB262 and MB277 (see Supplementary Table 6 for the sequences of oligonucleotides used in GCR strain construction). The partial loss of nuclear function *pif1-m2* allele was used instead of *pif1Δ* because *pif1-m2* cells are mitochondrial proficient[3]. *RRM3* was deleted with *HIS3MX6* using pFA6a-His3MX6 (ref. 41) and oligonucleotides MB30 and MB31. *SGS1* was deleted with the *S. pombe his5$^+$* gene using pUG27 (ref. 40) and oligonucleotides MB32 and MB33 (Supplementary Table 6). Strains containing 'inserts' (Supplementary Table 5) were made by deleting *PRB1* with *LEU2* marked cassettes using oligonucleotides KP321f and KP321r (Supplementary Table 8). The *LEU2* marked cassettes were derived from pRS415-based plasmids[30] containing the designated inserts cloned into the XbaI and BamHI sites (Supplementary Table 8).

Biochemical methods. Full-length *S. cerevisiae* Pif1 and Sgs1 and *E. coli* RecQ were purified as previously described[14,16,31]. *In vitro* analyses of independent protein preparations revealed little to no prep-to-prep variability and that these preparations had similar biochemical activities (that is, ssDNA binding and Y-structure DNA unwinding, see below) to previously published values[14,16,31].

Bacterial Pif1 helicases were cloned as follows. E. Allen-Vercoe, C. Parker, R. Johnson and H. L. Ayala-del-Rio provided genomic DNA from *Bacteroides* sp. 2_1_16, *Campylobacter jejuni* subsp. jejuni NCTC 11168, *E. coli* phage rv5 and *Psychrobacter* sp. PRwf-1, respectively. A pUC19-based plasmid containing the gene encoding the *Bdellovibrio bacteriovorus* HD100 Pif1 helicase (Supplementary Table 8) was a gift from E. Sockett. PCR primers were designed to amplify the Pif1-like helicase genes from the above mentioned organisms (see Supplementary Table 6) with iProof HF Master Mix (BioRad). PCR products were then digested and ligated into a modified pET21d vector (pMB116; Supplementary Table 8) such that they were in-frame with an amino-terminal 4× Strep-tag II sequence and a C-terminal 6× His tag. Additional cloning details and nucleotide sequences are available on request.

Expression plasmids were transformed into Rosetta 2(DE3) pLysS cells and selected for at 37 °C on Luria–Bertani (LB) medium supplemented with 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol. Fresh transformants were used to inoculate one or more 5-ml LB cultures supplemented with antibiotics and incubated at 30 °C for ~6 h with agitation. These starter cultures were then diluted 1:100 in ZYP-5052 auto-induction medium containing 1× trace metals mix[42], 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol, and incubated at 30 °C with agitation to attenuance (*D*) at 600 nm of >3 (~18 h). Cells were collected by centrifugation for 10 min in a GS-3 rotor at 4,225*g* and 4 °C. Cell pellets were weighed and frozen at −80 °C before lysis or for long-term storage.

The cells were thawed at room temperature and resuspended in 2 ml g^{-1} cell pellet buffer A (50 mM Na-HEPES, pH 8, 10% (v/v) glycerol, 300 mM NaCl and 5 mM MgCl$_2$) supplemented with 1× protease inhibitor cocktail (Sigma), 20 μg ml^{-1} DNase I, and 2.5 μg ml^{-1} RNase A. Cells were lysed for 10 min at room temperature by adding methyl-6-*O*-(*N*-heptylcarbamoyl)-α-D-glucopyranoside (HECAMEG; Sigma) to a final concentration of 0.05% (w/v) and 1× BugBuster (Novagen) or FastBreak (Promega) with gentle stirring. Subsequent steps were performed at 4 °C.

The soluble fraction was clarified by centrifugation for 30 min in an SA-600 rotor at 20,200*g* followed by filtering the supernatant through a 0.22-μm membrane. This mixture was then loaded onto a Strep-Tactin Sepharose gravity column (IBA) pre-equilibrated in buffer A. The column was washed with four column volumes each of buffer W1 (50 mM Na-HEPES, pH 8, 10% (v/v) glycerol, 500 mM NaCl, 5 mM MgCl$_2$ and 0.05% (v/v) HECAMEG), W2 (50 mM Na-HEPES, pH 8, 10% (v/v) glycerol, 300 mM NaCl, 5 mM MgCl$_2$, 0.05% (v/v)

HECAMEG and 5 mM ATP), and W3 (50 mM Na-HEPES, pH 8, 10% (v/v) glycerol, 300 mM NaCl, 5 mM MgCl$_2$ and 0.01% (v/v) HECAMEG). Protein was eluted with three column volumes of buffer W3 supplemented with 1 mM desthiobiotin. Column fractions were examined on 10% SDS–PAGE gels run at 20 V cm^{-1} and stained with Coomassie blue R-250 (BioRad).

Peak fractions were pooled and loaded onto a His60 Ni gravity column (Clontech) pre-equilibrated in buffer W3. The column was washed with five column volumes of buffer W3 supplemented with 25 mM imidazole, and protein was eluted with five column volumes of W3 supplemented with 250 mM imidazole. Fractions were analysed by SDS–PAGE as above, and peak fractions were pooled and extensively dialysed against storage buffer (50 mM Na-HEPES, pH 8, 30% (v/v) glycerol, 50 mM NaCl, 150 mM sodium acetate, pH 7.6, 25 mM (NH$_4$)$_2$SO$_4$, 5 mM magnesium acetate, 1 mM dithiothreitol and 0.01% (v/v) HECAMEG) using 30-kilodalton (kDa) (Slide-A-Lyzer; Pierce) or 50-kDa (Tube-O-DIALYZER; G-Biosciences) molecular mass cut-off membranes. Protein concentration and purity in the final dialysates were determined on SYPRO orange (Sigma)-stained SDS–PAGE gels using known amounts of a standard protein for comparison. In all cases, protein purity was ≥95%.

For some protein preparations, the N-terminal 4× Strep-II tag was removed by PreScission Protease (GE Healthcare) digestion (2 U protease per ml protein at 4 °C overnight) before His60 column chromatography. In all cases, removal of the tag had little effect on subsequent protein purity and no effect on the *in vitro* activities examined. However, tag cleavage occasionally resulted in precipitation of a considerable portion of the protein. Thus, recombinant proteins containing both N- and C-terminal tags were used for all experiments shown.

For preparation of substrates, various *S. cerevisiae* G4 motifs were chosen from the >500 identified G4 motifs in the budding yeast genome[32] (see Supplementary Tables 2 and 3 for sequences). Oligonucleotides of G4 motifs were synthesized by IDT. The concentrations of all oligonucleotides were estimated using extinction coefficients provided by the manufacturer. G4 DNA structures were formed *in vitro* as described[33]. Formation of G4 structures was confirmed by non-denaturing PAGE. After G4 structure formation, the substrates were 5'-labelled with T4 polynucleotide kinase (NEB) and [γ-^{32}P]ATP, purified by 7% non-denaturing PAGE, and visualized using phosphoimaging.

In all biochemical assays, 100 pM radioactively labelled DNA was used, unless noted otherwise, and the reaction buffers used were previously described for *S. cerevisiae* Pif1 (ref. 14) Sgs1 (ref. 16) and *E. coli* RecQ[31]. In brief, protein–DNA binding was analysed using a BioDot SF apparatus (Bio-Rad) by the double-filter binding method[34]. Reactions were set up as for helicase assays, but ATP was omitted. The reactions were incubated on ice for 30 min, filtered through the membranes, and then the membranes were washed with additional reaction buffer. The membranes were dried and analysed by phosphoimaging. Pif1, Sgs1 and RecQ helicase activity assays were performed essentially as described previously[14,16,31]. The helicase activity of non-yeast Pif1 enzymes was also measured as described for *S. cerevisiae* Pif1 previously[14]. WRN helicase assays were performed as described[35]. For protein titrations, reactions were incubated for 30 min at helicase at 25 °C Pif1), 30 °C (Sgs1), or 37 °C (RecQ, WRN, and non-yeast Pif1). In time-course experiments, 100 pM ScPif1, 10 nM BacPif1, 10 nM Sgs1 or 50 nM EcRecQ was added to the reaction; 100 pM Sgs1 or EcRecQ displayed only basal levels of unwinding in our G4 unwinding assays. For single-cycle conditions, we used a 500× excess of either G4 DNA or ssDNA as a protein trap. The excess trap DNA was added together with ATP to start the reactions.

The data were fit with rectangular hyperbolic curves using GraphPad Prism 5 and equation (1):

$$Y = \frac{B_{max} \cdot X}{K + X} \qquad (1)$$

in which *X* is the helicase concentration or time (as indicated), *Y* is either DNA binding or unwinding (as indicated), B_{max} is the maximum level of binding or unwinding (as indicated), and *K* is the midpoint of the curve. When a log$_{10}$-scale *x* axis is used, the hyperbolic curve assumes a sigmoidal shape.

GCR assays. GCR assays were cloned and performed essentially as described[36] (primer sequences for cloning are listed in Supplementary Table 6). In brief, sets of five or more 5-ml cultures of each *S. cerevisiae* GCR strain (Supplementary Table 5) were grown to saturation in YPD medium at 30 °C for 36–48 h. A final dilution of 1×10^{-7} of each culture was plated on YPD and incubated at room temperature for 4 days to determine the viable cell count. Cells (1.5 or 2 ml) from each culture were pelleted, resuspended in sterile water, plated on drop-out medium lacking uracil and arginine (US Biologicals) supplemented with 1 g l^{-1} 5-FOA and 60 mg l^{-1} canavanine sulphate (FOA+Can), and incubated at 30 °C for ~4 days. GCR rates were calculated using the FALCOR web server and MMS maximum likelihood method[37] and normalized to wild-type rate of 10^{-10} GCR events per cell division. The rates presented in Table 1 are the mean ± s.d. of

≥3 experiments per strain. We define GCR clones as colonies that grew on the FOA+Can plates. Such FOA- and Can-resistant (FOA^R Can^R) clones were selected for post-GCR analyses (below).

G4 inserts were sequenced from samples of genomic DNA from FOA^R Can^R clones prepared using a MasterPure Yeast DNA Purification kit (Epicentre) following the manufacturer's instructions. The oligonucleotides used for sequencing the Watson and Crick strands are: MB540, 5′-CAATAGGCCGAAA TCGGCAAAATCCC-3′, and MB537, 5′-CTCCTATGTTGTGTGGAATTG TGAGCGG-3′, respectively, which amplified a ~1-kb region containing the inserts. The results were analysed using the Biology WorkBench tools (http://workbench.sdsc.edu/) and classified into four different categories, as indicated in Supplementary Fig. 5: (1) no change, the G4 inserts were identical to the starting strain in the GCR clones; (2) recombination, the G4 insert was replaced with a partially homologous sequence from chromsome VII; (3) mutation and deletion, several of the guanines responsible for forming the G4 structures were either mutated to other residues or deleted; and (4) mutation/deletion/insertion, the G4 inserts experienced a variety of events, including substitution mutations and short deletions and/or insertions.

Suppression analyses of the *pif1-m2 rrm3Δ*+G4 GCR phenotype were performed by transforming strain KP326 (Supplementary Table 5) with *TRP1*-marked plasmids containing C-terminally 3× Flag-tagged helicase genes expressed under control of the *RRM3* promoter (Supplementary Table 8). Three independent colonies were used to inoculate 5 ml synthetic complete medium lacking tryptophan (SC−Trp) and grown on a roller drum for ~48 h at 30 °C. The $D_{660\,nm}$ for each culture was determined with a spectrophotometer, and the cells were pelleted by centrifugation and resuspended to $D_{660\,nm} = 10$ in sterile H$_2$O. Then, a repeat pipettor was used to spot 10-µl samples of each strain 50 times on a FOA+Can plate, and the plates were incubated at 30 °C for 4 days. This process was repeated ≥3 times for each strain. When colonies appeared on the FOA+Can plates, the number of colonies per 10-µl spot was counted, and the average number of colonies in the 50 spots per plate was calculated. The mean (± s.d.) of these values from the ≥3 plates per strain was determined and reported in the right column of Fig. 4c.

Western and Southern blotting. Cell extracts for western blotting were prepared as described previously[43]. In brief, cells were grown overnight in SC−Trp liquid medium at 30 °C with aeration. Then, 1 ml of $D_{600\,nm} = 2.5$ cells was collected, resuspended in 200 µl 0.1 N NaOH, incubated at room temperature for 5 min, pelleted, resuspended in 50 µl SDS–PAGE sample buffer, boiled for 3 min, and pelleted again. Subsequently, 6 µl of the supernatants was loaded onto an 8% (37.5:1 polyacrylamide:bis-acrylamide) SDS–PAGE gel and run at 20 V cm^{-1}. The proteins were transferred to a nitrocellulose membrane at 4 °C and blocked with 5% non-fat milk in TBST at room temperature using standard protocols. The blot was probed with a monoclonal anti-Flag antibody (F1804, Sigma) and visualized with a horseradish peroxidase (HRP)-conjugated secondary antibody and ECL detection reagents (GE Healthcare). The blot was then stripped and reprobed with an anti-tubulin antibody (G094, ABM) to verify equivalent protein loading.

For telomere blots, cells containing a plasmid with either a helicase gene (pMB267, 270, 282 and 292) or no insert (empty vector control; pMB13) (Supplementary Table 8) were transformed into the *PIF1 pif1-m2* diploid (KP448). The diploids were sporulated, and the *pif1-m2* spore clones carrying the plasmids were recovered. Genomic DNA was isolated from cells from restreaks 1–6 after sporulation (corresponding to approximately 25 generations per restreak) using a MasterPure Yeast DNA Isolation kit (Epicentre Biotechnologies), digested overnight with PstI and XhoI (NEB), and telomere length was determined by Southern analysis as described previously[44]. Results from DNA from restreaks 1–3 are shown in Fig. 5d, but the same results were also observed after 6 restreaks.

When colonies arose after GCR events, single colonies were restreaked onto FOA+Can plates, and genomic DNA was isolated from the survivors using a MasterPure Yeast DNA Isolation kit. The DNA was digested overnight at 37 °C with AlwN1 and run on a 0.7% agarose gel. The DNA was transferred to Hybond membranes (GE Life Sciences) by capillary action and blotted using the 400-bp

CIN8 PCR product as a probe. The *CIN8* PCR probe hybridizes to both sides of the AlwNI cut site in *CIN8* (see Fig. 3a), generating bands of 3.2 kb and 6.9 kb in the original strain (PRE; Fig. 4d–f). In most of the FOA^R Can^R strains (Fig. 3), the centromere-proximal 3.2-kb band is retained, indicating that any sequence loss does not extend to this region. Retention of the 6.9-kb band in FOA^R Can^R strains indicates that *CAN1* function has been lost in the absence of an overt DNA rearrangement. In contrast, bands of <6.9 kb are indicative of a GCR event, with fuzzy bands showing sites of telomere addition. In theory, the two bands in Fig. 4d that are between the 6.9 and 3.2 kb bands (in the fifth and seventh lanes from the left) could be either telomere additions of rearrangements. However, based on sequencing results and multiplex PCR, such bands are not likely to be telomere additions.

Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) of asynchronous yeast cells growing in rich medium was performed as described[6,20] and analysed using an iCycleriQ Real-Time PCR detection system (Bio-Rad Laboratories). Rrm3 was C-terminally tagged with 13 Myc epitopes[39]. An anti-Myc monoclonal antibody (Clontech 631206) was used as the anti-serum in ChIP. The amount of DNA in the immunoprecipitate was normalized to the amount in input samples. The ChIP experiment was analysed by qPCR in duplicate or triplicate to obtain an average value for each sample. The ChIP experiment was repeated ≥3 times at each locus. For each qPCR experiment, the amount of signal in the Rrm3 immunoprecipitate was normalized to input and to the immunoprecipitated signal from *ARO1*, a sequence that contains no candidate G4 DNA motif and that has low Rrm3 association[20].

Multiplex PCR. In brief, genomic DNA isolated from *S. cerevisiae* strains before and after GCR events was analysed by multiplex PCR using the primer pairs in Supplementary Table 7 and the following cycling parameters: initial denaturation for 5 min at 95 °C, 35 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 45 s, and a final extension at 72 °C for 10 min. PCR products (10 µl per reaction) were run at 90 V on 2.5% agarose gels containing ethidium bromide and visualized by ultraviolet transillumination (for primer sequences, see Supplementary Table 7).

qPCR. The indicated strains were grown in FOA+Can liquid media for 12 h and then transferred to YEPD for 12 h until reaching a concentration of $D_{600\,nm} = 0.5$, and total RNA was isolated using a Quick-RNA MiniPrep kit, including the DNase I treatment, as described by the manufacturer (Zymo Research). cDNA was synthesized from 200 ng DNase I-treated RNA using an iScript One-Step RT–PCR kit with SYBR Green (Bio-Rad) and analysed qPCR using a Bio-Rad CFX96 real-time system. The following primers were used: *URA3* cDNA, 5′-GTTCGACTGATGAGCTATTGAAACT-3′ and 5′-CGACAGTACCCTCATA ACTGAAATC-3′; *CAN1* cDNA 5′-AATATACATCGGGCGTTTAC-3′ and 5′-TCAGCAAGCATCAATAATCC-3′; *ACT1* cDNA, 5′-GTAACATCGTTATGT CCGGTGGTAC-3′ and 5′-CCAAGATAGAACCACCAATCCAGAC-3′. The cycling parameters were: 50 °C for 10 min, 95 °C for 5 min, and 40 cycles of 95 °C for 10 s followed by 57 °C (*ACT1*), 50 °C (*CAN1*), or 55 °C (*URA3*) for 30 s. The data were analysed by the $2^{-\Delta\Delta Ct}$ method[45].

39. Azvolinsky, A., Dunaway, S., Torres, J., Bessler, J. & Zakian, V. A. The *S. cerevisiae* Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev.* **20,** 3104–3116 (2006).
40. Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D. & Hegemann, J. H. A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30,** e23 (2002).
41. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae. Yeast* **14,** 953–961 (1998).
42. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
43. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16,** 857–860 (2000).
44. Runge, K. W. & Zakian, V. A. Introduction of extra telomeric DNA sequences into *Saccharomyces cerevisiae* results in telomere elongation. *Mol. Cell. Biol.* **9,** 1488–1497 (1989).
45. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25,** 402–408 (2001).

LETTER

Shear–driven dynamo waves at high magnetic Reynolds number

S. M. Tobias[1] & F. Cattaneo[2]

Astrophysical magnetic fields often display remarkable organization, despite being generated by dynamo action driven by turbulent flows at high conductivity[1,2]. An example is the eleven-year solar cycle, which shows spatial coherence over the entire solar surface[3–5]. The difficulty in understanding the emergence of this large-scale organization is that whereas at low conductivity (measured by the magnetic Reynolds number, Rm) dynamo fields are well organized, at high Rm their structure is dominated by rapidly varying small-scale fluctuations. This arises because the smallest scales have the highest rate of strain, and can amplify magnetic field most efficiently. Therefore most of the effort to find flows whose large-scale dynamo properties persist at high Rm has been frustrated. Here we report high-resolution simulations of a dynamo that can generate organized fields at high Rm; indeed, the generation mechanism, which involves the interaction between helical flows and shear, only becomes effective at large Rm. The shear does not enhance generation at large scales, as is commonly thought; instead it reduces generation at small scales. The solution consists of propagating dynamo waves, whose existence was postulated more than 60 years ago[6] and which have since been used to model the solar cycle[7].

The evolution of a magnetic field B in a fluid of electrical conductivity σ is governed by the (dimensionless) induction equation:

$$\frac{\partial B}{\partial t} = \nabla \times (u \times B) + \frac{1}{\mathrm{Rm}} \nabla^2 B \qquad (1)$$

where u is the velocity, and $\mathrm{Rm} = U\lambda\mu_0\sigma$ is a dimensionless measure of the conductivity. Here U and λ are a characteristic velocity amplitude and length scale, respectively, and Rm is typically huge in astrophysics[1]. The kinematic dynamo problem consists of finding velocities u for which equation (1) has exponentially growing solutions[2]. For the corresponding large-scale dynamo problem, these solutions must also be characterized by a spatial scale of $L > \lambda$ and possibly evolve on timescales long compared with the turnover time of the velocity.

One way to identify how organization may arise is to average equation (1), which gives the evolution of the large-scale part of the field. The resulting mean-field equation contains transport coefficients describing the average induction and diffusion, which are themselves determined by Rm and the flow. The average induction can be non-zero only if the underlying turbulence lacks reflectional symmetry[6,8]. Exponentially growing solutions of the mean-field equations can be found provided that the average induction can overcome the average diffusion. For this approach to be useful, the solutions of the averaged equations must correspond to the averages of the solutions of the true equation (1). At small Rm, this is ensured as then diffusion can control the growth of fields on small scales. However at high Rm, problems arise. In a turbulent flow, consisting of a range of eddies, the rate of strain increases with decreasing scale, so the smallest scales are the most efficient at amplifying the field. The resulting magnetic field is dominated by small-scale fluctuations with little organization on the large scales.

One possible way to remedy this is to introduce a large-scale flow, typically a shear, in such a way as to enhance the mean induction. This procedure appears to work at small to moderate Rm (refs 9–12), but not at large Rm because ultimately the small-scale generation overwhelms everything. Here we take a radically different approach. Instead of trying to boost the large-scale dynamo properties, we use the shear to reduce the efficiency of the small scales.

Our dynamo model is based on a velocity made of two parts: a large-scale shear of amplitude V_0, which varies on the scale of the domain, and a superposition of small-scale cellular flows[13,14] whose characteristic length scale is about ten times smaller and whose reflectional symmetry can be controlled (see Supplementary Information). The flow has the following important properties. It is 2.5-dimensional, in the sense that it has all three components—but only depends on the coordinates x and y, say. The resulting dynamo problem is then separable, with solutions of the form $B = \hat{B}(x,y,t) \exp ik_z z$ (here t is time, and k_z the vertical wavenumber). This reduces the problem from three to two dimensions, and therefore makes it possible to reach high Rm (ref. 13). By analogy with real turbulence, we have constructed the flow so that the velocity amplitude decreases with scale, while the shear rate and turnover frequency increase. Furthermore, the velocity has a scale-dependent renewal time, comparable with the local turnover time. The
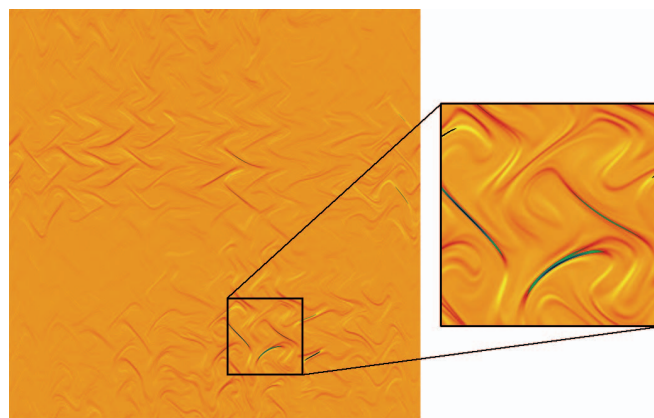
Figure 1 | Small-scale dynamo solution. Colour density plot of the amplitude of the x component of the magnetic field density, B_x, in the plane $z = 0$ for a domain of horizontal extent $2\pi \times 2\pi$. This figure has been normalized between plus and minus 1. For this and all subsequent calculations, the Rm associated with the largest-scale cellular flows (with wavenumber $k = 8$) is of the order of 2,500. Adequate resolution for this calculation was provided by a grid with 2,048 collocation points in each direction. In the z direction, the dependence of the field is sinusoidal with characteristic wavenumber k_z. Here $k_z = 2.5$; this is close to the value for which the growth rate is maximal for this flow, corresponding to magnetic structures that are roughly twice as tall as they are wide. We should note that at such high values of Rm, the maximum of the growth-rate curve is fairly broad in k_z, that is, the growth rate is locally insensitive to k_z. Indeed, changing k_z by a factor of two leads to no change in the growth rate to within the accuracy of the measurement of the growth rate. Inset, magnification of the region boxed in the main panel, displaying the fine details of the small-scale magnetic field.

[1]Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK. [2]The Computation Institute and Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA.
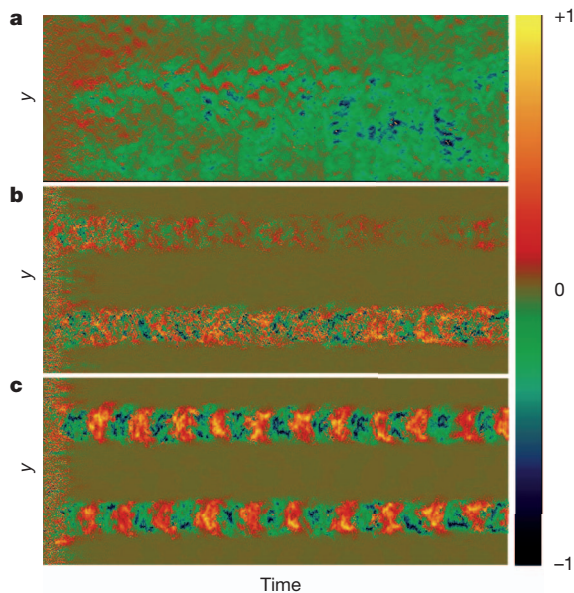
Figure 2 | Diagrams of activity versus time. Colour density plots of B_x averaged in x as a function of time (horizontal axis) and y (vertical axis). Each panel is obtained by averaging the magnetic field component in the direction of the shear ($B_x(x, y, 0, t)$) and plotting the results suitably normalized to remove the exponential growth. **a**, The solution shown in Fig. 1, with $V_0 = 0$ and a helical flow only. Clearly there is no evidence of any type of temporal or spatial organization—everything varies on the short turnover time. **b**, Solution for a shear with $V_0 = 5$ and a flow with no net helicity. Magnetic activity is confined in two bands localized near the regions where the velocity gradients are strongest ($y = \pi/2$ and $y = 3\pi/2$). Within each band there is not much temporal coherence or generation of shear-aligned flux. **c**, Solution for $V_0 = 5$ and a strongly helical flow. Again magnetic activity is confined to the regions of strong shear, but now large-scale spatio-temporal organization is apparent. In this case, in each wave there is substantial shear-aligned flux.

individual cellular flows are the circularly polarized velocity introduced in the study of small-scale dynamo action at high Rm (ref. 13). These are known to be fast dynamos, in the sense that the growth rate (s) is comparable with the turnover frequency and remains independent of Rm as Rm $\rightarrow \infty$. Furthermore, these flows are 'quick dynamos'[14] in that they reach their asymptotic growth rate quickly (that is, even at moderate Rm).

The typical solution for a representative case with no shear ($V_0 = 0$) is given in Fig. 1. The magnetic field grows exponentially, with a growth time comparable with that of the turnover time of the smallest eddies, and is concentrated at small scales, with filamentary structures of length comparable to that of the small-scale velocity and much smaller width controlled by diffusion[15]. For our choice of Rm, the

growth rate has reached its asymptotic value and the overall magnetic pattern changes on the turnover time.

We investigate the effects of a shear, and discuss two representative cases; one in which the velocity has no net helicity (that is, it is reflectionally symmetric), and the other one in which it is strongly helical. The key result is given in Fig. 2, which shows the magnetic activity as a function of latitude (y) and time. Figure 2a is for the case with no shear, and the magnetic field is all at small scales. Figure 2b shows the case with shear but no net helicity. The effect has been to concentrate the field in two bands in the regions of high shear. However, within each band the magnetic field is highly fluctuating both in space and time. In Fig. 2c there is strong helicity and shear; magnetic structures are generated that are organized both in space and time, and the solution shows a cyclic behaviour with a well-defined period of several turnover times.

It is important to determine why the presence of the shear brings about such a dramatic change in the helical case. We argue that the interaction of the shear and helical flow is not to boost inductive processes on the large scale, but rather to limit the growth of the fluctuations, so that the large-scale structure can manifest itself. Supporting evidence comes from the behaviour of the average induction and growth rate for the helical case as a function of shear amplitude (Fig. 3). Clearly the shear does not enhance the average (shear-aligned) induction, in fact quite the opposite[16]. So why then does the large-scale field manifest itself? We believe that the crucial role of the shear is to reduce the efficiency of the fast-dynamo processes at small scales, supported by the fact that the dynamo growth rate actually decreases with increasing shear amplitude. This reduction could be achieved, for instance, by decreasing the stretching properties of the flow as measured by the Lyapunov exponents, say, or increasing the efficiency of the diffusion by bringing oppositely directed field lines closer together[17].

We further investigate the spatio-temporal structure of the emergent organized field. The solution consists of two magnetic structures located near the extrema of the shear. Each varies in the z direction as $\exp(ik_z z)$, with a phase $\phi(t)$, which is plotted in Fig. 4a. This shows that the solution corresponds to two waves counter-propagating in the z direction. We identify these large-scale structures as Parker dynamo waves[6] because they propagate perpendicular both to the shear and to the direction of the variation of the shear, the propagation direction changes sign with the shear, and finally their period decreases with shear amplitude (Fig. 4b).

The emergence of dynamo waves at such high Rm ($\sim$2,500 for the cellular flow) is remarkable, leading one to ask how robust this mechanism could be. Because the underlying flow has the 'quick' property, we believe that this mechanism will remain effective even at higher Rm. The other important question is whether this mechanism will remain operative nonlinearly, when the generated field acts back on the flow and the velocity takes the form of a turbulent cascade. The
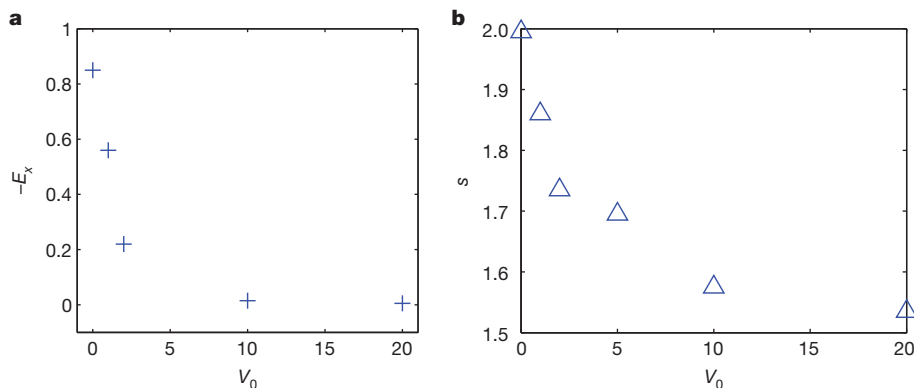
Figure 3 | Effect of the shear. **a**, Absolute value of the x-component of the electromotive force as a function of V_0, showing that the shear decreases the importance of the average induction. This is calculated by imposing a mean magnetic field in the x direction[18]; this quantity encodes the efficiency of the average induction processes and contributes to the growth of the large-scale field. **b**, Growth-rate s as a function of V_0, showing that the shear reduces the efficiency of small-scale dynamo action. The growth rate is calculated by fitting a straight line to the logarithm of the total magnetic energy as a function of time.
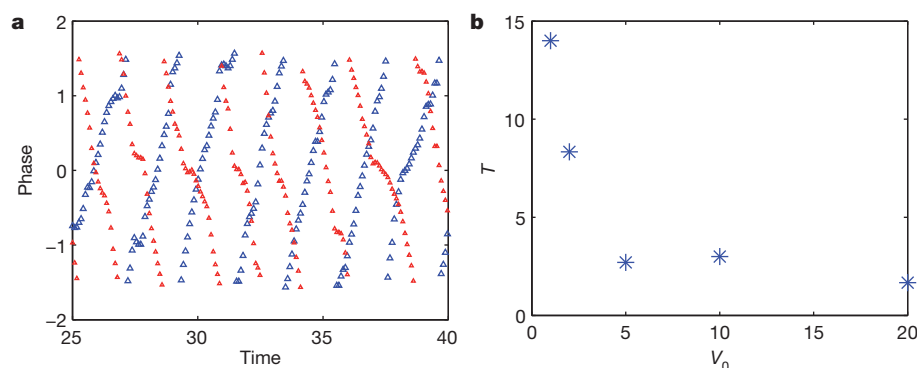
Figure 4 | Evidence for dynamo waves. a, Principal value of the phase as a function of time for the two separate magnetic structures shown separately as blue and red triangles. These indicate that the oscillations are in the form of counter-propagating waves. **b**, Period of the dynamo waves (*T*) as a function of shear amplitude. The period decreases with increasing shear amplitude.

smallest scales will be unaffected by the shear (and indeed rotation), and will saturate very quickly and at low amplitude. We conjecture that the dynamo at larger scales, associated with the mechanism described here, will continue on this background of small-scale fluctuations to generate organized fields. This mechanism is probably astrophysically relevant in any situation where there is rotation, leading both to strong shear and helical turbulence. In particular, it is most natural to apply some of these ideas to the shear layer at the base of the solar convection zone known as the tachocline, which is believed to be the seat of the solar dynamo[7].

1. Parker, E. N. *Cosmical Magnetic Fields: Their Origin and Their Activity* (Clarendon, 1979).
2. Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge Univ. Press, 1978).
3. Mestel, L. *Stellar Magnetism* (Clarendon, 1999).
4. Tobias, S. M. The solar dynamo. *Phil. Trans. R. Soc. Lond. A* **360,** 2741–2756 (2002).
5. Stix, M. *The Sun: An Introduction* (Springer, 2004).
6. Parker, E. N. Hydromagnetic dynamo models. *Astrophys. J.* **122,** 293–314 (1955).
7. Tobias, S. & Weiss, N. in *The Solar Tachocline* (eds Hughes, D. W., Rosner, R. & Weiss, N. O.) 319–350 (Cambridge University Press, 2007).
8. Steenbeck, M., Krause, F. & Rädler, K.-H. Berechnung der mittleren LORENTZ-Feldstärke für ein elektrisch leitendes Medium in turbulenter, durch CORIOLIS-Kräfte beeinflußter Bewegung. *Z. Naturforsch. A* **21,** 369–376 (1966).
9. Yousef, T. A. *et al.* Generation of magnetic field by combined action of turbulence and shear. *Phys. Rev. Lett.* **100,** 184501 (2008).
10. Käpylä, P. J. & Brandenburg, A. Turbulent dynamos with shear and fractional helicity. *Astrophys. J.* **699,** 1059–1066 (2009).
11. Sridhar, S. & Singh, N. K. The shear dynamo problem for small magnetic Reynolds numbers. *J. Fluid Mech.* **664,** 265–285 (2010).
12. Hughes, D. W. & Proctor, M. R. E. The effect of velocity shear on dynamo action due to rotating convection. *J. Fluid Mech.* **717,** 395–416 (2013).
13. Galloway, D. J. & Proctor, M. R. E. Numerical calculations of fast dynamos in smooth velocity fields with realistic diffusion. *Nature* **356,** 691–693 (1992).
14. Tobias, S. M. & Cattaneo, F. Dynamo action in complex flows: the quick and the fast. *J. Fluid Mech.* **601,** 101–122 (2008).
15. Finn, J. M. & Ott, E. Chaotic flows and fast magnetic dynamos. *Phys. Fluids* **31,** 2992–3011 (1988).
16. Courvoisier, A. & Kim, E.-J. Kinematic α effect in the presence of a large-scale motion. *Phys. Rev. E* **80,** 046308 (2009).
17. Cattaneo, F. & Tobias, S. M. Interaction between dynamos at different scales. *Phys. Fluids* **17,** 127105 (2005).
18. Courvoisier, A., Hughes, D. W. & Tobias, S. M. α Effect in a family of chaotic flows. *Phys. Rev. Lett.* **96,** 034503 (2006).

LETTER

Flux-freezing breakdown in high-conductivity magnetohydrodynamic turbulence

Gregory Eyink[1,2,3,4], Ethan Vishniac[5], Cristian Lalescu[1], Hussein Aluie[1,6], Kalin Kanov[7], Kai Bürger[8], Randal Burns[4,7], Charles Meneveau[3,4] & Alexander Szalay[2,4]

The idea of 'frozen-in' magnetic field lines for ideal plasmas[1] is useful to explain diverse astrophysical phenomena[2], for example the shedding of excess angular momentum from protostars by twisting of field lines frozen into the interstellar medium. Frozen-in field lines, however, preclude the rapid changes in magnetic topology observed at high conductivities, as in solar flares[2,3]. Microphysical plasma processes are a proposed explanation of the observed high rates[4–6], but it is an open question whether such processes can rapidly reconnect astrophysical flux structures much greater in extent than several thousand ion gyroradii. An alternative explanation[7,8] is that turbulent Richardson advection[9] brings field lines implosively together from distances far apart to separations of the order of gyroradii. Here we report an analysis of a simulation of magnetohydrodynamic turbulence at high conductivity that exhibits Richardson dispersion. This effect of advection in rough velocity fields, which appear non-differentiable in space, leads to line motions that are completely indeterministic or 'spontaneously stochastic', as predicted in analytical studies[10–13]. The turbulent breakdown of standard flux freezing at scales greater than the ion gyroradius can explain fast reconnection of very large-scale flux structures, both observed (solar flares and coronal mass ejections) and predicted (the inner heliosheath, accretion disks, γ-ray bursts and so on). For laminar plasma flows with smooth velocity fields or for low turbulence intensity, stochastic flux freezing reduces to the usual frozen-in condition.

Spontaneous stochasticity[10–12] can be understood using a simple model equation, $dr/dt = \delta u_r = Ar^h$, for the evolution of the separation, $r = x_2 - x_1$, of a pair of advected objects at respective positions x_1 and x_2. Here $\delta u_r = u(x_2) - u(x_1)$ is the relative velocity, and power-law scaling with an exponent satisfying $0 < h < 1$ implies a 'rough' velocity field with δu_r vanishing more slowly than linearly for decreasing r. This corresponds to a turbulent energy spectrum, $E(k) \propto k^{-1-2h}$ (ref. 14, section 4.5). An example is Kolmogorov scaling with $h = 1/3$, which is observed in the solar wind above the ion gyroradius[15] and in the interstellar medium[16]. Solving the model equation gives $r^2 = [r_0^{1-h} + (1-h)At]^{2/(1-h)}$, where r_0 is the pair separation at time zero. The long-time limit for $h = 1/3$ yields the $r^2 \propto t^3$ law of Richardson[9]. Notably, however, this law is also obtained in the limit $r_0 \to 0$, that is, when the two particles start or end at the same point. This violates the expectation of determinacy in classical dynamics. There is no contradiction, because the usual theorems on unique solutions of initial-value problems for differential equations assume that $h \geq 1$, whereas we consider flows where $0 < h < 1$.

The phenomena observed in the previous simple model arise in various physical limits in real turbulent flows. For example, a standard way[10–12] to model the effect of molecular diffusivity, λ, on a fluid tracer is to add a Brownian motion $W(t)$ such that the tracer's position, $x(t)$, is governed by the stochastic equation $dx = u(x,t)dt + \sqrt{2\lambda}dW(t)$ with initial condition $x(t_0) = x_0$. If the viscosity, v, of a smooth, laminar flow
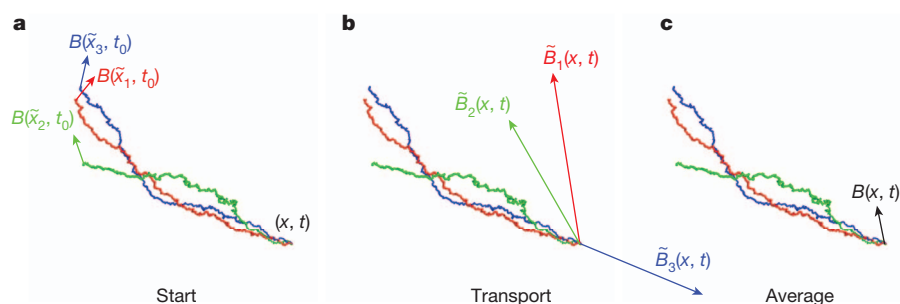
Figure 1 | Stochastic flux-freezing for resistive MHD. Magnetic fields that solve the resistive induction equation are frozen-in to stochastic trajectories, $\tilde{x}(\tau)$, that satisfy $d\tilde{x} = u(\tilde{x},\tau)d\tau + \sqrt{2\lambda}d\tilde{W}(\tau)$. The field vectors, which start at $\tau = t_0$, are transported to $\tau = t$ along these random trajectories as 'virtual' magnetic fields, stretched and rotated by the plasma flow. The ensemble of virtual field vectors that arrive at the same final point, (x,t), must be averaged (that is, 'glued' together by resistivity) to give the physical magnetic field $B(x,t)$ at that point[17]. To find the stochastic trajectories that arrive at (x,t), it is most efficient to integrate backwards in time. We use the velocity field, u, from the database to solve $d\tilde{x} = u(\tilde{x},\tau)d\tau + \sqrt{2\lambda}d\tilde{W}(\tau)$, starting with $\tilde{x}(t) = x$ and integrating back to the initial time, t_0. This step generates the N-sample ensemble of stochastic trajectories $\tilde{x}_n(\tau)$, $n = 1,\ldots,N$ (**a**). We then retrieve from the database the magnetic field, B, at the random locations $\tilde{x}_n(t_0)$, $n = 1,\ldots,N$, and transport these vectors back along the stochastic trajectories to the point (x,t) using the frozen-in equation $d\tilde{B}/d\tau = \tilde{B} \cdot \nabla u$ (**b**). This requires retrieving the velocity gradient, ∇u, from the database at each forward time step. Finally, the virtual magnetic field vectors, $\tilde{B}_n(x,t)$, $n = 1,\ldots,N$, are averaged over the N samples to recover the archived physical magnetic field, $B(x,t)$, by choosing N to be sufficiently large (**c**). For more discussion, see Supplementary Information.

[1]Department of Applied Mathematics & Statistics, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [2]Department of Physics & Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [3]Department of Mechanical Engineering, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [4]Institute for Data Intensive Engineering & Science, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [5]Department of Physics and Engineering Physics, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5E2, Canada. [6]Los Alamos National Laboratory, T-Division and Center for Nonlinear Studies, Los Alamos, New Mexico 87545, USA. [7]Department of Computer Science, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [8]Fakultät für Informatik, Technische Universität München, Boltzmannstraße 3, D-85748 Garching bei München, Germany.

is held fixed as $\lambda \rightarrow 0$, then the noise term vanishes and the solution becomes deterministic. However, if instead the ratio ν/λ (the Prandtl number) is fixed by microscopic physics as $\nu, \lambda \rightarrow 0$, which corresponds physically to the Kolmogorov spectral range becoming ever longer, then the noise term again vanishes but particle trajectories now have mean squared separations $\langle r^2 \rangle \propto t^3$ and remain indeterminate and random. This spontaneous stochasticity[10–12] should be present for any sufficiently long power-law range extending down to dissipation scales.

Magnetic fields in resistive magnetohydrodynamics (MHD) are an example of such a diffusing tracer, satisfying the induction equation $\partial B/\partial t = \nabla \times (u \times B) + \lambda \nabla^2 B$, where $\lambda = c^2/4\pi\sigma$ is the magnetic diffusivity (σ, conductivity). Magnetic field lines can be regarded as being frozen in along stochastic trajectories[17] (Fig. 1). For a smooth laminar flow, the trajectories wander diffusively only a distance $\sim \sqrt{\lambda t}$ in time

$$\frac{|E_{\text{Ohm}}|}{E'_{\text{mot}}}$$
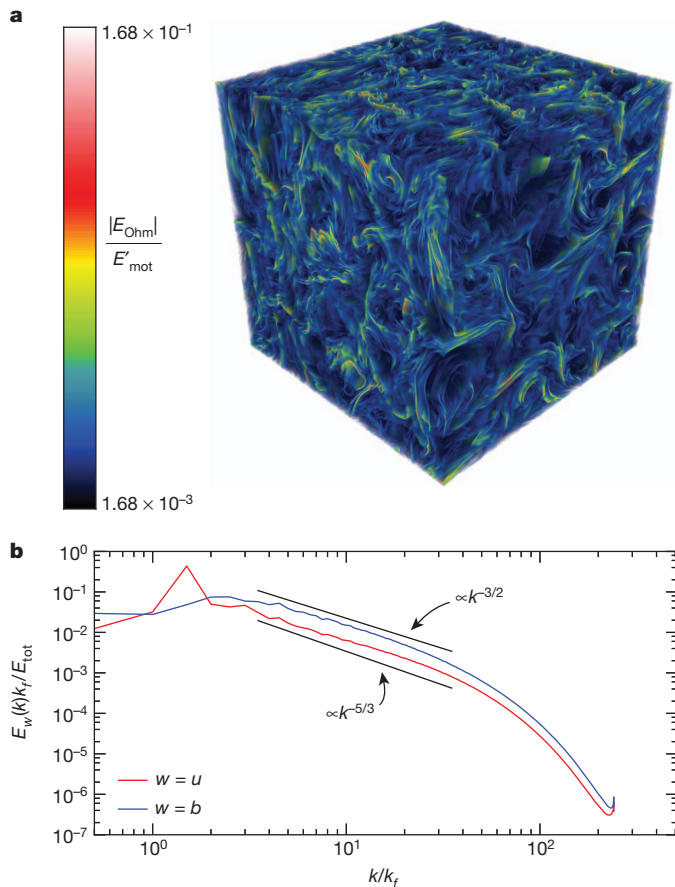
Figure 2 | The archived MHD turbulence flow. The data were generated by a simulation of the incompressible MHD equations on a space grid of $1{,}024^3$ points. The magnetic Prandtl number is 1, with $\nu = \lambda = 1.1 \times 10^{-4}$. The MHD momentum equation was driven by a body force F of Taylor–Green form at wavenumbers with magnitude $k_f = 2k_0$, twice the lowest value. It is important that no external source of electric field was added to the induction equation, so the stirring to generate turbulence does not directly break standard flux freezing. Once the simulation reached a statistical steady state, 1,024 time slices of data on the $1{,}024^3$ grid were stored. The 56 terabytes of data are public and can be accessed remotely through a modern web-services interface[24,25]. **a**, Ohmic electric field $E_{\text{Ohm}} = J/\sigma = c\nabla \times B/4\pi\sigma$ normalized by the r.m.s. value, E'_{mot}, of the motional field, $E_{\text{mot}} = -\nu \times B/c$, for one $1{,}024^3$-point time slice of the archived data. The colour scale covers a range from 0.1 to 10 times the r.m.s. value, 1.68×10^{-2}, of the normalized Ohmic field. The Ohmic electric field is negligible compared with the motional field, except in the most intense current sheets. **b**, Energy spectra normalized by total energy, E_{tot}, for the velocity (E_u, red) and magnetic field (E_b, blue) of the flow, each with about a decade of $-3/2$-power-law range. See Supplementary Information for more details.

t, and standard flux freezing[1] is recovered as $\lambda \rightarrow 0$. In a turbulent MHD flow, however, Richardson advection of field lines can allow trajectories to remain random as $\nu \rightarrow 0$ and $\lambda \rightarrow 0$ together. This has been proved mathematically to occur in the Kazantsev–Kraichnan model of a turbulent dynamo[18], and evidence for it has been observed in numerical simulations of a kinematic dynamo with hydrodynamic turbulence[19,20]. However, such effects have not previously been seen in nonlinear MHD turbulence[21].

We observe Richardson relative advection of field lines in a new, high-resolution data set generated by numerical solution of the incompressible MHD equations (Fig. 2). Turbulence was driven as usual by a force in the momentum equation, representing energy input from hydrodynamic instabilities or external stirring (for example a supernova in the interstellar medium). The conductivity, σ, is large, as measured by the magnetic Reynolds number, $\text{Re}_b = 4\pi\sigma b' L_b/c^2 = 764$, where b' is the root mean squared (r.m.s.) magnetic field strength and L_b is the magnetic integral length, and the flow has an inertial range of about one decade in wavenumber. The spectral exponent is closer to $-3/2$ than to $-5/3$, as in other MHD simulations at comparable Reynolds numbers. This fact has motivated theories to explain the $-3/2$ spectrum[22], corresponding to roughness exponent $h = 1/4$. The spectrum of MHD turbulence at much higher, astrophysically relevant Reynolds numbers may be steeper[23]. For our study, any exponent satisfying $0 < h < 1$ suffices. The output of our MHD simulation for one eddy turnover time has been deposited in a publicly accessible online database cluster[24,25]. It is crucial to have the entire time history available. Although the physical process is a coalescence of field lines forwards in time, numerical study of flux freezing in resistive MHD is only practical by tracking stochastic trajectories backwards in time (Fig. 1).

Our results reveal how MHD dynamics generates the magnetic field $B(x,t)$ at a point (x,t) from the magnetic field vectors at the earlier time t_0. We plot (Fig. 3a) the stochastic trajectories that arrive at representative point (x,t), together with the usual deterministic trajectory. The starting locations of these trajectories at time t_0 are the points whose magnetic vectors contribute substantially to $B(x,t)$. The archived value for the magnetic field at point (x,t) is recovered[17] by transporting the magnetic vectors at time t_0 to time t along the stochastic trajectories and resistively averaging (Supplementary Movie 1), with small errors for enough trajectories (Fig. 3b). This is a stringent test of the accuracy of our numerical procedures and of the quality of the archived data. However, despite the high conductivity of the flow, conventional flux freezing along the deterministic trajectory fails very badly.

It is clear from Fig. 3 that the stochastic trajectories are rapidly dispersed over space backwards in time. Figure 4 shows quantitative results on the growth of the relative separation, $r(t) = x_2(t) - x_1(t)$, of distinct members of the 'cloud' of $N = 1{,}024$ particles sampled at time t, averaged over $P = 512$ diagnostic points at the final time, $t_f = 2.56$, in the flow. We plot using the backward time variable, $t_* = t_f - t$. MHD turbulence is expected to be anisotropic[26], with the local mean magnetic field providing a distinguished direction. We therefore calculate the local field at each time by averaging over the field vectors sampled instantaneously by the cloud, and decompose the separation vector, r, into components, $r_{||}$ and $r_\perp$, respectively parallel and perpendicular to the local field. The standard estimate, $r^2 \approx 4\lambda t_*$, is valid only up to one resistive time. For larger t_*, the mean squared separation of field lines changes to a super-diffusive $t_*^{8/3}$ dependence, which is a clear signature of turbulent Richardson dispersion. Consistent with the $-3/2$ spectrum, $h = 1/4$ in $t_*^{2/(1-h)}$ yields the observed $t_*^{8/3}$ growth. We note that dispersion parallel to the field is greater than is dispersion perpendicular to the field, in agreement with prior observations at shorter times[21], but a $t_*^{8/3}$ power law is found in both directions for large (backward) times.

The crucial point is resistivity independence of backward dispersion. We find from our data in Fig. 4 that field lines in one large-eddy turnover time, L_u/u', arrive at the final point from distances of about $L_u/3$, one-third of the velocity integral length. If resistivity independence holds, then this remains true as $\lambda \rightarrow 0$ and standard flux freezing
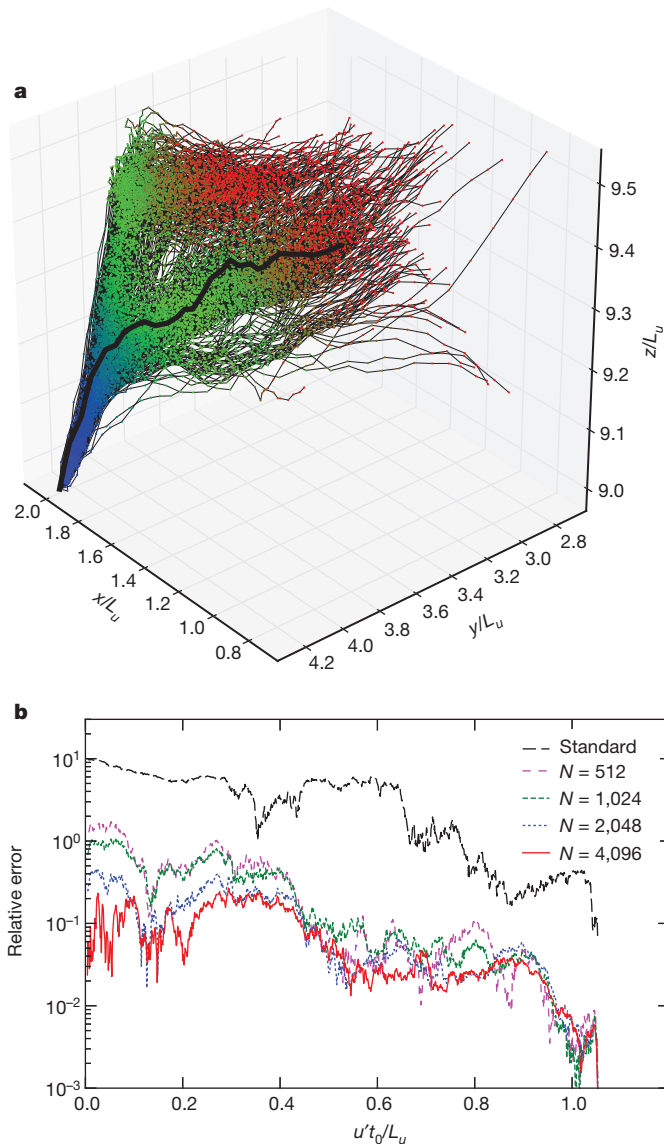
Figure 3 | Stochastic flux freezing in MHD turbulence. a, Stochastic trajectories that arrive at point $x_f = (1.09, 2.42, 5.03)$ in the archived flow at time $t_f = 2.56$. This final point is not in the vicinity of an intense current sheet, having a small Ohmic electric field just 1.3% of the r.m.s. motional field. The points along the stochastic trajectories are colour-coded red, green and blue from earlier to later times. The clouds of points indicate the spatial regions at those times that contribute significantly to $B(x_f, t_f)$ when calculated as described in Fig. 1. We also estimate this magnetic field using conventional flux freezing, by transporting the initial magnetic field along the deterministic trajectory satisfying $dx/dt = u(x, t)$ and $x(t_f) = x_f$ (black). **b**, Errors for the magnetic field calculated from stochastic flux freezing, relative to the archived magnetic field, $B(x_f, t_f)$. The errors for increasing N are small, of similar order as the interpolation error in calculating the velocity gradient. In contrast, the relative error for standard flux freezing (black) is very large. The results shown here for one point are representative (Supplementary Information).

is orders of magnitude in error at very high magnetic Reynolds numbers. In the previous study[20] of turbulent dynamo this resistivity independence could be checked by varying the Prandtl number, but the current MHD database has Pr = 1 only. On the other hand, the resistivity independence is a direct consequence of the Richardson theory. To verify that we are seeing an MHD analogue of Richardson dispersion, we check well-known consequences for the probability density functions (PDFs) of the separation vector, $P(r, t)$. These are predicted to be 'self-similar', that is, to collapse to a single curve when separations and PDFs are rescaled by the r.m.s. dispersion, with the curve given by
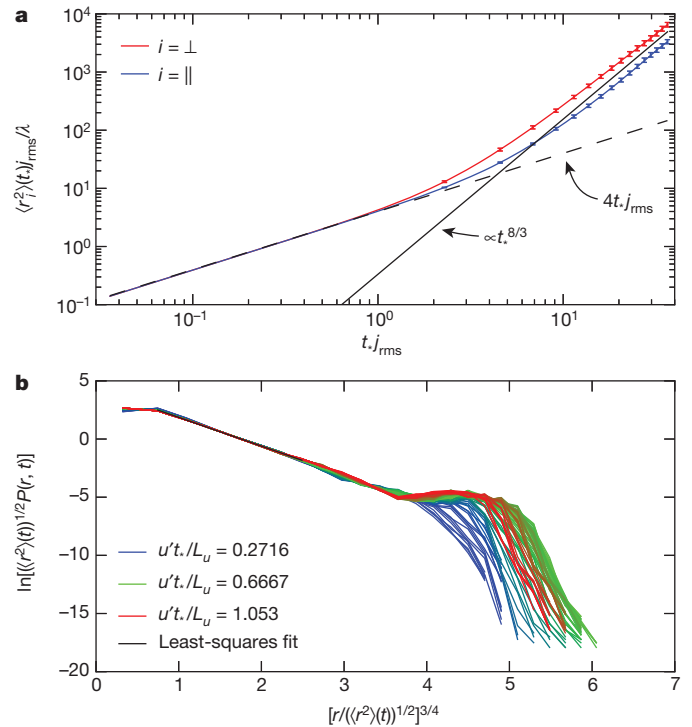
Figure 4 | Evidence of Richardson dispersion. a, Mean squared dispersion of field lines backwards in time with the variable $t_* = t_f - t$, in directions both parallel (red) and perpendicular (blue) to the local magnetic field. Times are normalized by the inverse r.m.s. current, $1/j_{rms}$, and distances are normalized by the resistive length, $\sqrt{\lambda/j_{rms}}$. Error bars, s.e.m. The dashed line shows the conventional diffusive estimate, $\langle r^2(t_*)\rangle = 4\lambda t_*$, and the solid line is $\propto t_*^{8/3}$. The long-time power laws are $\langle r_i^2(t)\rangle \approx g_i L_u^2 (u' t_*/L_u)^{8/3}$, $i = \|, \perp$, with L_u the velocity integral length. The constant $g_\| = 0.086$ in the direction parallel to the local field is greater than the constant $g_\perp = 0.035$ in the perpendicular direction. Both are substantially smaller than the corresponding constant in hydrodynamic turbulence, $g \approx 0.5$. To verify that we are seeing an MHD analogue of Richardson dispersion, we check for self-similarity of the PDFs of the pair separation. **b**, PDFs at 96 times in the interval where the $t_*^{8/3}$ dependence occurs, rescaled by r.m.s. values, with good collapse to the stretched-exponential form, $P \propto \exp(-Cr^{3/4})$, at less than four times the r.m.s. separation. The power in the exponent corresponds to $h = 1/4$, consistent with scaling of the energy spectrum and pair dispersion.

a stretched exponential, $P \propto \exp(-Cr^{1-h})$ (ref. 27). These predictions are confirmed by our data (Fig. 4b).

Spontaneous stochasticity of field line motion underlies the prediction of fast magnetic reconnection in the presence of MHD turbulence[7,8]. Although $h = 1/3$ was assumed in refs 7, 8, the predictions there depend only weakly upon the h value[7]. A strong dependence on mean magnetic field is expected, however[8,21]. The r.m.s. line dispersion in the Alfvén time, L_u/v_A, is predicted[8] to be smaller by a factor $M_T = v_T/v_A$ in the presence of a mean magnetic field, where v_A is the Alfvén velocity based on the mean field and v_T is the velocity of the turbulent eddies with strong nonlinear interactions, generally the r.m.s. velocity or just slightly less. This theory predicts that flux structures of scale L should reconnect with inflow velocity v_T and a turbulent reconnection layer of thickness $\Delta \approx LM_T$ (refs 7, 8). These predictions were verified in MHD reconnection simulations with background turbulence[28], but the detailed turbulent mechanism breaking the frozen-in constraint could not be studied there.

We note that MHD is a valid description of astrophysical plasmas only on scales above the ion gyroradius, ρ_i, and that the microscopic mechanism that dominates the 'slippage' of field lines relative to the plasma may not be the resistive electric field but other terms in the generalized Ohm's law[4,29]. However, Richardson implosive coalescence

of the field lines applies in a turbulent MHD plasma no matter the precise mechanism, as long as lines are separated by distances in the long inertial range above ρ_i. Overall reconnection rates are thus insensitive to the microscopic mechanism of slippage. If the layer thickness, Δ, is of order ρ_i, as is often the case in magnetospheric reconnection, then these ideas do not apply. However, most astrophysical reconnection is of the opposite sort.

For example, current sheets resulting from coronal mass ejections have heights, L, of order 10^{10} cm and thicknesses, Δ, of order 10^9 cm but gyroradii, ρ_i, of order 10^3 cm (ref. 30). The plasma flow is observed to be turbulent, perhaps because of magnetic energy released by initial reconnection due to microscale kinetic mechanisms[4–6]. Observations[30] indicate that $v_T \approx 10^7$ cm s^{-1} and $v_A \approx 10^8$ cm s^{-1}, explaining the thick current sheet with $\Delta \approx 0.1L$ (Supplementary Information). Our results thus solve the 'scale problem', explaining how microscale mechanisms of line slippage can be accelerated to reconnect rapidly such enormous structures.

1. Alfvén, H. On the existence of electromagnetic-hydrodynamic waves. *Ark. Mat. Astron. Fys.* **29,** 1–7 (1942).
2. Kulsrud, R. *Plasma Physics for Astrophysics* (Princeton Univ. Press, 2005).
3. Priest, E. R. & Forbes, T. G. *Magnetic Reconnection: MHD Theory and Applications* (Cambridge Univ. Press, 2000).
4. Che, H., Drake, J. F. & Swisdak, M. A current filamentation mechanism for breaking magnetic field lines during reconnection. *Nature* **474,** 184–187 (2011).
5. Daughton, W. *et al.* Role of electron physics in the development of turbulent magnetic reconnection in collisionless plasmas. *Nature Phys.* **7,** 539–542 (2011).
6. Moser, A. L. & Bellan, P. M. Magnetic reconnection from a multiscale instability cascade. *Nature* **482,** 379–381 (2012).
7. Lazarian, A. & Vishniac, E. T. Reconnection in a weakly stochastic field. *Astrophys. J.* **517,** 700–718 (1999).
8. Eyink, G. L., Lazarian, A. & Vishniac, E. T. Fast magnetic reconnection and spontaneous stochasticity. *Astrophys. J.* **743,** 51 (2011).
9. Richardson, L. F. Atmospheric diffusion shown on distance-neighbour graph. *Proc. R. Soc. Lond. A* **110,** 709–737 (1926).
10. Bernard, D., Gawȩdzki, K. & Kupiainen, A. Slow modes in passive advection. *J. Stat. Phys.* **90,** 519–569 (1998).
11. E, W. & Vanden Eijnden, E. Generalized flows, intrinsic stochasticity, and turbulent transport. *Proc. Natl Acad. Sci. USA* **97,** 8200–8205 (2000).
12. Chaves, M., Gawȩdzki, K., Horvai, P., Kupiainen, A. & Vergassola, M. Lagrangian dispersion in Gaussian self-similar velocity ensembles. *J. Stat. Phys.* **113,** 643–692 (2003).
13. Eyink, G. L. Turbulent diffusion of lines and circulations. *Phys. Lett. A* **368,** 486–490 (2007).
14. Frisch, U. *Turbulence* 52–56 (Cambridge Univ. Press, 1995).
15. Leamon, R. J., Smith, C. W., Ness, N. F., Matthaeus, W. H. & Wong, H.-K. Observational constraints on the dynamics of the interplanetary magnetic field dissipation range. *J. Geophys. Res.* **103,** 4775–4787 (1998).
16. Chepurnov, A. & Lazarian, A. Extending the big power law in the sky with turbulence spectra from Wisconsin Hα Mapper data. *Astrophys. J.* **710,** 853–858 (2010).
17. Eyink, G. L. Stochastic line motion and stochastic flux conservation for nonideal hydromagnetic models. *J. Math. Phys.* **50,** 083102 (2009).
18. Eyink, G. L. Fluctuation dynamo and turbulent induction at small Prandtl number. *Phys. Rev. E* **82,** 046314 (2010).
19. Sawford, B. L., Yeung, P. K. & Hackl, J. F. Reynolds number dependence of relative dispersion statistics in isotropic turbulence. *Phys. Fluids* **20,** 065111 (2008).
20. Eyink, G. L. Stochastic flux freezing and magnetic dynamo. *Phys. Rev. E* **83,** 056405 (2011).
21. Busse, A., Müller, W.-C., Homann, H. & Grauer, R. Statistics of passive tracers in three-dimensional magnetohydrodynamic turbulence. *Phys. Plasmas* **14,** 122303 (2007).
22. Boldyrev, S. Spectrum of magnetohydrodynamic turbulence. *Phys. Rev. Lett.* **96,** 115002 (2006).
23. Beresnyak, A. Spectral slope and Kolmogorov constant of MHD turbulence. *Phys. Rev. Lett.* **106,** 075001 (2011).
24. Li, Y. *et al.* A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J. Turbul.* **9,** N31 (2008).
25. The JHU Turbulence Database Cluster. http://turbulence.pha.jhu.edu (2012).
26. Goldreich, P. & Sridhar, S. Toward a theory of interstellar turbulence. 2: Strong Alfvénic turbulence. *Astrophys. J.* **438,** 763–775 (1995).
27. Hentschel, H. G. E. & Procaccia, I. Relative diffusion in turbulent media: the fractal dimension of clouds. *Phys. Rev. A* **29,** 1461–1470 (1984).
28. Kowal, G., Lazarian, A., Vishniac, E. T. & Otmianowska-Mazur, K. Numerical tests of fast reconnection in weakly stochastic magnetic fields. *Astrophys. J.* **700,** 63–85 (2009).
29. Schekochihin, A. A. *et al.* Astrophysical gyrokinetics: kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas. *Astrophys. J.* **182** (suppl.), 310–377 (2009).
30. Ciaravella, A. & Raymond, J. C. The current sheet associated with the 2003 November 4 coronal mass ejection: density, temperature, thickness, and line width. *Astrophys. J.* **686,** 1372–1382 (2008).

Supplementary Information is available in the online version of the paper.

Author Contributions All of the authors made significant contributions to this work. H.A. carried out the simulations of MHD turbulence. K.K., R.B., A.S. and C.M. were primarily responsible for the construction of the MHD database and online analysis tools. G.E. designed the study and developed the numerical algorithms for stochastic flux freezing. C.L. generated the simulation results using the database. K.B. developed the visualization of the archived MHD data. G.E., E.V., C.L. and C.M. analysed the simulation results and were primarily responsible for writing the paper. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.E. (eyink@jhu.edu).

LETTER

All–angle negative refraction and active flat lensing of ultraviolet light

Ting Xu[1,2], Amit Agrawal[1,3], Maxim Abashin[1,2], Kenneth J. Chau[4] & Henri J. Lezec[1]

Decades ago, Veselago[1] predicted that a material with simultaneously negative electric and magnetic polarization responses would yield a 'left-handed' medium in which light propagates with opposite phase and energy velocities—a condition described by a negative refractive index. He proposed that a flat slab of left-handed material possessing an isotropic refractive index of −1 could act like an imaging lens in free space. Left-handed materials do not occur naturally, and it has only recently become possible to achieve a left-handed response using metamaterials, that is, electromagnetic structures engineered on subwavelength scales to elicit tailored polarization responses. So far, left-handed responses have typically been implemented using resonant metamaterials composed of periodic arrays of unit cells containing inductive–capacitive resonators and conductive wires. Negative refractive indices that are isotropic in two[2] or three[3] dimensions at microwave frequencies have been achieved in resonant metamaterials with centimetre-scale features. Scaling the left-handed response to higher frequencies, such as infrared or visible, has been done by shrinking critical dimensions to submicrometre scales by means of top-down nanofabrication[4]. This miniaturization has, however, so far been achieved at the cost of reduced unit-cell symmetry, yielding a refractive index that is negative along only one axis. Moreover, lithographic scaling limits have so far precluded the fabrication of resonant metamaterials with left-handed responses at frequencies beyond the visible[5]. Here we report the experimental implementation of a bulk metamaterial with a left-handed response to ultraviolet light. The structure, based on stacked plasmonic waveguides[6], yields an omnidirectional left-handed response for transverse magnetic polarization characterized by a negative refractive index. By engineering the structure to have a refractive index close to −1 over a broad angular range, we achieve Veselago flat lensing, in free space, of arbitrarily shaped, two-dimensional objects beyond the near field. We further demonstrate active, all-optical modulation of the image transferred by the flat lens.

The electric and magnetic polarization responses of a linear, isotropic medium are described using two scalar constitutive parameters - the complex permittivity, $\underline{\varepsilon} = \varepsilon' + i\varepsilon''$, and the complex permeability, $\underline{\mu} = \mu' + i\mu''$. The physics of time-harmonic electromagnetic wave propagation in such a medium can be derived from Maxwell's equations and distilled into the form of plane-wave solutions. Plane waves are characterized by a wavevector, k, and a time-averaged Poynting vector, S, which respectively give the directions of phase velocity and energy flow in the medium. In an isotropic medium, k and S are always collinear, but are either parallel or antiparallel, following the causality condition[7] $\mathrm{sgn}(k \cdot S) = \mathrm{sgn}(\alpha)$, where $\alpha = \varepsilon'|\mu| + \mu'|\varepsilon|$. A general index of refraction given by $n = \mathrm{sgn}(\alpha)|\mathrm{Re}(\sqrt{\mu\varepsilon})|$ can then be introduced, which describes both the magnitude of $\overline{k}$ with respect to its value in free space ($|k| = |n|k_0 = |n|\omega/c$, where ω is the frequency and c is the speed of light) as well as the relative direction of k and S. For a material in which $\alpha > 0$—called 'right-handed' because k and S are parallel

($k \cdot S > 0$)—the index of refraction is positive. Plane waves in right-handed media propagate with wavefronts streaming along the direction of energy flow, similar to a wave in free space. The archetypal right-handed medium is a 'double-positive' material such as air or glass, for which ε' and μ' are simultaneously positive. In the contrasting case of a material in which $\alpha < 0$—called 'left-handed' because k and S are antiparallel ($k \cdot S < 0$)—the index of refraction is negative. Plane waves in left-handed media propagate as an exotic 'backwards wave' with wavefronts travelling in a direction opposite to that of energy flow. The classical left-handed medium, first introduced by Veselago, is the 'double-negative' one, for which ε' and μ' are simultaneously negative.

Plane-wave refraction at the boundary of a left-handed medium can be modelled using Snell's law, under the convention that its refractive index takes negative values. Conservation of phase and energy flow applied to a plane wave crossing the boundary between a medium with index n_1 and another with index n_2 yields $n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$, a generalized form of Snell's law in which n_1 and n_2 have arbitrary signs and θ_1 and θ_2 are the directed angles of the optical ray path with respect to the normal on each side of the interface. If n_1 and n_2 have opposite signs, then θ_1 and θ_2 have opposite signs and the ray deflects across the boundary to the same side of the normal—a phenomenon known as negative refraction. If n_1 and n_2 are equal in magnitude but opposite in sign, the ray path is perfectly mirrored about the interface. In the configuration of a flat slab, this type of negative refraction occurs at both interfaces and it becomes possible to focus a collection of rays from a point source to another point on the other side of the slab. As suggested by Veselago[1], a lens in free space could be realized by simply using a flat slab of material with an isotropic refractive index of −1.

The current limitations of top-down nanofabrication hinder straightforward implementation of metamaterials that are able to display a three-dimensionally isotropic, negative-index response at optical frequencies, using the canonical system of volumetric arrays of discrete, shape-based resonators with critical features and spacing of deep subwavelength dimensions[5] (that is, dimensions much smaller than half the wavelength). Metamaterials based on plasmonic waveguides consisting of alternating planar metal (M) and dielectric (D) layers offer an attractive alternative to achieving a high-frequency left-handed response[8–10]. This left-handed response is derived from a transverse-magnetic-polarized (TMP) backwards electromagnetic mode naturally sustained by the waveguide in a frequency range between the bulk plasmon resonance frequency, ω_P, of the metal—typically located in the ultraviolet—and the surface plasmon resonance frequency, ω_{SP}, of the metal–dielectric interface, which can be positioned within the visible by proper choice of constituent materials. Using this approach, a metal–dielectric–metal (MDM) plasmonic waveguide with an ultrathin dielectric core was employed to achieve a two-dimensionally isotropic, negative refractive index at green and blue frequencies, as evidenced by observation of in-plane negative refraction[11]. Recently it was theoretically shown that a plasmonic-waveguide-based system could be devised

[1]Center for Nanoscale Science and Technology, National Institute of Standard and Technology, Gaithersburg, Maryland 20899, USA. [2]Maryland Nano-Center, University of Maryland, College Park, Maryland 20742, USA. [3]Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York 13244, USA. [4]School of Engineering, The University of British Columbia, Kelowna, British Columbia V1V 1V7, Canada.

in which the left-handed response is extended to an extra dimension normal to the plane of the layers[6]. The proposed metamaterial, which is composed of a stack of strongly coupled plasmonic waveguides each having an MDMDM layer sequence, was designed to have an in-plane mode symmetry that allows propagation of a TMP backwards wave in three dimensions at a visible wavelength of 400 nm.

Here we design a plasmonic waveguide metamaterial based on repeated MDMDM layered unit cells (Fig. 1a) to have an all-angle left-handed response in the ultraviolet characterized by a refractive index close to −1. With Ag (metal) and TiO$_2$ (dielectric) as the constituent materials, we use the transfer matrix method[12] to produce a three-dimensional equifrequency contour (EFC) at the ultraviolet wavelength $\lambda_0 = 363.8$ nm for transverse magnetic polarization (magnetic field parallel to the plane of the layers). The EFC maps the angular dependence of the wavevector in the metamaterial and indicates the direction of the Poynting vector (which is normal to the EFC and points in the direction of its displacement as a function of frequency). The individual layer thicknesses of the MDMDM layers comprising each unit cell (33, 28, 30, 28 and 33 nm, respectively) are optimized to achieve a broad-angle refractive index of −1, as evidenced by an EFC
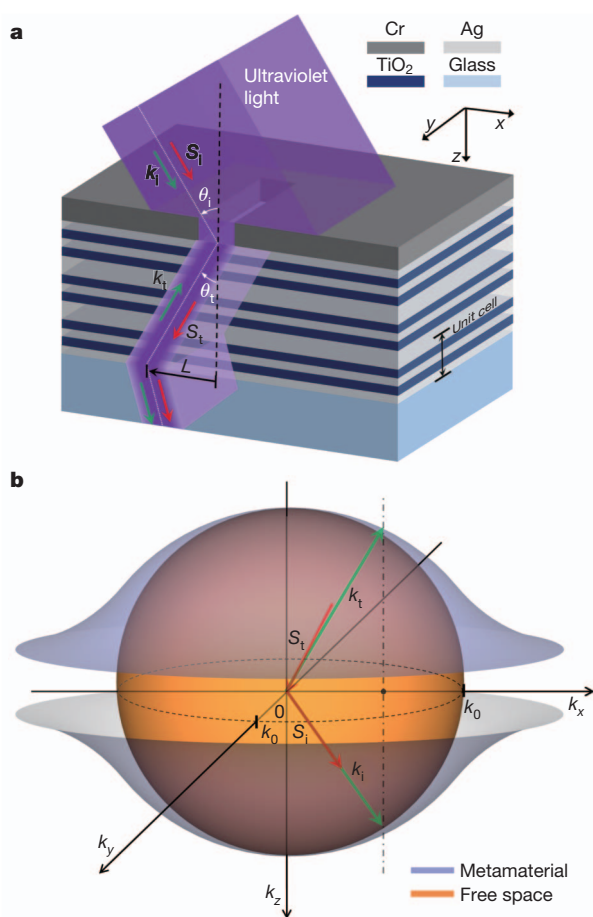
Figure 1 | Ultraviolet bulk metamaterial. a, Schematic of negative refraction of ultraviolet light from air into a coupled plasmonic waveguide metamaterial formed by three vertically stacked MDMDM unit cells and coated with a beam-defining mask consisting of a rectangular aperture in an opaque Cr film. S_i and S_t, incident and transmitted time-averaged Poynting vectors, respectively; k_i and k_t, incident and transmitted wavevectors, respectively. **b,** Calculated three-dimensional EFCs for an ultraviolet plane wave of free-space wavelength $\lambda_0 = 363.8$ nm in the as-designed metamaterial structure and in free space (k_0, free-space wavevector). The process of negative refraction for both the Poynting vector and the wavevector at the air–metamaterial interface is illustrated for an incident angle $\theta_i = 40°$; the conservation of the tangential wavevector component is represented by the black dash–dot line.

that has a near-spherical shape with a radius of approximately k_0 (Fig. 1b) and moves inward with increasing frequency[13] (such that $k \cdot S < 0$). For a plane wave incident from free space onto the metamaterial at an incident angle θ_i, continuity of phase and conservation of energy at the interface dictate that the transmitted backwards wave has a wavevector k_t (pointing towards the interface) and a Poynting vector S_t (pointing away from the interface) on the same side of the normal as the incident ray path, conditions describing negative refraction of phase and power. Because the metamaterial EFC is not perfectly spherical, k_t and S_t are not exactly collinear and Snell's law, which applies to phase refraction, cannot be rigorously applied to refraction of power. To describe power refraction from free space into the metamaterial in a manner analogous to Snell's law, we introduce a power refractive index $n_S \equiv \sin(\theta_i)/\sin(\theta_{t,S})$, where $\theta_{t,S}$ is the directed angle between the path of power flow with respect to the normal.

A slab of the designed metamaterial is fabricated by sputtering three MDMDM unit cells (metal, Ag; dielectric, TiO$_2$) onto a transparent glass substrate, to a total thickness of approximately 450 nm, followed by patterning of an illumination aperture consisting of a 600-nm-wide rectangular opening in an opaque Cr film (Fig. 2a). To measure power refraction from free space into the metamaterial, we illuminate the aperture-masked, air-facing surface of the metamaterial at various incidence angles (θ_i) with continuous-wave radiation at $\lambda_0 = 363.8$ nm, and extract the angular deflection of beam power ($\theta_{t,S}$) from the lateral shift, L, of the most intense part of the exiting beam relative to the aperture centre. For transverse magnetic polarization, we observe negative power refraction of the ultraviolet beam over a broad range of incident angles spanning $10° \leq \theta_i \leq 60°$ (Fig. 2b). Power refraction switches from negative to positive when the polarization of the ultraviolet beam is rotated to a transverse electric configuration (electric field parallel to the plane of the layers), a result expected owing to the absence of TMP plasmon modes to confer left-handed electromagnetic behaviour.

The power refractive index measured for transverse magnetic polarization is approximately −1 over a broad range of angles and closely matches expected values based on the curvature of the EFC of the designed metamaterial (Fig. 2c). There is also good agreement between measured power refraction and two virtual experiments performed with a finite-difference time-domain (FDTD) simulator solving Maxwell's equation: one which calculates the angular deflection of a beam through an aperture-bearing metamaterial akin to that in the experiments, and another which calculates the unit-cell-averaged Poynting vector in a bare metamaterial subject to plane-wave illumination (Fig. 2c). We extract the effective constitutive parameters of the bare metamaterial by applying a parameter retrieval technique to FDTD-simulated reflection and transmission data[14]. The effective refractive index, n_{eff}, is approximately −1 over a large angular range and agrees with the refractive index expected from the metamaterial EFC (Fig. 2d), confirming the realization of the designed omnidirectional left-handed response. The metamaterial has a double-negative response characterized by simultaneously negative values of ε' and μ' (Supplementary Fig. 1). In contrast with other metamaterial implementations that achieve visible-frequency left-handed behaviour from a single negative response[15] (in which one of ε' and μ' is negative), the double-negative response here enables low-loss wave propagation (evidenced by a relatively high figure of merit; Supplementary Fig. 1) beneficial for applications such as imaging.

Veselago flat lensing is a special refractory phenomenon that occurs when a flat slab of thickness d has an isotropic refractive index of −1 and is immersed in free space. In Fig. 3a, we depict the simplest form of Veselago flat lensing, in which diverging rays from a point object positioned on one of the faces of the slab are refracted at the opposite face and focused to a point image a distance d beyond the slab. Experimental demonstrations of Veselago flat lensing have so far only been possible at microwave frequencies, for which bulk left-handed metamaterials with two- or three-dimensionally isotropic, negative
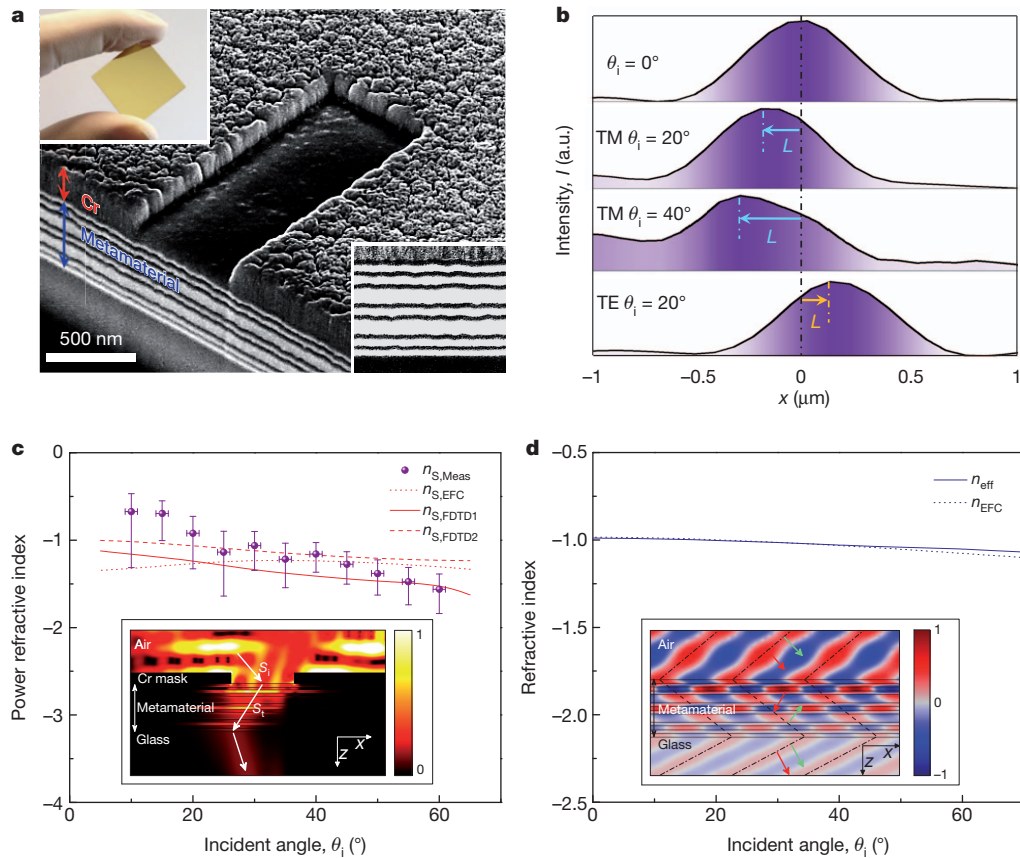
Figure 2 | Experimental and simulated results for refraction of power and phase. a, Scanning electron microscope image of the fabricated sample used in the refraction experiment, showing the beam-defining aperture, and sectioned by focused-ion-beam (FIB) milling to reveal the internal metamaterial structure. Left inset, glass slide uniformly coated with 450-nm-thick metamaterial; right inset, magnified cross-section of the metamaterial layers. **b**, Intensity profile of transmitted beam imaged with an optical microscope at the glass–metamaterial interface and integrated over aperture length, at various incident angles θ_i, for transverse magnetic (TM) and transverse electric (TE) polarizations. **c**, Power refractive index of the metamaterial, as a function of θ_i, for TMP light. The purple points show $n_{S,Meas}$, the experimentally derived value of n_S, for which the horizontal error bars correspond to the mechanical uncertainty ($\pm 1°$) in the angle of illumination and the vertical error bars represent the uncertainty in the measured lateral shift due to the pixel resolution of the recording charge-coupled-device camera. The red dotted line shows $n_{S,EFC}$, the value of n_S derived from the metamaterial's EFC. The red solid

line shows $n_{S,FDTD1}$, the value of n_S derived from FDTD simulations in which an ultraviolet beam is diffracted through an aperture-bearing metamaterial akin to that in the experiments. The red dashed line shows $n_{S,FDTD2}$, the value of n_S derived from FDTD simulations in which a plane wave is incident on a bare metamaterial and the time-averaged Poynting vector is averaged over a unit cell. Inset, simulated profile of the time-averaged Poynting vector for the aperture-bearing metamaterial illuminated at $\theta_i = 40°$. **d**, Refractive index of the bare metamaterial, as a function of θ_i, for TMP light. The blue solid line shows n_{eff}, the refractive index derived from effective parameter retrieval applied to FDTD-simulated reflection and transmission data. The blue dotted line shows n_{EFC}, the refractive index derived from the metamaterial EFC. Inset, simulated magnetic field distribution for the metamaterial slab illuminated with a plane wave incident at $\theta_i = 40°$. The dash–dot lines show phase fronts in free space and glass, the dashed lines show effective phase fronts in the metamaterial inferred from Snell's law (on the basis of n_{eff}), and the red and green arrows represent the Poynting vector and the wavevector, respectively.

refractive indices are available[3,16–18]. Here we exploit the broad-angle negative refractive index achieved with our left-handed metamaterial to demonstrate Veselago flat lensing in the ultraviolet, extending the experimental realization of this effect across the electromagnetic spectrum and into a frequency range with important applications such as fluorescence imaging and photolithography. Although our metamaterial has a left-handed response only for transverse magnetic polarization, we show that imaging with circular polarization enables a flat slab of the metamaterial to perform three-dimensional imaging of two-dimensional objects beyond the near field ($d > \lambda_0$), closely approximating ideal, far-field flat lensing discussed by Veselago.

The flat lens consists of a free-standing film of the stacked-waveguide metamaterial described and characterized above. To probe the performance of the lens, imaging apertures of various shapes are fabricated directly on the lens surface. The apertures are illuminated at normal incidence with coherent ultraviolet light ($\lambda_0 = 363.8$ nm), replicating the schematic set-up in Fig. 3a. The lensing performance of the structure is first characterized by using it to image a slit aperture of width $w_s = 180$ nm, illuminated by TMP light (magnetic field oriented

along the slit). The lateral intensity profile of the transmitted light on the opposite side of the metamaterial slab is recorded with an optical microscope as a function of distance, z, relative to the exit surface of the slab (Fig. 3b), where $z > 0$ corresponds to a position outside the slab. A curve fit to the full-width at half-maximum (FWHM) of the intensity profile reveals a beam waist located at a position $z_f = 390$ nm, which is interpreted as the focal distance of the flat lens relative to the lens exit surface. The measured focal distance is comparable to the theoretical value of 450 nm expected for a flat lens consisting of an ideally isotropic left-handed medium of refractive index $n = -1$ and thickness $d = 450$ nm. In addition, z_f is consistent with the numerical value of 360 nm predicted by an FDTD simulation of the stacked-waveguide metamaterial lens bearing a slit aperture (Fig. 3b, inset). The experimentally measured beam width at z_f has a FWHM value of $w_{f,Meas} \approx 370$ nm, which is limited by the resolution of the optical microscope used for the observation. The FDTD simulations yield a minimum beam width of $w_{f,FDTD} = 200$ nm, which is close to the theoretical resolution limit of the implemented flat lens (Supplementary Fig. 2).
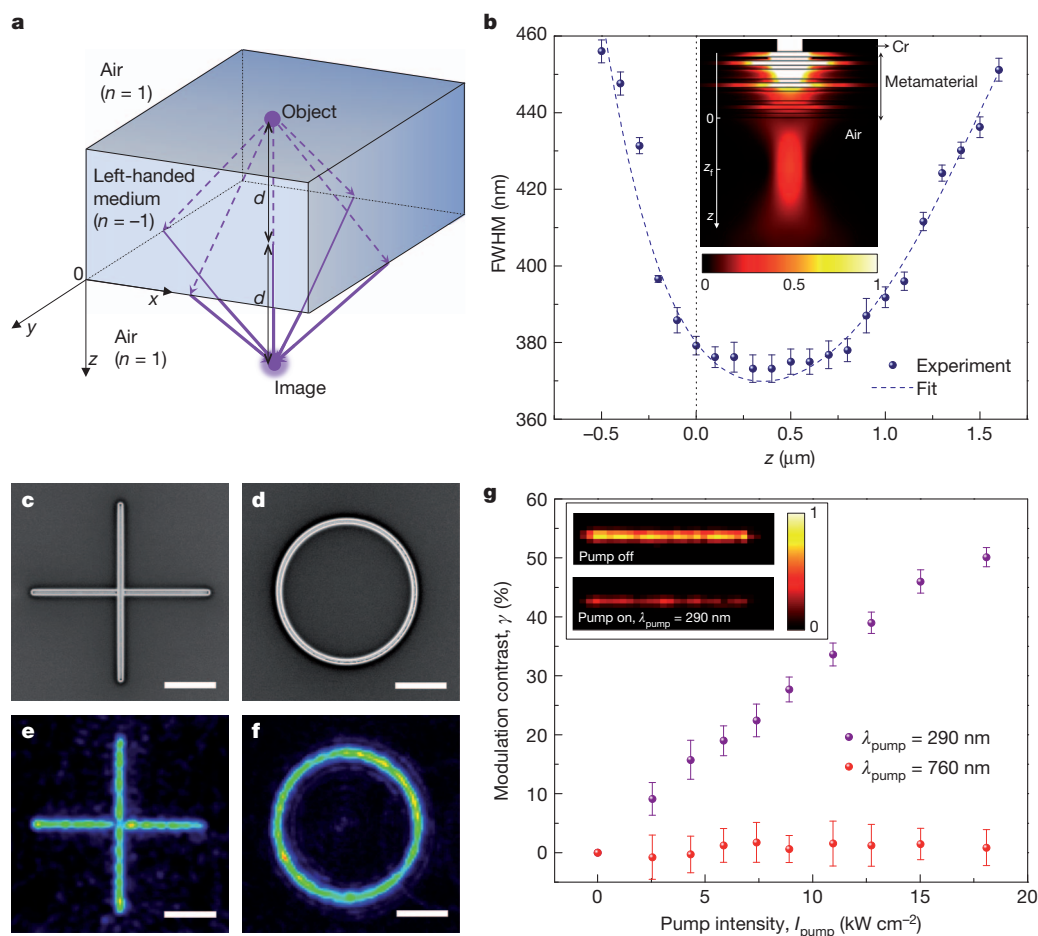
Figure 3 | Ultraviolet flat lensing. a, Schematic diagram of imaging with a Veselago flat lens (of index $n = -1$ and thickness d). Optical rays diverging into the bulk of the lens from a given constituent point source on the lens surface are each negatively refracted at the exit surface of the lens, with equal and opposite ray angles on either side of the boundary, as dictated by Snell's law. This refraction process then leads all rays to converge to a point image located in free space at a distance from the exit surface equal to the thickness, d, of the lens. **b,** FWHM of experimental lateral intensity profile of the metamaterial flat-lens image of a 180-nm-wide slit aperture, as recorded with an optical microscope, plotted against distance, z, relative to the exit surface of the slab. The black dotted line indicates the exit surface of the lens ($z = 0$). Errors, standard deviation for repeated measurements (four in total). The dashed line shows a

polynomial fit. Inset, profile of the time-averaged Poynting vector resulting from FDTD simulation of the experimental configuration. The simulated transmitted beam is characterized by a waist at $z_f = 360$ nm. **c, d,** Scanning electron microscope images of slit apertures located on the metamaterial flat lens surface, forming cross (**c**) and ring (**d**) objects. **e, f,** Corresponding images produced by the flat lens under illumination by circularly polarized ultraviolet light, as recorded using an optical microscope focused at z_f. Scale bars, 3 μm. **g,** Intensity modulation contrast, γ, as a function of pump intensity I_{pump} at wavelengths of 290 (purple) and 760 nm (red). Pump pulse length, 150 fs; repetition rate, 80 MHz. Errors, standard deviation for repeated measurements (four in total). Inset, recorded intensity profiles of slit image at z_f with pump off (top) and pump on (bottom). Pump intensity, 18 kW cm^{-2}.

Although image transfer with no loss of spatial frequencies beyond the diffraction limit (super-resolution) was derived for the ideal case of a lossless, non-dispersive Veselago flat lens with $\varepsilon = \mu \equiv -1$ (ref. 19), imaging performance under more realistic conditions of loss, dispersion or even infinitesimal departure from rigorous equality between ε and μ has been shown to be significantly degraded[20,21]. In particular, near-complete loss of super-resolution is expected when a lossy, dispersive left-handed Veselago flat lens of super-wavelength thickness ($d \geq \lambda_0$, as here) is used for image transfer beyond the near field[16,22–24]. Though image transfer with a finite amount of super-resolution has been experimentally demonstrated in recent years at ultraviolet frequencies comparable to that of the present experiment using a non-left-handed 'poor-man's' superlens consisting of a thin, flat Ag film[25] ($d \approx \lambda_0/12$), this electrostatic-limit imaging mode is possible only over near-field distances, a situation which is conceptually different from that of the present study.

The ability of the fabricated flat lens to image two-dimensional objects of arbitrary shapes is demonstrated using cross (Fig. 3c) and ring (Fig. 3d) slit apertures. The aperture width is set to a constant value of 180 nm ($w_s < \lambda_0/2$) to suppress transverse-electric-polarized light,

which is not focused by the lens. To maintain imaging of the full specified shapes of the two-dimensional aperture objects, circularly polarized illumination is used to probe evenly all radial directions of each object. The resulting images of the objects formed by the flat lens are shown in Fig. 3e (cross) and Fig. 3f (ring), as recorded by the optical microscope focused at z_f. The salient characteristics of the images, including point-to-point replication of features of the two-dimensional objects, are consistent with those expected from a Veselago flat lens imaging beyond the near field.

In parallel with their exotic refractive properties, actively tunable metamaterials offer the possibility of dynamic control of light[26]. Here we demonstrate that the intensity of an ultraviolet image transferred by the metamaterial flat lens (for continuous-wave illumination at $\lambda_0 = 363.8$ nm) can be dynamically varied by illumination of the exit surface of the lens with a pulsed pump beam ($\lambda_{pump} = 290$ nm) tuned above the bandgap of the as-deposited constituent TiO$_2$ (characterized in Supplementary Fig. 3). As the pump intensity, I_{pump}, is increased, the transmitted intensity, I_{ON}, is observed to decrease (Fig. 3g). The corresponding intensity modulation contrast, $\gamma = (I_{OFF} - I_{ON})/I_{OFF}$ (where I_{OFF} is the transmitted intensity in the absence of pumping),

displays a quasilinear dependence on I_{pump} and increases to a value of $\gamma \approx 50\%$ at the maximum explored pump intensity below the damage threshold of the sample. Conversely, pumping the metamaterial to similar intensities at a wavelength tuned below the TiO_2 bandgap ($\lambda_{pump} = 760$ nm) yields no significant transmission modulation ($\gamma \approx 0\%$). These observations support the hypothesis that pump-induced transmission suppression results from absorption by free carriers generated in the TiO_2 (refs 27, 28). The achieved modulation contrast exceeds that of electro-absorption-based solid-state devices operating in the ultraviolet[29,30].

We have experimentally demonstrated a bulk metamaterial with an all-angle negative index of refraction, and have used it to achieve Veselago flat lensing and all-optical modulation in the ultraviolet. The fabrication-friendly planar architecture of this metamaterial shows promise for building large-area left-handed optical elements for manipulating light at frequencies beyond the visible.

METHODS SUMMARY

To fabricate the structure used for the refraction experiment, sequential direct-current and radio-frequency sputtering in the same sputter chamber were used to deposit alternating layers of Ag and TiO_2 on a transparent glass slide (deposition rates: $R_{Ag} \approx 3.6 \, \text{Å s}^{-1}$, $R_{TiO_2} \approx 0.35 \, \text{Å s}^{-1}$). The resulting metamaterial was then sputter-coated with a 120-nm-thick Cr layer, with an aperture etched through the Cr layer using FIB milling. Finally, two reference alignment marks consisting of 100-nm-wide, 1-μm-long slits through the full thickness of the metamaterial down to the glass substrate were also patterned by FIB milling in the vicinity of the aperture. The structure used for the flat-lensing experiment was fabricated by sputter deposition of a metamaterial stack onto a Si_3N_4 membrane previously coated with a 120-nm-thick Cr layer. FIB milling was then used from the membrane side to remove the Si_3N_4 locally and pattern various slit-aperture objects through the Cr layer.

For both refraction and lensing experiments, accurate referencing of the focal plane of the optical microscope (with a ×100 objective lens) relative to the exit surface of the metamaterial was achieved by focusing the diverging ultraviolet light emitted by the alignment-mark slits on the exit surface. In addition, the illuminated alignment marks were also used as a reference for accurate calibration of the lateral shift of the transmitted beam under refraction at various angles of incidence. For the active flat-lensing experiment, a ×40, deep-ultraviolet objective lens was used to record the probe beam and simultaneously focus the pump beam onto the sample surface. After each pump-on experiment, the transmitted intensity of the probe beam was rechecked to exclude the possibility of the flat lens sample being damaged during the pump process.

1. Veselago, V. G. The electrodynamics of substances with simultaneously negative values of ε and μ. *Sov. Phys. Usp.* **10,** 509–514 (1968).
2. Shelby, R. A., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292,** 77–79 (2001).
3. Rudolph, S. M. & Grbic, A. A broadband three-dimensionally isotropic negative-refractive-index medium. *IEEE Trans. Antenn. Propag.* **60,** 3661–3669 (2012).
4. Shalaev, V. M. Optical negative-index metamaterials. *Nature Photon.* **1,** 41–48 (2007).
5. Soukoulis, C. M. & Wegener, M. Past achievements and future challenges in the development of three dimensional photonic metamaterials. *Nature Photon.* **5,** 523–530 (2011).
6. Verhagen, E., Waele, R. D., Kuipers, L. & Polman, A. Three-dimensional negative index of refraction at optical frequencies by coupling plasmonic waveguides. *Phys. Rev. Lett.* **105,** 223901 (2010).
7. Depine, R. A. & Lakhtakia, A. A new condition to identify isotropic dielectric-magnetic materials displaying negative phase velocity. *Microw. Opt. Technol. Lett.* **41,** 315–316 (2004).
8. Shvets, G. Photonic approach to making a material with a negative index of refraction. *Phys. Rev. B* **67,** 035109 (2003).
9. Shin, H. & Fan, S. All-angle negative refraction for surface plasmon waves using a metal-dielectric-metal structure. *Phys. Rev. Lett.* **96,** 073907 (2006).
10. Alù, A. & Engheta, N. Optical nanotransmission lines: synthesis of planar left-handed metamaterials in the infrared and visible regimes. *J. Opt. Soc. Am. B* **23,** 571–583 (2006).
11. Lezec, H. J., Dionne, J. A. & Atwater, H. A. Negative refraction at visible frequencies. *Science* **316,** 430–432 (2007).
12. Yeh, P. *Optical Waves in Layered Media* 118–143 (Wiley, 1988).
13. Foteinopoulou, S. & Soukoulis, C. M. Negative refraction and left-handed behavior in two-dimensional photonic crystals. *Phys. Rev. B* **67,** 235107 (2003).
14. Menzel, C. *et al.* Retrieving effective parameters for metamaterials at oblique incidence. *Phys. Rev. B* **77,** 195328 (2008).
15. Xiao, S. *et al.* Yellow light negative index metamaterials. *Opt. Lett.* **34,** 3478–3480 (2009).
16. Houck, A. A., Brock, J. B. & Chuang, I. L. Experimental observations of a left-handed material that obeys Snell's law. *Phys. Rev. Lett.* **90,** 137401 (2003).
17. Grbic, A. & Eleftheriades, G. V. Overcoming the diffraction limit with a planar left-handed transmission-line lens. *Phys. Rev. Lett.* **92,** 117403 (2004).
18. Iyer, A. K. & Eleftheriades, G. V. Free-space imaging beyond the diffraction limit using a Veselago-Pendry transmission-line metamaterial superlens. *IEEE Trans. Antenn. Propag.* **57,** 1720–1727 (2009).
19. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85,** 3966–3969 (2000).
20. Ziolkowski, R. W. & Heyman, E. Wave propagation in media having negative permittivity and permeability. *Phys. Rev. E* **64,** 056625 (2001).
21. Garcia, N. & Nieto-Vesperinas, M. Left-handed materials do not make a perfect lens. *Phys. Rev. Lett.* **88,** 207403 (2002).
22. Loschialpo, P. F., Smith, D. L., Forester, D. W., Rachford, F. J. & Schelleng, J. Electromagnetic waves focused by a negative-index planar lens. *Phys. Rev. E* **67,** 025602 (2003).
23. Rao, X. S. & Ong, C. K. Subwavelength imaging by a left-handed material superlens. *Phys. Rev. E* **68,** 067601 (2003).
24. Nieto-Vesperinas, M. Problem of image superresolution with a negative-refractive-index slab. *J. Opt. Soc. Am. A* **21,** 491–498 (2004).
25. Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308,** 534–537 (2005).
26. Hess, O. *et al.* Active nanoplasmonic metamaterials. *Nature Mater.* **11,** 573–584 (2012).
27. Carp, O., Huisman, C. L. & Reller, A. Photoinduced reactivity of titanium dioxide. *Prog. Solid State Chem.* **32,** 33–177 (2004).
28. Chen, X. & Mao, S. S. Titanium dioxide nanomaterials: synthesis, properties, modifications, and applications. *Chem. Rev.* **107,** 2891–2959 (2007).
29. Zhang, X. Y., Dhawan, A., Wellenius, P., Suresh, A. & Muth, J. F. Planar ZnO ultraviolet modulator. *Appl. Phys. Lett.* **91,** 071107 (2007).
30. Kao, C., Bhattacharyya, A., Thomidis, C., Paiella, R. & Moustakas, T. D. Electroabsorption modulators based on bulk GaN films and GaN/AlGaN multiple quantum wells. *J. Appl. Phys.* **109,** 083102 (2011).

LETTER

Enantiomer–specific detection of chiral molecules via microwave spectroscopy

David Patterson[1], Melanie Schnell[2,3] & John M. Doyle[1]

Chirality plays a fundamental part in the activity of biological molecules and broad classes of chemical reactions, but detecting and quantifying it remains challenging[1]. The spectroscopic methods of choice are usually circular dichroism and vibrational circular dichroism, methods that are forbidden in the electric dipole approximation[2]. The resultant weak effects produce weak signals, and thus require high sample densities. In contrast, nonlinear techniques probing electric-dipole-allowed effects have been used for sensitive chiral analyses of liquid samples[3–7]. Here we extend this class of approaches by carrying out nonlinear resonant phase-sensitive microwave spectroscopy of gas phase samples in the presence of an adiabatically switched non-resonant orthogonal electric field; we use this technique to map the enantiomer-dependent sign of an electric dipole Rabi frequency onto the phase of emitted microwave radiation. We outline theoretically how this results in a sensitive and species-selective method for determining the chirality of cold gas-phase molecules, and implement it experimentally to distinguish between the S and R enantiomers of 1,2-propanediol and their racemic mixture. This technique produces a large and definitive signature of chirality, and has the potential to determine the chirality of multiple species in a mixture.

Our approach to determining chirality depends only on the parity conserving Hamiltonian of an asymmetric top in an external electric field[8]: the three rotational constants A, B and C and the corresponding dipole moment component magnitudes $|\mu_a|$, $|\mu_b|$ and $|\mu_c|$ determine the rotational energy levels of such a molecule, and further specification of the signs of μ_a, μ_b and μ_c fully determines its chirality (Fig. 1). The sign of any two of the three dipole moment components μ_a, μ_b and μ_c is arbitrary and changes with the choice of axes, whereas the sign of the combined quantity $\mu_a\mu_b\mu_c$ is axis independent and changes sign with enantiomer. Equivalently, any measurement of the combined quantity $\mu_a\mu_b\mu_c$, and indeed any measurement of chirality, is time-even and parity-odd. In the context of direct molecular chirality detection, parity violating energy shifts arising from the weak interaction have been predicted[9–12] and could in principle distinguish enantiomers; but the resultant very small frequency shifts have never been observed and are not relevant to the techniques demonstrated here.

The Rabi frequency describing an electric dipole transition between rotational states of a chiral molecule differs in sign for opposite enantiomers[13], and we demonstrate here theoretically and experimentally that it can be mapped onto the phase of emitted radiation by applying an electric field that varies during the molecular coherence time. This switched electric field, E_x, combined with an applied microwave field, E_z, induces $\hat{y}$-polarized oscillations of the molecules' dipole, causing them to emit radiation polarized along the y axis, orthogonal to both E_x and E_z (Fig. 2). This radiation is detected and amplified, with the extracted signal phase ϕ differing by 180° between left- and right-handed (S and R) enantiomers and indicating the dominant enantiomer (that is, the sign of the enantiomeric excess), while the signal magnitude V indicates the magnitude of the enantiomeric excess. With known molecular constants A, B, C, μ_a, μ_b and μ_c, the enantiomer-dependent

signal can be accurately predicted via numerical integration of the well-known asymmetric-top Hamiltonian (Fig. 2b).

We validate our technique using R-, S- and racemic 1,2-propanediol. This molecule was chosen because the relevant molecular constants are well characterized ($A = 8{,}572.05$ MHz, $B = 3{,}640.10$ MHz, $C = 2{,}790.96$ MHz, $\mu_a = 1.2$ D, $\mu_b = 1.9$ D, $\mu_c = 0.36$ D; ref. 14) and because it is readily available commercially in enantiopure form. 1,2-propanediol and larger molecules occupy a large number of quantum states at room temperature, which dilutes the signal obtained from a single rotational level (for example, more than 5,000 occupied states for 1,2-propanediol). To reduce the number of occupied states and increase the resonant polarizability of the sample, we therefore substantially cool the molecular gas using techniques developed previously and discussed in detail elsewhere[15]: the warm molecules are injected into a cryogenic buffer gas cell thermally anchored to a closed-cycle pulse-tube refrigerator and cooled to a temperature of $\sim$7 K. However, other sources of cold molecules could also be used. A natural choice would be a pulsed supersonic beam source (rotational temperature $\sim$2 K), which is frequently used as a cold molecule beam source for Fourier transform microwave spectroscopy[16].

The experimental set-up, sketched in Fig. 3, has two walls of the cryogenic cell formed by mirrors that comprise a tunable plano-concave Fabry–Pérot microwave cavity. The cavity is used to excite and detect the molecules, with each of its transverse and longitudinal spatial modes supporting two (degenerate) modes of orthogonal polarization. These modes can be separately addressed via waveguides attached to the planar mirror and coupled to the cavity via apertures (denoted A and B in
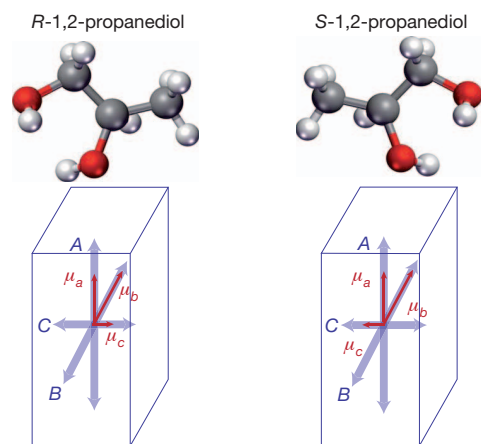

Figure 1 | The two enantiomers of chiral 1,2-propanediol. The Hamiltonian of a rigid chiral molecule in an external electric field is enantiomer-dependent. In any chiral molecule, opposite enantiomers have the same rotational constants A, B and C, and the same magnitude of dipole moment components $|\mu_a|$, $|\mu_b|$ and $|\mu_c|$, but the sign of the combined quantity $\mu_a\mu_b\mu_c$ is distinct for each enantiomer, independent of choice of axes. The displayed orientation of the molecules is for illustrative purposes.

[1]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Center for Free-Electron Laser Science, D-22607 Hamburg, Germany. [3]Max-Planck-Institut für Kernphysik, D-69117 Heidelberg, Germany.
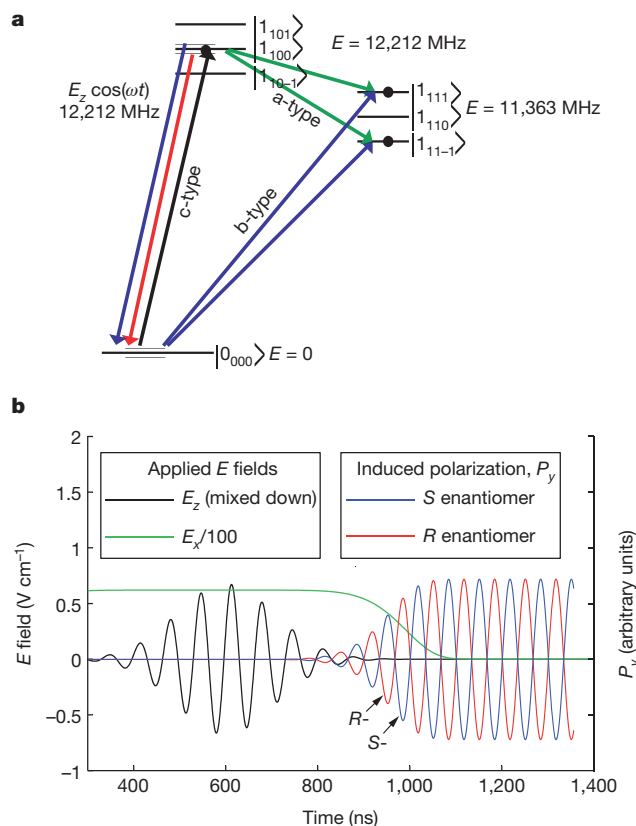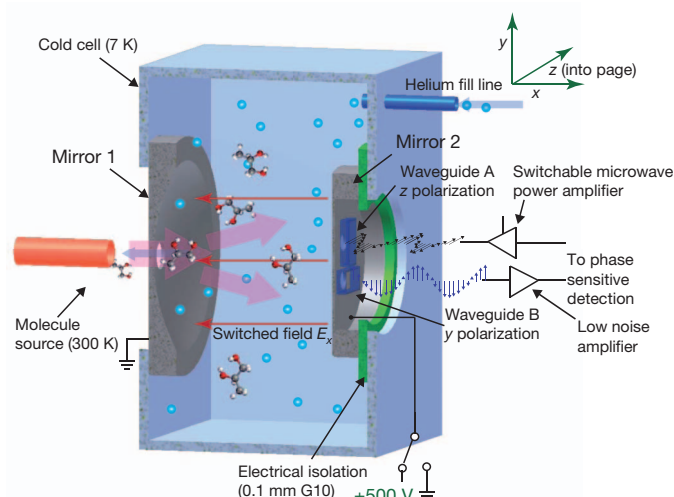
Figure 3 | A cryogenic buffer gas Fourier transform microwave spectrometer that provides enantiomer-specific detection. Molecules are introduced into a cryogenic cell from a room-temperature tube held close to an aperture in mirror 1. Microwave radiation is introduced via waveguides via coupling apertures in mirror 2. This mirror can be rapidly switched between ±500 V and ground, applying a time-varying electric field in the $\hat{x}$ direction; it is insulated from the cell by G10 fibreglass. The molecules are polarized initially in the $\hat{z}$ direction via a linearly polarized microwave pulse coupled from waveguide A. The resulting enantiomer-specific radiation is coupled out of the cavity via waveguide B, which is oriented to detect $\hat{y}$-polarized microwaves.

Figure 2 | 1,2-propanediol as a rigid rotor. a, The relevant rotational level structure of 1,2-propanediol. Each state is designated with $|J_{k_{-1}k_1m}\rangle$. Molecules initially in the ground state are prepared in a superposition of $|0_{000}\rangle$ and $|1_{100}\rangle$ via a c-type microwave transition. A change in the electric field E_x mixes in components of $|1_{101}\rangle$ and $|1_{10-1}\rangle$ with complex phases proportional to μ_a. Allowed electric dipole radiation between these admixed states and the ground state produces an oscillating electric field in the laboratory $\hat{y}$ direction via a b-type transition. In the relevant limit of small E_z and E_x, this electric field E_y is proportional to $\mu_a\mu_b\mu_c$ and therefore changes sign with enantiomer. The applied field E_x of 65 V cm^{-1} induces a 500 kHz Stark splitting in the $|1_{10m}\rangle$ states that is not relevant to the dynamics due to the short (~200 ns) duration of the field. Admixing with the $|1_{01}\rangle$ state at 6,431 MHz (not shown) also contributes a small amount to the chiral signature. **b,** A simulation of 1,2-propanediol in applied electric fields E_z (microwave) and E_x. The frequency of E_z is 12,212 MHz, corresponding to the $|0_{00}\rangle \rightarrow |1_{10}\rangle$ c-type transition. In the figure, this field is mixed down to 15 MHz for clarity. Also shown is the simulated molecular polarization P_y for each enantiomer, induced by the change in the time-varying electric field E_x. P_y oscillates at the same 12,212 MHz frequency as E_z, and is also shown mixed-down to 15 MHz for visibility. The enantiomer-dependent phase of P_y is evident. To within experimental uncertainty, the sequence of applied fields shown here corresponds to the fields used to produce the enantiomer-dependent signals shown in Fig. 4.

Fig. 3). A time-varying electric field E_x parallel to the cavity axis can be applied by rapidly changing the voltage of the planar mirror ($V_{mirror} = \pm 500$ V, $E_x = \pm 65$ V cm^{-1}); this voltage is controlled with high-voltage switches with a switching time of ~100 ns.

In the experiments, molecules enter the cell continuously from a warm (300 K) feed tube and immediately start to cool through collisions with the cold helium buffer gas. By the time they diffuse into the central region of the cell, they have reached a rotational temperature of ~7 K. The cold molecules remain in the gas phase as they diffuse through the cell for several milliseconds, until they arrive at a cold cell wall and freeze there. Our detected signal comes only from the cold gas-phase molecules. The experimental sequence of applied electric fields begins with the application of E_x (see Fig. 3). The cavity is then driven with a strong, linearly polarized microwave field $E_z(t) = \hat{z}E_{mw}\cos(\omega t)$. Here t is time, $\hat{z}$ is the unit vector along z, and

E_{mw} is the field strength of the applied microwave field in the z direction. The excitation frequency ω is tuned to the $|0_{00}\rangle \rightarrow |1_{10}\rangle$ rotational transition of the ground-state conformer of 1,2-propandiol at 12,212 MHz. (We use the $|J_{k_{-1}k_1}\rangle$ nomenclature of ref. 8, where k_{-1} and k_1 indicate the quantum numbers of the limiting prolate and oblate symmetric top respectively; when m is specified, we use $|J_{k_{-1}k_1m}\rangle$).

The maximum magnitude of E_z and the pulse length τ are adjusted to make $|\Omega|\tau \lesssim \pi/2$ for all molecules, where Ω is the Rabi frequency. This $\hat{z}$-polarized microwave pulse induces an oscillating electric dipole polarization, P_z, in the molecular ensemble. The electric field E_x is set to zero 200 ns after the end of the microwave pulse; E_x is switched over about 100 ns—rapidly compared to the molecular dephasing but slowly compared to ω. The change in E_x induces a sizable fraction of the oscillating molecular dipole to radiate at the same frequency ω with $\hat{y}$ polarization. Under our experimental conditions, the chirality-dependent polarization P_y has about 10% of the amplitude of P_z. The induced field E_y, which has an enantiomer-dependent phase, is amplified and recorded. The molecules continue to radiate in this manner until they rethermalize rotationally via collisions with helium atoms, typically after about 5 μs. This completes the experimental cycle, which can be started again by turning E_x back on. As in conventional Fourier transform microwave spectroscopy, molecules in distinct parts of the cavity radiate constructively into the original cavity mode used to polarize the sample. This important advantage is retained here because modes with orthogonal polarizations share an identical spatial structure.

Figure 4 shows averaged signals for the S and R enantiomers of 1,2-propanediol, and for a racemic mixture. Each signal represents the difference between a time trace taken with $E_x = +65$ V cm^{-1} and a trace taken with $E_x = -65$ V cm^{-1}. As predicted, the S and R enantiomers show a clear 180° phase shift. The opposite enantiomers and a racemic mixture are resolved with exceptionally high confidence. Our Hamiltonian integration calculations (shown in Fig. 2) are in excellent agreement with experimental data.

We note that the common spectroscopic methods of circular dichroism and vibrational circular dichroism can resolve spectral signatures sufficiently to deconvolve a chiral mixture that is comparatively simple[17],

whereas our method has the ability to identify an enantiomer amid a complex mixture of many other (possibly chiral) components. Although the chiral signature could be misread as for a different molecule, this would only occur in the very unlikely case of a second molecule having a rotational transition within the frequency range of the observed free-induction decay (in this case $12,212.14 \pm 0.02$ MHz). We expect that straightforward improvements to the electronics will allow us to resolve enantiomers present in concentrations as low as one part per thousand in a complex mixture, and that combining our method with broadband chirped-pulse spectroscopy techniques[18] will enable rapid and simultaneous species-specific and enantiomer-specific identification of many chiral components in a complex mixture.

Any coherent method of enantiomer-specific analysis can in principle be converted into a method of enantiomer purification[13,19–21], which could be realized here by adding a third, phase-controlled field pulse E_y; appropriate choice of the phase of E_z, E_x and E_y would transfer molecules of only one enantiomer to an excited state that could then be selectively addressed by additional fields.

We have applied our method to a single molecular species, but our simulations have identified applicable microwave transitions in a wide range of molecules. The only essential requirement is that the target molecule has at least one conformer with non-zero dipole moment components μ_a, μ_b and μ_c. Larger molecules, including most chiral molecules, typically exhibit several stable conformers and it is overwhelmingly likely that at least one of them meets this criterion. Our simulations of the enantiomer-dependent signals for molecules as diverse as carvone, nicotine and alanine predict a strong signal in every case, and we therefore expect that the upper size limit of molecules amenable to this technique is set by the limitations of gas-phase microwave spectroscopy (which has been used to characterize neurotransmitters, amino acids and drugs[22–24]). Furthermore, the absolute configuration of previously uncharacterized molecules could be resolved by using the technique presented here. The relevant molecular constants of 1,2-propanediol and the molecules listed above were calculated from the microwave spectra of cold molecular samples; such spectra are also measurable in our apparatus.

METHODS SUMMARY

Enantiopure R-1,2-propanediol and S-1,2-propanediol (Sigma Aldrich) were used. The free-induction decay signals were recorded using standard Fourier transform microwave spectroscopy techniques[16], with a typical polarization pulse length of 200 ns. The signal was amplified, mixed down to an intermediate frequency of 20.2 MHz, further amplified, and digitized by a fast signal averager (Agilent U1084). Each data point in Fig. 4b represents 140,000 averages, about 20 s of experimental time, and 1 mg of 1,2-propanediol. The apparatus used here could identify the absolute configuration of an unknown enantiomer with 99% accuracy in less than 50 ms.

Full Methods and any associated references are available in the online version of the paper.

1. Busch, K. W. & Busch, M. A. (eds) *Chiral Analysis* (Elsevier, 2006).
2. Fischer, P. & Hache, F. Nonlinear optical spectroscopy of chiral molecules. *Chirality* **17,** 421–437 (2005).
3. Fischer, P., Wiersma, D. S., Righini, R., Champagne, B. & Buckingham, A. D. Three-wave mixing in chiral liquids. *Phys. Rev. Lett.* **85,** 4253–4256 (2000).
4. Li, Y. & Bruder, C. Dynamic method to distinguish between left- and right-handed chiral molecules. *Phys. Rev. A* **77,** 015403 (2008).
5. Rhee, H., Choi, J.-H. & Cho, M. Infrared optical activity: electric field approaches in time domain. *Acc. Chem. Res.* **43,** 1527–1536 (2010).
6. Hiramatsu, K. *et al.* Observation of Raman optical activity by heterodyne-detected polarization-resolved coherent anti-Stokes Raman scattering. *Phys. Rev. Lett.* **109,** 083901 (2012).
7. Hirota, E. Triple resonance for a three-level system of a chiral molecule. *Proc. Jpn Acad. B* **88,** 120–128 (2012).
8. Townes, C. & Schawlow, A. *Microwave Spectroscopy* (Dover Publications, 1975).
9. Darquié, B. *et al.* Progress toward the first observation of parity violation in chiral molecules by high-resolution laser spectroscopy. *Chirality* **22,** 870–884 (2010).
10. Quack, M. How important is parity violation for molecular and biomolecular chirality? *Angew. Chem. Int. Edn* **41,** 4618–4630 (2002).
11. Schnell, M. & Küpper, J. Tailored molecular samples for precision spectroscopy experiments. *Faraday Discuss.* **150,** 33–49 (2011).
12. Quack, M., Stohner, J. & Willeke, M. High-resolution spectroscopic studies and theory of parity violation in chiral molecules. *Annu. Rev. Phys. Chem.* **59,** 741–769 (2008).
13. Jacob, A. & Hornberger, K. Effect of molecular rotation on enantioseparation. *J. Chem. Phys.* **137,** 044313 (2012).
14. Lovas, F. J. *et al.* Microwave spectrum of 1,2-propanediol. *J. Mol. Spectrosc.* **257,** 82–93 (2009).
15. Patterson, D. & Doyle, J. M. Cooling molecules in a cell for FTMW spectroscopy. *Mol. Phys.* **110,** 1757–1766 (2012).
16. Balle, T. J. & Flygare, W. H. Fabry-Perot cavity pulsed Fourier transform microwave spectrometer with a pulsed nozzle particle source. *Rev. Sci. Instrum.* **52,** 33–45 (1981).
17. Guo, C. *et al.* Determination of enantiomeric excess in samples of chiral molecules using Fourier transform vibrational circular dichroism spectroscopy: simulation of real-time reaction monitoring. *Anal. Chem.* **76,** 6956–6966 (2004).
18. Brown, G. G. *et al.* A broadband Fourier transform microwave spectrometer based on chirped pulse excitation. *Rev. Sci. Instrum.* **79,** 053103 (2008).
19. Král, P., Thanopulos, I., Shapiro, M. & Cohen, D. Two-step enantio-selective optical switch. *Phys. Rev. Lett.* **90,** 033001 (2003).
20. Thanopulos, I., Král, P. & Shapiro, M. Theory of a two-step enantiomeric purification of racemic mixtures by optical means: the D₂S₂ molecule. *J. Chem. Phys.* **119,** 5105–5116 (2003).
21. Gerbasi, D., Brumer, P., Thanopulos, I., Král, P. & Shapiro, M. Theory of the two step enantiomeric purification of 1,3 dimethylallene. *J. Chem. Phys.* **120,** 11557–11563 (2004).
22. Pate, B. H. & De Lucia, F. C. (eds) Special issue: Broadband molecular rotational spectroscopy. *J. Mol. Spectrosc.* **280,** 1–2 (2012).
23. Mata, S., Peña, I., Cabezas, C., López, J. & Alonso, J. A broadband Fourier transform microwave spectrometer with laser ablation source: the rotational spectrum of nicotinic acid. *J. Mol. Spectrosc.* **280,** 91–96 (2012).
24. Grabow, J.-U. *et al.* Rapid probe of the nicotine spectra by high-resolution rotational spectroscopy. *Phys. Chem. Chem. Phys.* **13,** 21063–21069 (2011).

Figure 4 | Enantiomer-dependent microwave radiation. a, The enantiomer-dependent free-induction decay traces for S-1,2-propanediol (blue trace), R-1,2-propanediol (red) and racemic 1,2-propanediol (black). As predicted, opposite enantiomers show a $180°$ phase difference. **b,** Repeated measurements of enantiomeric excess (horizontal axis) and of total molecule number (vertical axis) for S-, R- and racemic 1,2-propanediol. Each data point represents about 20 s of experimental time, and 1 mg of sample. Enantiomers are clearly resolved.

METHODS

Numerical simulations. For a given sequence of applied fields, the simulation proceeds as follows: the matrix representations of the field-free Hamiltonian H_0 and the three Cartesian components of the dipole moment operator $\hat{\mu}_j$ ($j = x,y,z$) are calculated following ref. 8. The off-diagonal elements in these matrices are enantiomer dependent. The matrices are calculated in the oblate symmetric top basis with an arbitrary cut-off of rotational quantum number $J \leq J_{max}$. $J_{max} = 2$ for the simulation results shown in Fig. 2.

Molecules are assumed to start in an eigenstate of $\hat{H}_0$, $\psi(t=0) = \psi_{Jk_{-1}k_1 m}$. For each time step Δt the electric fields to be applied at time t are calculated, and the Hamiltonian $\hat{H}(t) = \hat{H}_0 - E_x(t)\mu_x - E_y(t)\mu_y - E_z(t)\mu_z$ for an asymmetric top subject to those fields is calculated. Ψ is then evolved, according to:

$$\psi(t + \Delta t) = e^{\frac{-i\hat{H}(t)\Delta t}{\hbar}}\psi(t)$$

The time step Δt is chosen to be significantly faster than any frequency in the simulation. The induced laboratory-frame polarizations $P_j = \langle \mu_i \rangle$ are then calculated. $P_y(t)$, which has an enantiomer-dependent phase for our choice of applied fields, is plotted in Fig. 2. To simulate a thermal ensemble, this procedure is repeated for all initial states Ψ_0 up to $J = J_{max}$ and the resulting polarizations, weighted by the Boltzmann factor $e^{-\varepsilon/k_B T}$ for each initial state, are summed to give the experimentally observable ensemble polarization. The validity of the simulations was checked by comparing predicted levels, line strengths and Stark shifts with the PGOPHER simulation package[25]. All simulations were written and performed in MATLAB.

Experimental details. Enantiopure R-1,2-propanediol and S-1,2-propanediol (Sigma Aldrich) were used. In the experiments, a flow rate of 1.5×10^{18} helium atoms s^{-1} into the cell provides an estimated in-cell helium density of 10^{14} atoms cm^{-3}. Approximately 5×10^{17} 1,2-propanediol molecules s^{-1} are sprayed towards the cell aperture; we estimate the 1,2-propanediol density within the cell to be about 10^{12} molecules cm^{-3}. The experimental repetition rate of 7 kHz is set (and limited by) the maximum switching frequency of our high-voltage switches. Without this limitation, or when E_x is set to zero, the rate is limited only by the molecular rethermalization rate of 200 kHz. The applied electric field E_x is ± 65 V cm^{-1} (± 500 V on mirror 2). This high voltage is switched via Behlke HTS 151 high-voltage MOSFET switches. The polarizing pulse is typically 200 ns in duration, and our maximum microwave field E_z is estimated to be 0.5 V cm^{-1} from measurements of the Rabi frequency Ω. The cavity was run in the TEM$_{11,n=6}$ mode; the more natural choice of the TEM$_{00}$ mode could not be used because it was highly damped by loss through the molecule input aperture.

Mirror 2 of the cavity was mounted on flexible bellows and could be moved axially by about 1 cm, tuning the cavity (see Fig. 3 for cavity details). This tuning was accomplished via three thermally isolating, flexible shafts connected via rotary feedthroughs to knobs outside the Dewar. The aluminium cavity has a measured finesse of 105, a length of 8.1 cm, and a radius of curvature on its spherical mirror of 22.5 cm. The input aperture for the molecules has a diameter of 1.1 cm, and the coupling apertures A and B have diameters of 0.8 cm. The waveguides connected via apertures A and B are WR-62, operating from 12 to 18 GHz. E_y is detected after waiting ~2 µs for the cavity to ring down. The signal is amplified by a low-noise amplifier (Pasternack PE1524) connected immediately outside the vacuum chamber with no protection diodes or switches. The amplified signal is mixed down to an intermediate frequency of 20.2 MHz, further amplified, and digitized by a fast signal averager (Agilent U1084). Each data point in Fig. 2b represents 140,000 averages, about 20 s of experimental time, and 1 mg of 1,2-propanediol. Our signal to noise ratio is sufficient to identify the absolute configuration of an unknown enantiomer with 99% accuracy in less than 50 ms.

The chirality-dependent signal E_y is proportional to $E_x E_z$. In order to cancel some systematic offsets in the detected microwave field, we exploit the change of sign of E_y with E_x by subtracting traces taken with equal and opposite values of E_x. This systematic offset in E_y arises from crosstalk between what are nominally orthogonal cavity modes. Even with this cancellation, the racemic 1,2-propanediol sample produces a small but statistically significant non-zero signal. A plausible explanation for this is imperfect reversal of our electric field; distinct MOSFET switches, resistor networks and power supplies are used for the positive-going and negative-going applied fields, and small errors in the voltage or timing could result in artefacts similar to those observed.

25. Western, C. M. *PGOPHER, a Program for Simulating Rotational Structure* Version 7.1 (Univ. Bristol, 2010).

Weakened stratospheric quasibiennial oscillation driven by increased tropical mean upwelling

Yoshio Kawatani[1] & Kevin Hamilton[2]

The zonal wind in the tropical stratosphere switches between prevailing easterlies and westerlies with a period of about 28 months[1]. In the lowermost stratosphere, the vertical structure of this quasibiennial oscillation (QBO) is linked to the mean upwelling[2–4], which itself is a key factor in determining stratospheric composition. Evidence for changes in the QBO have until now been equivocal, raising questions as to the extent of stratospheric circulation changes in a global warming context. Here we report an analysis of near-equatorial radiosonde observations for 1953–2012, and reveal a long-term trend of weakening amplitude in the zonal wind QBO in the tropical lower stratosphere. The trend is particularly notable at the 70-hectopascal pressure level (an altitude of about 19 kilometres), where the QBO amplitudes dropped by roughly one-third over the period. This trend is also apparent in the global warming simulations of the four models in the Coupled Model Intercomparison Project Phase 5 (CMIP5) that realistically simulate the QBO. The weakening is most reasonably explained as resulting from a trend of increased mean tropical upwelling in the lower stratosphere. Almost all comprehensive climate models have projected an intensifying tropical upwelling in global warming scenarios[5–7], but attempts to estimate changes in the upwelling by using observational data have yielded ambiguous, inconclusive or contradictory results[8–10]. Our discovery of a weakening trend in the lower-stratosphere QBO amplitude provides strong support for the existence of a long-term trend of enhanced upwelling near the tropical tropopause.

Unlike the annual or diurnal cycles, the QBO has no direct periodic astronomical forcing, but its effects on the zonal wind display a remarkable degree of regularity and persistence. Each QBO zonal wind cycle studied has had the same characteristic pattern of descending shear zones, and individual cycles have had periods ranging between 21 and 36 months (Supplementary Fig. 2). The zonal-wind QBO near the Equator has a characteristic vertical amplitude structure; it is near zero just below the equatorial tropopause ($\sim$100 hPa) and then increases rapidly with height in the lowermost stratosphere to a peak at around 20 hPa.

The mean wind accelerations during the QBO are thought to be largely forced by interactions of the mean zonal wind with vertically propagating internal waves generated in the tropical troposphere[1,11] (Methods). This mechanism has been incorporated into simple dynamical models[2,12,13] and has been shown to operate in comprehensive global climate models[11,14,15]. The period and vertical structure of the oscillation depend on the wave fluxes entering the tropical stratosphere and also on the mean tropical upwelling, which is associated with the mean meridional circulation of the stratosphere, or Brewer-Dobson circulation. The wave-induced QBO descending phases are opposed by the mean upwelling and, in models at least, the penetration of the QBO is very strongly dependent on the strength of the mean upwelling[2].

A number of earlier studies have looked at whether the period of the QBO exhibits systematic long-term trends or dependence on

such factors as solar activity or the El Niño/Southern Oscillation (ENSO)[16–21]. There does not seem to be a clear long-term trend in the period of the QBO, and we still lack a detailed explanation of the cycle-to-cycle variability of the QBO period apparent in the observations.

Whereas the variability of the period has been studied extensively, there are apparently no earlier studies on the question of systematic long-term QBO amplitude changes, possibly because a simple inspection of the raw time-series data (Supplementary Fig. 2) indicates that cycle-to-cycle amplitudes are much more stable than the period. The issue of possible long-term changes in amplitude arose from our recent global modelling study of the response of the simulated QBO to increased greenhouse gas concentrations[3,4]. We found a projected small increase in the mean period, but one that would be hard to detect amid the large cycle-to-cycle variability in the observed record. Also apparent was a significant reduction in QBO amplitude, particularly near the lowest stratospheric levels. In the model, the reduction in the QBO amplitude is clearly caused by an enhancement of the mean upwelling rate of the tropospheric air into the stratosphere (Supplementary Fig. 1).

Global models forced with scenarios of rising greenhouse gas concentrations project large increasing trends of mean upwelling in the tropical upper troposphere and lower stratosphere[7] and such trends will play an important role in long-term changes in stratospheric water vapour and other aspects of stratospheric chemistry[22]. Trends in stratospheric water vapour and other stratospheric trace constituents in
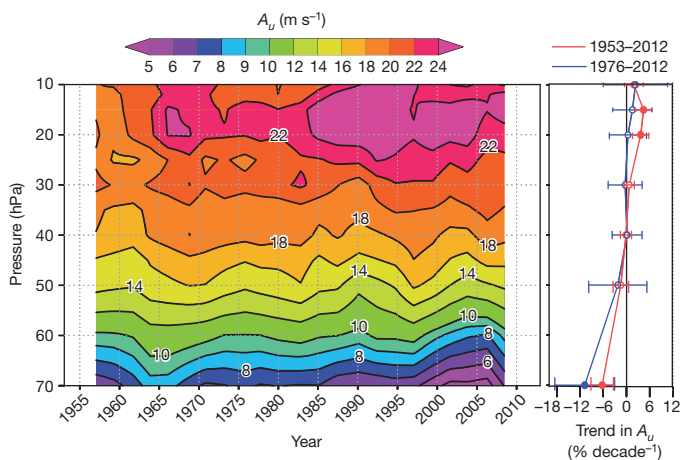

Figure 1 | Time–height section of the three-cycle mean amplitude of the observed QBO, and the height dependence of the amplitude trend over time. In the main panel, the horizontal axis represents the central date of the three-cycle mean of A_u; contour intervals are 1 m s^{-1} and 2 m s^{-1} for values less and greater than 10 m s^{-1}, respectively. The right panel shows the linear regression trends in the QBO amplitude (in per cent per decade) over time with height from 1953 to 2012 (red) and 1976 to 2012 (blue). Filled circles indicate that the trend at that level is different from zero with a statistical significance $P < 0.05$. Error bars denote $P < 0.05$ confidence intervals in the trend estimates.

[1]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, 3173-25 Showamachi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan. [2]International Pacific Research Center, SOEST, University of Hawaii at Manoa, 1680 East West Road, Honolulu, Hawaii 96822, USA.

turn can make a significant contribution to the global mean radiative forcing of climate[23]. In this work, we analyse the observed prevailing equatorial wind record to identify any significant long-term trends of the QBO amplitude in the lower stratosphere that may also indicate changes in the mean upwelling.

The data analysed here are monthly mean values of operational balloon-borne radiosonde observations from near-equatorial stations[24] for January 1953 through to the end of December 2012. This data set is pieced together from three separate near-equatorial stations: Canton Island (2.8° S, 171.7° W) for the period January 1953 to August 1967, Gan (0.7° S, 73.2° E) for September 1967 to December 1975, and Singapore (1.4° N, 104.0° E) for January 1976 to present (Methods). The QBO amplitude in the lower stratosphere is thought to be almost zonally symmetric, and the discontinuity associated with changing station location should be small[25].

At each height, we define a time-varying measure of the amplitude of the QBO, A_u (see Methods). A_u is defined over overlapping chunks of the record spanning 3 cycles of the QBO (roughly a 7 year running mean) and so smooths out cycle-to-cycle variability that might be associated with, for example, ENSO[20]. The time–height section of A_u is plotted in Fig. 1. The red curve in the right panel shows the linear regression trend in A_u at each level over the whole record, while the blue curve shows results calculated using only the Singapore (January 1976 and later) data. Trends are expressed in terms of percentage per decade at each level. An overall trend towards a much weaker QBO at 70 hPa and a slightly stronger QBO at 30 hPa and higher is evident in Fig. 1. The trend at 70 hPa is statistically significant ($P < 0.05$) in the record for the entire period and also when only post-1975 data are

considered. The linear trend at 70 hPa implies a reduction in the mean amplitude of about one-third over 1953–2012. The data set we analysed to produce Fig. 1 has been widely applied in QBO studies, and is notable for the long record from stations very close to the Equator. There are also data available from other very low-latitude stations with reasonably long records and relatively few gaps. Analysis of the records from 5 other such stations, as well an extended record from Singapore, reveals negative QBO amplitude trends in each case (Supplementary Fig. 3). We believe that this is the first report of a long-term trend in the observed QBO amplitude.

Although the linear trend at 70 hPa is statistically significant, shorter-timescale variations in A_u are also evident. Notably, one of the deepest penetrations of the mean easterlies into the lowermost stratosphere occurred in boreal spring 2010 (although this easterly phase persisted for an unusually short time at 70 hPa; see Supplementary Fig. 2). This one exceptional event accounts for the increase in A_u at 70 hPa in the final point plotted in Fig. 1.

We have used the same approach to examine trends in the equatorial zonal-mean wind QBO in the CMIP5 forced climate simulations. Most comprehensive global models do not simulate anything resembling the real QBO in the tropical stratosphere, but the most recent phase of the CMIP is notable for having four models that do have a reasonable QBO. These models include a parameterization of small-scale gravity waves (Methods). We analysed a single realization of the CMIP5 'historical runs' (forced with observed greenhouse gas and aerosol concentrations) from 1900–2005 along with the continuation for the 'rcp4.5 scenario' projections of climate forcing for the remainder of the twenty-first century[26]. The main panels of Fig. 2 show A_u calculated from the equatorial zonal-mean zonal wind in each of the four models. The black curves in the right panels show the linear regression trends based on all 200 years of model data and include $P < 0.05$ confidence estimates. With all 200 years of data the trends are well determined, but there is enough shorter-period variability that the trends over a period of only 60 years (that is, the observed record) are much more variable. A simple way to look at this aspect of the variability is provided by the red curves, which show the trend in A_u using model results only from the 1953–2012 period.

Over the full 200 year record, all four model simulations show a gradual, but quite persistent, weakening trend for A_u, with occasional strong variations on decadal scales. The weakening is seen up to 10–15 hPa, but is very noticeable (particularly in percentage terms) at 70 hPa. By the end of the twenty-first century, the amplitude at 70 hPa is projected by these models to be severely reduced from that at the beginning of the twentieth century.

Although each model shows consistent trends in the 200-year and 60-year results at 70 hPa (statistically significant at $P < 0.05$ in all negative trends, except at $P < 0.06$ in MIROC-ESM-CHEM and insignificant in MPI-ESM-MR in 60-year results; see open or filled circles in Fig. 2), there is more difference, even in sign, among the trends at higher levels. This suggests that the small increasing trends seen at 30–10 hPa in observations (Fig. 1) may not be representative of the very long-term global warming driven changes. The positive trend in A_u at 15–20 hPa is statistically significant from 1953 to 2012, but is not from 1976 to 2012 (that is, using only the Singapore data), and is much smaller in percentage terms than the negative trend at 70 hPa. The radiosonde record analysed here is from three near-equatorial stations at different longitudes, and so the pre-1976 results may have spurious changes introduced by the changes in station locations. This is probably a bigger issue for the middle stratospheric levels (as the QBO amplitude there has some modest ($\sim$10%) longitudinal variation) than at 70 hPa (where the QBO is expected to be almost zonally-symmetric)[25].

Figure 3 shows the individual model time series of A_u at 70 hPa compared with the observations. The decreasing trend in the observations is larger than that of any of the models, but the overall behaviour in terms of quasi-decadal variability imposed on a very long-term gradual decrease is quite similar in the observations and the model time series.
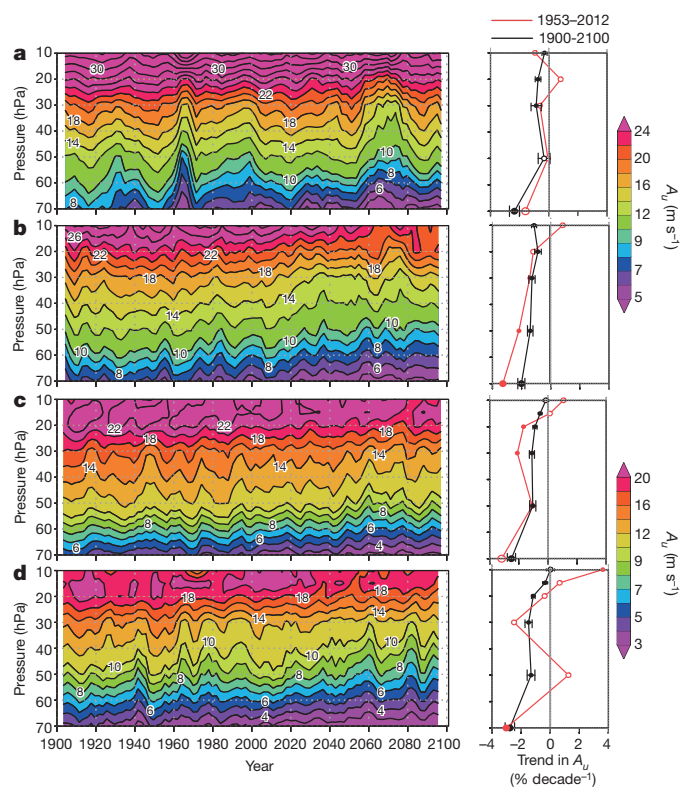

Figure 2 | Time–height section of the three-cycle mean amplitude of the simulated QBO, and height dependence of amplitude trend over time. In the main panel, we show the modelled zonal mean amplitude of the QBO over the Equator for MPI-ESM-MR (**a**), HadGEM2-CC (**b**), MIROC-ESM-CHEM (**c**) and MIROC-ESM (**d**) for about 200 years. Contour intervals are the same as in Fig. 1. The use of colour is the same as in Fig. 1 for **a** and **b**, but shifted 2 m s^{-1} downward for **c** and **d**. The right panels are as Fig. 1 but for 1953 to 2012 (red) and the twentieth to the twenty-first century (black). Other features are as in Fig. 1.

The penetration of the QBO into the lowermost stratosphere is controlled by the mean vertical upwelling near the Equator[2–4]. The CMIP5 model data archive does not include the information needed to calculate the transformed-Eulerian mean (TEM) circulation (which is a good approximation to the Lagrangian mean upwelling[27]). Instead, we have examined trends in the Eulerian vertical velocity, which may roughly approximate the trends in TEM circulation near the Equator (Supplementary Fig. 5). These trends at 70 hPa are shown for the models in Fig. 3. The mean vertical velocity was not available for the Hadley Centre HadGEM2-CC model, so results from other versions of the Hadley Centre model are presented. In each case, an increasing trend in the mean upwelling velocity is apparent. The increasing trend is strong in the upper troposphere and lower stratosphere region, whereas the trend at 10–30 hPa is smaller and much more variable in all models (see Supplementary Fig. 6).

The present analysis shows that both models and the observations display a robust weakening trend in 70 hPa QBO amplitude, but it is reasonable to ask as well whether the observed and simulated changes in QBO period agree. It seems that no trend in the period of the QBO can be detected in the 60 years of observations. The simulated model trends are inconsistent in sign among the models, and are small

enough that they would be very hard to detect in just a 60-year record (Supplementary Figs 7 and 8).

We also analysed 200 years of output from the CMIP5 'pre-industrial' runs with fixed climate forcing (Supplementary Fig. 9). The trends in the QBO amplitude and mean upwelling in these runs are extremely small, indicating that the trends seen in Figs 2 and 3 are due to the forced global warming.

Efforts have been made to characterize the robust aspects of model projected changes expected from anthropogenic climate forcing as a 'fingerprint' pattern that can be observationally evaluated[28]. This idea has been generally applied to the spatial patterns of mean warming but could be generalized to include other aspects of the climate system. Our present results suggest that the decrease of QBO amplitude in the lowermost stratosphere is a consistent projection of global models and represents a subtle, but telling, part of the 'fingerprint' of the expected response of the climate system to anthropogenic climate forcing. Although the possible effects of changes in observing practices on radiosonde measurements of wind speeds have been tested less rigorously than for measurements of temperatures, our results are unlikely to be significantly affected by such potential inhomogeneities (Supplementary Information section 4) and they support a significant decline of QBO amplitude over the past six decades.

METHODS SUMMARY

Dunkerton and Delisi[29] suggested a simple approach to characterizing the QBO amplitude. They calculated the standard deviation, σ_u, for each level of the monthly mean time series after the mean seasonal cycle over the entire record had been removed. They then estimated the amplitude as $A_u = \sqrt{2}\sigma_u$, assuming that almost all variability in the monthly means in the tropical stratosphere can reasonably be attributed to the QBO. We apply a similar technique to search for trends. We calculate A_u at each level for overlapping time chunks spanning three consecutive QBO cycles. The beginning and end of each chunk were based on the dates on which winds at 30 hPa were observed to change from westerly to easterly. The wind observations from January 1953 to December 2012 were used to define the mean seasonal cycle for the period, which was then subtracted from the raw data. The resulting series were then smoothed by taking 5-month running means (for example, Supplementary Fig. 4). Then, A_u was calculated for the first 3-cycle chunk using these deseasonalized and slightly smoothed data for the periods December 1953–February 1960, February 1956–March 1962, and so on.

Full Methods and any associated references are available in the online version of the paper.
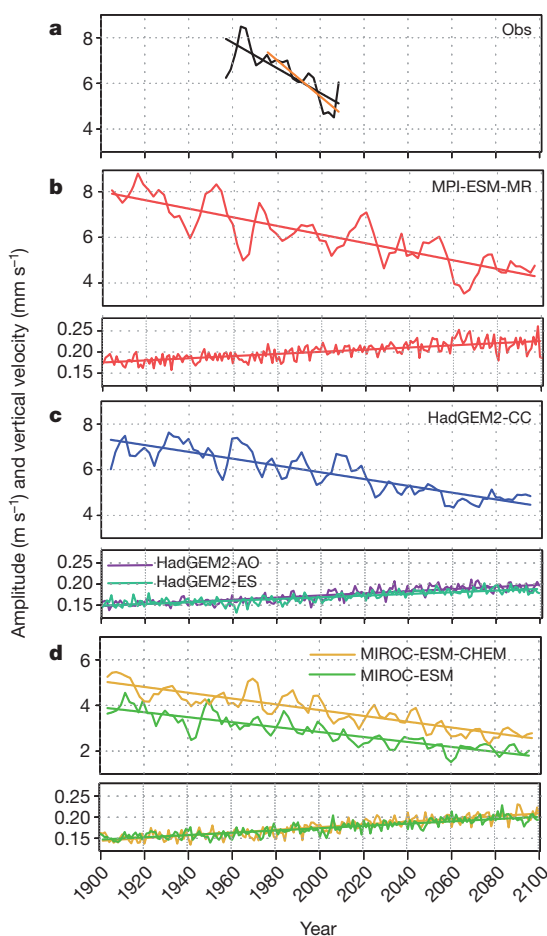
Figure 3 | Time variation in the three-cycle mean amplitude of the observed and modelled QBO and in modelled annual mean upward velocity at 70 hPa over 15° S–15° N. a, Radiosonde observations (black); **b**, MPI-ESM-MR (red); **c**, HadGEM2-CC (blue), HadGEM2-AO (purple) and HadGEM2-ES (light blue); **d**, MIROC-ESM-CHEM (yellow) and MIROC-ESM (green). Panel **a** and the upper parts of panels **b–d** show the QBO amplitude (in m s⁻¹); the lower parts of panels **b–d** display the mean upward velocity (in mm s⁻¹). The amplitude range of the vertical axis is shifted 2 m s⁻¹ downward for **d**. The solid line is the linear regression. The orange solid line in **a** shows regression only using data from Singapore. The linear regression trends shown are all statistically significant at $P < 0.05$.

1. Baldwin, M. P. *et al.* The quasi-biennial oscillation. *Rev. Geophys.* **39,** 179–229 (2001).
2. Saravanan, R. A multiwave model of the quasi-biennial oscillation. *J. Atmos. Sci.* **47,** 2465–2474 (1990).
3. Kawatani, Y., Hamilton, K. & Watanabe, S. The quasi-biennial oscillation in a double CO_2 climate. *J. Atmos. Sci.* **68,** 265–283 (2011).
4. Watanabe, S. & Kawatani, Y. Sensitivity of the QBO to mean tropical upwelling under a changing climate simulated by an Earth system model. *J. Meteorol. Soc. Jpn* **90A,** 351–360 (2012).
5. Butchart, N. *et al.* Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. *Clim. Dyn.* **27,** 727–741 (2006).
6. Garcia, R. R. & Randel, W. J. Acceleration of the Brewer–Dobson circulation due to increases in greenhouse gases. *J. Atmos. Sci.* **65,** 2731–2739 (2008).
7. McLandress, C. & Shepherd, T. G. Simulated anthropogenic changes in the Brewer-Dobson circulation, including its extension to high latitudes. *J. Clim.* **22,** 1516–1540 (2009).
8. Engel, A. *et al.* Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geosci.* **2,** 28–31 (2009).
9. Stiller, G. P. *et al.* Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period. *Atmos. Chem. Phys.* **12,** 3311–3331 (2012).
10. Young, P. J. *et al.* Changes in stratospheric temperatures and their implications for changes in the Brewer-Dobson circulation, 1979–2005. *J. Clim.* **25,** 1759–1772 (2012).
11. Kawatani, Y. *et al.* The roles of equatorial trapped waves and internal inertia–gravity waves in driving the quasi-biennial oscillation. Part I: Zonal mean wave forcing. *J. Atmos. Sci.* **67,** 963–980 (2010).
12. Lindzen, R. S. & Holton, J. R. A theory of the quasi-biennial oscillation. *J. Atmos. Sci.* **25,** 1095–1107 (1968).
13. Plumb, R. A. The interaction of two internal waves with the mean flow: implications for the theory of the quasi-biennial oscillation. *J. Atmos. Sci.* **34,** 1847–1858 (1977).

14. Takahashi, M. Simulation of the stratospheric quasi-biennial oscillation using a general circulation model. *Geophys. Res. Lett.* **23,** 661–664 (1996).
15. Hamilton, K., Wilson, R. J. & Hemler, R. Atmosphere simulated with high vertical and horizontal resolution versions of a GCM: improvement in the cold pole bias and generation of a QBO-like oscillation in the tropics. *J. Atmos. Sci.* **56,** 3829–3846 (1999).
16. Quiroz, R. S. Period modulation of the stratospheric quasi-biennial oscillation. *Mon. Weath. Rev.* **109,** 665–674 (1981).
17. Geller, M. A., Shen, W., Zhang, M. & Tan, W. W. Calculations of the stratospheric quasi-biennial oscillation for time-varying wave forcing. *J. Atmos. Sci.* **54,** 883–894 (1997).
18. Salby, M. & Callaghan, P. Connection between the solar cycle and the QBO: the missing link. *J. Clim.* **13,** 2652–2662 (2000).
19. Hamilton, K. On the quasi-decadal modulation of the stratospheric QBO period. *J. Clim.* **15,** 2562–2565 (2002).
20. Taguchi, M. Observed connection of the stratospheric quasi-biennial oscillation with El Niño–Southern Oscillation in radiosonde data. *J. Geophys. Res.* **115,** D18120, http://dx.doi.org/10.1029/2010JD014325 (2010).
21. Garfinkel, C. I. & Hartmann, D. L. Effects of the El Niño – Southern Oscillation and the quasibiennial oscillation on polar temperatures in the stratosphere. *J. Geophys. Res.* **112,** D19112, http://dx.doi.org/10.1029/2007JD008481 (2007).
22. Butchart, N. & Scaife, A. A. Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate. *Nature* **410,** 799–802 (2001).
23. Solomon, S. *et al.* Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science* **327,** 1219–1223 (2010).
24. Naujokat, B. An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. *J. Atmos. Sci.* **43,** 1873–1877 (1986).
25. Hamilton, K., Hertzog, A., Vial, F. & Stenchikov, G. Longitudinal variation of the stratospheric quasi-biennial oscillation. *J. Atmos. Sci.* **61,** 383–402 (2004).
26. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
27. Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, 1987).
28. Karoly, D. J. *et al.* An example of fingerprint detection of greenhouse climate change. *Clim. Dyn.* **10,** 97–105 (1994).
29. Dunkerton, T. J. & Delisi, D. P. Climatology of the equatorial lower stratosphere. *J. Atmos. Sci.* **42,** 376–396 (1985).

Supplementary Information is available in the online version of the paper.

Author Contributions Y.K. and K.H. analysed the data. K.H. and Y.K. wrote the manuscript.

METHODS

Radiosonde data. We used monthly mean values of operational balloon-borne radiosonde observations provided by the Free Berlin University (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/)[24]. The monthly mean zonal wind components are calculated for the levels 70, 50, 40, 30, 20, 15 and 10 hPa (7 levels) for full periods. Since 1987, levels 60, 45, 35, 25 and 12 hPa are additionally available (making a total of 12 levels). In drawing the time–height section of the QBO amplitude (main panel of Fig. 1), 12-level data linearly interpolated from 7-level data are used from 1953 to 1986 and original 12-level data are used after 1987. In drawing the trend of the QBO amplitude (right panel of Fig. 1), 7-level data, which cover the full analysed period, are used.

Calculating QBO amplitude. Dunkerton and Delisi[29] suggested a simple approach of calculating the standard deviation, σ_u, for each level of the monthly mean time series after the mean seasonal cycle over the entire record had been removed. They then estimated the amplitude as $A_u = \sqrt{2}\sigma_u$, assuming that almost all the variability in the monthly means in the tropical stratosphere can reasonably be attributed to the QBO. This was then used to compute the long-term mean amplitude of the QBO.

Here, we apply a similar technique to search for trends in shorter records. Specifically, we calculate A_u at each level for overlapping time chunks spanning three consecutive QBO cycles. The beginning and end of each chunk were based on the dates on which winds at 30 hPa were observed to change from westerly to easterly; these dates are December 1953, February 1956, April 1958, March 1960, April 1962, March 1965, July 1967, January 1970, May 1972, February 1974, June 1976, March 1979, May 1981, July 1983, June 1986, November 1988, May 1991, October 1993, November 1995, January 1998, May 2000, February 2003, March 2005, March 2007, June 2009 and October 2011.

The wind observations from January 1953 to December 2012 were used to define the mean seasonal cycle for the period, which was then subtracted from the raw data. The resulting series were then smoothed by taking 5-month running means (for example, Supplementary Fig. 4). Then, A_u was calculated for the first 3-cycle chunk using these deseasonalized and slightly smoothed data from

December 1953–February 1960, February 1956–March 1962, ..., and finally March 2005–Sep 2011. The 3-cycle mean (roughly a 7-year mean) could attenuate the influence of any QBO modulation by the El Niño/Southern Oscillation[20].

CMIP5 models. MPI-ESM-MR was developed at the Max Planck Institute for Meteorology, Germany; HadGEM2-CC, HadGEM2-AO and HadGEM2-ES at the Met Office Hadley Centre, United Kingdom; MIROC-ESM and MIROC-ESM-CHEM at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute (AORI), University of Tokyo, and National Institute for Environmental Studies (NIES), Japan. More detailed information is given in Supplementary Information section 10.

The QBO is driven by vertically propagating internal waves: large-scale equatorial waves (zonal wavelengths $\geq$10,000 km), medium-scale gravity waves (zonal wavelengths $\sim$1,000–10,000 km) and small-scale gravity waves (zonal wavelengths approximately $\leq$1,000 km) generated in the tropical troposphere[1,11]. The effects of small-scale gravity waves are parameterised (that is, non-stationary gravity wave parameterization) because CMIP5 models cannot resolve most of the small-scale gravity waves. The modelled QBO is driven by a combination of explicitly resolved and parameterised waves. This issue is discussed in Supplementary Information section 10, and the raw time series of the equatorial monthly mean zonal-mean zonal wind for a 200-year simulation are shown in Supplementary Figs 10–13.

Statistical analysis. To test significance, a two-sided Student's t-test is used. The number of QBO cycles is 25 for 1953–2012 and 15 for 1976–2012 for the radiosonde record; 82 for 1900–2100 and 22 for 1953–2012 for MPI-ESM-MR; 93 for 1900-2099 and 25 for 1953–2012 for HadGEM2-CC; 90 for 1900–2100 and 26 for 1953–2012 for MIROC-ESM-CHEM; and 93 for 1900–2100 and 27 for 1953–2012 for MIROC-ESM. As our analysis uses 3-cycle segments, these numbers divided by 3 are used as the number of degrees of freedom for calculating statistical significance. For the vertical velocity we analysed annual mean values over the available records and took the sample size to be the number of years (that is, 201 for MPI-ESM-MR, MIROC-ESM-CHEM and MIROC-ESM, and 200 for HadGEM2-CC, HadGEM2-AO and HadGEM2-ES).

LETTER

Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination

Petr Znamenskiy[1,2] & Anthony M. Zador[2]

The neural pathways by which information about the acoustic world reaches the auditory cortex are well characterized, but how auditory representations are transformed into motor commands is not known. Here we use a perceptual decision-making task in rats to study this transformation. We demonstrate the role of corticostriatal projection neurons in auditory decisions by manipulating the activity of these neurons in rats performing an auditory frequency-discrimination task. Targeted channelrhodopsin-2 (ChR2)[1,2]-mediated stimulation of corticostriatal neurons during the task biased decisions in the direction predicted by the frequency tuning of the stimulated neurons, whereas archaerhodopsin-3 (Arch)[3]-mediated inactivation biased decisions in the opposite direction. Striatal projections are widespread in cortex and may provide a general mechanism for the control of motor decisions by sensory cortex.

After reaching the cortex, auditory information is processed and relayed to a number of cortical and subcortical targets by distinct, largely non-overlapping populations of pyramidal neurons. These targets include parietal cortex, secondary auditory areas, inferior colliculus and striatum, regions that have distinct roles in perception and behaviour. How these outputs help to establish associations between sounds and motor responses is unknown.

We sought to test the hypothesis that the projection from the auditory cortex to the striatum carries acoustic information that drives behavioural choices during auditory discrimination. The striatum receives topographic inputs from throughout the cortex. Striatal regions that receive input from motor and prefrontal cortices have been implicated in a wide range of cognitive processes, including decision-making[4], action selection[5] and reinforcement learning[6,7]. Through downstream structures of the basal ganglia, the striatum influences the activity in the motor thalamus[8] as well as superior colliculus[9], a structure that has been implicated in driving behavioural choices in two-alternative choice tasks[10]. Plasticity of corticostriatal connections may enable them to encode the arbitrary stimulus–response associations that are acquired in such tasks[11].

In contrast, the function of striatal regions that receive direct input from sensory cortical areas is less well established. The striatum is one of the major long-range targets of the auditory cortex[12]. The auditory cortex projects to a specific region of the striatum[13]. This striatal region does not receive input from any other cortical area and contains neurons sensitive to auditory stimuli[14,15]. We set out to test whether the auditory corticostriatal projection drives choices during auditory discrimination.

We first developed a novel auditory discrimination task, inspired by experiments in primate middle temporal area (area MT)[16], that was designed to exploit the tonotopic organization of the auditory system[17]. We trained rats to discriminate low- and high-frequency 'cloud-of-tones' stimuli in a two-alternative choice task[18] (Fig. 1a). On each trial the stimulus consisted of a train of short overlapping pure tones distributed over a three-octave range (5–40 kHz). Subjects were required to choose between a left and a right reward port depending on whether the stimulus contained mostly low-frequency (5–10 kHz) or high-frequency (20–40 kHz) tones (Fig. 1b). Rats were free to withdraw and

report their choice at any time after the onset of the stimulus. Rats readily learned this task. Performance varied smoothly with stimulus difficulty, and approached 100% on the easiest stimuli (Fig. 1c).

We used tetrode recordings to characterize the activity of individual neurons in the auditory cortex and the auditory striatum while rats performed the task. Auditory striatal neurons have been characterized previously in anaesthetized and passively listening animals[14,15] but not during behaviour. The firing rates of neurons in both auditory cortex and striatum varied monotonically with the stimulus (Supplementary Fig. 1a, b, e, f). 'Ideal observer' analysis showed that stimulus selectivity of cortical and striatal neurons was similar (Supplementary Fig. 1c, d, g, h). Although individual neurons rarely matched the performance of the rats, many neurons in both areas could discriminate the stimulus above chance levels (Supplementary Fig. 1d, h). These observations indicate that neurons in the striatum preserve the information available to perform the cloud-of-tones task.

We then tested whether manipulating cortical activity with ChR2 (refs 1, 2) could affect subjects' choices in the task. We proposed that activation of auditory cortical neurons in different regions of the tonotopic map would result in choice biases, the directions of which would depend on the frequency–response association that the rat had been trained to make. We implanted arrays of optical fibres coupled to tetrodes (Supplementary Fig. 2), which enabled us to characterize the frequency preference of neurons near the fibre tip. We activated ChR2-expressing auditory cortex neurons on a subset of trials, and
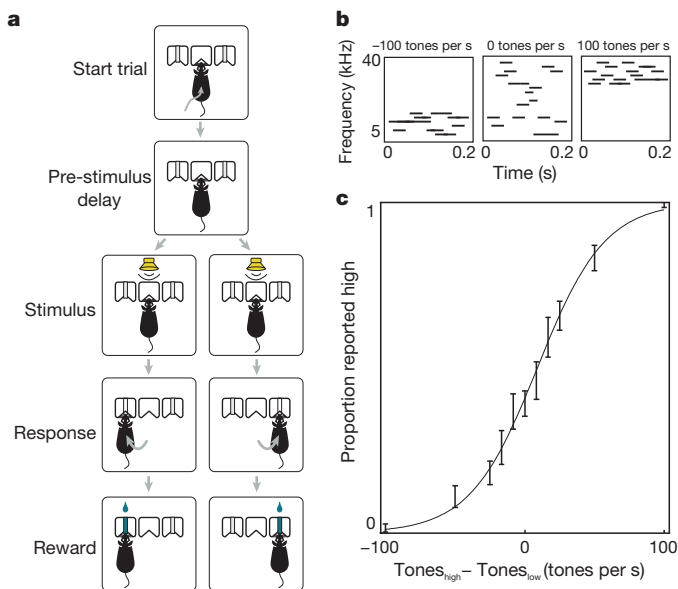

Figure 1 | Cloud-of-tones task. a, Format of the behavioural task. **b,** Example stimulus spectrograms at −100, 0 and +100 tones per s. **c,** Psychometric curve from a single rat (error bars, 95% confidence interval).

[1]Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

compared performance on stimulated trials to that on control trials without light activation. To minimize behavioural adaptation to photostimulation, we limited the number of stimulation trials to 25% of the total trials, and rewarded the animal as in control trials, according to the frequency content of the stimulus. To quantify the effects of photostimulation on subjects' behaviour, we fit a logistic regression model to the subjects' choices (see Methods), estimating what change to the auditory stimulus would produce a behavioural effect equivalent to that of photostimulation.

We first expressed ChR2 non-specifically in the primary auditory cortex using adeno-associated virus (AAV). ChR2 expression was distributed throughout layers II to V and was present, although to a lesser extent, in layer VI (Supplementary Fig. 3a, b). In some sessions, activation of ChR2-expressing neurons induced substantial biases (Supplementary Fig. 3c). However, contrary to our expectations the direction of the bias was not predicted by the tuning of the neurons near the fibre (Supplementary Fig. 3f). Moreover, this non-specific activation not only biased subjects' choices but also interfered with their ability to perform the task, as measured by the slope of the psychometric curve (Supplementary Fig. 3d, e).

The lack of correspondence between the direction of stimulation-evoked choice biases and frequency tuning was surprising and has several possible explanations. First, these choice biases may result from activation of intracortical projection neurons from other frequency bands of the auditory cortex. Second, they may arise from preferential activation of local interneurons in response to synchronous activation of a large number of cortical cells. Third, these biases may be a consequence of stimulation of competing output pathways, projecting to target areas with different functional roles. Consistent with this third possibility, stimulation resulted in predominantly contralateral biases (36 out of 57 sites, $P = 0.015$, Supplementary Fig. 3g, h), suggesting the activation of a lateralized motor pathway. These results suggested that the auditory cortex plays a role in the task, but that the method of non-specific activation of diverse neuronal populations used in these experiments does not provide sufficient experimental control over the activity of the projection neurons driving choices in this behaviour. We therefore proposed that targeted stimulation of corticostriatal neurons might produce more systematic effects.

We used two complementary strategies to test the role of the neuronal subpopulation in auditory cortex that projects to the striatum. In the first strategy, we targeted ChR2 specifically to corticostriatal neurons using a herpes simplex virus-1 (HSV-1) engineered to express Cre recombinase[19–21] (Fig. 2a). This virus, when injected into the auditory striatum, was transported retrogradely along the axons and drove Cre expression in corticostriatal neurons. We then injected a Cre-dependent AAV driving expression of ChR2 into the auditory cortex, resulting in expression of ChR2 only in neurons infected by both the HSV-1–Cre and the AAV-ChR2. Most (84%) of ChR2-expressing neurons were located in layer V (Fig. 2c, d), consistent with previous results[22].

Pulses of blue light delivered through fibres in the cortex reliably drove action potentials in a small population of presumed corticostriatal neurons (Fig. 2b; only 4 of 201 cells responded in more than 50% of trials). Across sites, stimulation consistently biased subjects' choices towards the choice port associated with the preferred frequency of the stimulation site (Fig. 2e, f). The biases were not significant in individual sessions ($P < 0.05$ in only 1 out of 33 sessions), partly because the number of stimulated trials per session was typically small (70 ± 24 ($\pm$ s.d.) trials). In contrast to non-specific stimulation, contralateral and ipsilateral biases were observed with similar frequency (12 out of 33 contralateral sites, $P = 0.31$, signed-rank test) and subjects' psychophysical performance was not impaired (Supplementary Fig. 4a). Choice biases were not observed in uninjected controls (Fig. 2g and Supplementary Fig. 5a), indicating that they were driven by the ChR2-mediated activation of corticostriatal neurons and not non-specific effects of light delivery. Thus, activation of corticostriatal neurons biased behaviour in a manner predicted by their frequency tuning.
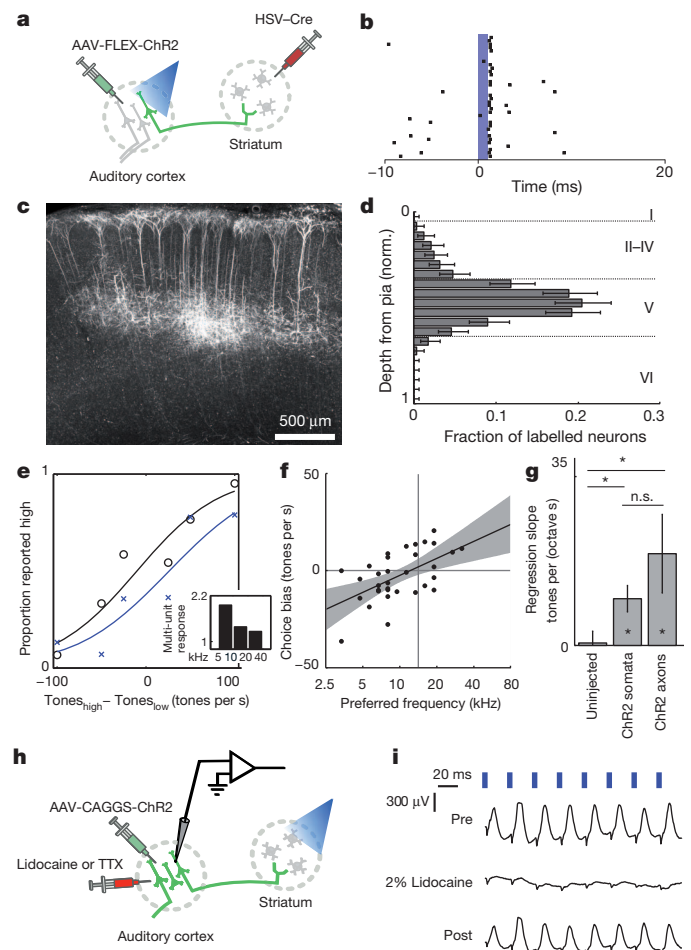

Figure 2 | ChR2 stimulation of corticostriatal neurons biases subjects' choices. a, Strategy targeting ChR2 expression to corticostriatal neurons using retrograde HSV–Cre and AAV-FLEX-ChR2, where ChR2 expression is controlled by the Cre-dependent FLEX switch. **b**, A corticostriatal neuron identified on the basis of its short latency light-evoked response (approximately 1 ms). Blue bar indicates the timing of the light pulse (10 mW, 1 ms duration). **c**, ChR2 expression in corticostriatal neurons. **d**, Depth distribution of ChR2 expression. Dashed lines mark the approximate location of layer boundaries. Error bars, 95% confidence intervals. **e**, Psychometric performance during a behavioural session in control trials (black) and during stimulation of corticostriatal neurons (blue). $P = 0.02$ for stimulation-evoked choice bias. Inset shows frequency tuning of multi-unit responses normalized to baseline, binned by octave. **f**, Across the population, the direction and magnitude of choice biases evoked by stimulation depend on the frequency preference of the stimulation site ($n = 33$ sites from 2 rats, $P = 0.0006$ for linear regression of choice bias and frequency preference). Grey shading, 95% confidence interval for regression line. **g**, Dependence of choice bias on preferred frequency of the stimulation site in uninjected controls, and during specific stimulation of corticostriatal neurons and their axons. *$P < 0.05$; n.s., not significant, error bars are standard error through bootstrap. P values within bars compare the frequency dependence of choice bias to 0. **h**, Strategy for the inactivation of recurrent cortical activity during stimulation of striatal axons. TTX, tetrodotoxin. CAGGS, chicken β-actin promoter and cytomegalovirus early enhancer. **i**, Lidocaine infusion into the auditory cortex reversibly abolishes antidromic light-evoked local field potential responses.

The behavioural effects of photostimulation could arise either directly through excitation of striatal neurons by their cortical inputs, or indirectly through excitation of other output pathways of the auditory cortex through recurrent connections. Excitation of corticostriatal neurons resulted in general suppression of local cortical activity (Supplementary Fig. 6), favouring the hypothesis that behavioural effects of photostimulation are mediated by long-range rather than local outputs of these neurons.

To test this hypothesis directly, we used a second strategy. This strategy did not rely on cortical stimulation and therefore enabled us to test whether stimulation of corticostriatal neurons could bias subjects' choices when recurrent cortical activity was blocked. We stimulated the axons of corticostriatal neurons in the striatum while inhibiting the recurrent cortical excitation pharmacologically (Fig. 2h). To define the frequency tuning of stimulated fibres, we exploited the topography of corticostriatal projections[23]. We confirmed the topographic organization of the striatal projection of the auditory cortex in rats through retrograde (Supplementary Fig. 7) tracing. Owing to this topography, focal light stimulation in the striatum excites corticostriatal axons arising from a restricted region of the tonotopic map. We used striatal multi-unit activity to characterize the frequency preference of the striatal stimulation site.

Stimulation of corticostriatal axons biased subjects' choices towards the choice port associated with the preferred frequency of the stimulation site (Fig. 2g and Supplementary Fig. 5b). Stimulation also had a modest effect on the subjects' performance at sites with large choice biases, possibly owing to a ceiling effect (Supplementary Fig. 4b, c). Thus, both direct activation of corticostriatal somata in the cortex, as well as antidromic activation of corticostriatal terminals in the striatum, biased the rats' choices in a manner predicted from the frequency tuning of the activated neurons.

We next tested whether stimulation of the axons of corticostriatal neurons could bias subjects' choices in the absence of cortical recurrent activity. We selected striatal sites whose stimulation produced significant biases ($P < 0.05$) in subjects' choices and repeated stimulation after pharmacologically inactivating the ipsilateral auditory cortex. Cortical inactivation (2% lidocaine, $n = 2$ sites; 125 μM tetrodotoxin, $n = 5$ sites) blocked antidromic light-evoked local field potentials (Fig. 2i). Photostimulation of corticostriatal axons still biased choices in the absence of cortical activity ($P = 0.016$, signed-rank test, Supplementary Figs 8 and 9), demonstrating that these biases are mediated directly by long-range outputs of corticostriatal neurons.

We next examined the effects of stimulation of corticostriatal neurons on rats' response times. Although individual rats varied in how they prioritized response speed versus response accuracy (Supplementary Fig. 10a–f), most rats used in the stimulation and inactivation experiments showed a small but significant increase (in the order of 20 to 50 ms) in response time on challenging trials, consistent with that reported previously in rats[18]. We quantified the shift of the chronometric curve produced by stimulation of corticostriatal neurons (Supplementary Fig. 10g, h). Across sessions, stimulation shifted chronometric curves, mimicking the effect of adding acoustic evidence favouring the choice associated with the preferred frequency of the stimulation site (Supplementary Fig. 10i, j).

So far, our results establish that activation of the corticostriatal pathway is sufficient to bias subjects' choices during an auditory discrimination task, but do not establish whether corticostriatal activity contributes to the formation of decisions under normal conditions, in the absence of ChR2 activation. If corticostriatal spikes are important under normal conditions, then suppressing these spikes during the task should lead to an 'anti-bias'; that is, a bias in the direction opposite to that induced by ChR2 activation. To test this hypothesis, we targeted the expression of the light-activated proton pump Arch[3] to corticostriatal neurons using the HSV-1-based approach described above (Fig. 3a). Rather than attempting to inactivate all corticostriatal cells—a technically difficult feat whose interpretation may be obscured by the compensatory plasticity common to lesion studies—we sought instead to silence corticostriatal neurons within a restricted region of the tonotopic map.

Pulses of green light in Arch-expressing animals inhibited spiking of putative corticostriatal neurons (Fig. 3b). As predicted, inactivation of corticostriatal neurons biased subjects' choices away from the reward port associated with the frequency band of the inactivation site (Fig. 3c) but did not affect subjects' sensitivity quantified as the slope of the
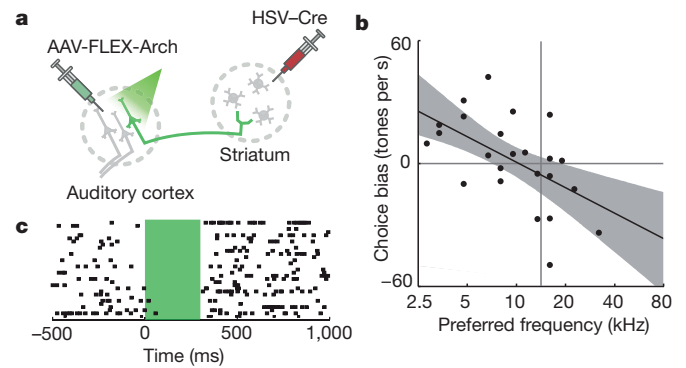

Figure 3 | Arch-mediated inactivation of corticostriatal neurons anti-biases subjects' choices. a, Strategy targeting Arch expression to corticostriatal neurons. **b,** Activity of a presumed corticostriatal neuron inhibited by pulses of green light. Green shading shows the timing of the light pulse (300 ms, 5 mW). **c,** Inactivation biased choices away from the direction associated with the preferred frequency of the inactivation site ($n = 24$ sites from 3 rats, $P = 0.0043$ for linear regression of choice bias and frequency preference). Grey shading, 95% confidence interval for regression line.

psychometric curve (Supplementary Fig. 4d). A similar trend was reflected in the subjects' chronometric functions but did not reach significance (Supplementary Fig. 11k, l). Pulses of green light did not affect the behaviour of uninjected control animals (Supplementary Fig. 5c). These results indicate that corticostriatal neurons in the auditory cortex transmit signals used by rats to make decisions driven by sounds.

These experiments provided us with an opportunity to quantify the contribution of the output of single corticostriatal cells to behaviour. Earlier studies showed that subjects can be trained to detect the artificial stimulation of as few as 6 to 197 neurons[24] or even a single neuron[25]. Our experiments allow us to estimate the contribution of single neurons to behaviour during normal perception.

We tagged Arch with green fluorescent protein (Arch–GFP) to estimate the number of neurons silenced in each experiment (Fig. 4b). Using *in vivo* recordings of putative corticostriatal cells, we estimated that the efficiency of inactivation decayed with a space constant of approximately 564 μm (Fig. 4a and Supplementary Fig. 11). We quantified the number of Arch-expressing neurons within a radius of 1 mm, or approximately two space constants, from the inactivation fibre for each session. The magnitude of choice biases depended on the number of corticostriatal neurons expressing Arch near the fibre (Fig. 4c). Robust choice biases were observed for sites with 1,040 to 2,444 Arch-expressing neurons within 1 mm of the fibre, comprising 0.35 to 0.83% of neurons in that
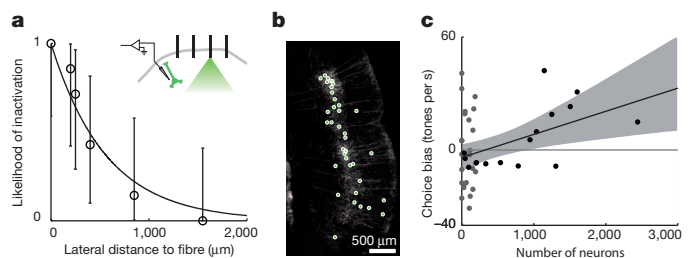

Figure 4 | Estimation of numbers of neurons affected by Arch inactivation. a, The likelihood of silencing for Arch-expressing corticostriatal neurons as a function of distance from the fibre ($n = 7$). Error bars, 95% binomial confidence intervals. **b,** Somata of neurons expressing Arch-GFP (green circles) identified in a fluorescence image of the auditory cortex at the end of an inactivation experiment. **c,** Effects of inactivation at sites tuned <10 kHz or >20 kHz were correlated with the number of Arch-expressing neurons near the fibre ($n = 41$ sites from 5 rats, $P = 0.021$ for linear regression of choice bias and neuron number). Grey points are data for 2 out of 5 rats, for which Arch expression was low throughout the auditory cortex (on average 72 ± 69 cells per site). Grey shading, 95% confidence interval for regression line.

volume. Our observation that inactivation of this minute fraction of cells targeted to corticostriatal neurons could affect behaviour suggests a privileged role of these cells in auditory discrimination.

The size of the neuronal pool accounting for behavioural performance in psychophysical tasks is much smaller than the number neurons which carry signals relevant to the task[26]. The redundancy arising from correlated neuronal firing may restrict the effective size of the pool[26], limiting psychophysical performance. Our results suggest that the finite bandwidth provided by a small number of projection neurons that drive decisions constitutes an additional constraint limiting the effective pool size and therefore psychophysical performance.

We have shown that the striatal projection of the auditory cortex conveys signals that drive behavioural choices during performance of an auditory discrimination task. Although our results do not exclude the participation of other parallel pathways, the ubiquity of corticostriatal projections in cortex may provide a general mechanism for control of motor decisions by sensory context[27].

METHODS SUMMARY

Male Long Evans rats (Taconic Farms) were trained in a two-alternative choice discrimination task. Animal procedures were approved by the Cold Spring Harbour Laboratory Animal Care and Use Committee and carried out in accordance with National Institutes of Health standards.

Full Methods and any associated references are available in the online version of the paper.

1. Nagel, G. et al. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proc. Natl Acad. Sci. USA 100, 13940–13945 (2003).
2. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. Nature Neurosci. 8, 1263–1268 (2005).
3. Chow, B. Y. et al. High-performance genetically targetable optical neural silencing by light-driven proton pumps. Nature 463, 98–102 (2010).
4. Ding, L. & Gold, J. I. Caudate encodes multiple computations for perceptual decisions. J. Neurosci. 30, 15747–15759 (2010).
5. Kimchi, E. Y. & Laubach, M. Dynamic encoding of action selection by the medial striatum. J. Neurosci. 29, 3148–3159 (2009).
6. Reynolds, J. N., Hyland, B. I. & Wickens, J. R. A cellular mechanism of reward-related learning. Nature 413, 67–70 (2001).
7. Pasupathy, A. & Miller, E. K. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433, 873–876 (2005).
8. Beckstead, R. M., Domesick, V. B. & Nauta, W. J. Efferent connections of the substantia nigra and ventral tegmental area in the rat. Brain Res. 175, 191–217 (1979).
9. Hopkins, D. A. & Niessen, L. W. Substantia nigra projections to the reticular formation, superior colliculus and central gray in the rat, cat and monkey. Neurosci. Lett. 2, 253–259 (1976).
10. Felsen, G. & Mainen, Z. F. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. Neuron 60, 137–148 (2008).
11. Kreitzer, A. C. & Malenka, R. C. Striatal plasticity and basal ganglia circuit function. Neuron 60, 543–554 (2008).
12. Allen Brain Institute. Allen Mouse Connectivity Atlas http:// connectivity.brain-map.org (2012).
13. McGeorge, A. J. & Faull, R. L. The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience 29, 503–537 (1989).
14. Bordi, F. & LeDoux, J. Sensory tuning beyond the sensory system: an initial analysis of auditory response properties of neurons in the lateral amygdaloid nucleus and overlying areas of the striatum. J. Neurosci. 12, 2493–2503 (1992).
15. Bordi, F., LeDoux, J., Clugnet, M. C. & Pavlides, C. Single-unit activity in the lateral nucleus of the amygdala and overlying areas of the striatum in freely behaving rats: rates, discharge patterns, and responses to acoustic stimuli. Behav. Neurosci. 107, 757–769 (1993).
16. Salzman, C. D., Britten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. Nature 346, 174–177 (1990).
17. Sally, S. L. & Kelly, J. B. Organization of auditory cortex in the albino rat: sound frequency. J. Neurophysiol. 59, 1627–1638 (1988).
18. Uchida, N. & Mainen, Z. F. Speed and accuracy of olfactory discrimination in the rat. Nature Neurosci. 6, 1224–1229 (2003).
19. Lilley, C. E. et al. Multiple immediate-early gene-deficient herpes simplex virus vectors allowing efficient gene delivery to neurons in culture and widespread gene delivery to the central nervous system in vivo. J. Virol. 75, 4343–4356 (2001).
20. Lima, S. Q., Hromádka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. PLoS ONE 4, e6099 (2009).
21. Ciocchi, S. et al. Encoding of conditioned fear in central amygdala inhibitory circuits. Nature 468, 277–282 (2010).
22. Reiner, A., Jiao, Y., Del Mar, N., Laverghetta, A. V. & Lei, W. L. Differential morphology of pyramidal tract-type and intratelencephalically projecting-type corticostriatal neurons and their intrastriatal terminals in rats. J. Comp. Neurol. 457, 420–440 (2003).
23. Reale, R. A. & Imig, T. J. Auditory cortical field projections to the basal ganglia of the cat. Neuroscience 8, 67–86 (1983).
24. Huber, D. et al. Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. Nature 451, 61–64 (2008).
25. Houweling, A. R. & Brecht, M. Behavioural report of single neuron stimulation in somatosensory cortex. Nature 451, 65–68 (2008).
26. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12, 4745–4765 (1992).
27. Jiang, H., Stein, B. E. & McHaffie, J. G. Physiological evidence for a trans-basal ganglia pathway linking extrastriate visual cortex and the superior colliculus. J. Physiol. (Lond.) 589, 5785–5799 (2011).

METHODS

Viral production. AAV-CAGGS-ChR2-Venus plasmid was used for non-specific cortical stimulation. AAV-CAGGS-FLEX-ChR2-tdTomato virus was used for targeted stimulation of corticostriatal neurons. AAV-FLEX-Arch–GFP construct was generated by subcloning Arch–GFP from an AAV-CAG-Arch–GFP plasmid into AAV-FLEX backbone from AAV-EF1a-FLEX-ChR2–YFP. The plasmid DNA was prepared using standard Maxiprep protocols (Qiagen). AAV serotype 2/9 was packaged by the University of North Carolina viral core at a titre of 1 to 2×10^{12} particles per ml. HSV–mCherry-IRES–iCre construct (for targeting Arch expression) and HSV–iCre-2A-Venus construct (for targeting ChR2 expression) were packaged at titres of 2.4×10^{10} and 3×10^{9} transducing units per ml.

Animal subjects. Animal procedures were approved by the Cold Spring Harbour Laboratory Animal Care and Use Committee and carried out in accordance with National Institutes of Health standards. Male Long Evans rats (Taconic Farms) were housed with free access to food, but were water restricted after the start of behavioural training. Water was available during task performance (24 µl for each correct trial) and freely available for 15 to 30 min after the end of each behavioural sessions and for at least 1 h on days when behavioural sessions were not conducted.

Viral injection. Three- to five-week-old rats were anaesthetized with a mixture of ketamine (60 mg kg^{-1} of body weight) and medetomidine (0.24 mg kg^{-1}) and placed in a stereotaxic apparatus. To target the auditory cortex, part of the temporalis muscle was resected to expose the temporoparietal suture, which was used as a landmark to target injections. For optogenetic experiments, four injections were made unilaterally spanning primary auditory cortex at 0.5, 1.5, 2.5 and 3.5 mm from the rostral edge of parietal bone and 1.2 mm from its ventral edge. A small craniotomy was made for each injection and a glass micropipette was inserted perpendicular to the surface of the brain. Two injections were made at the depths of 400 and 800 µm, expelling approximately 250 nl of virus at each depth. To target the auditory striatum, two small craniotomies were made 2.0 and 2.5 mm caudal of Bregma and 4.5 mm lateral of the midline. Injections were made at depths between 3.5 and 6 mm, 0.5 mm between injection sites, approximately 100 nl per site. Injections were carried out by delivering brief pulses of pressure using Picospritzer II (Parker), each pulse delivering approximately 2 nl at 0.2 Hz. Rats were monitored during their recovery from surgery and returned to group housing.

Behavioural training. The cloud-of-tones stimulus consisted of stream of 30-ms overlapping pure tones presented at 100 Hz (that is, with 10 ms between tone onsets). Eighteen possible tone frequencies were logarithmically spaced between 5 and 40 kHz, a range in which rats' hearing thresholds are low and relatively constant[28]. For each trial either the low (5 to 10 kHz) or the high (20 to 40 kHz) octave was selected as the target octave. Stimulus strength r determined the difference in the rate of high and low octave tones in the stimulus. Tones were drawn from the target octave with a probability of $(1 + 2r/100)/3$. The stream of tones continued until the rat withdrew from the centre port.

After reaching a weight of 200 to 250 g, rats were placed on a water deprivation schedule and behavioural training commenced. The rats were placed in a soundproof behavioural chamber and presented with three choice ports. The rats were trained to first poke into the centre port, wait for the onset of the auditory stimulus and select one of the other two ports to receive a water reward (24 µl). Rats were trained to carry out this sequence using the following procedure. During the first phase of training, water was delivered at the correct choice port as soon as the stimulus was played. The duration of the pre-stimulus delay, during which the rat was required to remain in the centre port, was drawn from an exponential distribution whose mean was gradually increased from 0.05 s to 0.3 s. The next phase of training required the rat to poke at the correct choice port to trigger water delivery. However, the rat was allowed to correct his choice if it made a mistake. When rats had learned to perform the discrimination so that they consistently scored higher than expected by chance, they were required to make the correct choice on the first attempt. Error trials were punished with a 4-s timeout (2 s during recording sessions). Initially, rats were trained to discriminate tone-cloud stimuli composed of tones either entirely in the high octave or entirely in the low octave ($r = 100$) and were gradually introduced to increasingly difficult stimuli.

Sound intensity of individual tones was constant during each trial. To discourage subjects from using loudness differences in discrimination, tone intensity was randomly selected on each trial from a uniform distribution 45 to 75 dB (SPL) during training. During manipulation and recording sessions, sound intensity was kept constant at 60 dB.

Electrophysiology and optogenetics. Custom-built optical fibre and tetrode arrays were assembled in-house. Each array carried 6 multimodal optical fibres 62.5 µm in diameter with a 50-µm core. The fibre tips were sharpened to a point using a diamond wheel to improve tissue penetration and increase the angle of the light exit cone. Each fibre was glued to a tetrode and the tetrode tip was cut to terminate within approximately 100 µm of the fibre tip. The fibre and tetrode assemblies were mounted on individually movable microdrives. The tetrodes were

gold-plated to an impedance of 300 kΩ at 1 kHz and the tetrode and fibre tips were coated with DiI fluorescent dye (Life Technologies) to assist with the identification of fibre tracks in brain tissue.

To implant the fibre and tetrode array, rats were anaesthetized with a mixture of ketamine (40 mg kg^{-1} of body weight) and medetomidine (0.16 mg kg^{-1}) and placed in a stereotaxic apparatus. A craniotomy was made over the target area (for auditory cortex, 3.5 to 6.0 mm caudal of Bregma and 6.5 to 7.0 mm lateral from the midline; for auditory striatum, 2.5 to 3.5 mm caudal of Bregma and 4 to 5 mm lateral from the midline). All rats, with the exception of one Arch-expressing and one uninjected control animal, were implanted in the left hemisphere. The dura was removed and the implant was placed over the target area and fixed in place with dental acrylic. The tetrodes were then lowered until the first action potentials were encountered.

To characterize the frequency tuning of stimulation and inactivation sites, the rats were placed in a soundproof chamber and pure tones were played in free field at approximately 0.5 Hz. Tone frequencies spanned from 1 to 64 kHz and were played in a random order at 30, 50 or 70 dB-SPL. Only sites that significantly responded to sounds ($P < 0.01$, signed-rank test comparing firing rate 5 to 55 ms after sound onset to 0 to 50 ms preceding sound onset) were included in the analysis of stimulation and inactivation experiments. To determine the preferred frequency, firing rates in the window from 5 to 55 ms after sound onset were computed for each frequency at 70 dB. The resulting tuning curve was smoothed with a half-octave sliding window. The peak of the smoothed tuning curve was selected as the preferred frequency. Sites tuned to frequencies more than one octave outside the range used in the task (below 2.5 kHz) were excluded from analysis.

For optogenetic manipulations, laser light was coupled into a FC/PC parch cord using a FibrePort Collimator (Thor Labs). Laser power was adjusted to produce the desired output at the end of the patch cord. A single implanted fibre was selected for manipulation and coupled to the patch cord. For ChR2 activation, 473-nm laser light (10 mW) was delivered in 1-ms pulses (except at nine sites corticostriatal stimulation sites, at which we used 0.5 ms pulses) at 40 Hz while the rat remained in the centre port. For Arch, 530-nm laser light (50 mW) was delivered in 1-ms pulses (except at six sites at which we used 0.5-ms pulses) at 100 Hz, which produced an average power of 5 mW. To decrease the ability of the rat to detect the stimulation light, a mask light-emitting diode (LED) of a wavelength similar to that of the laser was placed above the centre port in the behaviour chamber. The mask LED was activated on control as well as stimulation trials in the same temporal pattern as the laser. Manipulation trials were randomly interleaved among control trials. The optical fibre was advanced approximately 300 µm between manipulation sessions.

For action potential recordings, signals were filtered 600 to 6,000 Hz and recorded using the Neuralynx Cheetah 32 system and Cheetah data acquisition software.

Pharmacological inactivation. Two PEEK tubing cannulas (Plastics One) were implanted into the auditory cortex with a separation of approximately 1 mm. Six stereotrodes were implanted alongside the cannulas, spanning 3 to 4 mm of the auditory cortex, to confirm the efficiency of inactivation. One hour before the start of the behavioural session, the animal was briefly anaesthetized with 2% isofluorane and 0.4 µl of drug (during inactivation sessions) or 9 g l^{-1} NaCl (during control sessions) was injected at the rate of 0.08 µl min^{-1} in both cannulas.

Analysis of behaviour data. Although each trial had a target frequency distribution, which determined which choice port was to be rewarded, as rats sampled the stimulus for a finite period of time, the frequency distribution they experienced might have been substantially different from the target. Therefore, we determined the rate of presentation of high and low tones (f_{high} and f_{low}), that is, the number of tones divided by the reaction time, that the rats actually heard on each trial. To quantify rats' behaviour, psychometric curves were fit with a logistic regression model:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 \left(f_{high} - f_{low} \right) \qquad (1)$$

where p is the fraction of choices towards the port associated with high frequencies. Parameters β_0 and β_1 measure the bias and slope of the psychometric curve, respectively. To quantify the effects of stimulation or inactivation, we extended the model to incorporate the effects of these manipulations:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 \left(f_{high} - f_{low} \right) + \beta_2 S + \beta_3 S \left(f_{high} - f_{low} \right)$$

where S is 1 on manipulation trials and 0 on control trials. The choice bias evoked by the manipulation is $\beta_2 / (\beta_1 + \beta_3)$ tones per s. In experiments in which ChR2 or Arch expression was specifically targeted to corticostriatal neurons, the interaction term β_3 was not significantly different from 0 and was omitted in estimates of choice bias:

$$\ln\frac{p}{1-p} = \beta_0 + \beta_1\left(f_{high} - f_{low}\right) + \beta_2 S$$

The light-evoked choice bias is then β_2/β_1 tones per s. Confidence intervals and P values for individual manipulation sessions were estimated from standard errors of regression coefficients.

To quantify the effects of stimulus strength $|f_{high} - f_{low}|$ on response times t_r, we first used linear regression:

$$t_r = t_0 + \frac{m \times |f_{high} - f_{low}|}{100}$$

Regression slope m estimates the change in response time between ambiguous and easiest trials. To fit subjects' chronometric functions on individual sessions, we calculated the re-centred stimulus strength f', such that subjects responded with equal probability to either choice port at $f' = 0$, using bias β_0 and slope β_1 estimates from the psychometric fit in (1):

$$f' = f_{high} - f_{low} + \beta_0/\beta_1$$

We then fit subjects' response times t_r to the following function[29]:

$$t_r = \frac{k}{f'}\tanh(\beta_1 f') + t_m$$

Parameter t_m can be thought of as the motor component of the response time, and k sets the magnitude of modulation of response times by the stimulus. We found the values of t_m and k that minimized the mean square error of t_r predictions using MATLAB's *fminsearch* function. We measured the shift of the chronometric function produced by stimulation q, adding it to the response time model:

$$t_r = \frac{k}{f' + qS}\tanh(\beta_1(f' + qS)) + t_m$$

To avoid local minima, we re-ran *fminsearch* 100 times using random starting parameter values and selected the set of parameter estimates with the smallest mean square error. Sites for which the fitting procedure failed to converge or produced extreme estimates of the chronometric shift (greater than 200 tones per s) were excluded from further analysis.

In summary plots, linear regression using the ordinary least squares method was used to evaluate the relationship between frequency tuning and stimulation-evoked choice biases. Confidence intervals for regression fits were estimated using bootstrap resampling. A small number of sites (2 out of 59 using non-specific cortical stimulation; 4 out of 56 using striatal axonal stimulation) produced extreme choice biases (greater than 200 tones per s) outside the stimulus range tested in the task and were excluded from regression analysis. Including these sites did not alter the conclusion that striatal axonal stimulation tended to bias subjects in the direction associated with the preferred frequency of the stimulation site ($P = 0.03$, signed-rank test).

As we expected that the effects of inactivation would only be detectable when a large fraction of the local population of corticostriatal neurons were silenced, we excluded sites with low Arch expression near the inactivation site (greater than 1,000 cells) in the analysis of behavioural effects of Arch-inactivation in Fig. 3c. For two fibres in one of the rats we could not locate the tracks in histological sections and precisely estimate cell counts. Excluding the behavioural effects of inactivation at these fibres from Fig. 3c did not affect our conclusions ($P = 0.014$ for linear regression of choice bias and frequency tuning). Furthermore, the relationship between inactivation choice bias and tuning was unchanged whether we selected sites with greater than 250 neurons ($P = 0.012$) or greater than 500 neurons ($P = 0.0052$) near the fibre.

Analysis of electrophysiology data. To isolate single units, spikes were manually clustered using MClust (MClust-3.5, http://redishlab.neuroscience.umn.edu/MClust/MClust.html). Neurometric functions were computed using the first 200 ms of the auditory response and only included trials in which the rat remained in the centre port for at least that period of time. We selected neurons whose firing rate during that epoch was greater than or equal to 0.5 Hz. We used leave-one-out cross-validation to determine neurometric thresholds and frequency preference for each neuron. Specifically, for each trial we used ROC (receiver operating characteristic) analysis, including firing rates on all other trials in the recording session, to select a firing rate threshold that best discriminated the frequency content of the auditory stimulus and to determine whether the neuron prefers high or low frequency stimuli. Trials in which the firing rate was greater than or equal to the discrimination threshold were scored as reporting the preferred frequency of the neuron. The neuronal choices were then fit with a logistic regression model (see equation (1)).

To calculate peristimulus time histograms (PSTHs), neuronal firing rates were smoothed with a causal half-Gaussian kernel ($\sigma = 5$ ms). Confidence intervals were derived through bootstrap resampling.

For each Arch-expressing corticostriatal neuron that we encountered, we delivered light at different fibres along the array to estimate the maximum distance at which light delivery could reduce the neuron's firing by 50%. We took this distance to be the mean of the distances to the furthest fibre for which inactivation was greater than 50% and the closest fibre for which inactivation was less than 50%. The likelihood of inactivation at a given distance was estimated as the fraction of neurons inactivated at that point.

Histology and cell-count analysis. At the end of the experiment, rats were deeply anaesthetized with ketamine and medetomidine. Small electrolytic lesions were made by passing 30 μA direct cathodal current through each tetrode for approximately 10 s, marking the final position of the tetrode tip. The rats were then perfused with 4% paraformaldehyde (PFA), their brains were extracted and post-fixed in 4% PFA overnight. The brains were cut into 100-μm sections and mounted using Vectashield (Vector Laboratories) for confocal microscopy. To ensure that HSV-mediated labelling was confined to corticostriatal neurons, we verified that viral expression was absent in adjacent brain structures.

To quantify the depth distribution of opsin expression, we measured the distance from each fluorescent cell soma to the pia as a fraction of total cortical thickness. To estimate the number of neurons affected by optogenetic manipulations, confocal stacks 50 μm in depth were acquired from alternate sections. The sections were registered using rigid registration maximizing the cross-correlation of fluorescence images of adjacent sections. The locations of ChR2- or Arch-expressing neurons were identified manually. Fibre tracks were identified with the help of electrolytic lesions and DiI labelling. Using the locations of the ends of the tracks, we estimated that processing resulted in approximately 10% shrinkage of the tissue. We estimated the location of the fibre tip during each inactivation session and counted the number of expressing neurons within 1 mm from the fibre at 35 sites illuminated with 5 mW light. At 6 sites illuminated at 2.5 mW we reduced the radius according to square root of the power to 0.7 mm. Excluding these sites from the analysis did not affect our conclusions ($P = 0.036$, linear regression of cell count and choice bias). As we only identified cells within a 50-μm stack every 200 μm, our estimate of the total number of manipulated cells is four times this count.

28. Kelly, J. B. & Masterton, B. Auditory sensitivity of the albino rat. *J. Comp. Physiol. Psychol.* **91,** 930–936 (1977).
29. Palmer, J., Huk, A. & Shadlen, M. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5,** 376–404 (2005).

LETTER

Non–redundant coding of aversive odours in the main olfactory pathway

Adam Dewan[1]*, Rodrigo Pacifico[1]*, Ross Zhan[1], Dmitry Rinberg[2]† & Thomas Bozza[1]

Many species are critically dependent on olfaction for survival. In the main olfactory system of mammals, odours are detected by sensory neurons that express a large repertoire of canonical odorant receptors and a much smaller repertoire of trace amine-associated receptors (TAARs)[1–4]. Odours are encoded in a combinatorial fashion across glomeruli in the main olfactory bulb, with each glomerulus corresponding to a specific receptor[5–7]. The degree to which individual receptor genes contribute to odour perception is unclear. Here we show that genetic deletion of the olfactory *Taar* gene family, or even a single *Taar* gene (*Taar4*), eliminates the aversion that mice display to low concentrations of volatile amines and to the odour of predator urine. Our findings identify a role for the TAARs in olfaction, namely, in the high-sensitivity detection of innately aversive odours. In addition, our data reveal that aversive amines are represented in a non-redundant fashion, and that individual main olfactory receptor genes can contribute substantially to odour perception.

There are 15 *Taar* genes in the mouse, 14 of which are expressed in the main olfactory pathway and serve a chemosensory function[8]. All of the *Taar* genes are located in a single gene cluster on mouse chromosome 10 with no interspersed genes[9] (Fig. 1a). To determine how the TAARs contribute to odour perception, we used *in vivo* trans-allelic recombination to generate a mouse strain ('ΔT2-9') in which all 14 olfactory *Taar* genes (from *Taar2* through to *Taar9*) are deleted (Fig. 1a). Homozygous ΔT2-9 mice breed normally, show no apparent health issues or behavioural deficits, and exhibit the same weight and locomotor activity as wild-type littermates (Supplementary Fig. 1).

To examine the functional consequences of removing the TAARs, we performed *in vivo* optical imaging of odour-evoked responses from glomeruli in the olfactory bulbs of anaesthetized mice. This was done by crossing ΔT2-9 mice to 'OMP-spH' mice in which the genetically encoded activity reporter synaptophluorin (spH) is expressed from the olfactory marker protein (*Omp*) locus and thus in all glomeruli[10]. We compared odorant responses in ΔT2-9 homozygous mice with those in control mice that retain two intact *Taar* gene clusters.

Glomeruli in the olfactory bulb receive axonal inputs from sensory neurons that express the same odorant receptor or *Taar* gene, and sensory neurons that express a majority of the TAARs project to a cluster of glomeruli in the dorsal-caudal olfactory bulb[11,12] (Fig. 1b). Consistent with our previous observations, low concentrations of structurally diverse amines robustly activated a small subset of dorsal glomeruli in control mice (Fig. 1c–e), and glomeruli with specific response profiles could be recognized across individual animals[11,13]. All of these high-sensitivity amine responses were abolished in homozygous ΔT2-9 mice, while responses to non-amine odorants persisted (Fig. 1c–e). These results demonstrate that all of the high-sensitivity amine responses derive from glomeruli corresponding to *Taar* genes.

Urine is a rich source of amines that could be exploited for intra- and interspecific chemical communication. It has been reported that the urine of predator cats contains high concentrations of β-phenylethylamine (PEA), an odorant that specifically activates TAAR4 in cultured cells[8,14]

and TAAR4-expressing olfactory sensory neurons[13]. Using our *in vivo* imaging method, we observe that TAAR4 glomeruli are activated by PEA and by the volatiles from the urine of an adult puma (Fig. 1e). Responses to puma urine in the dorsal bulb were abolished in homozygous ΔT2-9 mice (Fig. 1e). Therefore, the most sensitive amine/urine-responsive glomeruli in the dorsal bulb correspond to the TAARs. We note that it is possible that glomeruli outside our imaging area (in the ventral bulb) respond to amines at the concentrations tested.

Many amines share a characteristic, offensive odour. In fact, two primary amines, PEA and isopentylamine (IPA), have been reported to elicit innate aversion in mice[14,15]. We therefore tested whether the amines that activate TAAR glomeruli are aversive and whether the TAARs mediate this aversion. Wild-type, heterozygous and homozygous ΔT2-9 littermates (*n* = 504 mice) were tested in a two-chamber place preference assay where they could choose to occupy an odorized or a non-odorized compartment (Fig. 2a). Odorants were diluted in water and presented in partially enclosed dishes so that mice could smell the stimuli without direct contact with the odour source. Under these conditions, all mice strongly avoided the well-characterized aversive odorant trimethylthiazoline (TMT, 2% in water), which is derived from the anal gland of the red fox[15–17]. In contrast, negative control odours, water, ethyl vanillin and peanut butter oil, did not elicit aversion (Fig. 2b).

Using this assay, we observe that wild-type and heterozygous ΔT2-9 mice exhibit aversion to structurally diverse amines including PEA, IPA, *N*-methylpiperidine (NMP) and cadaverine (CAD) when tested at multiple concentrations. Notably, the aversion elicited by low concentrations of amines was TAAR-dependent as it was abolished in homozygous ΔT2-9 mice (Fig. 2c). We note that concentrated amines (100% PEA and 10% IPA), which are highly pungent to humans, were aversive to mice regardless of genotype (Fig. 2c). Thus, mice are averse to certain amines, and aversion to low concentrations of amines is dependent on the TAARs.

To determine whether the TAARs are required for aversion to natural stimuli that contain ethologically relevant concentrations of amines, we tested for avoidance of the odour of predator cat urine, which is enriched in PEA[14]. Wild-type and heterozygous ΔT2-9 mice were averse to puma urine. The aversion to urine was abolished in homozygous ΔT2-9 mice, which lack the olfactory TAARs (Fig. 2b, c). Taken together, the data indicate that the TAAR family is required for innate aversive responses to volatile amines at naturally occurring concentrations.

Next, we examined the functional impact of removing a single *Taar* gene from the receptor repertoire. To do this, we used a gene-targeted mouse strain (ΔT4-YFP) in which the TAAR4 coding sequence is replaced with that of yellow fluorescent protein (YFP)[11](Fig. 1a). TAAR4 responds selectively and robustly to PEA and urinary volatiles from predator cats when expressed in cultured cells and in native olfactory sensory neurons[8,13,14]. Using our *in vivo* imaging assay, we find that low concentrations of PEA or volatiles from puma urine

[1]Department of Neurobiology, Northwestern University, Evanston, Illinois 60208, USA. [2]HHMI, Janelia Farm Research Campus, Ashburn, Virginia 20165, USA. †Present address: Neuroscience Institute, New York University Langone Medical Center, New York, New York 10016, USA.
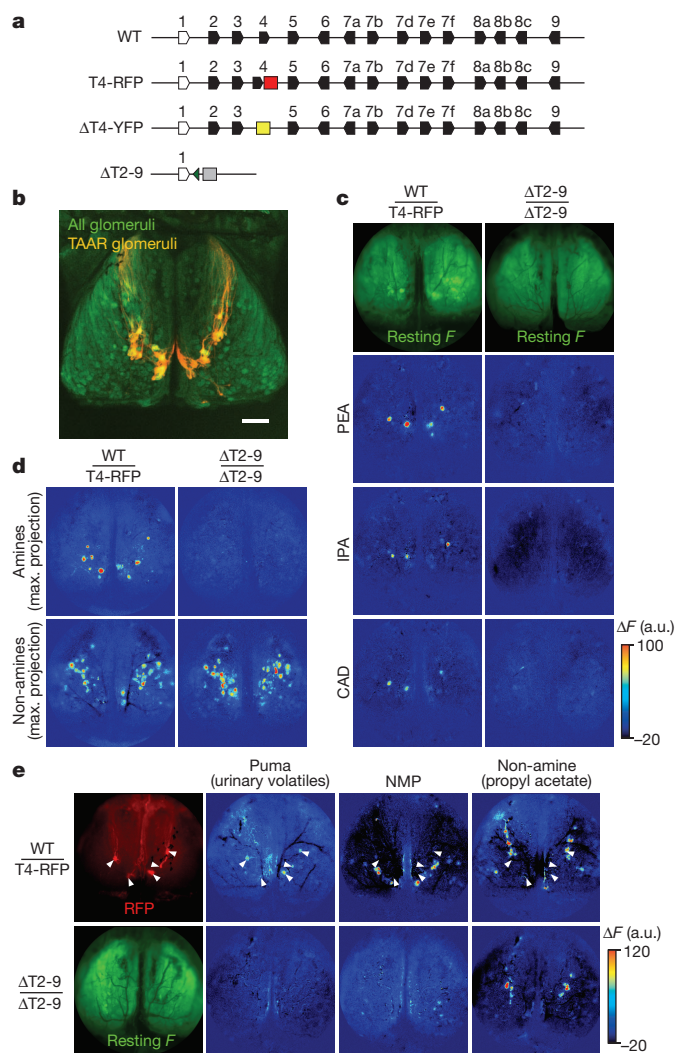*These authors contributed equally to this work.

Figure 1 | Deleting the olfactory TAARs abolishes high-sensitivity amine and predator odour responses in the dorsal olfactory bulb. a, Diagram of the *Taar* gene cluster and targeted alleles. Olfactory *Taars* (black) and non-olfactory *Taar1* (white) are shown (polygons reflect gene orientation). T4-RFP: the tau::mCherry marker (red) is inserted downstream of *Taar4*. ΔT4-YFP: the *Taar4* coding sequence is replaced with Venus YFP (yellow). ΔT2-9: all olfactory TAAR genes are deleted and the *Taar9* coding sequence is replaced with the odorant receptor S50 or CFP (cyan fluorescent protein; grey box; green triangle indicates *loxP* site; see Methods). WT, wild type. **b**, Dorsal view of the olfactory bulbs from a double heterozygous ΔT4-YFP;OMP-spH mouse in which all glomeruli express spH (green) and TAAR glomeruli are labelled (yellow). Anterior is up. Scale bar, 500 μm. **c**, Imaging of odour evoked activity in the olfactory bulbs of a heterozygous T4-RFP mouse (left panels) and a homozygous ΔT2-9 mouse (right panels). Top panels show resting spH fluorescence. Pseudocoloured panels show fluorescence changes in response to β-phenylethylamine (PEA; 2 nM vapour concentration, v.c.), isopentylamine (IPA; 24 nM v.c.) and cadaverine (CAD; 550 nM v.c.). Data are displayed as Δ*F* in arbitrary units (see Methods). Maximum response is 8.3% Δ*F/F*. **d**, Maximum response projection for amines or non-amines in a heterozygous T4-RFP mouse (left) and a homozygous ΔT2-9 mouse (right). Amine stimuli: PEA (2 nM v.c.), IPA (24 nM v.c.), CAD (550 nM v.c.), *N*-methylpiperidine (NMP; 7.5 μM v.c.) and trimethylamine (4 μM v.c.). Non-amine stimuli: propyl acetate (19 μM v.c.), phenetole (860 nM v.c.), 2-heptanone (5.3 μM v.c.) and isopropyl tiglate (1.5 μM v.c.). Maximum response, 7.4% Δ*F/F*. **e**, Responses to urinary volatiles in heterozygous T4-RFP (top) and homozygous ΔT2-9 (bottom) mice. Locations of red fluorescent protein (RFP)-labelled TAAR4 glomeruli (top left) are indicated (arrowheads). SpH fluorescence is shown in a ΔT2-9 mouse (bottom left). Pseudocoloured panels show responses to puma urine (undiluted headspace vapour), NMP (150 nM v.c.) and propyl acetate (19 μM v.c.). Maximum response, 5.4% Δ*F/F*.
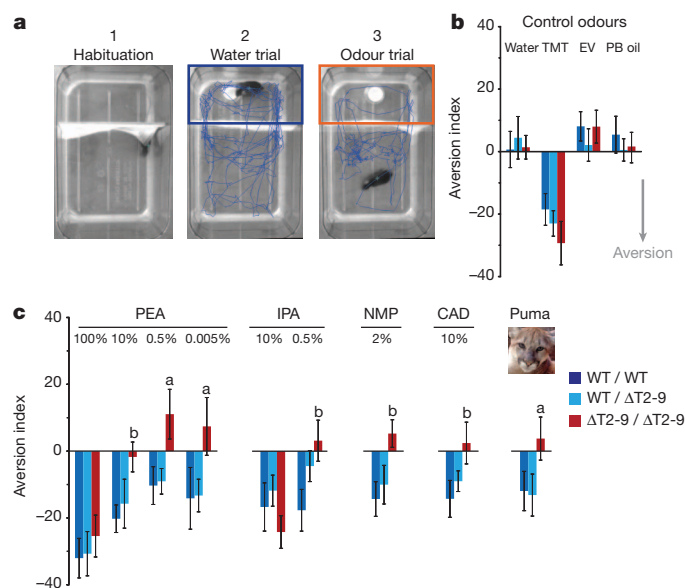
Figure 2 | Deletion of all olfactory *Taar* genes abolishes aversion to low concentrations of structurally diverse amines and predator urine. a, Video images of the behavioural testing chamber. Mice move between odorized (top) and non-odorized (bottom) compartments separated by a curtain. Blue traces show the location of the mouse during single three-minute trials. Panels represent three stages of one experiment—habituation to the chamber, exposure to water, and exposure to a test stimulus (odour or water). Aversion index = Time$_{odour}$ − Time$_{water}$. **b**, Aversion index values for wild-type, heterozygous and homozygous ΔT2-9 cluster deletion mice (key in **c**). Odorants are 2% trimethylthiazoline (TMT), 0.5% ethyl vanillin (EV) and undiluted peanut butter oil (PB oil). Data are mean ± s.e. ($n = 20$–25 mice per genotype, per odorant). **c**, Aversion index values for PEA, IPA, NMP, CAD and puma urine. Concentrations are given as percentage dilution of pure odorant in water. Puma urine was undiluted. Data are mean ± s.e. ($n = 20$–25 mice per genotype, per odorant). Statistical significances for pairwise comparisons are indicated: 'a', homozygous mice differ from wild type and heterozygous, 'b', homozygous mice differ from wild type ($P < 0.05$, generalized linear mixed model). Wild-type and heterozygous mice did not differ statistically for any odour, and the aversion response did not differ with sex ($P = 0.669$).

preferentially activate TAAR4 glomeruli in the dorsal bulb (Fig. 3a). Genetic deletion of *Taar4* specifically eliminated these high-sensitivity responses to PEA and puma urine volatiles without influencing the activation of neighbouring glomeruli by other amines (Fig. 3a).

To determine whether behavioural aversion to PEA and predator urine is mediated by TAAR4, we tested wild-type, heterozygous and homozygous ΔT4-YFP littermates ($n = 245$ mice) in the two-chamber place preference test described above (Fig. 2a). As expected, wild-type and heterozygous ΔT4-YFP mice exhibited robust aversion to low concentrations of PEA, IPA and NMP, as well as to urine from two predator species, puma and Canadian lynx. In contrast, homozygous ΔT4-YFP mice (which lack TAAR4) exhibited no avoidance of PEA or predator urine odours (Fig. 3b). The loss of aversion was odour-specific, as homozygous ΔT4-YFP mice still avoided IPA and NMP, amines that activate other TAAR glomeruli (Fig. 3b). Thus, removing a single *Taar* gene, *Taar4*, abolishes aversion to PEA and urinary volatiles.

Taken together, our data reveal that the *Taar* gene family contributes significantly to the perception of amines in mice. Moreover, the aversive quality of amines and predator urine is encoded in the olfactory system in a non-redundant manner, because removal of even a single *Taar* gene can have a significant impact on the aversive response. The vomeronasal system of mice mediates innate avoidance of proteins found in the urine of predator species[18,19], a process that may require direct contact with the stimulus. The TAARs may contribute to predator avoidance over longer distances by mediating aversion to trace concentrations of urinary volatiles. In this regard, it is interesting to note that the response
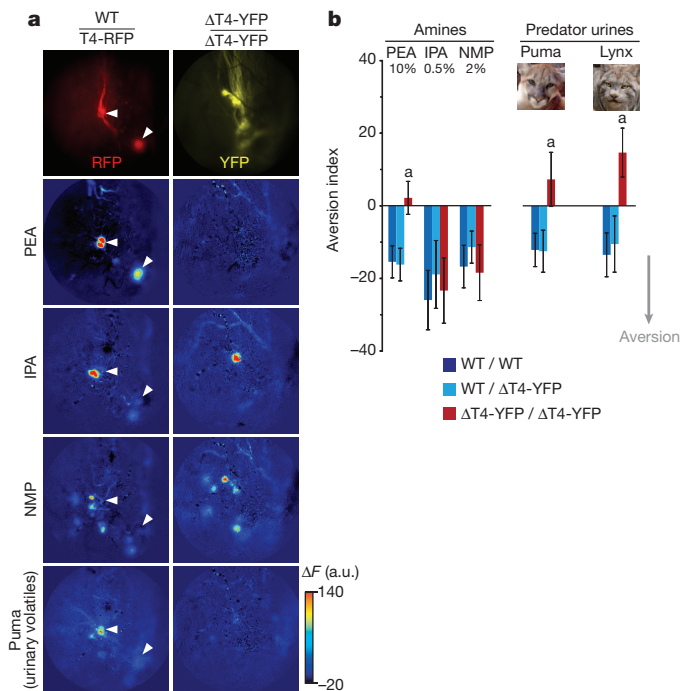
Figure 3 | Deletion of a single *Taar* gene abolishes aversion to a specific amine and to natural predator odours. a, *In vivo* imaging of the left caudomedial olfactory bulb in a heterozygous T4-RFP mouse and a homozygous ΔT4-YFP mouse (anterior is up, medial is right). TAAR4 glomeruli are labelled in T4-RFP mice (top left panel, red) and their locations are indicated with arrowheads in subsequent panels. All dorsal TAAR glomeruli, except for those corresponding to the deleted *Taar4* gene, are labelled in homozygous ΔT4-YFP mice (top right panel, yellow)[11]. Pseudocoloured panels show odour-evoked fluorescence changes in response to PEA (2 nM v.c.), IPA (24 nM v.c.), NMP (150 nM v.c.) and puma urine (undiluted headspace vapour). Response maps are displayed as ΔF in arbitrary units (see Methods). Maximum response, 7.5% $\Delta F/F$. **b,** Aversion index values for wild-type, heterozygous and homozygous ΔT4-YFP mice. Negative values indicate avoidance. Odorants are PEA, IPA, NMP, puma urine and Canadian lynx urine. Concentrations given as percentage dilution of pure odorant in water. Predator urines were undiluted. Data are mean ± s.e. ($n = 20$–25 mice per genotype, per odorant). Statistical significances for pairwise comparisons: 'a', homozygous mice differ from both wild-type and heterozygous mice ($P < 0.05$, generalized linear mixed model). Wild-type and heterozygous mice did not differ statistically for any odour, and the aversion response did not differ with sex ($P = 0.639$).

thresholds of TAAR4 sensory neurons and glomeruli to the aversive predator cue, PEA, are the lowest so far observed in the main olfactory system[13].

We note that the TAARs may not function solely as detectors of predator-derived and aversive odours. Although we show that mouse TAARs contribute to the detection and avoidance of several aversive amines, recent data indicate that TAAR5 mediates attraction to its preferred ligand, trimethylamine—a socially relevant metabolite that is enriched in male mouse urine[20]. However, it should be noted that this amine elicits robust aversion in rats. The TAAR repertoire is also evolutionarily retained in many vertebrate species, including humans[3,9]. Apart from their role in aversion or attraction, our view is that the TAARs are retained in many species because they are required more generally for high-sensitivity amine detection. The behavioural response to this input may be context and species specific.

It is generally thought that odour representations in the main olfactory bulb are highly distributed and redundant, with each input channel (glomerulus) making a small contribution to the representation of a given odour[21]. In this view, single receptor deletions should have little effect at the level of behaviour. Behavioural deficits have been induced by the genetic removal of receptors in specialized olfactory pathways.

Mutant mice lacking a subset of vomeronasal receptors, which map to the accessory olfactory bulb, display deficits in aggression and mating[22]. Mice lacking the gene for guanylyl cyclase D, which is expressed in sensory neurons that project to atypical necklace glomeruli, show deficits in social transmission of food preference[23,24]. In contrast, the TAARs are mapped to a subset of the ~2,000 typical main olfactory bulb glomeruli that are thought to represent odorants in a combinatorial fashion. In spite of this, removal of even a single TAAR results in a measurable deficit in odour-guided behaviour. Our data suggest that the representations of general odours in the main olfactory system may be less redundant than previously thought, a fact that may shed light on how vertebrates retain large numbers of chemosensory receptor genes over evolutionary time.

METHODS SUMMARY

All procedures were approved by the Northwestern University Animal Care and Use Committee. Generation of the ΔT2-9^{S50}, T4-RFP and ΔT4-YFP alleles was previously described[11]. The TAAR cluster deletion allele ΔT2-9^{CFP} was similarly generated by *in vivo* Cre-mediated *trans*-allelic recombination[11,25]. Imaging of odour-evoked activity from the olfactory bulbs was performed as previously described with some modifications[10,11]. Recordings were made from 8 to 12 week-old male and female heterozygous OMP-spH mice that were heterozygous for the T4-RFP allele ($n = 13$), homozygous for the ΔT2-9^{S50} allele ($n = 9$), or homozygous for the ΔT4-YFP allele ($n = 4$). Data are expressed as ΔF to account for the properties of the resting spH fluorescence[10,26]. Response images are averages of two or more trials, or the first trial in cases where significant adaptation was observed.

Aversion was assessed in 5–7-week-old male and female F$_1$ mice from heterozygous crosses. A total of 504 ΔT2-9^{CFP} and 238 ΔT4-YFP mice were used for 1,138 experiments (20–25 mice of each genotype for each odour). The assay was similar to that used in previous studies[14,15] and was performed under dim red light. A curtain divided the cage into odorized (1/3) and non-odorized (2/3) chambers. After two days of habituation to the experimental protocol (with no odour), mice were re-habituated to the chamber for 3 min, exposed to a water stimulus for 3 min, and then to the test stimulus for 3 min. The location of the mouse was tracked using Limelight 3.0 software (Actimetrics). The aversion index was calculated as the difference between the times spent in the odorized chamber with an odorant as a stimulus and with water as a stimulus. Data were analysed with a generalized linear mixed model with genotype, odour and sex as fixed factors and a random-subjects factor using SPSS (IBM). Pairwise contrasts were used to compare genotypes for each odour with a least-square-differences adjustment for multiple comparisons.

Full Methods and any associated references are available in the online version of the paper.

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65,** 175–187 (1991).
2. Liberles, S. D. Trace amine-associated receptors are olfactory receptors in vertebrates. *Ann. NY Acad. Sci.* **1170,** 168–172 (2009).
3. Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Rev. Genet.* **9,** 951–963 (2008).
4. Touhara, K. & Vosshall, L. B. Sensing odorants and pheromones with chemosensory receptors. *Annu. Rev. Physiol.* **71,** 307–332 (2009).
5. Wilson, R. I. & Mainen, Z. F. Early events in olfactory processing. *Annu. Rev. Neurosci.* **29,** 163–201 (2006).
6. Kauer, J. S. & White, J. Imaging and coding in the olfactory system. *Annu. Rev. Neurosci.* **24,** 963–979 (2001).
7. Hildebrand, J. G. & Shepherd, G. M. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu. Rev. Neurosci.* **20,** 595–631 (1997).
8. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442,** 645–650 (2006).
9. Lindemann, L. *et al.* Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* **85,** 372–385 (2005).
10. Bozza, T., McGann, J. P., Mombaerts, P. & Wachowiak, M. *In vivo* imaging of neuronal activity by targeted expression of a genetically encoded probe in the mouse. *Neuron* **42,** 9–21 (2004).
11. Pacifico, R., Dewan, A., Cawley, D., Guo, C. & Bozza, T. An olfactory subsystem that mediates high-sensitivity detection of volatile amines. *Cell Reports* **2,** 76–88 (2012).

12. Johnson, M. A. *et al.* Neurons expressing trace amine-associated receptors project to discrete glomeruli and constitute an olfactory subsystem. *Proc. Natl Acad. Sci. USA* **109,** 13410–13415 (2012).
13. Zhang, J., Pacifico, R., Cawley, D., Feinstein, P. & Bozza, T. Ultrasensitive detection of amines by a trace amine associated receptor. *J. Neurosci.* **33,** 3228–3239 (2013).
14. Ferrero, D. M. *et al.* Detection and avoidance of a carnivore odor by prey. *Proc. Natl Acad. Sci. USA* **108,** 11235–11240 (2011).
15. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450,** 503–508 (2007).
16. Vernet-Maury, E., Polak, E. H. & Demael, A. Structure-activity relationship of stress-inducing odorants in the rat. *J. Chem. Ecol.* **10,** 1007–1018 (1984).
17. Fendt, M., Endres, T., Lowry, C. A., Apfelbach, R. & McGregor, I. S. TMT-induced autonomic and behavioral changes and the neural basis of its processing. *Neurosci. Biobehav. Rev.* **29,** 1145–1156 (2005).
18. Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141,** 692–703 (2010).
19. Ben-Shaul, Y., Katz, L. C., Mooney, R. & Dulac, C. *In vivo* vomeronasal stimulation reveals sensory encoding of conspecific and allospecific cues by the mouse accessory olfactory bulb. *Proc. Natl Acad. Sci. USA* **107,** 5172–5177 (2010).
20. Li, Q. *et al.* Synchronous evolution of an odor biosynthesis pathway and behavioral response. *Curr. Biol.* **23,** 11–20 (2013).
21. Kauer, J. S. Contributions of topography and parallel processing to odor coding in the vertebrate olfactory pathway. *Trends Neurosci.* **14,** 79–85 (1991).
22. Del Punta, K. *et al.* Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419,** 70–74 (2002).
23. Munger, S. D. *et al.* An olfactory subsystem that detects carbon disulfide and mediates food-related social learning. *Curr. Biol.* **20,** 1438–1444 (2010).
24. Munger, S. D., Leinders-Zufall, T. & Zufall, F. Subsystem organization of the mammalian sense of smell. *Annu. Rev. Physiol.* **71,** 115–140 (2009).
25. Wu, S., Ying, G., Wu, Q. & Capecchi, M. R. Toward simpler and faster genome-wide mutagenesis in mice. *Nature Genet.* **39,** 922–930 (2007).
26. McGann, J. P. *et al.* Odorant representations are modulated by intra- but not interglomerular presynaptic inhibition of olfactory sensory neurons. *Neuron* **48,** 1039–1053 (2005).

METHODS

Gene targeting. The TAAR cluster deletion allele ΔT2-9^{CFP} was generated by Cre-mediated *trans*-allelic recombination *in vivo*[25]. We used two targeted alleles that introduce *loxP* sites into the 5′ and 3′ ends of the cluster—*au1::Taar1-loxP-IRES-tau::Venus* (aT1-YFP) in which a loxP site is inserted just downstream of the *Taar1* coding sequence, and *Cerulean→Taar9-loxP* (ΔT9-CFP) in which the *Taar9* coding sequence is replaced with that of Cerulean CFP followed by *loxP*. An *HPRT-Cre* strain (129S1/Sv-Hprt$^{tm1(cre)Mnn}$/J; Jax 004302; ref. 27) was used to mediate recombination in aT1-YFP/ΔT9-CFP compound heterozygotes as described[11].

***In vivo* imaging.** For glomerular imaging, mice were anaesthetized with sodium pentobarbital as described[11], or with urethane (1 g kg^{-1} i.p.; Sigma) and chlorprothixene hydrochloride (10 mg kg^{-1}), and given atropine sulphate (5.4 mg kg^{-1}; Med-Pharmex). The bone overlying the bulbs was thinned using a dental drill. Glomeruli were imaged using a custom Nikon epifluorescence microscope and a 4$\times$ (0.2 NA) objective. Light excitation was provided using a 200 W metal-halide lamp (Prior Scientific) attenuated by neutral density filters and standard filter sets for mCherry (49008; Chroma), YFP (86001 JP3, Chroma), or green fluorescent protein (GFP) (96343, Nikon).

Odorants were applied using a custom-made, flow dilution olfactometer and controller (LASOM, RPMetrix). Amines were diluted in water and subsequently by flow dilution. Predator urine volatiles were applied from the undiluted headspace concentration. Images were acquired at 25 Hz over 20 s (encompassing a 4 s pre-stimulus period and a 4 s odorant pulse) using a NeuroCCD-SM256 camera and Neuroplex software (RedShirtImaging). Blank trials were subtracted from odour trials before analysis to compensate for photobleaching. Response maps were obtained by subtracting a 3 s temporal average preceding the stimulus from a 3 s temporal average encompassing the response peak. Responses are expressed as ΔF to account for the fact that the background spH fluorescence is not correlated with the pool of indicator that reports neuronal activity[10,26]. Stimuli were presented at least twice in a given experiment. Images were processed and analysed in Neuroplex (RedShirtImaging) and Image J (http://imagej.nih.gov/ij) software. Vapour concentrations were estimated using published vapour pressures (US EPA, Estimation Programs Interface Suite, v 4.0).

Behavioural analysis. ΔT4-YFP and ΔT2-9^{CFP} littermates were housed in same-sex groups of 2–5 individuals. All animals were maintained in a reverse 12/12 h light-dark cycle and provided with food and water *ad libitum*. Cages were changed daily to prevent adaptation to amines that are present in mouse urine. Mice of all genotypes were tested between 5 and 7 weeks of age under low intensity red light during the nocturnal phase.

The experimental protocol consisted of three parts: handling (2 days), pre-trials (2 days) and experimental trials (1–3 days). Handling habituated the mice to the experimenter and consisted of placing each mouse individually onto the experimenter's cupped, gloved hands for five minutes and allowing them to roam this small area freely. Pre-trials were identical to experimental trials (see below) except that no odorants were used. Pre-trials functioned to eliminate the novelty of the odour delivery and experimental chamber. Experiments were performed in clean, autoclaved 30 $\times$ 18 $\times$ 12 cm cages. A disposable curtain isolated one-third of the cage with minimal air transfer between sections. This smaller section, or 'odorized' compartment, was topped with a thin piece of clear acrylic, which functioned to minimize the loss of odour. Each mouse was introduced to the larger section of the cage and allowed to habituate for 3 min. At the end of this time period, a 3.5 cm covered Petri dish containing 20 µl of water on filter paper was introduced to the odorized compartment. The top of the Petri dish was perforated to allow odorants to escape, but to prevent direct contact with the stimulus. Mice were allowed to interact with this Petri dish for 3 min. The Petri dish was then removed and another identical Petri dish with 20 µl of an odorant on filter paper was added. The experiment was terminated after another 3 min.

The acrylic top was cleaned with 70% isopropyl alcohol and the cages were washed and autoclaved. The mice were video recorded and their location tracked using Limelight 3.0 software (Actimetrics). The aversion index was calculated as the difference between the time spent in the odorized chamber when an odour (or water) was present and the time spent in the odorized chamber when water was present. Data from mice that showed a very strong preference for either chamber (those that spent <2% or >98% of the trial duration in the odorized chamber) during the initial water trial were discarded.

Mice were naïve to each stimulus and were tested only once for a given odour. Monomolecular stimuli consisted of three control odorants, water, trimethylthiazoline (2% in water) and a saturated solution of ethyl vanillin (0.5% w/v, in water), as well as 4 amines that activate TAAR glomeruli, β-phenylethylamine (100%, 10%, 0.5% and 0.005%), isopentylamine (10% and 0.5%), N-methylpiperidine (2%) and cadaverine (10%). Complex odorants were undiluted peanut butter oil, puma urine (from *Puma concolor*) and lynx urine (from *Lynx canadensis*). Predator urines were collected at the Philadelphia Zoo, shipped frozen and stored at −80 °C. Odour concentrations represent what was placed in the Petri dish and are expressed as percent dilution in water. The saturated vapour concentrations (maximum possible odorant concentration) for the stimuli are as follows: 100% phenylethylamine = 22 µM; 10% isopentylamine = 243 µM; 10% cadaverine = 5 µM; 2% N-methylpiperidine = 30 µM. The actual stimulus concentrations in the odorized chamber are probably much lower than these theoretical maxima. We note that the concentrations in the behavioural and imaging experiments are difficult to compare given the differences in odour presentation methods.

27. Tang, S. H., Silva, F. J., Tsark, W. M. & Mann, J. R. A cre/*loxP*-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis* **32**, 199–202 (2002).

LETTER

Sema3A regulates bone–mass accrual through sensory innervations

Toru Fukuda[1]*, Shu Takeda[1]*, Ren Xu[2,3]*, Hiroki Ochi[1], Satoko Sunamura[1], Tsuyoshi Sato[4], Shinsuke Shibata[5], Yutaka Yoshida[6], Zirong Gu[6], Ayako Kimura[2,7], Chengshan Ma[2,3], Cheng Xu[2,3], Waka Bando[8], Koji Fujita[2,3], Kenichi Shinomiya[2], Takashi Hirai[2], Yoshinori Asou[2], Mitsuhiro Enomoto[2], Hideyuki Okano[5], Atsushi Okawa[2,3] & Hiroshi Itoh[8]

Semaphorin 3A (Sema3A) is a diffusible axonal chemorepellent that has an important role in axon guidance[1–5]. Previous studies have demonstrated that $Sema3a^{-/-}$ mice have multiple developmental defects due to abnormal neuronal innervations[6,7]. Here we show in mice that Sema3A is abundantly expressed in bone, and cell-based assays showed that Sema3A affected osteoblast differentiation in a cell-autonomous fashion. Accordingly, $Sema3a^{-/-}$ mice had a low bone mass due to decreased bone formation. However, osteoblast-specific Sema3A-deficient mice ($Sema3a_{col1}^{-/-}$ and $Sema3a_{osx}^{-/-}$ mice) had normal bone mass, even though the expression of Sema3A in bone was substantially decreased. In contrast, mice lacking Sema3A in neurons ($Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice) had low bone mass, similar to $Sema3a^{-/-}$ mice, indicating that neuron-derived Sema3A is responsible for the observed bone abnormalities independent of the local effect of Sema3A in bone. Indeed, the number of sensory innervations of trabecular bone was significantly decreased in $Sema3a_{synapsin}^{-/-}$ mice, whereas sympathetic innervations of trabecular bone were unchanged. Moreover, ablating sensory nerves decreased bone mass in wild-type mice, whereas it did not reduce the low bone mass in $Sema3a_{nestin}^{-/-}$ mice further, supporting the essential role of the sensory nervous system in normal bone homeostasis. Finally, neuronal abnormalities in $Sema3a^{-/-}$ mice, such as olfactory development, were identified in $Sema3a_{synasin}^{-/-}$ mice, demonstrating that neuron-derived Sema3A contributes to the abnormal neural development seen in $Sema3a^{-/-}$ mice, and indicating that Sema3A produced in neurons regulates neural development in an autocrine manner. This study demonstrates that Sema3A regulates bone remodelling indirectly by modulating sensory nerve development, but not directly by acting on osteoblasts.

Sema3A, the first member to be identified from the large semaphorin family, is a well-known chemorepellent[1,2,5]. Sema3A, which is secreted from target tissue, forms steep gradients and causes the inhibition of axonal outgrowth and cell migration in a concentration-dependent manner[4,5]. This repulsive signal is essential for neural development[3]. Indeed, $Sema3a^{-/-}$ mice have several neuronal defects, including abnormalities in the cerebral cortex, olfactory system and axon projections of dorsal root ganglia (DRG) neurons[6–8]. More recently, Sema3A has been implicated in additional physiological functions, such as immunoregulation, angiogenesis and cancer development[5,9–12]. Osteoblasts and osteoclasts express semaphorin family proteins, and recent reports suggested that these proteins regulate bone remodelling locally by acting as paracrine signalling molecules[13–15]. To address the role of Sema3A in bone remodelling, we first analysed the expression in bone of: $Sema3a$; $Nrp1$, encoding a Sema3A receptor; and $Plxna1, 2, 3$ and 4, encoding Sema3A co-receptors[1]. $Sema3a$ was ubiquitously

expressed in many tissues including bone (Fig. 1a), indicating its diverse physiological action. The expression of $Sema3a$ increased during osteoblast differentiation (Fig. 1b) and decreased during osteoclast differentiation (Fig. 1c). Plxna1, 2 and 3 and Nrp1 were expressed in osteoblasts (Fig. 1d and Supplementary Fig. 1), indicating that Sema3A may have a role in osteoblast differentiation. Therefore, we analysed $Sema3a^{-/-}$ mice. $Sema3a^{-/-}$ mice showed a 25% decrease in bone mass due to a decrease in bone formation without an overt change in bone resorption at 3, 6 and 12 months of age, although a non-significant increase was observed in a serum bone resorption marker (Fig. 1e–g and Supplementary Figs 2b, 3), demonstrating that Sema3A accelerates bone formation. To determine whether the defect in bone formation in $Sema3a^{-/-}$ mice was cell autonomous, we isolated osteoblasts from wild-type and $Sema3a^{-/-}$ mice. In accordance with in vivo observations, $Sema3a^{-/-}$ osteoblasts showed a defect in differentiation, as observed
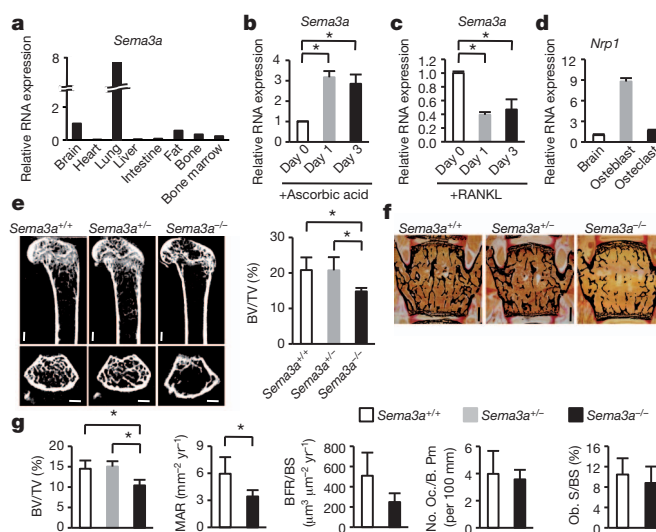

Figure 1 | $Sema3a^{-/-}$ mice have low bone mass. a–d, Gene expression analysis of $Sema3a$ (**a–c**) and its receptor (**d**) ($n = 4$). **e–g**, Analysis of $Sema3a^{-/-}$ mice. Microcomputed tomography (micro-CT) analysis (**e**), histological analysis (**f, g**) in 3-month-old mice (male, $n = 8$ for $Sema3a^{+/+}$, $n = 6$ for $Sema3a^{+/-}$, $n = 10$ for $Sema3a^{-/-}$). Bone volume/tissue volume (BV/TV; %), mineral apposition rate (MAR; mm yr^{-1}), bone formation rate/bone surface (BFR/BS; mm^3 mm^{-2} yr^{-1}), osteoclast number/bone perimeter (No. Oc./B. Pm) and osteoblast surface/bone surface (Ob. S/BS; %). Note a decrease of bone mass in $Sema3a^{-/-}$ mice. Scale bars, 500 μm. Error bars, mean ± standard deviation (s.d.). *$P < 0.05$.

[1]Department of Internal Medicine, School of Medicine, Keio University, Shinanomachi 35, Shinjyuku-ku, Tokyo 160-8582, Japan. [2]Department of Orthopedic Surgery, Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. [3]Global Center of Excellence (GCOE) Program, International Research Center for Molecular Science in Tooth and Bone Diseases, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. [4]Department of Oral and Maxillofacial Surgery, Saitama Medical University, Morohongo 38, Moroyama-machi, Iruma-gun, Saitama 350-0495, Japan. [5]Department of Physiology, School of Medicine, Keio University, Shinanomachi 35, Shinjyuku-ku, Tokyo 160-8582, Japan. [6]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [7]Hard Tissue Genome Research Center, Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. [8]Section of Nephrology, Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, Keio University, Shinanomachi 35, Shinjyuku-ku, Tokyo 160-8582, Japan.
*These authors equally contributed to this work.

by a decrease in alkaline phosphatase activity and the expression of osteoblast markers (Supplementary Fig. 2c). We next determined whether treatment with Sema3A affected osteoblast and osteoclast differentiation and proliferation. Sema3A accelerated osteoblastic differentiation when added exogenously in culture media (Supplementary Fig. 2d). The knockdown of *Sema3a* in osteoblasts hampered their differentiation, and treatment with Sema3A rescued the defects in osteoblast differentiation in *Sema3a* knocked-down osteoblasts (Supplementary Fig. 2e). The expression of a dominant-negative form of Rac1 inhibited Sema3A-dependent osteoblast differentiation (Supplementary Fig. 2f), which indicates that Sema3A regulates osteoblast differentiation through the activation of Rac1, a signalling molecule required for the action of Sema3A on the collapse of the growth cone[16]. Sema3A treatment did not affect osteoblast proliferation (Supplementary Fig. 4). By contrast, the treatment of osteoclast precursors with Sema3A caused a decrease in osteoclast differentiation, as shown by a decrease in the number of multinucleated TRAP-positive osteoclasts and the expression of osteoclast marker genes, such as cathepsin K (Supplementary Fig. 2g). These results suggested that Sema3A functions in an autocrine fashion to stimulate osteoblast differentiation and inhibit osteoclast differentiation, thereby increasing bone mass.

To determine whether Sema3A affected bone metabolism in an autocrine manner *in vivo*, we generated osteoblast-specific Sema3A-deficient mice using $\alpha 1$(I)-Cre mice ($Sema3a_{col1}^{-/-}$ mice)[17]. $Sema3a_{col1}^{-/-}$ mice developed normally without any gross abnormality. The expression of *Sema3a* in the bones of $Sema3a_{col1}^{-/-}$ mice was decreased (Supplementary Fig. 5a), and the osteoblasts isolated from $Sema3a_{col1}^{-/-}$ mice showed a defect in osteoblast differentiation *in vitro* (Supplementary Fig. 5b, c), as observed for $Sema3a^{-/-}$ osteoblasts isolated from $Sema3a^{-/-}$ mice (Supplementary Fig. 2c). However, $Sema3a_{col1}^{-/-}$ mice had normal bone formation and bone mass (Fig. 2a–c and Supplementary Fig. 5d). Moreover, another osteoblast-specific knockout mouse created using the osx-Cre mouse[18], which deletes Sema3A in early osteoprogenitors, did not develop any bone abnormalities ($Sema3a_{osx}^{-/-}$ mice) (Supplementary Fig. 5e). Taken together, these results indicate that the lack of Sema3A in osteoblasts is not the sole cause of bone abnormalities in $Sema3a^{-/-}$ mice.

Bone remodelling has been shown to be regulated by the central and peripheral nervous systems[19]. Thus, we examined the necessity of neuron-derived Sema3A in the maintenance of normal bone mass by deleting Sema3A in neurons using synapsin-I-Cre mice[20] or nestin-Cre mice[21] ($Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice, respectively). $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice developed normally without any growth retardation or perinatal lethality, which were observed in $Sema3a^{-/-}$

mice[6]. Surprisingly, $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice showed significant reductions in bone mass in both males and females, in both vertebrae and long bones, and in both trabecular and cortical bones; that is, in all of the bones in the entire body at any stages analysed (Fig. 3b–d and Supplementary Figs 6, 7b–d, f–h). This low bone mass phenotype was accompanied by decreased bone formation and osteoblast differentiation *in vivo* (Fig. 3d, e). $Sema3a_{synapsin}^{-/-}$ primary osteoblasts did not show overt abnormalities in osteoblast differentiation *in vitro* (Supplementary Fig. 8a, b), whereas $Sema3a_{nestin}^{-/-}$ primary osteoblasts exhibited a defect (Supplementary Fig. 8c, d), indicating a decreased differentiation state of the osteoblast precursors, some of which are known to be nestin positive[22]. Interestingly, expression of *Sema3a* in the bones of $Sema3a_{synapsin}^{-/-}$ mice was decreased (Fig. 3a), suggesting that neurons in the bones are responsible for a substantial amount of Sema3A expression. $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice showed a decrease in bone mass (Fig. 3b–d and Supplementary Figs 6a, b, 7b–d, f–h) whereas $Sema3a_{col1}^{-/-}$ mice had normal bone mass (Fig. 2a–c and Supplementary Fig. 5d) in the presence of comparable decreases in the expression of *Sema3a* in bones in $Sema3a_{col1}^{-/-}$, $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice (Fig. 3a and Supplementary Figs 5a, 7a). This indicated that the decreased expression of *Sema3a* in bone was not the cause of the bone abnormality. In line with this observation, $Sema3a^{+/-}$ mice, in which the expression of *Sema3a* in bone was decreased to a similar extent to in $Sema3a_{col1}^{-/-}$, $Sema3a_{synapsin}^{-/-}$ and $Sema3A_{nestin}^{-/-}$ mice (Fig. 3a and Supplementary Figs 2a, 7a), had normal bone mass (Fig. 1e, f).

Because Sema3A is a well-known axon guidance molecule, especially as a chemorepellent, we tested whether the development of the nervous system was affected in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice. Surprisingly, nerve innervations were significantly decreased in the bones of both mice, as shown by a decrease in neurofilament-positive fibres (Fig. 4a and Supplementary Figs 9, 10). Dopamine β-hydroxylase (DBH)-positive sympathetic nerve fibres, which inhibit bone mass accrual[19], were not significantly affected (Fig. 4b and Supplementary Fig. 10), and these fibres did not colocalize with neurofilament-positive fibres (Supplementary Fig. 11b). Neurofilament-positive fibres in bone are considered to be calcitonin-gene-related peptide (CGRP)-positive sensory fibres[23], and indeed, CGRP-positive and transient receptor potential cation channel subfamily V member 1 (TRPV1)-positive fibre innervations were decreased in $Sema3a_{synapsin}^{-/-}$ and
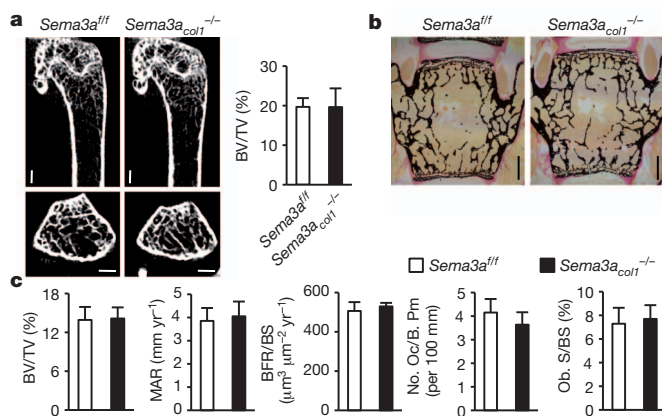
Figure 2 | Osteoblast-derived Sema3A does not affect bone mass *in vivo*. a–c, Analysis of osteoblast-specific Sema3A-knockout mice. a, Micro-CT analysis of the femurs of 3-month-old $Sema3a_{col1}^{-/-}$ mice and their littermates (male, $n = 7$). b, c, Histological analyses of vertebrae (male, $n = 7$). Note an absence of bone abnormalities in $Sema3a_{col1}^{-/-}$ mice. Scale bars, 500 μm. Error bars, mean ± s.d.
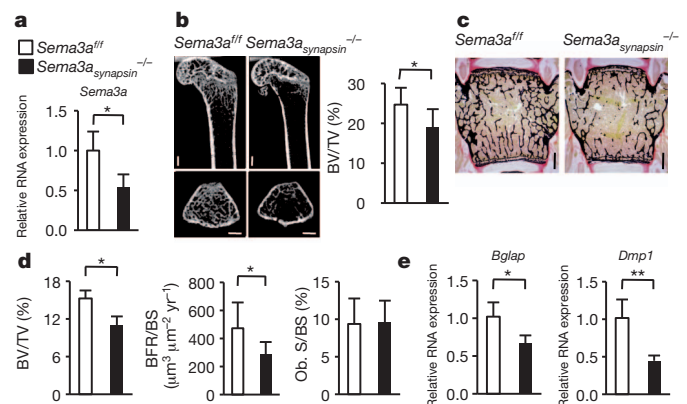
Figure 3 | Neuron-derived Sema3A is indispensable for normal bone mass *in vivo*. a–e, Sema3A was conditionally inactivated in neurons using synapsin-I-Cre mice ($Sema3a_{synapsin}^{-/-}$ mice) at 3 months. a, Quantitative real-time PCR analysis (male, $n = 4$). Note a significant decrease of *Sema3a* expression in $Sema3a_{synapsin}^{-/-}$ mouse femurs. b–d, Micro-CT analysis of femurs (b) or histological analyses of vertebrae (c, d) from $Sema3a_{synapsin}^{-/-}$ mice (male, $n = 8$ for $Sema3a^{f/f}$, $n = 6$ for $Sema3a_{synapsin}^{-/-}$). Note a decrease of bone mass in $Sema3a_{synapsin}^{-/-}$ mice. Scale bars, 500 μm. e, Expression of bone marker genes in femurs (male, $n = 4$). Note a decrease in osteoblast differentiation in $Sema3a_{synapsin}^{-/-}$ mice. Error bars, mean ± s.d. *$P < 0.05$, **$P < 0.01$.

$Sema3a_{nestin}^{-/-}$ mice (Fig. 4c and Supplementary Figs 9c, 10, 11a), and both CGRP- and TRPV1-postive fibres were also neurofilament positive (Supplementary Fig. 11b). In addition, the innervation of these fibres into cortical bone was also reduced (Supplementary Fig. 11c), demonstrating that the projections of peripheral nerve fibres into bone tissue are uniformly decreased, in agreement with all of the bones being affected in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice. To examine rigorously disturbances in nerve innervation in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice, we performed three different experiments. First, $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice were intercrossed with $Sox10$-Venus mice, which express Venus fluorescent protein under the control of a $Sox10$ promoter[24]. Venus-positive sensory nerve innervation was significantly decreased in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice (Supplementary Fig. 12). Second, the retrograde fluorescent tracer, Fluoro-Gold, was injected into the bone marrow of $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice, resulting in fewer neurons being labelled in the DRG (Supplementary Fig. 13). Last, sensory evaluation revealed that $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice were anaesthetic to all of the tested stimuli, indicating functional sensory deficits in these mice (Supplementary Fig. 14). Taken together, these results demonstrated the disruption of nerve innervation in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice. Importantly, a longitudinal analysis showed that bone mass and nerve innervation into bone were decreased at all analysed time points in $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice (that is, from embryo to 3 months old; Figs 3b–d, 4a, c and Supplementary Figs 6a, b, 7b–d, f–h, 15), which further indicated that bone mass accrual during development required normal nerve innervation into bone.

We also confirmed that the projections of CGRP-positive sensory fibres into bone were similarly decreased in $Sema3a^{-/-}$ mice (Fig. 4c), whereas DBH-positive sympathetic nerve fibres were unaffected (Fig. 4b). By contrast, the projections of peripheral nerve fibres into bone tissues were not affected in $Sema3a_{col1}^{-/-}$, $Sema3a_{osx}^{-/-}$ or $Sema3a^{+/-}$ mice (Supplementary Fig. 16), which had normal bone mass, indicating that the abnormal projections of sensory fibres, not the expression of Sema3A in bone, are responsible for the bone abnormalities in $Sema3a^{-/-}$, $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice.
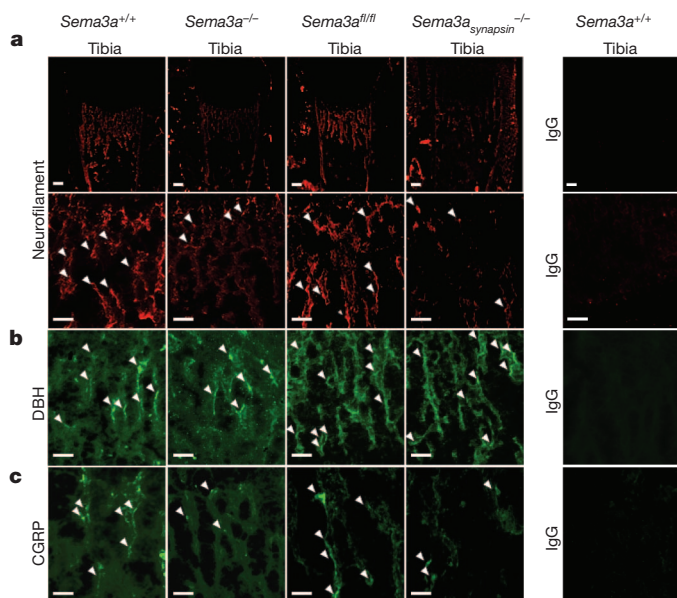
Figure 4 | Neuron-derived Sema3A is indispensable for sensory nerve innervation into bone. a–c, Longitudinal section of tibiae from $Sema3a^{-/-}$, $Sema3a_{synapsin}^{-/-}$ and control mice at postnatal day 3 ($n = 3$), immunostained for neurofilament (**a**, red), DBH (**b**, green) or CGRP (**c**, green) fibres (arrowheads). Representative pictures are shown from at least three separate experiments. Note a decrease of nerve innervation in $Sema3a^{-/-}$ or $Sema3a_{synapsin}^{-/-}$ mice (**a, c**). Scale bars, 100 µm.

To address formally the causal relationship between abnormal sensory development and low bone mass, we intercrossed Sema3A-floxed mice with nestin-CreERT mice[25] expressing a tamoxifen-inducible Cre recombinase under the control of a nestin promoter, and injected tamoxifen after birth to bypass the effect of Sema3A on sensory nervous system development ($Sema3a_{nestin}^{ERT-/-}$ mice). Tamoxifen treatment did not significantly affect nerve innervation or bone mass accrual in $Sema3a_{nestin}^{ERT-/-}$ mice, which indicated that Sema3A expression during development before birth is essential for nerve innervation (Supplementary Fig. 17)—a prerequisite for normal bone mass accrual. Second, wild-type mice treated with capsaicin lost bone mass and showed decreased sensory innervation[26] (Supplementary Fig. 18). However, $Sema3a_{nestin}^{-/-}$ mice, which had diminished bone mass and sensory innervation in the absence of capsaicin treatment (Supplementary Figs 7b–d, f–h, 9, 11a), did not exhibit a further loss in bone mass with capsaicin treatment (Supplementary Fig. 18). Collectively, these results suggested that abnormal development of the sensory nervous system or disruption of the capsaicin-sensitive TRPV1-positive sensory nerves led to low bone mass. Moreover, $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice also showed reduced bone regeneration and defective nerve innervation into newly synthesized bone after femoral bone marrow ablation (Supplementary Fig. 19), indicating that Sema3A also participated in bone modelling in adult mice via the modulation of nerve innervation.

To investigate the role of the Sema3A signalling pathway in nerve innervation in bone, we analysed $Plxna4^{-/-}$ mice[27]. $Plxna4^{-/-}$ mice developed a low bone mass phenotype and showed decreased nerve innervation in bone (Supplementary Fig. 20), reminiscent of the $Sema3a_{synapsin}^{-/-}$ and $Sema3a_{nestin}^{-/-}$ mice phenotypes. These results suggested that the Sema3A–Plxna4 pathway is important for nerve innervation in bone.

We next determined whether other neuronal abnormalities in $Sema3a^{-/-}$ mice were also observed in $Sema3a_{synapsin}^{-/-}$ mice. The abnormal development of the olfactory system and spinal cord[6], reduced thickness of the cerebral cortex[6], and altered pattern of sympathetic innervation in the heart[28] that are characteristics of $Sema3a^{-/-}$ mice were identified in $Sema3a_{synapsin}^{-/-}$ mice (Supplementary Fig. 21). These results suggested that, *in vivo*, neuron-derived Sema3A contributes to normal nervous system development; that is, Sema3A works as an autocrine factor for neuronal development.

We demonstrate that neuron-derived Sema3A regulates bone remodelling through the modulation of sensory nerve projections into bone. A potential direct action of Sema3A on osteoblasts has been suggested recently using global Sema3A knockout mice and a supraphysiological dose of Sema3A (ref. 13). However, given the absence of bone abnormalities in $Sema3a_{col1}^{-/-}$ and $Sema3a_{osx}^{-/-}$ mice, Sema3A must physiologically regulate bone metabolism primarily through the modulation of sensory nerve innervation, *in vivo*, although this does not rule out the possibility that osteoblast-derived Sema3A has a role in bone metabolism[13].

Although Sema3A has strong inhibitory effects on axonal elongation *in vitro* when added exogenously[23], we demonstrated that Sema3A, which is produced in neurons, is indispensable in the normal development of peripheral neurons *in vivo*. The co-expression of Npn1 and Sema3A in motor neurons regulates axon sensitivity to environmental Sema3A sources during motor axon pathfinding[29]. Given that Sema3A is expressed in DRG sensory nerves (Supplementary Fig. 22) and that $Sema3a_{synapsin}^{-/-}$ mice exhibit phenotypic similarities with $Sema3a^{-/-}$ mice (Supplementary Fig. 21), Sema3A expression in neurons may be a fundamental component of proper neural development.

This is, to our knowledge, the first study to demonstrate that sensory nerves have an essential role in bone remodelling. The sympathetic nervous system is known to inhibit bone mass accrual, yet a 'sensing' system is required to achieve bone homeostasis. Therefore, an intriguing question is whether any interaction exists between 'osteo-anabolic' afferent sensory nerves and 'osteo-catabolic' efferent sympathetic nerves.

Clinically, familial dysautonomia patients, who are characterized by the loss of unmyelinated axons including sensory nerves, suffer from osteoporosis[30], supporting the importance of sensory nerves in bone homeostasis. Further studies are necessary to determine whether other bone degenerative diseases, including osteoporosis, involve a loss of sensory nervous system integrity.

METHODS SUMMARY

To delete Sema3A in osteoblasts and neurons, we mated Sema3A-floxed mice[7] with α1(I)-Cre[17] mice or osx-Cre[18] mice, or synapsin-I-Cre[20] mice or nestin-Cre[21] mice. Histomorphometry and microcomputed tomography analysis were performed as previously described. Primary osteoblasts and osteoclasts were cultured as previously described. Fluorescent immunohistochemistry was performed as previously described. Detailed experimental information is described in Methods.

Full Methods and any associated references are available in the online version of the paper.

1. Kruger, R. P., Aurandt, J. & Guan, K. L. Semaphorins command cells to move. *Nature Rev. Mol. Cell Biol.* **6,** 789–800 (2005).
2. Tamagnone, L. & Comoglio, P. M. To move or not to move? Semaphorin signalling in cell migration. *EMBO Rep.* **5,** 356–361 (2004).
3. Pasterkamp, R. J. & Giger, R. J. Semaphorin function in neural plasticity and disease. *Curr. Opin. Neurobiol.* **19,** 263–274 (2009).
4. Tran, T. S., Kolodkin, A. L. & Bharadwaj, R. Semaphorin regulation of cellular morphology. *Annu. Rev. Cell Dev. Biol.* **23,** 263–292 (2007).
5. Roth, L. *et al.* The many faces of semaphorins: from development to pathology. *Cell. Mol. Life Sci.* **66,** 649–666 (2009).
6. Behar, O., Golden, J. A., Mashimo, H., Schoen, F. J. & Fishman, M. C. Semaphorin III is needed for normal patterning and growth of nerves, bones and heart. *Nature* **383,** 525–528 (1996).
7. Taniguchi, M. *et al.* Disruption of semaphorin III/D gene causes severe abnormality in peripheral nerve projection. *Neuron* **19,** 519–530 (1997).
8. Nakamura, F. *et al.* Increased proximal bifurcation of CA1 pyramidal apical dendrites in *sema3A* mutant mice. *J. Comp. Neurol.* **516,** 360–375 (2009).
9. Tian, L., Rauvala, H. & Gahmberg, C. G. Neuronal regulation of immune responses in the central nervous system. *Trends Immunol.* **30,** 91–99 (2009).
10. Potiron, V., Nasarre, P., Roche, J., Healy, C. & Boumsell, L. Semaphorin signaling in the immune system. *Adv. Exp. Med. Biol.* **600,** 132–144 (2007).
11. Adams, R. H. & Eichmann, A. Axon guidance molecules in vascular patterning. *Cold Spring Harb. Perspect. Biol.* **2,** a001875 (2010).
12. Neufeld, G. & Kessler, O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nature Rev. Cancer* **8,** 632–645 (2008).
13. Hayashi, M. *et al.* Osteoprotection by semaphorin 3A. *Nature* **485,** 69–74 (2012).
14. Negishi-Koga, T. *et al.* Suppression of bone formation by osteoclastic expression of semaphorin 4D. *Nature Med.* **17,** 1473–1480 (2011).
15. Takegahara, N. *et al.* Plexin-A1 and its interaction with DAP12 in immune responses and bone homeostasis. *Nature Cell Biol.* **8,** 615–622 (2006).
16. Hall, A. & Lalli, G. Rho and Ras GTPases in axon growth, guidance, and branching. *Cold Spring Harb. Perspect. Biol.* **2,** a001818 (2010).
17. Dacquin, R., Starbuck, M., Schinke, T. & Karsenty, G. Mouse α1(I)-collagen promoter is the best known promoter to drive efficient Cre recombinase expression in osteoblast. *Dev. Dyn.* **224,** 245–251 (2002).
18. Rodda, S. J. & McMahon, A. P. Distinct roles for Hedgehog and canonical Wnt signaling in specification, differentiation and maintenance of osteoblast progenitors. *Development* **133,** 3231–3244 (2006).
19. Takeda, S. Central control of bone remodelling. *J. Neuroendocrinol.* **20,** 802–807 (2008).
20. Zhu, Y. *et al.* Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. *Genes Dev.* **15,** 859–876 (2001).
21. Okada, S. *et al.* Conditional ablation of Stat3 or Socs3 discloses a dual role for reactive astrocytes after spinal cord injury. *Nature Med.* **12,** 829–834 (2006).
22. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466,** 829–834 (2010).
23. Mach, D. B. *et al.* Origins of skeletal pain: sensory and sympathetic innervation of the mouse femur. *Neuroscience* **113,** 155–166 (2002).
24. Shibata, S. *et al.* *Sox10*-Venus mice: a new tool for real-time labeling of neural crest lineage cells and oligodendrocytes. *Mol. Brain* **3,** 31 (2010).
25. Imayoshi, I., Ohtsuka, T., Metzger, D., Chambon, P. & Kageyama, R. Temporal regulation of Cre recombinase activity in neural stem cells. *Genesis* **44,** 233–238 (2006).
26. Offley, S. C. *et al.* Capsaicin-sensitive sensory neurons contribute to the maintenance of trabecular bone integrity. *J. Bone Miner. Res.* **20,** 257–267 (2005).
27. Suto, F. *et al.* Plexin-A4 mediates axon-repulsive activities of both secreted and transmembrane semaphorins and plays roles in nerve fiber guidance. *J. Neurosci.* **25,** 3628–3637 (2005).
28. Ieda, M. *et al.* Sema3a maintains normal heart rhythm through sympathetic innervation patterning. *Nature Med.* **13,** 604–612 (2007).
29. Moret, F., Renaudot, C., Bozon, M. & Castellani, V. Semaphorin and neuropilin co-expression in motoneurons sets axon sensitivity to environmental semaphorin sources during motor axon pathfinding. *Development* **134,** 4491–4501 (2007).
30. Maayan, C., Bar-On, E., Foldes, A. J., Gesundheit, B. & Pollak, R. D. Bone mineral density and metabolism in familial dysautonomia. *Osteoporos. Int.* **13,** 429–433 (2002).

Supplementary Information is available in the online version of the paper.

Author Contributions T.F. conducted most of the experiments. R.X., H. Ochi, A.K., Z.G., Y.Y., C.M., C.X., T.H., Y.A. and M.E. conducted mice analyses. T.S., K.F., W.B. and S. Sunamura conducted *in vitro* experiments. S. Shibata and H. Okano generated mutant mice. A.O., H.I. and K.S. discussed the project. S.T. wrote most of the manuscript. S.T. designed and supervised the project.

METHODS

Animals. Sema3A-knockout mice (RBRC01104) and floxed Sema3A mice (RBRC01106) were provided by RIKEN BRC. Transgenic mice expressing Cre recombinase under the control of the α1(I)-collagen promoter (col1-Cre)[17], osterix promoter (osx-Cre)[18], synapsin I promoter (synapsin-I-Cre)[20] and nestin promoter (nestin-Cre)[21], were mated with floxed mice to obtain conditional knockout mice. nestin-CreERT mice were described previously[25]. For neonatal activation of CreERT, 83.5 mg kg^{-1} tamoxifen (Sigma) in corn oil (Sigma) was intraperitoneally injected into lactating mothers once a day for five consecutive days from the day of delivery. For histological experiments, corn-oil-treated mice were used as controls. At 8 weeks, mice were subjected to histological experiments. For capsaicin treatment, mice were deprived of drinking water for 6 h before capsaicin injection to prevent pulmonary oedema. Capsaicin (Sigma) and vehicle (10:10:80 (v/v) of Tween-80, ethanol, saline) were freshly prepared each time, capsaicin was sonicated in vehicle until homogenously suspended, and mice were injected subcutaneously above the dorsal spine. Four-week-old mice received two rounds of capsaicin or vehicle treatment. In each round, mice received daily injections for three consecutive days (10 mg kg^{-1} on days 1 and 2, and 15 mg kg^{-1} on day 3), and each round was seven days apart. At 8 weeks, mice were subjected to histological experiments. To obtain $Sema3a_{nestin}^{-/-}$;$Sox10$-Venus transgenic mice, $Sema3a_{nestin}^{f/+}$;$Sox10$-Venus mice[24] were mated with $Sema3a^{f/f}$ mice. $Plxna4^{-/-}$ mice were previously described[28].

We maintained all the mice under a 12 h light–dark cycle with *ad libitum* access to regular food and water, unless specified. All animal experiments were performed with the approval of the Animal Study Committee of Keio university school of medicine and conformed to relevant guidelines and laws.

Cell culture. *In vitro* primary osteoblast cultures were established as previously described[31,32]. Briefly, calvarial osteoblast cells were isolated from 4-day-old mice by enzymatic digestion in α-minimal essential medium (α-MEM) with 0.5 mg ml^{-1} collagenase-P (Roche) and 0.05% trypsin. For osteoblastic differentiation, cells (3×10^4 cells per cm^2) were cultured in osteogenic medium (0.1 mg ml^{-1} ascorbic acid, 10 mM β-glycerophosphate). Medium was changed every 2 days. After 7 days, osteoblastic differentiation was confirmed by the measurement of ALP activity and expression of osteoblastic marker genes. Results are representative of more than four individual experiments. The murine preosteoblast cell line MC3T3-E1 was cultured in α-MEM. Cells were seeded (3×10^4 cells per cm^2) and treated with osteogenic medium with or without recombinant mouse Sema3A Fc chimaera (R&D systems). Medium was changed every 2 days. After 7 days, ALP measurement was performed. Results are representative of more than four individual experiments. *In vitro* osteoclast differentiation was accomplished as previously described[33]. Briefly, bone marrow cells (3×10^5 cells per cm^2) of 6- to 8-week-old mouse femurs were cultured in α-MEM supplemented with FBS in the presence of human macrophage colony-stimulating factor (M-CSF, 10 ng ml^{-1}; R&D Systems) for 2 days and then differentiated into osteoclasts using human RANKL (50 ng ml^{-1}; Peprotech) and M-CSF for 3 days. Subsequently, the differentiation of osteoclasts was evaluated by TRAP staining. Results are representative of more than four individual experiments. The proliferation assay for osteoblasts was performed using the Cell Counting kit-8 (DOJINDO), according to the manufacturer's instructions. Results are representative of more than four individual experiments.

Transfection and infection. For the Sema3A knockdown study, MC3T3-E1 cells were seeded (3×10^4 cells per cm^2) and transfected with 20 nM of short interfering (si)Sema3A (Invtrogen) or siControl (Invtrogen) using HiPerFect (Qiagen) according to the manufacturer's instructions. siRNA sequences were 5′-AAUAGUUGUUG GUUCCCGGAAGACG-3′ and 5′-CGUCUUCCGGGAACCAACAACUAUU-3′ (for *Sema3a*). After transfection, cells were cultured with osteoblastic differentiation media. Culture medium was changed every 2 days. Recombinant mouse Sema3A Fc chimaera (R&D systems) was added every 2 days at indicated concentrations. After 7 days, osteoblastic differentiation was confirmed by the measurement of ALP activity. Results are representative of more than four individual experiments. To confirm Sema3A intracellular signalling, MC3T3-E1 cells (3×10^4 cells per cm^2) were infected with adenovirus containing a dominant-negative form of Rac1 (T17N) or enhanced green fluorescent protein (eGFP). Infected cell were cultured in osteogenic medium with or without Sema3A (2 μg ml^{-1}). Medium was changed every 2 days. After 7 days, ALP measurement was performed. Results are representative of more than four individual experiments.

Histological and histomorphometric analyses. We injected mice with calcein (25 mg kg^{-1}; Sigma) and stained undecalcified sections of the lumbar vertebrae using von Kossa and TRAP as previously described[34]. We performed static and dynamic histomorphometric analyses using the Osteomeasure Analysis System (Osteometrics) following nomenclature defined by the American Society for Bone and Mineral research as previously described[34]. Bone volume/tissue volume (BV/TV; %), bone formation rate/bone surface (BFR/BS; μm^3 μm^{-1} yr^{-1}), mineral apposition rate (MAR; mm yr^{-1}) osteoblast surface/bone surface (Ob. S/BS; %) and osteoclast number/bone perimeter (No. Oc./B. Pm) were analysed. Oc. S/BS and No. Oc./B. Pm were calculated on TRAP staining slices.

Micro-CT analyses. We obtained two-dimensional images of the distal femurs by micro-CT analysis (μCT; Comscan).

Measurement of deoxypyridinoline cross-links. We measured serum deoxypyridinoline cross-links (DPD) with the MicroVue tDPD kit (QUIDEL) according to the manufacturer's instructions.

Quantitative real-time PCR analysis. Total RNA from tissues and cultured cells was extracted using TRizol reagent (Invitrogen) and reverse transcription was performed by ReverTra Ace qPCR RT Kit (TOYOBO) according to the manufacturer's instructions. We performed quantitative analysis of gene expression using the Mx3000P real-time PCR system (Agilent Technologies). We used *Gapdh* expression as an internal control. The following primers were used: *Sema3a* sense, 5′-TGGGC TGGTTCACTGGGATTGC-3′ and *Sema3a* antisense, 5′-CTGGAGCTGTTGG CCAAGCCAT-3′; *Alpl* sense, 5′-ACACCTTGACTGTGGTTACTGCTGA-3′ and *Alpl* antisense, 5′-CCTTGTAGCCAGGCCCGTTA-3′; *Bglap* sense, 5′-TCTGAC AAAGCCTTCATGTCCA-3′ and *Bglap* antisense, 5′-CGGTCTTCAAGCCATA CTGGTC-3′; *Dmp1* sense, 5′-CCCAGAGGGACAGGCAAATA-3′ and *Dmp1* antisense, 5′-TCCTCCCCACTGTCCTTCTT-3′; *Plxna1* sense, 5′-GAGTGCAA GGAAGCTTTTGC-3′ and *Plxna1* antisense, 5′-TCCTCAATCCCAGGAAACAG-3′; *Plxna2* sense, 5′-GGGGGCTATCAATCGTGTCT-3′ and *Plxna2* antisense, 5′-AGGG CTGTACAATGAGGGGT-3′; *Plxna3* sense, 5′-GCCAACCCACCTCGGCAGAC-3′ and *Plxna3* antisense, 5′-GCAGAGGGGTGTTGATGCAGGG-3′; *Plxna4* sense, 5′-A TCTAGAGTGGCGACAAGGAAG-3′ and *Plxna4* antisense, 5′-TGGAGACAGTGG AGTTGTTCAC-3′; *Nrp1* sense, 5′-GAAGCACCGAGAAAACAAGG-3′ and *Nrp1* antisense, 5′-TTGCCTTCGAACGACTTAGC-3′; *Gapdh* sense, 5′-ACCCAGAAGAC TGTGGATGG-3′ and *Gapdh* antisense, 5′-CACATTGGGGGTAGGAACAC-3′.

Immunohistochemistry. Bones were directly embedded into OCT after dissection. Brains and spinal cords were fixed in 4% paraformaldehyde (PFA) in 0.1 M phosphate buffered saline (PBS) for 24 h and dehydrated in 30% sucrose in 0.1 M PBS for 48 h. Brain, lumbar spinal cord and olfactory bulbs were dissected and embedded in 1.5% low-melting-temperature agarose (BM Equipment) in PBS. Brains were cut into 25-μm-thick coronal sections using a cryostat and slices from every 150 μm were used. Lumbar spinal cords were cut into 25-μm-thick axial sections. To characterize the expression of neuronal markers in the brain, series of sections from each experimental group were stained with neuronal markers. After treatment with 0.2% Triton X-100 for 5 min, sections were blocked with 5% NGS and incubated with monoclonal antibodies, SMI31 (a marker for neurofilament M+H, 1:500, COVANCE), polyclonal antibodies, CGRP (1:800, BioMol), TRPV1 (1:300, Transgenic), tyrosine hydroxylase (TH) (1:500 Chemicon), DBH (1:500 Chemicon) for 24 h as previously described[35]. Primary antibodies were visualized with goat anti-mouse IgG Alexa-594 (1:500, Molecular Probes) and goat anti-rabbit IgG Alexa-488 (1:400, Molecular Probes).

For quantification of innervations of nerve fibres, the percentage of neurofilament-positive (red) area to the area corresponding to the secondary spongiosa of the right proximal tibial metaphyses was calculated by Image J software (http://rsbweb.nih.gov/ij/).

Retrograde labelling of tibia afferents. To assess tibia afferents, retrograde tracing with Fluoro-Gold (FG: Fluorochrome) was performed in wild-type and $Sema3a_{nestin}^{-/-}$ mice. A total of 2 μl of 4% FG was injected into the right tibial tuberosity with a Hamilton syringe. The DRGs were transcardially perfused with a solution of 4% PFA in PBS 1 week after the FG injection and carefully dissected. The L3 and L4 DRGs were cut into 10-μm-thick transverse sections perpendicular to the spinal nerve using a cryostat. The sections were incubated with NeuroTrace green fluorescent Nissl stain (Invitrogen). The numbers of FG-positive neurons were counted in the section of middle portion of DRG.

Sensory behavioural tests. The mice were placed in plexiglass boxes, which were $9.5 \times 21 \times 25$ cm in size, to become acclimated to the testing environment. These boxes were then placed on an elevated perforated plastic surface for a minimum of 30 min before all behavioural tests. A blind observer conducted the behavioural testing.

In the tactile threshold test, mechanical sensitivity was measured by applying a series of calibrated von Frey filaments (0.02–8 g) to the plantar aspect of the hind paw. Each filament was applied once to each mouse. Beginning with the 1 g filament, each filament was applied perpendicular to the hind paw for 4–6 s. A brisk withdrawal of the hind paw indicated a positive response, and a lack of withdrawal indicated a negative response. This filament testing is repeated a maximum of two additional times, and a 2/3 response to the filament indicated an overall positive response. If the mouse demonstrated an overall positive response, the next lower force filament was applied as described above. In case an overall positive response was not observed (0/3 or 1/3 responses), the next greater force filament was applied as described above. Once the crossover threshold could be determined (that is,

from response to no response, or vice versa), the responses to the next five filaments were recorded to determine the median withdrawal threshold.

Response to acetone was tested using a plastic tube connected to a 1 ml syringe and 100 μl of acetone was applied to the plantar surface of the foot without touching the skin. Acetone was applied five times to each paw at an interval of at least 30 s, and the number of brisk foot withdrawals in response to the acetone application was recorded.

Response to noxious heat stimulus was tested in the following manner. A response latency to a noxious heat source was measured with a hot plate apparatus (NISSIN, Japan). Mice were placed on a surface heated to 50 or 55 °C, and the amount of time between placement on the apparatus and a hindpaw lick or jump was recorded. To avoid tissue damage due to prolonged exposure to the heated surface, we used a cut-off time of 45 s.

Bone marrow ablation. Mice were anaesthetized intraperitoneally with 400 mg kg^{-1} trichloroacetaldehyde monohydrate (Chloral Hydrate; Wako Pure Chemical). After removal of the hair from both hind limbs, the bone marrow of both femora on each animal was ablated as described previously[36]. Briefly, bilateral longitudinal incisions were made on the knees of each mouse to expose the femoral condyle by dislocation of the patella. A hole was made at the intercondylar notch of the femur using a dental drill. A 0.6 mm diameter Kirschner wire was inserted in the proximal end of the femur to confirm completion of marrow ablation by radiography. After removal of the Kirschner wire, the dislocated patella was reposed and the skin was sutured. Mice were killed at day 7 after ablation and bone histological analyses were performed.

***In situ* hybridization.** Spinal cords were fixed in 4% PFA in PBS and dehydrated in 30% sucrose in PBS. Frozen sections (20 μm) were hybridized with digoxigenin (DIG)-labelled antisense riboprobes corresponding to a partial cDNA of mouse *Sema3a* (3,852–4,378 bp) at 60 °C overnight. For detection, signals were developed using anti-digoxigenin antibody conjugated with alkaline phosphatase. After antibody treatment, sections were incubated with NBT/BCIP (Roche).

Statistical analysis. We performed statistical analysis using Tukey–Kramer and Student's *t*-test. Values were considered statistically significant at $P < 0.05$. All data are represented as mean ± s.d. Results are representative of more than four individual experiments.

31. Kimura, A. *et al.* Runx1 and Runx2 cooperate during sternal morphogenesis. *Development* **137,** 1159–1167 (2010).
32. Inose, H. *et al.* A microRNA regulatory mechanism of osteoblast differentiation. *Proc. Natl Acad. Sci. USA* **106,** 20794–20799 (2009).
33. Fujita, K. *et al.* Vitamin E decreases bone mass by stimulating osteoclast fusion. *Nature Med.* **18,** 589–594 (2012).
34. Sato, S. *et al.* Central control of bone remodeling by neuromedin U. *Nature Med.* **13,** 1234–1240 (2007).
35. Kusano, K. *et al.* Enhancement of sciatic nerve regeneration by adenovirus-mediated expression of dominant negative RhoA and Rac1. *Neurosci. Lett.* **492,** 64–69 (2011).
36. Okuda, N. *et al.* ED-71, a novel vitamin D analog, promotes bone formation and angiogenesis and inhibits bone resorption after bone marrow ablation. *Bone* **40,** 281–292 (2007).

LETTER

Immune surveillance by CD8αα+ skin–resident T cells in human herpes virus infection

Jia Zhu[1,2]*, Tao Peng[2,3]*, Christine Johnston[3], Khamsone Phasouk[2], Angela S. Kask[1], Alexis Klock[1], Lei Jin[1], Kurt Diem[1], David M. Koelle[1,2,3,4,5], Anna Wald[1,2,3,6], Harlan Robins[7] & Lawrence Corey[1,2,3,4]

Most herpes simplex virus 2 (HSV-2) reactivations in humans are subclinical and associated with rapid expansion and containment of virus. Previous studies have shown that CD8+ T cells persist in genital skin and mucosa at the dermal–epidermal junction (DEJ)— the portal of neuronal release of reactivating virus—for prolonged time periods after herpes lesions are cleared[1,2]. The phenotype and function of this persistent CD8+ T-cell population remain unknown. Here, using cell-type-specific laser capture microdissection, transcriptional profiling and T-cell antigen receptor β-chain (TCRβ) genotyping on sequential genital skin biopsies, we show that CD8αα+ T cells are the dominant resident population of DEJ CD8+ T cells that persist at the site of previous HSV-2 reactivation. CD8αα+ T cells located at the DEJ lack chemokine-receptor expression required for lymphocyte egress and recirculation, express gene signatures of T-cell activation and antiviral activity, and produce cytolytic granules during clinical and virological quiescent time periods. Sequencing of the TCR β-chain repertoire reveals that the DEJ CD8αα+ T cells are oligoclonal with diverse usage of TCR variable-β genes, which differ from those commonly described for mucosa-associated invariant T cells and natural killer T cells. Dominant clonotypes are shown to overlap among multiple recurrences over a period of two-and-a-half years. Episodes of rapid asymptomatic HSV-2 containment were also associated with a high CD8 effector-to-target ratio and focal enrichment of CD8αα+ T cells. These studies indicate that DEJ CD8αα+ T cells are tissue-resident cells that seem to have a fundamental role in immune surveillance and in initial containment of HSV-2 reactivation in human peripheral tissue. Elicitation of CD8αα+ T cells may be a critical component for developing effective vaccines against skin and mucosal infections.

Clinical studies have shown that 50–80% of HSV reactivations are subclinical and of short duration (<6 h)[3–5], and that CD8+ T cells not only infiltrate selectively to the site of viral reactivation, but also persist locally at the DEJ for months after resolution of herpes lesions[1,2]. To define the phenotype, function and diversity of these DEJ CD8+ T cells during virological quiescence as well as early containment of reactivation, we used cell-type-specific laser capture microdissection (LCM) to identify and select individual CD8+ T cells from the DEJ (DEJ CD8), dermis near to blood vessels (BV CD8), HSV-2-affected area, and contralateral HSV-2-unaffected genital tissue (control CD8) (Supplementary Fig. 1a, b) to examine the persistent nature, antiviral signature and TCR repertoire of these cells from sequential skin biopsies. The purity of captured CD8 cells was validated by the abundance of *CD8A* and the absence of *CD4* gene expression (Supplementary Fig. 1c). Detailed immunofluorescent staining demonstrated that DEJ CD8+ T cells expressed CD3ε and TCRβ, but not TCRγδ (Supplementary Fig. 2). These cells did not express co-receptors and markers for dendritic cells or natural killer (NK) cells or non-conventional T cells such as NKT cells

and mucosa-associated invariant T (MAIT) cells (Supplementary Fig. 3). Thus, the CD8+ T cells that persist at the DEJ are CD8+ TCRαβ T cells.

Because tissue-based CD8+ T cells and memory precursors have been shown to express varied levels of CD8α and CD8β[6,7], we measured gene expression of these co-receptors by quantitative TaqMan PCR. DEJ CD8 cells almost exclusively expressed *CD8A* messenger RNA. *CD8B* was detected in only one out of eight and three out of ten subjects biopsied at 2 and 8 weeks post-healing (w.p.h.) from the HSV-2-affected area, respectively. By contrast, BV CD8 cells from the same 8 w.p.h. biopsies expressed high levels of *CD8B* transcripts in nine out of ten patients (Fig. 1a). Cell-surface expression of CD8α and CD8β chains circumstantiated these transcriptional patterns (Fig. 1b, c). CD8α+β− cells predominated at the DEJ from the time of active lesion to 2 and 8 w.p.h., ranging from 85 to 91% of the total DEJ CD8α+ cells, and the number of these cells remained relatively stable. By contrast, CD8α+β+ cells localized mainly to the dermis during HSV-2 ulceration and rapidly
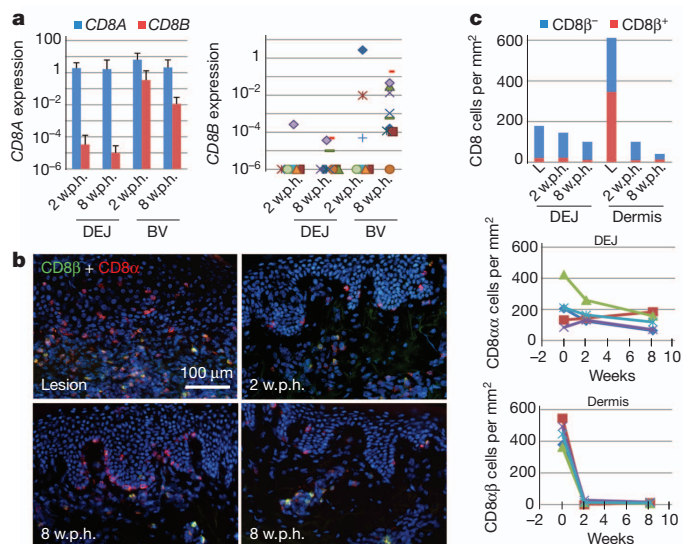

Figure 1 | CD8αα+, but not CD8α+β+, T cells persist at the DEJ in human HSV-2 infection. a, Average expression of *CD8A* and *CD8B* genes (left) and individual expression of *CD8B* gene (right) in DEJ CD8 and BV CD8 cells laser captured from tissues of 2 (n = 8) and 8 (n = 10) w.p.h. biopsies, relative to *ACTB*. Each symbol represents one subject. Error bars, s.d. **b**, Differential localization of CD8αα+ (red) to the DEJ and CD8α+β+ (yellow) to the dermis in active lesions and 2 w.p.h. and 8 w.p.h. tissue (high- and low-density area, bottom left and bottom right, respectively). **c**, Top, quantification of CD8αα+ and CD8α+β+ T cells at the DEJ and dermis over time. Middle, individual and average (light blue) kinetics of DEJ CD8αα+ T cells. Bottom, dermal CD8α+β+ T cells over time (n = 4). L denotes lesions (week 0).

[1]Department of Laboratory Medicine, University of Washington, Seattle, Washington 98195, USA. [2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [3]Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [4]Department of Pathobiology, University of Washington, Seattle, Washington 98195, USA. [5]Benaroya Research Institute, Seattle, Washington 98104, USA. [6]Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA. [7]Program in Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA.
*These authors contributed equally to this work.

diminished after completely healing, resulted in a 98% reduction in cell density during the first 2 w.p.h. at a rate (169 cells per mm^2 per week) almost ten times higher than that of CD8$\alpha^+\beta^-$ cells at the DEJ (17.5 cells per mm^2 per week) for the same period. When detected in either 2 or 8 w.p.h. biopsy tissue, CD8$\alpha^+\beta^+$ cells were mainly located near the blood vessels. These data indicate, at both the RNA and protein level, that CD8$^+$ T cells that selectively persist and preferentially localize at the DEJ have a CD8$\alpha\alpha$ homodimer phenotype.

To investigate the potential mechanism for persistence of CD8$\alpha\alpha^+$ T cells at the DEJ, we compared the transcriptional profiles of DEJ CD8 versus BV CD8 cells isolated from the same tissue biopsy obtained at 2 and 8 w.p.h. Gene expression of chemokine receptors and lymphocyte-trafficking receptors was different between DEJ CD8 and BV CD8 cells. We observed a downregulation of *CCR7*, *CCR8*, *CXCR4*, *CXCR6*, *CXCR7* and *S1PR1* transcripts in DEJ CD8 compared to BV CD8 cells (Fig. 2a). Using TaqMan PCR, we found no detectable *CCR7* gene expression in DEJ CD8 cells from 8 w.p.h. tissue in nine out of ten subjects; an expression pattern also observed for *S1PR1*, which was undetectable in eight out of ten subjects (Fig. 2b). Comparatively, high levels of *CCR7* and *S1PR1* expression were detected in BV CD8 cells. It has been

shown that CCR7 is required for guiding lymphocytes exiting peripheral tissue and S1PR1 functions in lymphocyte recirculation and trafficking in the blood[8–11]. The absence of transcriptional activity of both of these genes in DEJ CD8 cells suggests that a potential mechanism for CD8$\alpha\alpha^+$ T-cell persistence at the DEJ is a lack of chemotactic requirements for exit.

We also compared expression profiles of genes critical for T-cell activation and antiviral and cytolytic function between captured DEJ CD8 and control CD8 cells from 8 w.p.h. biopsies[12], a time of presumptive virological quiescence. Upregulation of genes for TCR co-receptors (*CD8A* and *CD3G*) and co-stimulation (*ICOS* and *ITGAL* (also known as *LFA-1A*)), G1/S cell cycle transition (*CD27* and *CCND3*), cytolytic activity (*PRF1*, *GZMA*, *GZMB*, *GZMK* and *GZMH*) and cytokines/chemokines and their receptors (*IFNG*, *TNFA*, *RANTES* (also known as *CCL5*) and *IL-7RA* (also known as *IL7R*)) suggests that DEJ CD8 cells were actively engaging and eliminating infection (Fig. 2c). We performed quantitative PCR on a subset of these genes to delineate the expression patterns further (Fig. 2d). Expression of *RANTES* and *GZMA* was upregulated in DEJ CD8 cells from all nine subjects. *GZMB* and *IFNG* were induced in seven and six out of nine subjects, respectively, a significantly higher proportion compared to contralateral control samples (one out of nine) ($P < 0.01$, Fisher's exact test). Perforin protein expression was also detected in DEJ CD8 cells at 8 w.p.h. (Fig. 2e); no perforin granules were detected in the contralateral control CD8 cells (data not shown). The strongest perforin granule expression at the 8 w.p.h. time point was seen in DEJ cells from patients who had evidence of subclinical HSV-2 shedding (Fig. 2e). Co-staining of biopsy tissue showed that perforin granules were detected in DEJ CD8$\alpha^+\beta^-$ cells (Fig. 2f upper panels) but not in BV CD8$\alpha^+\beta^+$ T cells (Fig. 2f lower panels). The expression of cytolytic granules and mRNA patterns of T-cell activation and antiviral responses indicate that DEJ CD8$\alpha\alpha^+$ T cells are programmed and regulated to respond to virus infection.

We investigated the role of CD8$\alpha\alpha^+$ T cells in the early containment of HSV-2 reactivation during episodes of subclinical recurrence. HSV-2 DNA was detected in tissue sections in two of the ten 8 w.p.h. biopsies.
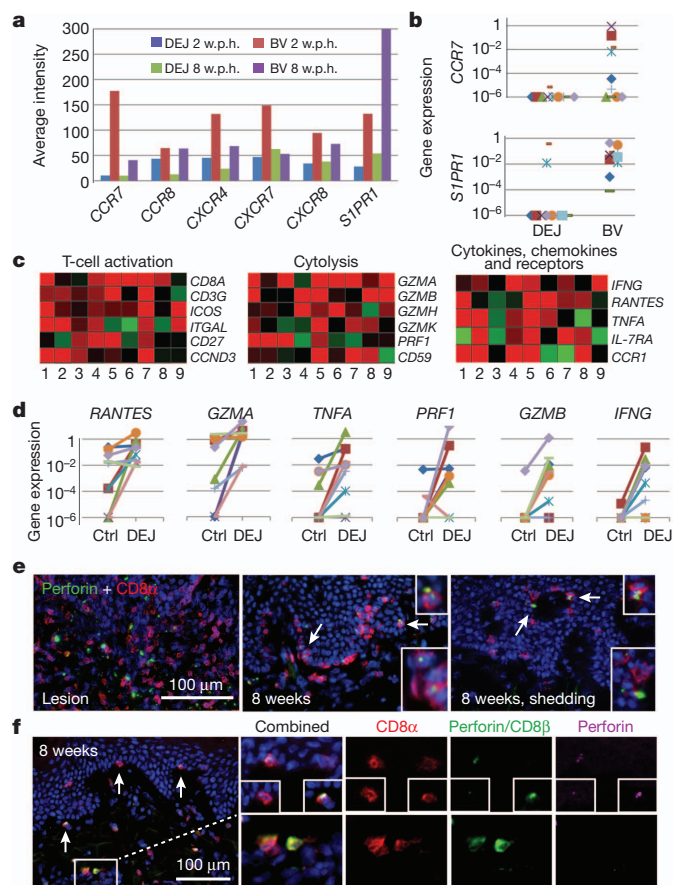

Figure 2 | Tissue residence and effector function of DEJ CD8$\alpha\alpha^+$ T cells in post-healed tissue. a, Downregulation of chemokine-receptor expression in DEJ CD8 compared with BV CD8 cells. Data are average signal intensities on the Illumina BeadArray ($n = 5$). **b**, *CCR7* (top) and *S1PR1* (bottom) expression by quantitative PCR. Each symbol represents one subject. **c**, Activated effector signatures of DEJ CD8 cells. Pair-wised comparison was performed between DEJ CD8 and control CD8 cells from the same individual at the 8 w.p.h. time point ($n = 9$). Red, upregulated; green, downregulated. **d**, Quantitative PCR analysis of antiviral gene expression ($n = 9$). Ctrl, control. Each symbol represents one subject. **e**, Perforin protein expression in DEJ CD8 cells of active lesion (left) and 8 w.p.h. with (right) and without (middle) virus shedding. **f**, Perforin protein expression in DEJ CD8$\alpha\alpha^+$ T cells but not BV CD8$\alpha^+\beta^+$ T cells. Perforin, yellow granules (both pink and green); CD8α, red; CD8β, green. Arrows in **e** and **f** denote enlarged area of perforin expression.
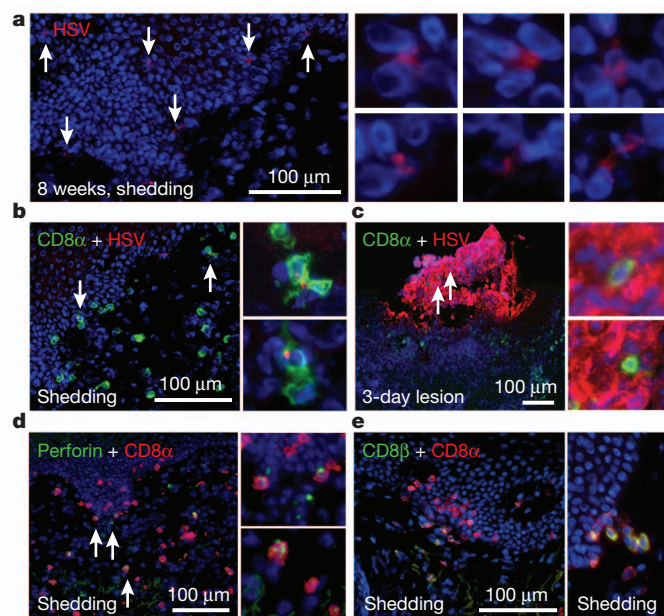

Figure 3 | DEJ CD8 cells in early containment during asymptomatic HSV-2 reactivation. a, HSV-2 antigen expression in keratinocytes during asymptomatic HSV-2 reactivation. Arrows denote enlarged area of infected cells showing on the right panels. **b, c**, CD8 effector-to-target ratio during asymptomatic (**b**) and lesional (**c**) HSV-2 reactivation. **d**, High level of CD8α^+ T cells expressing perforin granules during asymptomatic shedding. Arrows denote enlarged areas on the right. **e**, Focal accumulation of CD8$\alpha\alpha^+$ T cells (left, red) and recruitment of CD8$\alpha^+\beta^+$ T cells (right, yellow) to the DEJ.

DNA levels were significantly lower than those during a genital lesion (<20 copies per 5×10^4 cells versus 10^6 copies per 5×10^4 cells, respectively) and the biopsies concurrently showed no histologic evidence of ulceration or epithelial cell necrosis, indicating an episode of spontaneous asymptomatic shedding controlled by the host immune system[1,13,14]. Immunofluorescence staining revealed a punctate distribution of HSV-2 antigen in several keratinocytes dispersed in the epidermis (Fig. 3a). CD8[+] T cells and HSV-2-infected cells formed clusters in which two to four CD8[+] T cells were in direct contact with a virally infected cell (Fig. 3b). This high effector-to-target ratio during asymptomatic shedding is in contrast to that which occurs during clinical episodes of recurrent genital herpes (Fig. 3c). *In situ* quantum dot 655-conjugated peptide–major histocompatibility complex I (Qdot655-pMHCI) multimer staining identified HSV-2-specific CD8[+] T cells present in these shedding biopsies (Supplementary Fig. 4) along with increased numbers of perforin granules (Fig. 3d), providing evidence of antigen-specific early containment of virus infection. Furthermore, high-density CD8$\alpha\alpha^+$ T cells were found in foci at the DEJ and there was increased recruitment of CD8$\alpha^+\beta^+$ T cells to the DEJ (Fig. 3e). Again, we detected perforin granule expression in CD8$\alpha^+\beta^-$ T cells, suggesting these CD8$\alpha\alpha^+$ T cells are cytolytically active during early host antiviral control (Supplementary Fig. 5). These studies suggest that CD8$\alpha\alpha^+$ T cells are important effectors in early immune containment.

In order to understand the persistence, diversity and specificity of CD8$\alpha\alpha^+$ T cells at the DEJ further, we analysed the amino acid sequence in the hypervariable complementarity-determining region 3 (CDR3) of the T-cell receptor β chain (TCRB) in DEJ CD8 cells using high-throughput sequencing assays[15]. We collected DEJ CD8 cells using LCM from six biopsies obtained during a 2.5-year time period from an individual with long-standing recurrent HSV-2 who participated in a series of trials involving sequential skin biopsies during and after HSV-2 reactivation (Fig. 4a). The sequence pattern of DEJ CD8$\alpha\alpha^+$ T cells indicated an oligoclonal immune profile, with the top six most-abundant sequences making up 85 to 98% of the TCRB repertoire in each DEJ CD8 sample (Fig. 4a and Supplementary Table 1). Twelve TCRB CDR3 sequences were common among the six tissue biopsies (Fig. 4a and Supplementary Table 2): two sequences were detected in all six biopsies and one was present in five biopsies over the 2.5-year time period (Fig. 4b). Moreover, both the abundant sequences and the 12 commonly shared sequences (Supplementary Tables 1 and 2) involved diverse usage of variable-β-gene (VB) families; for example, 18 VB genes were found among the 27 unique TCRB amino acid sequences that predominated in DEJ CD8$\alpha\alpha^+$ T cells. Interestingly, evidence of constrained VB gene usage as described in CD8$\alpha\alpha^+$ MAIT cells was not seen[16,17]; Vβ2 and Vβ13 families not found in DEJ CD8 cells.

To examine the antigen specificity of DEJ CD8$\alpha\alpha^+$ T cells, we sequenced TCRB from tissue-infiltrating T cells during lesion healing, peripheral blood mononuclear cell (PBMC)-derived CD8$\alpha\alpha^+$ T cells and HSV-2-reactive CD8[+] T cells isolated directly *ex vivo* from PBMCs from the same individual[18]. Fifteen TCRB CDR3 sequences were shared by all three T-cell populations (Fig. 4c and Supplementary Table 3), and one of these sequences, CASRLDRGTLTDTQYF, was also detected in DEJ CD8 cells from two separate biopsies 21 months apart from distinct HSV-2 recurrences (March 2005, newly healed of episode no. 1; and December 2006, 4 w.p.h. of episode no. 3; Fig. 4a). These results provide evidence that DEJ CD8 cells are comprised of a CD8$\alpha\alpha^+$ T-cell

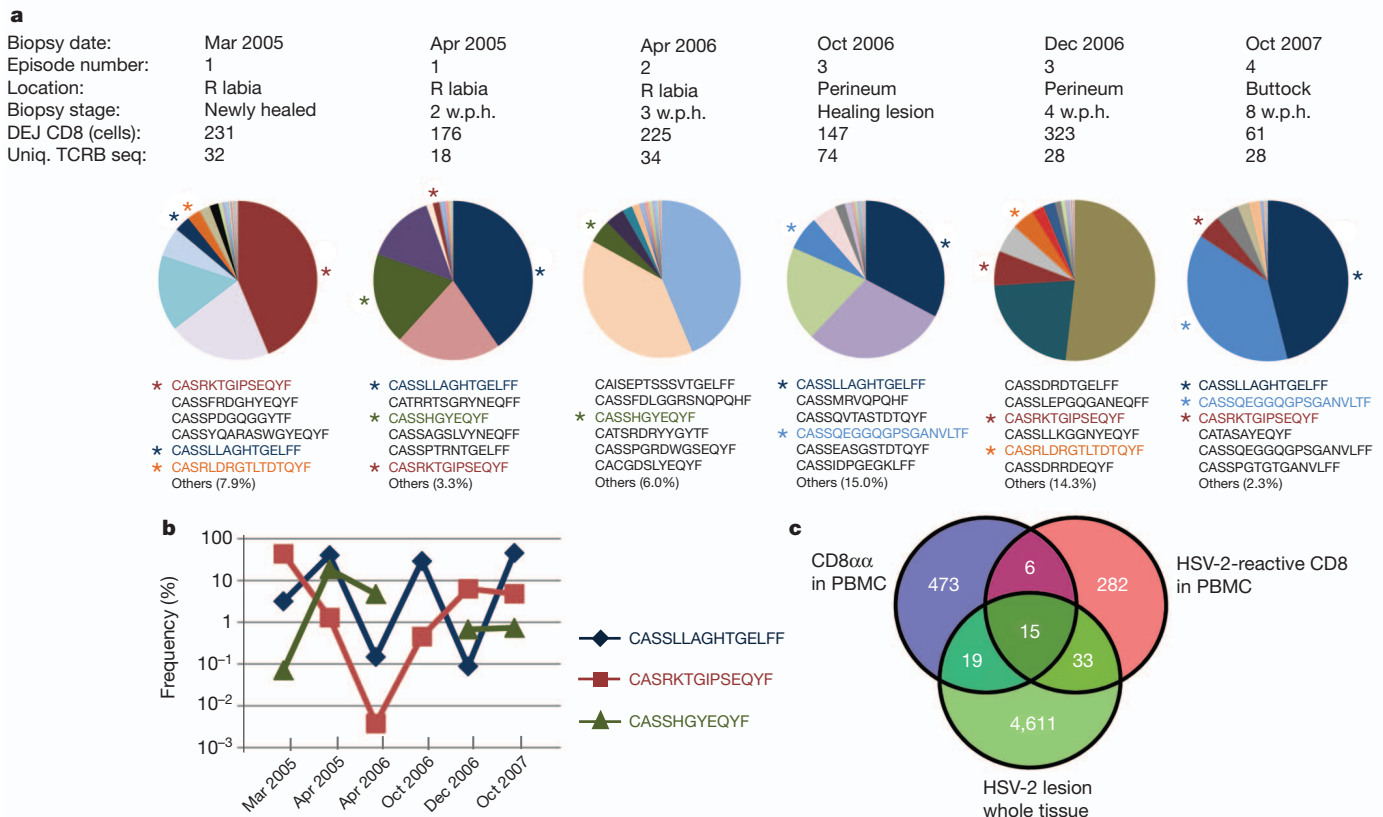**Figure 4 | Characteristics of DEJ CD8 cell TCRβ repertoire. a**, Oligoclonal distribution and dominant CDR3 sequences of the DEJ CD8 cell TCRβ repertoire. DEJ CD8 cells were laser captured from six biopsies obtained during and after multiple episodes of herpes recurrence from one individual. The pie charts illustrate the distribution of the unique CDR3 sequences detected in the DEJ CD8 TCRβ repertoire from each biopsy. Amino acid sequences of the top six clones are listed in the order of abundance. Shared amino acid sequences between biopsies are highlighted in the same colour as in the pie charts. Each colour represents one clone. **b**, Frequencies of the three most commonly shared DEJ CD8 clones. **c**, Venn diagram of overlapping TCRβ repertoires among three T-cell populations from the same individual, showing tissue-infiltrating T cells from a healing lesion biopsy, blood-PBMC-derived HSV-2-reactive CD8[+] T cells and PBMC-derived CD8$\alpha\alpha^+$ T cells. Clone tracking was performed at the nucleotide level.

population that responds to HSV-2 infection in genital tissue. The sequence CASRLDRGTLTDTQYF and the two commonly shared sequences in all six DEJ CD8 samples, CASSLLAGHTGELFF and CASRKTGIPSEQYF, were abundant in the lesion-healing biopsy sample. These sequences were ranked numbers 2, 6 and 12 in frequency among 4,678 unique TCR sequences detected, further supporting the important role of DEJ CD8$\alpha\alpha^+$ T cells in response to HSV-2 antigen in tissue.

Our studies demonstrate the ability to dissect cell-type-specific immune responses *in situ* in humans and provide several new insights into the tissue-resident immunity in genital skin and mucosa. Accumulating evidence indicates that antigen-specific CD8$^+$ T cells retain long-term residence at the peripheral site of infection, exhibiting superior protection over their blood counterparts against re-infection of invading pathogens[1,19–25]. Our human studies have shown that CD8$^+$ T cells persisting at the DEJ site of previous HSV-2 lesions possess a CD8$\alpha\alpha$ homodimer phenotype. The persistence of CD8$\alpha\alpha^+$ T cells contrasts with the rapid depletion of CD8$\alpha^+\beta^+$ T cells after lesion healing. These CD8$\alpha\alpha^+$ T cells lack chemotactic signals, such as CCR7 and S1PR1, distinguishing them from circulating CD8$\alpha^+\beta^+$ T cells.

The detailed characterization of cell-type-specific markers and TCR VB gene diversity indicate the CD8$\alpha\alpha^+$ T cells described in this study are not innate or semi-innate T cells; rather, they are components of the adaptive immune response retained in the barrier tissue as sentinel cells. CD8$\alpha\alpha$ homodimer expression has been associated with high-affinity antiviral effector T cells and proposed to preserve optimal effector cells at the portal of pathogen entry for long-lived mucosal memory[7,26]. Whether DEJ CD8$\alpha\alpha^+$ T cells are directed at a unique class of viral antigens is unknown and is of major interest.

The unique anatomic localization of DEJ CD8$\alpha\alpha^+$ T cells contiguous to the neural gateway of reactivating virus suggested these cells might recognize and contain HSV infection. Although the functional relevance remains to be proven, the activated gene signature, oligoclonal immune profile and HSV-2-specific T-cell response at sites of previous reactivation support the notion that DEJ CD8$\alpha\alpha^+$ T cells might mediate a rapid and effective peripheral immunity to control frequent HSV antigen exposure in human skin and genital tract. The prompt CD8 antiviral response at the site of virus release during asymptomatic HSV reactivation is in sharp contrast to the delayed CD8$^+$ T-cell infiltration during a lesion-forming herpes recurrence, in which CD8$^+$ T-cell infiltration occurs 2 to 3 days after initial infection, after NKT-cell and CD4$^+$ T-cell arrival[27,28]. Enhancement of the quantity and the function of tissue-resident CD8$\alpha\alpha^+$ T cells may potentially be a mechanism for improved immunotherapeutic treatment and for prevention of HSV-2 reactivation in humans. Similarly, the immunological role of these persisting T cells in other chronic infections of the skin and genital tract, such as HIV, should also be explored.

METHODS SUMMARY

All human genital skin biopsies were obtained according to a study protocol and informed consent procedures approved by the University of Washington Institutional Review Board. For a detailed description of all techniques used in this study, please see Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 5 January 2012; accepted 20 March 2013.
Published online 8 May 2013.

1. Zhu, J. *et al.* Virus-specific CD8$^+$ T cells accumulate near sensory nerve endings in genital skin during subclinical HSV-2 reactivation. *J. Exp. Med.* **204**, 595–603 (2007).
2. Zhu, J. *et al.* Persistence of HIV-1 receptor-positive cells after HSV-2 reactivation is a potential mechanism for increased HIV-1 acquisition. *Nature Med.* **15**, 886–892 (2009).
3. Wald, A., Zeh, J., Selke, S., Ashley, R. L. & Corey, L. Virologic characteristics of subclinical and symptomatic genital herpes infections. *N. Engl. J. Med.* **333**, 770–775 (1995).
4. Wald, A. *et al.* Reactivation of genital herpes simplex virus type 2 infection in asymptomatic seropositive persons. *N. Engl. J. Med.* **342**, 844–850 (2000).
5. Mark, K. E. *et al.* Rapidly cleared episodes of herpes simplex virus reactivation in immunocompetent adults. *J. Infect. Dis.* **198**, 1141–1149 (2008).
6. Hayday, A., Theodoridis, E., Ramsburg, E. & Shires, J. Intraepithelial lymphocytes: exploring the Third Way in immunology. *Nature Immunol.* **2**, 997–1003 (2001).
7. Huang, Y. *et al.* Mucosal memory CD8$^+$ T cells are selected in the periphery by an MHC class I molecule. *Nature Immunol.* **12**, 1086–1095 (2011).
8. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
9. Bromley, S. K., Thomas, S. Y. & Luster, A. D. Chemokine receptor CCR7 guides T cell exit from peripheral tissues and entry into afferent lymphatics. *Nature Immunol.* **6**, 895–901 (2005).
10. Debes, G. F. *et al.* Chemokine receptor CCR7 required for T lymphocyte exit from peripheral tissues. *Nature Immunol.* **6**, 889–894 (2005).
11. Chi, H. & Flavell, R. A. Cutting edge: regulation of T cell trafficking and primary immune responses by sphingosine 1-phosphate receptor 1. *J. Immunol.* **174**, 2485–2488 (2005).
12. Peng, T. *et al.* An effector phenotype of CD8$^+$ T cells at the junction epithelium during clinical quiescence of herpes simplex virus 2 infection. *J. Virol.* **86**, 10587–10596 (2012).
13. Schiffer, J. T. *et al.* Mucosal host immune response predicts the severity and duration of herpes simplex virus-2 genital tract shedding episodes. *Proc. Natl Acad. Sci. USA* **107**, 18973–18978 (2010).
14. Schiffer, J. T. *et al.* Frequent release of low amounts of herpes simplex virus from neurons: results of a mathematical model. *Sci. Transl. Med.* **1**, 7ra16 (2009).
15. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).
16. Tilloy, F. *et al.* An invariant T cell receptor α chain defines a novel TAP-independent major histocompatibility complex class Ib-restricted αβ T cell subpopulation in mammals. *J. Exp. Med.* **189**, 1907–1921 (1999).
17. Le Bourhis, L. *et al.* Mucosal-associated invariant T cells: unconventional development and function. *Trends Immunol.* **32**, 212–218 (2011).
18. Jing, L. *et al.* Cross-presentation and genome-wide screening reveal candidate T cells antigens for a herpes simplex virus type 1 vaccine. *J. Clin. Invest.* **122**, 654–673 (2012).
19. Clark, R. A. *et al.* Skin effector memory T cells do not recirculate and provide immune protection in alemtuzumab-treated CTCL patients. *Sci. Transl. Med.* **4**, 117ra7 (2012).
20. Gebhardt, T. *et al.* Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nature Immunol.* **10**, 524–530 (2009).
21. Gebhardt, T. *et al.* Different patterns of peripheral migration by memory CD4$^+$ and CD8$^+$ T cells. *Nature* **477**, 216–219 (2011).
22. Jiang, X. *et al.* Skin infection generates non-migratory memory CD8$^+$ T$_{RM}$ cells providing global skin immunity. *Nature* **483**, 227–231 (2012).
23. Khanna, K. M., Bonneau, R. H., Kinchington, P. R. & Hendricks, R. L. Herpes simplex virus-specific memory CD8$^+$ T cells are selectively activated and retained in latently infected sensory ganglia. *Immunity* **18**, 593–603 (2003).
24. Purwar, R. *et al.* Resident memory T cells (T$_{RM}$) are abundant in human lung: diversity, function, and antigen specificity. *PLoS ONE* **6**, e16245 (2011).
25. Wakim, L. M., Woodward-Davis, A. & Bevan, M. J. Memory T cells persisting within the brain after local infection show functional adaptations to their tissue of residence. *Proc. Natl Acad. Sci. USA* **107**, 17872–17879 (2010).
26. Trautmann, L. *et al.* Selection of T cell clones expressing high-affinity public TCRs within human cytomegalovirus-specific CD8 T cell responses. *J. Immunol.* **175**, 6123–6132 (2005).
27. Cunningham, A. L., Turner, R. R., Miller, A. C., Para, M. F. & Merigan, T. C. Evolution of recurrent herpes simplex lesions. An immunohistologic study. *J. Clin. Invest.* **75**, 226–233 (1985).
28. Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A. CD8$^+$ T lymphocyte mobilization to virus-infected tissue requires CD4$^+$ T-cell help. *Nature* **462**, 510–513 (2009).

Supplementary Information is available in the online version of the paper.

Author Contributions J.Z. and L.C. conceived the study and wrote the manuscript. J.Z. and T.P. developed the technology and analysed and interpreted the data. K.P., A.S.K., A.K., L.J. and K.D. performed the experiments. D.M.K. isolated HSV-2 reactive CD8$^+$ Tcells and peptide epitope used in the study. C.J. and A.W. directed human biopsy studies. H.R. contributed to TCR data analysis and interpretation. All authors contributed to the discussion.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.Z. (jiazhu@u.washington.edu) or L.C. (lcorey@fhcrc.org).

METHODS

Human biopsies. The study protocol, biopsy procedure and informed consent were approved by the University of Washington Institutional Review Board. Enrolled subjects had culture-proven symptomatic genital HSV-2 infection and were all HIV-1/2 seronegative. All genital skin biopsies were collected from HSV-2-seropositive immunocompetent adults at the time of active lesion and then at 2 and 8 weeks after complete lesion healing, and from contralateral unaffected (control) genital skin[1,2,12]. Tissue biopsies were snap frozen in optimum cutting temperature compound within 12 h of collection and stored at $-80\,^{\circ}$C until processing. Peripheral blood was collected at the same time points.

Immunofluorescence staining. Detection of T-cell markers, viral antigen and HSV-2-specific CD8^{+} T cells in genital tissue was performed as described[1,2]. Tyramide signal amplification (TSA; Invitrogen) was applied for the first antibody in dual staining, followed by incubation with CD8α–Alexa Fluor 647 (AF647). For triple staining, sequential TSA was performed by first applying antibody to perforin and then antibody to CD8β, followed by incubation with CD8α–AF647. Antibodies used in this study are listed in Supplementary Table 4.

CD8-specific LCM. Frozen tissue was sectioned into 8-µm slices and placed on a polyethylene naphthalate membrane slide (Carl Zeiss). Tissue sections were fixed in 100% ethanol, incubated with CD8α–AF647 on ice for 5 min, washed in PBS and dried with ethanol (75%, 95% and 100%) and xylene. CD8^{+} cells were identified and laser microdissected using either a PALM MicroBeam instrument (Carl Zeiss) or an Arcturus Veritas microdissection system (Molecular Devices). About 100 CD8^{+} cells were laser captured from each sample for transcriptional analysis, unless otherwise specified.

RNA extraction, complementary DNA amplification and hybridization to Illumina bead arrays. Total RNA from LCM-captured CD8 cells from each group were extracted using Arcturus PicoPure RNA Isolation kits (Life Technologies) and quantified by a NanoDrop 1000 Spectrophotometer. One nanogram of total RNA was then used to generate and amplify cDNA through whole-transcriptome amplification using WT-Ovation Pico RNA Amplification System (NuGEN). The cDNA were biotin labelled following a protocol from NuGEN and labelled cDNA (750 ng) were then used to hybridize to Illumina HumanRef8 v3 bead arrays in the Shared Resources at Fred Hutchison Cancer Research Center.

Bead array data analysis. Detailed analysis has been described previously[12]. In brief, raw data were imported to GenomeStudio (v2010.3, Illumina) for initial quality control. Normalized data were then exported to R and differentially expressed genes between DEJ CD8 and control CD8 or between DEJ CD8 and BV CD8 were selected. The differentially expressed genes were analysed further using an unsupervised hierarchical clustering method and using Spotfire DecisionSite for functional genomics (v9.1.2). Enriched functional categories and network analysis for the set of differentially expressed genes were performed using Ingenuity Pathway Analysis

(IPA 8.8). The GoMiner program was used to annotate all the 18,401 genes on HumanRef8 v3 bead arrays.

Microarray data accession numbers. Complete gene expression array data for DEJ CD8, control CD8 and BV CD8 can be accessed under Gene Expression Omnibus accession numbers GSE39625 and GSE44975.

TaqMan quantitative PCR. TaqMan PCR assays were performed using primer/probe sets obtained from Applied Biosystems (Supplementary Table 5). Relative expression to *ACTB* is shown for each gene. HSV-2 DNA detection in tissue was described previously[1,29].

DNA extraction and amplification. Genomic DNA from laser capture microdissected CD8^{+} T cells was extracted using Arcturus PicoPure DNA Extraction kits (Life Technologies). Recovered DNA was whole-genome amplified through multiple displacement amplification using REPLI-g Mini kit (Qiagen).

TCRβ sequencing. TCR β-chain sequencing was performed at Adaptive Biotechnologies using the ImmunoSEQ platform[15]. All TCRβ chains were PCR amplified from 400 ng of DNA from each sample. The library was sequenced using Illumina HiSeq and the resulting reads were processed and uploaded onto the ImmunoSEQ Analyzer secure relational database.

Purification of HSV reactive CD8^{+} T cells from PBMC *ex vivo*. We obtained bulk HSV-2 reactive CD8 T cells as described previously[18] and adapted the method for HSV-2. HSV-2-infected HeLa cell debris was added to cultured PBMC-derived dendritic cells, which were then incubated with negatively selected CD8 T cells from the same subject and stained with 7-aminoactinomycin D for viability assessment and with anti-CD8α, anti-CD3 and anti-CD137 conjugated to fluorochromes. The cells were sorted on a FACSAria III instrument and live, CD3^{+}, CD8α^{+} and CD137^{bright} cells were selected. Approximately 5,000 directly sorted cells were submitted for TCRβ sequencing.

Purification, sorting and deep sequencing of CD8αα^{+} T cells from PBMC. CD8αα T cells were purified from PBMC and labelled with antibodies of CD8α Pacific Blue, CD8β APC, CD3 APC-Cy7, TCRαβ^{+} Alexa 488 and CD4 phycoerythrin-Cy7. The cells were sorted using BD FACSAria. CD8αα T cells, described as CD3^{+} TCRαβ^{+} CD4^{-} CD8α^{+} CD8β^{-}, were collected and expanded using a rapid expansion protocol[30] in the presence of 50 IU ml^{-1} of recombinant human IL-2 and 1 ng ml^{-1} of recombinant human IL-7. A total of 890,000 CD8α^{+}β^{-} CD4^{-} T cells (98% pure) were collected and submitted to Adaptive Biotechnologies for TCRβ deep sequencing.

29. Magaret, A. S., Wald, A., Huang, M. L., Selke, S. & Corey, L. Optimizing PCR positivity criterion for detection of herpes simplex virus DNA on skin and mucosa. *J. Clin. Microbiol.* **45,** 1618–1620 (2007).
30. Riddell, S. R. *et al.* Restoration of viral immunity in immunodeficient humans by the adoptive transfer of T cell clones. *Science* **257,** 238–241 (1992).

LETTER

The TLR4 antagonist Eritoran protects mice from lethal influenza infection

Kari Ann Shirey[1], Wendy Lai[1], Alison J. Scott[2], Michael Lipsky[3], Pragnesh Mistry[1], Lioubov M. Pletneva[4], Christopher L. Karp[5]†, Jaclyn McAlees[5], Theresa L. Gioannini[6,7], Jerrold Weiss[6], Wilbur H. Chen[8], Robert K. Ernst[2], Daniel P. Rossignol[9]†, Fabian Gusovsky[9], Jorge C. G. Blanco[4] & Stefanie N. Vogel[1]

There is a pressing need to develop alternatives to annual influenza vaccines and antiviral agents licensed for mitigating influenza infection. Previous studies reported that acute lung injury caused by chemical or microbial insults is secondary to the generation of host-derived, oxidized phospholipid that potently stimulates Toll-like receptor 4 (TLR4)-dependent inflammation[1]. Subsequently, we reported that $Tlr4^{-/-}$ mice are highly refractory to influenza-induced lethality[2], and proposed that therapeutic antagonism of TLR4 signalling would protect against influenza-induced acute lung injury. Here we report that therapeutic administration of Eritoran (also known as E5564)—a potent, well-tolerated, synthetic TLR4 antagonist[3,4]—blocks influenza-induced lethality in mice, as well as lung pathology, clinical symptoms, cytokine and oxidized phospholipid expression, and decreases viral titres. CD14 and TLR2 are also required for Eritoran-mediated protection, and CD14 directly binds Eritoran and inhibits ligand binding to MD2. Thus, Eritoran blockade of TLR signalling represents a novel therapeutic approach for inflammation associated with influenza, and possibly other infections.

Influenza continues to evolve, and new antigenic variants emerge annually, giving rise to seasonal outbreaks. During annual influenza epidemics, 5–15% of the population is affected with upper respiratory tract infections, with hospitalization and death occurring mainly in the elderly and chronically ill. Annual influenza epidemics are estimated to result in 3–5 million cases of severe illness and 250,000–500,000 deaths yearly worldwide[5,6]. In addition, strains to which humans have no prior immunity may appear suddenly, and the resulting pandemics can be catastrophic, as illustrated by the 1918 'Spanish flu' that killed millions[7,8]. Logistical problems related to the prediction of future immunogenic epitopes, and production and distribution issues, often limit efficacy and/or vaccine availability. Moreover, increasing resistance to existing antiviral therapy, coupled with the need to administer these agents within 2–3 days after infection, limits their usefulness[9–11]. Thus, there is a critical need for a safe and effective therapeutic adjunct and/or alternative to influenza vaccines and antiviral agents.

The prototype TLR4 agonist, Gram-negative lipopolysaccharide (LPS), is a highly potent inflammatory stimulus that has been strongly implicated in Gram-negative septic shock, including acute respiratory distress syndrome caused by endothelial leak[12]. In 2008, a highly provocative paper[1] proposed that induction of acute lung injury, induced by acid aspiration, infection by respiratory viruses and bacteria, or exposure to their products (for example, inactivated H5N1 influenza), was mediated by a common signalling pathway: NADPH oxidase-dependent production of reactive oxygen species (ROS) generated oxidized host phospholipids—for example, oxidized 1-palmitoyl-2-arachidonoyl-phosphaticylcholine (OxPAPC)—that, in turn, potently stimulated TLR4. The authors proposed that the ensuing cytokine storm, and in particular, interleukin-6 (IL-6), mediated acute lung injury. Our subsequent findings that $Tlr4^{-/-}$

mice were protected from lethality caused by mouse-adapted influenza virus (A/PR/8/34, known as PR8)[2], led us to propose that blocking TLR4 therapeutically would protect against influenza infection. Here we provide compelling data showing that Eritoran (E5564; Eisai), a potent TLR4 antagonist[3,4], is highly protective when administered therapeutically to mice infected with a lethal dose of influenza.

Figure 1a illustrates our initial protocol. On 'day 0', 6–8-week old, female C57BL/6J mice were infected intranasally (i.n.) with a dose of PR8 that was determined to kill ~90% of mice (~LD_{90}). Two days later, the TLR4 signalling antagonist Eritoran (E5564) was administered daily for 5 consecutive days (200 µg per mouse; days 2–6) intravenously (i.v.). Each mouse was weighed and clinical symptoms (for example, lethargy, piloerection, ruffled fur, hunched posture, rapid shallow breathing, audible
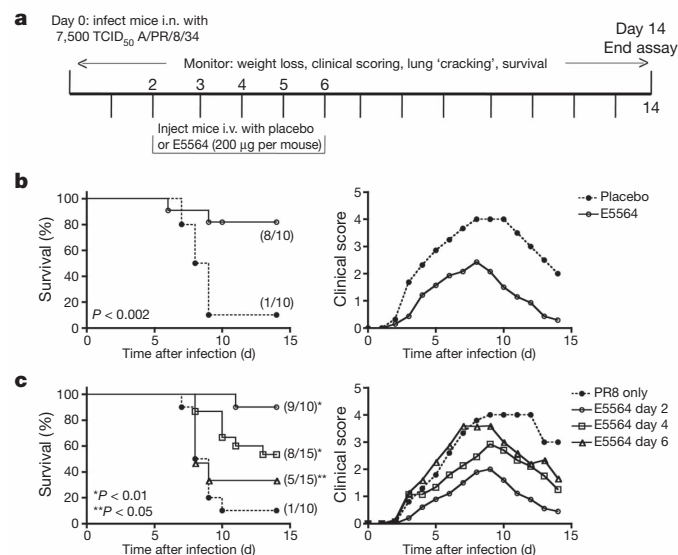

Figure 1 | Eritoran treatment protects mice from lethal influenza challenge. a, Basic experimental protocol used to test Eritoran in mice infected with influenza. C57BL/6J mice were infected with mouse-adapted influenza, strain PR8 (~7,500 $TCID_{50}$, i.n.; ~LD_{90}). Two days later, mice received placebo (vehicle only) or E5564 (Eritoran; 200 µg per mouse i.v.) once daily for 5 successive days (days 2 to 6). **b**, Mice were treated as shown in **a**. Left, survival was monitored daily ($P < 0.002$). Right, clinical scores (see Methods) were also measured daily. Each graph represents the combined results of 2 separate experiments, each with 5 mice per treatment group per experiment. **c**, Mice were infected as described in **a**, but treated with Eritoran starting on days 2, 4 or 6 after infection. Left, survival (day 2 and day 4, $P < 0.01$; day 6, $P \leq 0.05$); right, clinical scores. Results are combined results from 2–3 separate experiments, with 5 mice per treatment group per experiment.

[1]Department of Microbiology and Immunology, University of Maryland, Baltimore, Baltimore, Maryland 21201, USA. [2]Department of Microbial Pathogenesis, University of Maryland, Baltimore, Baltimore, Maryland 21201, USA. [3]Department of Pathology, University of Maryland, Baltimore, Baltimore, Maryland 21201, USA. [4]Sigmovir Biosystems, Inc., Rockville, Maryland 20850, USA. [5]Division of Molecular Immunology, Cincinnati Children's Hospital Research Foundation, Cincinnati, Ohio 45229, USA. [6]The Inflammation Program, University of Iowa, Coralville, Iowa 52241, USA. [7]Department of Veterans Affairs Medical Center, Iowa City, Iowa 52241, USA. [8]Center for Vaccine Development, University of Maryland, Baltimore, Baltimore, Maryland 21201, USA. [9]Eisai, Inc., Andover, Massachusetts 01810, USA. †Present addresses: The Bill & Melinda Gates Foundation, Seattle, Washington 98109, USA (C.L.K.); Taisho Pharmaceutical R&D, Inc., Morristown, New Jersey 07960, USA (D.P.R.).

rattling) were scored daily for 2 weeks. Eritoran-treated mice were significantly protected from influenza-induced lethality, whereas 90% of mice administered placebo (vehicle only) died (Fig. 1b, left). This was paralleled by more severe clinical scores (Fig. 1b, right) and weight loss (Supplementary Fig. 1A) in control mice versus Eritoran-treated mice. Treatment of mice with tenfold less Eritoran (20 µg per mouse) decreased survival to 40% (4/10 mice survived; data not shown). The efficacy of Eritoran treatment decreased from 90% survival in mice infected with 7,500 50% tissue culture infective dose ($TCID_{50}$) to 60% and 25% survival in mice infected with 10,000 and 20,000 $TCID_{50}$, respectively (Supplementary Fig. 1B). Treatment with Eritoran also protected mice infected with a lethal dose of the non-adapted 2009 human pandemic influenza strain A/California/07/2009 H1N1 (Supplementary Fig. 1C).

Additional experiments were performed in which Eritoran treatment was started on day 2, day 4 or day 6 after infection. Figure 1c (left) illustrates that 90% of PR8-infected mice that did not receive Eritoran treatment died, whereas mice that received Eritoran starting on days 2, 4 or 6 had statistically significant survival rates of 90%, 53% and 33%, respectively. The clinical scores for these same mice (Fig. 1c, right) illustrate that the later mice began Eritoran treatment, the more severe

their clinical scores. Mice that did not receive Eritoran showed the most severe symptoms and highest clinical scores. Similarly, weight loss was progressively greater the later treatment was initiated after infection (data not shown). Thus, in contrast to current antivirals that must be administered soon after infection, Eritoran treatment significantly enhanced survival, even when started as late as 6 days after infection.

Eritoran treatment also led to a statistically significant reduction in influenza-induced lung pathology. Groups of mice received saline only ('mock'), PR8 only ('PR8 untreated'), or were infected with PR8 and treated with Eritoran (PR8 + E5564) starting 2 days after infection (Fig. 1a), and were killed 7 days after infection. In mice infected with PR8 only, extensive lung damage (that is, highly basophilic lining epithelium in the bronchioles, focal degenerating cells undergoing apoptosis or necrosis, and extensive cellular infiltrates (neutrophils, monocytes and lymphocytes)) were observed; however, in infected, Eritoran-treated mice, ~80% of each lung section examined exhibited nearly normal lung architecture (similar to 'mock'), whereas in ~20% of each section, inflammatory infiltrates could be observed, although to a lesser extent than in the lungs of mice infected with PR8 only (Fig. 2a). These observations are supported by blinded histological scoring (Fig. 2b). By day
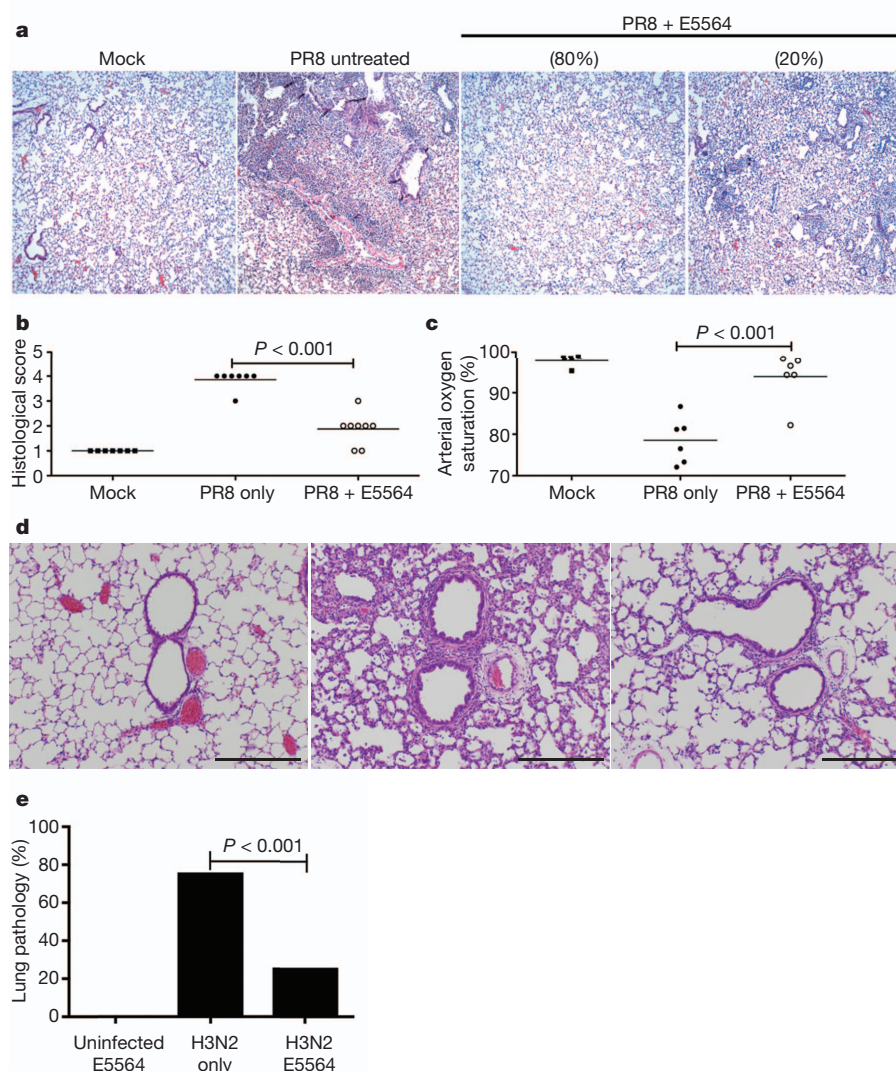
Figure 2 | Eritoran treatment inhibits influenza-induced lung pathology and lung function. a, Eritoran treatment improves lung pathology induced by PR8 infection. Representative haematoxylin and eosin (H&E)-stained sections were derived from mice treated as described in the text (4–5 mice per treatment group) killed on day 7 after infection. Original magnification, ×400. **b**, Histological scoring for PR8-infected mice with or without Eritoran treatment ± standard error of the mean (s.e.m.). **c**, Pulse oximetry data collected on day 6 after infection ± s.e.m. ($P < 0.001$, comparing PR8 only with mock-infected or PR8 plus E5564; 4–6 mice per treatment group). **d**, Eritoran treatment improves lung pathology of H3N2 infection of cotton rats. Scale bars, 500 µm. **e**, Histological scoring for H3N2-infected cotton rats with or without Eritoran treatment ($n = 5$ rats per treatment group; data shows mean ± s.e.m.).

14 after infection, Eritoran-treated mice showed normal lung architecture (data not shown). Previously, a direct correlation between pulse-oximetry measurements of oxygen saturation and lung pathology, which reflected the morbidity and survival of mice infected with PR8, was reported[13]. By day 6 after infection, the oxygen saturation levels observed in mock-infected and PR8-infected, Eritoran-treated mice were not significantly different. By contrast, PR8-infected mice demonstrated a significant oxyhaemoglobin desaturation to 78%, suggesting a functional consequence of the alveolar injury demonstrated histologically (Fig. 2c).

To determine whether the therapeutic effect of Eritoran extends to other animal models of human influenza infection, we performed experiments in cotton rats (*Sigmodon hispidus*)[14,15]. A/Wuhan/359/95 (H3N2), a human non-adapted strain of influenza, replicates in lungs of cotton rats on day 1 and produces peak lung pathology on day 4 after infection (Fig. 2d, middle; and Fig. 2e, H3N2 only). Animals treated with Eritoran after H3N2 challenge showed significant reductions in lung pathology on day 4 compared with animals treated with vehicle (Fig. 2d, right; and Fig. 2e, H3N2/E5564).

Figure 3a shows that by day 6 after infection, Eritoran treatment resulted in a statistically significant decrease in lung viral titres that was further decreased by day 7 (Fig. 3b) and day 8 (data not shown). This was not a direct effect of Eritoran on virus replication (Supplementary Table 1). By day 14, all surviving mice had completely cleared the virus, regardless of treatment. Because the effects of influenza infection are systemic[16], levels of the liver enzymes alanine aminotransferase (ALT) and aspartate aminotransferase (AST) were measured in sera. Although PR8 infection resulted in significant increases in ALT and AST levels, these were significantly lower in Eritoran-treated mice 6 days after infection (Supplementary Fig. 2).

Influenza infection potently induces expression of both pro- and anti-inflammatory genes *in vivo* that contribute to the inflammatory response[1,17]. To determine whether Eritoran also mitigates PR8-induced gene expression, groups of mice were infected on day 0, treated with Eritoran starting on day 2, and killed on day 2 (3 h after Eritoran treatment) and on days 4 and 6. Total RNA was extracted from lung homogenates and gene expression was measured by quantitative polymerase chain reaction with quantitative real-time reverse transcription (qRT–PCR). Eritoran-treated mice showed significantly blunted pro- and anti-inflammatory gene expression at each time point (Fig. 4a and Supplementary Fig. 3A; $P < 0.001$), as early as 3 h after treatment on day 2. Supplementary Fig. 3B shows that influenza induced type I, II and III interferon (IFN); however, only *Ifnb* and *Ifng* messenger RNA expression were blunted by Eritoran treatment. In addition, cotton rats infected with the non-adapted human Wuhan H3N2 strain and treated with Eritoran showed decreased lung expression of IL-6 and IL-10 (Fig. 4b).

Ifnb[−/−] mice are highly susceptible to PR8 and are not protected by the potent IFN-β inducer, 5,6-dimethylxanthenone-4-acetic acid[18]. In addition, endogenous IFN-β is required for normal basal and LPS-inducible expression of many proinflammatory cytokines in macrophages[19]. PR8-infected *Ifnb*[−/−] mice were not protected by Eritoran (Supplementary
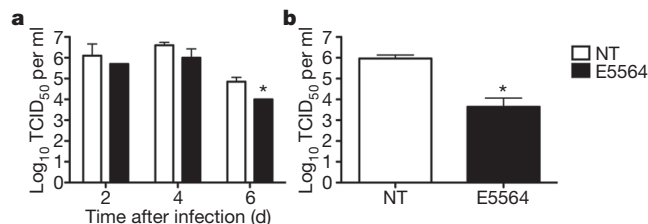
Figure 3 | Treatment with Eritoran reduces lung viral titres. a, b, Mice were infected with PR8 and then either left untreated (NT) or were treated with Eritoran (E5564) as described in Fig. 1a, and were killed on days 2, 4, 6 (**a**) and 7 (**b**) after infection. **a,** Results represent the combined results of 2 separate experiments (5 mice per treatment group per experiment); *$P < 0.05$. **b,** Day 7 after infection (7 mice per treatment group); *$P < 0.001$.
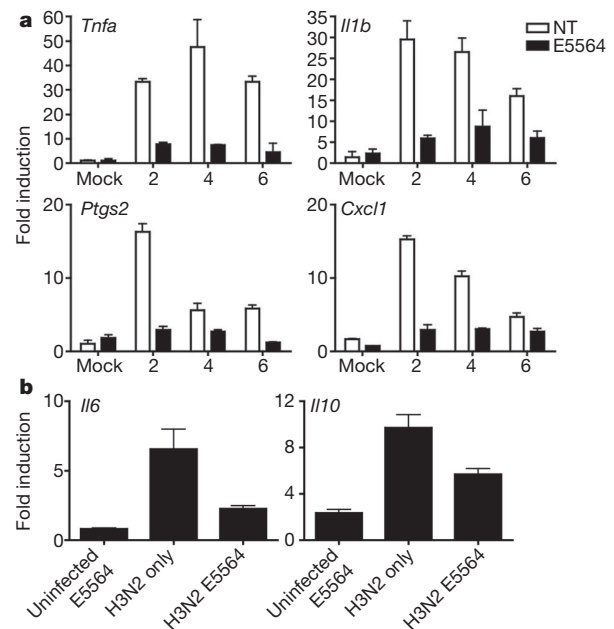
Figure 4 | Eritoran treatment suppresses influenza-induced cytokine gene expression. a, b, Mice (**a**) and cotton rats (**b**) were treated as described in Fig. 1a and killed on days 2 (3 h after treatment), 4 and 6 after infection (4 mice per treatment group per experiment; data are mean ± s.e.m. from 2 separate experiments; $P < 0.01$ at each time point). Lungs were processed for total RNA and subjected to qRT–PCR for detection of specific gene expression. NT, not treated.

Fig. 4). This indicates that although both pro-inflammatory and *Ifnb* gene expression are blunted in PR8-infected, Eritoran-treated wild-type mice (Fig. 4a, b), the elaboration of some IFN-β is required for antiviral protection in the absence of significant priming.

Most potent activation of TLR4 by LPS depends on LPS binding protein (LBP)-catalysed extraction and transfer of individual LPS molecules from aggregated LPS to CD14 and then from CD14 to MD2, followed by engagement and dimerization of TLR4. Crystallographic analysis revealed that Eritoran, like LPS lipid A monomer, binds in the deep hydrophobic pocket of MD2; however, in contrast to lipid A, Eritoran fails to induce TLR4 dimerization and activation, thus acting as a TLR4 antagonist[20].

We confirmed the finding[2] that PR8-induced lethality in mice was TLR4 dependent (Fig. 5a), but also found, surprisingly, that influenza mortality was not CD14 dependent, as indicated by the similar mortality resulting from PR8 infection in wild-type and *Cd14*[−/−] mice versus the resistance of *Tlr4*[−/−] mice (Fig. 5a). These findings suggest that TLR4-dependent influenza mortality is mediated by influenza-induced TLR4 agonists that, unlike LPS, do not depend on CD14 for engagement and activation of TLR4/MD2. By contrast, the protection induced in wild-type mice when Eritoran treatment is started on day 2 after infection is absent in *Cd14*[−/−] mice, indicating that the protective effect of Eritoran in this influenza infection model is CD14 dependent (Fig. 5a).

One possible explanation for the CD14 dependence of Eritoran-mediated protection is that engagement by Eritoran of MD2 and antagonism of TLR4 depends on CD14. To test this possibility, we assayed *in vitro* the capacity of Eritoran to bind CD14 and MD2 by measuring inhibition by Eritoran of LBP-dependent transfer of tritiated lipooligosaccharide ([³H]LOS; the LPS of *Neisseria*) to CD14 (Fig. 5b, left), as well as the transfer of [³H]LOS from CD14 to MD2 (Fig. 5b, right). As the inhibitory effect of Eritoran on the transfer of [³H]LOS to MD2 is 100 times more potent when LBP and soluble CD14 are present, this indicates that the interaction of Eritoran with MD2 occurs most efficiently after LBP-dependent interaction of Eritoran with CD14.

Although MD1, a homologue of MD2 that also binds lipid A[21], was considered as an alternative target for Eritoran, this possibility was
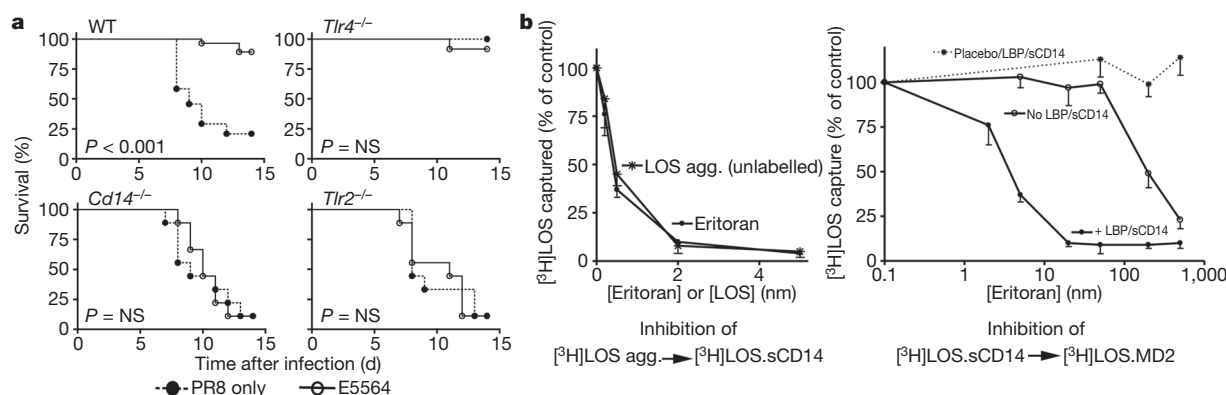
Figure 5 | Molecular requirements of Eritoran-induced protection. a, Wild-type (WT), $Tlr4^{-/-}$, $Cd14^{-/-}$, $Tlr2^{-/-}$ mice were either untreated (filled circles), or treated with E5564 (open circles) 2 days after infection, for 5 successive days. Wild-type data were combined from 5 separate experiments (5–6 mice per treatment group per experiment), $Tlr4^{-/-}$ data from 3 separate experiments (5–6 mice per treatment group per experiment), $Cd14^{-/-}$ mice (2 separate experiments; 5 mice per treatment group per experiment), and $Tlr2^{-/-}$ mice (2 separate experiments; 4–5 mice per treatment group per experiment). Wild-type: untreated versus E5564 treatment ($P < 0.0001$); $Tlr4^{-/-}$: untreated versus E5564 treatment ($P =$ non-significant (NS)); $Cd14^{-/-}$: untreated versus E5564 treatment ($P =$ NS); $Tlr2^{-/-}$: untreated versus E5564 treatment ($P =$ NS). **b**, Dose-dependent Eritoran inhibition of the transfer and binding of monomeric [^{3}H]LOS from aggregated (agg.) [^{3}H]LOS

to soluble CD14 (sCD14), and from [^{3}H]LOS.sCD14 to MD2. Samples containing [^{3}H]LOS aggregates (0.2 nM), His$_6$–sCD14 (~0.5 nM), and increasing concentrations as indicated of Eritoran or unlabelled LOS (left) or 2 nM [^{3}H]LOS.sCD14, approximately 2 nM His$_6$–MD2, and increasing concentrations of Eritoran (or placebo) ± LBP (50 pM) and sCD14 (2 nM) (right) were incubated for 30 min at 37 °C, followed by addition and incubation with NiFF Sepharose beads to capture His-tagged proteins. Formation of complexes of [^{3}H]LOS with His$_6$–sCD14 (left) or MD2 (right) was assayed by measuring co-capture of [^{3}H]LOS by NiFF Sepharose as previously described[25]. Data are expressed as per cent of co-capture of [^{3}H]LOS observed in the absence of added Eritoran. Data show mean ± s.e.m. of 3 separate experiments with duplicate samples for each dose.

dismissed because MD1 failed to substitute for MD2 to enable TLR4 signalling (Supplementary Fig. 5).

Diacylated lipids can act as TLR2 agonists. Hence, if oxidized host phospholipids (for example, OxPAPC) have a role in TLR4-dependent influenza, it is conceivable that these diacylated lipids could also engage TLR2-containing receptor complexes and that influenza-induced mortality might also be TLR2 dependent. However, $Tlr2^{-/-}$ mice showed a similar susceptibility to PR8 infection as wild-type mice (Fig. 5a), arguing against an important role for TLR2 in influenza-induced mortality. Remarkably, however, protection induced by Eritoran treatment beginning at day 2 after infection was absent in $Tlr2^{-/-}$ mice, suggesting a novel role for TLR2 in Eritoran action.

If the PR8-induced acute lung injury and mortality we have observed is due substantially to influenza-induced oxidized host phospholipids, as previously proposed[1], these compounds should show preferential activation of TLR4 (versus TLR2) and be inhibited by Eritoran. Commercially obtained OxPAPC comparably activated $Il6$ gene expression in wild-type and $Tlr2^{-/-}$ peritoneal macrophages, but was inhibited by ~84% in $Tlr4^{-/-}$ macrophages when compared to the wild-type response. Cellular activation by OxPAPC was also significantly inhibited by Eritoran, including the low level of TLR4-independent production of $Il6$ mRNA observed in $Tlr4^{-/-}$ macrophages (Supplementary Fig. 6). This indicates that the response to OxPAPC is predominantly TLR4 dependent and Eritoran sensitive.

To assess the effect of Eritoran on production of oxidized phospholipids during infection, matrix-assisted laser desorption ionization-imaging mass spectrometry (MALDI-IMS) was used to identify alterations in the lipid composition of mouse lungs after PR8 infection, with or without Eritoran treatment. Predicted oxidation products (mass to charge ratio (m/z) 650, 830) of 1-palmitoyl-2-arachidonayl-sn-glycero-3-phosphocholine (PAPC), or other polyunsaturated phosphatidylcholines, as described previously[1], were detected in greater abundance in PR8-infected versus mock-infected or infected and Eritoran-treated lungs (Supplementary Fig. 7). These results were confirmed by electrospray ionization–mass spectrometry (data not shown). Together, these results indicate that oxidized phosphatidylcholines are present in influenza-infected lungs and show that treatment with Eritoran reduces the relative abundance of these characteristic ions. Thus, Eritoran treatment of influenza-infected

mice not only blocks the cytokine storm exacerbated by influenza-induced endogenous TLR4 agonists, but also inhibits the production and accumulation of certain oxidized phospholipids in infected lungs, including OxPAPC and possibly others.

Our findings provide further support for a prominent role of influenza-induced oxidized host phospholipids in TLR4-dependent acute lung injury and mortality, and demonstrate a novel therapeutic potential of Eritoran towards this and, potentially, other respiratory infections. In the hypothetical model shown in Supplementary Fig. 8, we would predict, based our data, that Eritoran blocks oxidized-phospholipid-induced TLR4 signalling, thereby mitigating the cytokine storm and subsequent potentiation of phospholipid oxidation induced by ROS. Our failure to detect commensal bacteria or LPS in lung homogenates from PR8-infected mice suggests that the disease manifestations are not the result of secondary bacterial infection (data not shown). The TLR2 and CD14 independence of influenza mortality is consistent with the selective agonist properties of at least certain oxidized host phospholipids that we have observed and suggests that these compounds, unlike most LPS species, can engage MD2/TLR4 in the absence of CD14, perhaps by interacting with an alternative lipid acceptor/donor such as albumin[22,23]. By contrast, the CD14 and TLR2 dependence of the therapeutic action of Eritoran may reflect CD14-dependent engagement of MD2/TLR4 by Eritoran, precluding the action of endogenous TLR4 agonists generated after initiation of infection and early tissue alterations, as well as uncharacterized interactions of Eritoran with TLR2-containing receptors. This possibility is supported by the observation that mycobacterial lipoarabinomannan requires CD14 for TLR2 activation[24]. That influenza infection modulates TLR2 expression in many cell types[25,26], and synthetic TLR2 ligands provide prophylaxis against lethal influenza infection[27], strengthens the notion that TLR2-containing receptors may provide another target for Eritoran. These findings underscore the novel therapeutic utility of Eritoran, either used as a monotherapy or, more likely, as an adjunct to antiviral agents. The utility of Eritoran is further enhanced by its well-documented safety profile[28].

Many drugs targeting endotoxin or the endotoxin-induced cytokine storm were originally developed for use in sepsis, yet none have emerged as effective therapeutics for sepsis. Moreover, whereas many of these drugs failed to provide therapeutic protection in animal models of

endotoxicity or in clinical trials in septic humans, they were found to be of great benefit in other, more localized models of inflammation (such as rheumatoid arthritis and inflammatory bowel disease). Our data suggest that Eritoran might follow a similar development pathway. Eritoran demonstrated remarkable activity *in vitro* and *in vivo*[4,28] in endotoxin challenge studies when administered prophylactically or very early in the 'sepsis cascade'. A non-statistically significant improvement in mortality was also observed in a small phase 2 severe sepsis study[29]. The phase 2 study design and statistical power calculations were expanded to perform a single dose, placebo-controlled phase 3 severe sepsis study (the ACCESS study) in which 1,984 patients were treated in a 2:1 ratio of drug and placebo. The results of this study demonstrated that new treatment regimens including aggressive fluid resuscitation and early goal-directed therapy markedly reduced mortality in the placebo group; however, further benefit to septic patients was not observed with Eritoran treatment[30]. Thus, it is notable that in this study, Eritoran prevented mortality when administered up to 6 days after influenza infection of mice. This more practical timing of treatment may enable effective clinical treatment of influenza and, perhaps, other infectious agents.

METHODS SUMMARY

***In vivo* measurements.** Mice and cotton rats were infected i.n. as indicated with influenza virus, followed by administration of control or Eritoran i.v. Survival, clinical score, weight loss, cytokine gene expression, viral titres, OxPAPC abundance and ALT/AST levels were measured.

Cell culture measurements. Primary murine macrophage cultures were treated with TLR agonists or OxPAPC with or without Eritoran, and *Il6* mRNA was measured by qRT–PCR.

Biochemical measurements. The ability of Eritoran to inhibit binding of LPS to CD14 or the transfer of LPS from CD14 to MD2 was carried out as reported[22,23].

Full Methods and any associated references are available in the online version of the paper.

1. Imai, Y. *et al.* Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. *Cell* **133,** 235–249 (2008).
2. Nhu, Q. M. *et al.* Novel signaling interactions between proteinase-activated receptor 2 and Toll-like receptors *in vitro* and *in vivo*. *Mucosal Immunol.* **3,** 29–39 (2010).
3. Kalil, A. C. *et al.* Influence of severity of illness on the effects of eritoran tetrasodium (E5564) and on other therapies for severe sepsis. *Shock* **36,** 327–331 (2011).
4. Mullarkey, M. *et al.* Inhibition of endotoxin response by E5564, a novel Toll-like receptor 4-directed endotoxin antagonist. *J. Pharmacol. Exp. Ther.* **304,** 1093–1102 (2003).
5. Thompson, W. W. *et al.* Mortality associated with influenza and respiratory syncytial virus in the United States. *J. Am. Med. Assoc.* **289,** 179–186 (2003).
6. Thompson, W. W. *et al.* Influenza-associated hospitalizations in the United States. *J. Am. Med. Assoc.* **292,** 1333–1340 (2004).
7. Reid, A. H., Taubenberger, J. K. & Fanning, T. G. The 1918 Spanish influenza: integrating history and biology. *Microbes Infect.* **3,** 81–87 (2001).
8. Taubenberger, J. K., Reid, A. H., Janczewski, T. A. & Fanning, T. G. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Phil. Trans. R. Soc. Lond. B* **356,** 1829–1839 (2001).
9. Hurt, A. C., Holien, J. K., Parker, M., Kelso, A. & Barr, I. G. Zanamivir-resistant influenza viruses with a novel neuraminidase mutation. *J. Virol.* **83,** 10366–10373 (2009).
10. McKimm-Breschkin, J. L. *et al.* Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Virol.* **72,** 2456–2462 (1998).
11. Mishin, V. P., Hayden, F. G. & Gubareva, L. V. Susceptibilities of antiviral-resistant influenza viruses to novel neuraminidase inhibitors. *Antimicrob. Agents Chemother.* **49,** 4515–4520 (2005).
12. Goldblum, S. E., Ding, X., Brann, T. W. & Campbell-Washington, J. Bacterial lipopolysaccharide induces actin reorganization, intercellular gap formation, and endothelial barrier dysfunction in pulmonary vascular endothelial cells: concurrent F-actin depolymerization and new actin synthesis. *J. Cell. Physiol.* **157,** 13–23 (1993).
13. Verhoeven, D., Teijaro, J. T. & Farber, D. L. Pulse-oximetry accurately predicts lung pathology and the immune response during influenza infection. *Virology* **390,** 151–156 (2009).
14. Blanco, J. C. G. *et al.* Receptor characterization and suceptibility of cotton rats to avian and 2009 pandemic influenza virus strains. *J. Virol.* **87,** 2036–2045 (2013).
15. Ottolini, M. G. *et al.* The cotton rat provides a useful small-animal model for the study of influenza virus pathogenesis. *J. Gen. Virol.* **86,** 2823–2830 (2005).
16. Polakos, N. K. *et al.* Kupffer cell-dependent hepatitis occurs during influenza infection. *Am. J. Pathol.* **168,** 1169–1178 (2006).
17. Taubenberger, J. K. & Morens, D. M. The pathology of influenza virus infections. *Annu. Rev. Pathol.* **3,** 499–522 (2008).
18. Shirey, K. A. *et al.* The anti-tumor agent, 5,6-dimethylxanthenone-4-acetic acid (DMXAA), induces IFN-β-mediated antiviral activity *in vitro* and *in vivo*. *J. Leukoc. Biol.* **89,** 351–357 (2011).
19. Thomas, K. E., Galligan, C. L., Newman, R. D., Fish, E. N. & Vogel, S. N. Contribution of interferon-β to the murine macrophage response to the toll-like receptor 4 agonist, lipopolysaccharide. *J. Biol. Chem.* **281,** 31119–31130 (2006).
20. Kim, H. M. *et al.* Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist eritoran. *Cell* **130,** 906–917 (2007).
21. Yoon, S. I., Hong, M., Han, G. W. & Wilson, I. A. Crystal structure of soluble MD-1 and its interaction with lipid IVa. *Proc. Natl Acad. Sci. USA* **107,** 10990–10995 (2010).
22. Esparza, G. A., Teghanemt, A., Zhang, D., Gioannini, T. L. & Weiss, J. P. Endotoxin-albumin complexes transfer endotoxin monomers to MD-2 resulting in activation of TLR4. *Innate Immun.* **18,** 478–491 (2012).
23. Gioannini, T. L., Zhang, D., Teghanemt, A. & Weiss, J. P. An essential role for albumin in the interaction of endotoxin with lipopolysaccharide-binding protein and sCD14 and resultant cell activation. *J. Biol. Chem.* **277,** 47818–47825 (2002).
24. Means, T. K. *et al.* The CD14 ligands lipoarabinomannan and lipopolysaccharide differ in their requirement for Toll-like receptors. *J. Immunol.* **163,** 6748–6755 (1999).
25. Pauligk, C., Nain, M., Reiling, N., Gemsa, D. & Kaufmann, A. CD14 is required for influenza A virus-induced cytokine and chemokine production. *Immunobiology* **209,** 3–10 (2004).
26. Lee, R. M., White, M. R. & Hartshorn, K. L. Influenza A viruses upregulate neutrophil Toll-like receptor 2 expression and function. *Scand. J. Immunol.* **63,** 81–89 (2006).
27. Shinya, K. *et al.* Toll-like receptor pre-stimulation protects mice against lethal infection with high pathogenic influenza viruses. *Virol. J.* **8,** 97–101 (2011).
28. Wong, Y. N. *et al.* Safety, pharmacokinetics, and pharmacodynamics of E5564, a lipid A antagonist, during an ascending single-dose clinical study. *J. Clin. Pharmacol.* **43,** 735–742 (2003).
29. Tidswell, M. *et al.* Phase 2 trial of eritoran tetrasodium (E5564), a Toll-like receptor 4 antagonist, in patients with severe sepsis. *Crit. Care Med.* **38,** 72–83 (2010).
30. Cohen, J., Opal, S. & Calandra, T. Sepsis studies need new direction. *Lancet Infect. Dis.* **12,** 503–505 (2012).

Author Contributions K.A.S. and S.N.V. carried out the study design (with advice from J.C.G.B., D.P.R., J.W., R.K.E. and C.L.K.). K.A.S., W.L., J.C.G.B., L.M.P., A.J.S., T.L.G., J.M. and P.M. performed experiments. M.L. performed histological analysis. F.G., W.H.C. and D.P.R. provided crucial reagents and advice. K.A.S. and S.N.V. prepared the manuscript, with input and approval from all other co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.N.V. (svogel@som.umaryland.edu).

METHODS

Reagents. Eritoran (E5564) and its corresponding placebo (vehicle only) were provided by Eisai, Inc. Eritoran was prepared at $2.33\,mg\,ml^{-1}$ in sterile, endotoxin-free water and diluted for injection in sodium-bicarbonate-buffered 5% dextrose water. OxPAPC was purchased from Hycult Biotech. *Escherichia coli* K235 LPS was prepared as previously described[1]. Pam3Cys was purchased from EMC Microcollections.

Mice and cotton rats. Six- to eight-week old wild-type C57BL/6J, *Tlr2*$^{-/-}$ and *Cd14*$^{-/-}$ mice were purchased from The Jackson Laboratory. *Tlr4*$^{-/-}$ mice (provided by S. Akira) on a C57BL/6 background, were bred in a University of Maryland, Baltimore accredited facility. Inbred young adult (4–8 weeks old) cotton rats (*S. hispidus*) were bred at Sigmovir Biosystems. All animal experiments were conducted with Institutional Animal Care and Use Committee approval from the University of Maryland, Baltimore and Sigmovir Biosystems.

Virus. Mouse-adapted H1N1 influenza A/PR/8/34 virus (PR8; ATCC) was grown in the allantoic fluid of 10-day-old embryonated chicken eggs as described[2], and was provided by D. Farber. Non-adapted human influenza virus strain A/Wuhan/359/95 (H3N2), and A/Brisbane/59/07 (H1N1) was obtained and grown as previous described[3,4]. Non-adapted human influenza A/California/07/2009 strain (human pandemic H1N1) was provided by T. Ross.

Virus challenge and treatments. C57BL/6J wild-type, *Tlr4*$^{-/-}$, *Tlr2*$^{-/-}$ and *Cd14*$^{-/-}$ mice were infected with mouse-adapted influenza virus, strain A/PR/8/34 (PR8; ~7,500 TCID$_{50}$, i.n., 25 μl per nares) or the non-adapted human H1N1 strain, A/California/07/2009 (~10^7 TCID$_{50}$, i.n.). This dose was found in preliminary experiments to kill ~90% (PR8) or ~75% (human H1N1) of infected mice. Two days after infection, mice received either placebo or Eritoran (E5564; 200 μg per mouse in 100 μl, i.v.) daily (day 2 to day 6). In some experiments, some groups of mice were treated with Eritoran starting at day 4 or day 6 after infection and treated for 5 or 3 consecutive days, respectively. In some experiments, some mice were infected with ~10,000 or 20,000 TCID$_{50}$ PR8. Mice were monitored daily for survival, weight loss, and clinical signs of illness (for example, lethargy, piloerection, ruffled fur, hunched posture, rapid shallow breathing, audible crackling) for 14 days. A clinical score ranging from 0 (no symptoms) to 5 (moribund) was ascribed to each mouse daily[5]. In some experiments, mice were killed at the indicated times after infection to harvest serum for liver enzyme levels or lungs for analysis of gene expression, lung pathology, or viral titres.

Cotton rats were infected with non-adapted human influenza strain A/Wuhan/359/95 (H3N2; ~10^7 TCID$_{50}$ or p.f.u., i.n., 50 μl per nares). Two hours after infection, animals received E5564 (200 μg per rat in volume, retro-orbital) daily for 5 days (days 0–4). Animals were killed at day 4 post-infection to determine lung pathology and total RNA for measuring gene expression by real-time PCR.

Histopathology. Lungs were inflated and perfused, and fixed with 4% PFA. Fixed sections (8 μm) of paraffin-embedded lungs were stained with H&E. Slides were randomized, read blindly, and examined for tissue damage, necrosis, apoptosis and inflammatory cellular infiltration.

Pulse oximetry measurements. At 6 days after infection, per cent arterial oxygen saturation was measured on individual mice that were lightly anaesthetized with nebutol. An extra small collar sensor and a Small Animal Oximetry Restraint Device (Starr-Gate) was used as per the manufacturer's instructions.

Viral titration. Virus titres were obtained from supernatants of lung homogenates of PR8-infected mice that were either left untreated or treated with Eritoran and harvested on days 2, 4, 6 and 7 after infection, and expressed as TCID$_{50}$ per ml as described previously[6].

qRT–PCR. Total RNA isolation and qRT–PCR were performed as previously described[7]. Levels of mRNA for specific genes are reported as relative gene expression normalized to mock-infected lungs.

Liver enzyme levels. Serum was collected on day 7 from C57BL/6J wild-type mice that were either mock-infected with saline or infected with PR8, and were either left untreated or treated with Eritoran (E5564) starting on day 2 after infection. ALT and AST levels were measured using a Dimension Vista System Flex Reagent cartridge (Siemens Healthcare Diagnostics).

Macrophage cell cultures and treatment. Thioglycollate-elicited peritoneal macrophages from wild-type, *Tlr4*$^{-/-}$ and *Tlr2*$^{-/-}$ mice were enriched as described[8] after plating in 12-well (2×10^6 cells per well) tissue culture plates. Macrophages were pre-treated with E5564 ($10\,ng\,ml^{-1}$) for 1 h and then stimulated with LPS ($10\,ng\,ml^{-1}$), P3C ($300\,ng\,ml^{-1}$) or OxPAPC ($20\,μg\,ml^{-1}$) for 8 h.

MALDI-IMS. Mouse lung tissue was snap frozen by floating in an aluminium foil raft on liquid nitrogen. Tissues were cryosectioned (unfixed and unembedded) in 12 μm sections, each section was transferred to a conductive MALDI-IMS glass slide, heat-fixed until visibly dry (approximately 30–60 s), and desiccated for at least 1 h. A $12.5\,mg\,ml^{-1}$ solution of norharman MALDI matrix solvated in chloroform:methanol:water (1:2:0.8, v:v:v) was spray coated onto slides using the Bruker ImagePrep device. MS data was collected in positive mode (detection range: m/z 400–900), raster width 150 μm, 500 shots per raster, on a Bruker Daltonics AutoFlex Speed Matrix-Assisted Laser Desorption Ionization Time-of-Flight/Time-of-Flight Mass Spectrometer (MALDI-TOF/TOF MS; source) using flexControl software. Subsequently, the data were analysed using flexImaging software. After IMS analysis, the matrix was stripped in 70% methanol and tissue sections were restained by H&E for gross histological reference using standard protocols. All MALDI-IMS specific materials, equipment, instruments and software were obtained from Bruker Daltonics. Unless otherwise specified, all reagents were purchased from Sigma-Aldrich.

Statistics. Statistical differences between two groups were determined using an unpaired, two-tailed Student's *t*-test with significance set at $P < 0.05$. For comparisons between ≥3 groups, analysis was done by one-way ANOVA followed by a Tukey's multiple comparison test with significance determined at $P < 0.05$. For survival studies, a Log-Rank (Mantel-Cox) test was used.

LETTER

The microRNA miR–235 couples blast–cell quiescence to the nutritional state

Hidefumi Kasuga[1], Masamitsu Fukuyama[1], Aya Kitazawa[1], Kenji Kontani[1] & Toshiaki Katada[1]

The coordination of stem- and blast-cell behaviours, such as self-renewal, differentiation and quiescence, with physiological changes underlies growth, regeneration and tissue homeostasis[1–3]. Germline stem and somatic blast cells in newly hatched *Caenorhabditis elegans* larvae can suspend postembryonic development, which consists of diverse cellular events such as migration, proliferation and differentiation, until the nutritional state becomes favourable (termed L1 diapause[4–6]). Although previous studies showed that the insulin/insulin-like growth factor (IGF) signalling (IIS) pathway regulates this developmental quiescence[5–8], the detailed mechanism by which the IIS pathway enables these multipotent cells to respond to nutrient availability is unknown. Here we show in *C. elegans* that the microRNA (miRNA) miR-235, a sole orthologue of mammalian miR-92 from the oncogenic *miR-17-92* cluster[9,10], acts in the hypodermis and glial cells to arrest postembryonic developmental events in both neuroblasts and mesoblasts. Expression of *mir-235* persists during L1 diapause, and decreases upon feeding in a manner dependent on the IIS pathway. Upregulation of one of the miR-235 targets, *nhr-91*, which encodes an orthologue of mammalian germ cell nuclear factor, is responsible for defects caused by loss of the miRNA. Our findings establish a novel role of a miR-92 orthologue in coupling blast-cell behaviours to the nutritional state.

miRNAs are a class of small RNAs that repress expression of target messenger RNAs by base pairing with imperfect complimentary sequences in their 3′ untranslated regions (3′ UTRs)[11]. miRNAs have been found to control various functions of stem and blast cells such as pluripotency[12] and stage-specific cell fate[13,14]. Thus, we proposed that an miRNA has a critical role in regulating L1 diapause. Among the stem and blast cells that initiate postembryonic development during the L1 stage, we focused on reactivation of quiescent P neuroblasts owing to ease of detection; the process first begins with migration from the ventra-lateral sides towards the ventral midline (Fig. 1a). After screening 85 lines of miRNA mutants covering deletions of 105 individual miRNAs[15], only animals carrying *mir-235(n4504)* showed defects in arresting migration of P neuroblasts when cultured under nutritionally poor conditions after hatch (Fig. 1b, c and Supplementary Fig. 2). *n4504* deletes the entire stem-loop structure of the miR-235 miRNA, an orthologue of human miR-92 from the oncogenic *miR-17-92* cluster[9,10], and the putative promoter region of the adjacent protein-coding gene, *T09B4.7* (Supplementary Fig. 3). In fed larval stage (L)1 larvae, migrated P cells subsequently undergo several rounds of cell division, and some of their descendants then differentiate into motor neurons[16] (Supplementary Fig. 4a). Similarly, P cells in some starved *mir-235(n4504)* animals develop further to give rise to motor neurons (Supplementary Fig. 4b). In addition to P cells, starved *mir-235(n4504)* L1 larvae fail to maintain quiescence in M mesoblasts, the reactivation of which occurs around the same time as that of P cells in feeding wild-type larvae (Fig. 1a–c and Supplementary Fig. 2).

In addition to blast-cell reactivation, other postembryonic developmental events such as moulting, expression of *mlt-10* and *lin-42a*—both of which are required for proper moulting[17,18]—and upregulation of the developmental timer, *lin-4* miRNA[13,19], are also induced in

starved *mir-235(n4504)* animals (Fig. 1b–d and Supplementary Fig. 2). We noticed that P-cell migration, M-cell division and moulting apparently occur in that order in starved *mir-235(n4504)* animals, as the event/s later in the sequence never took place without the event/s
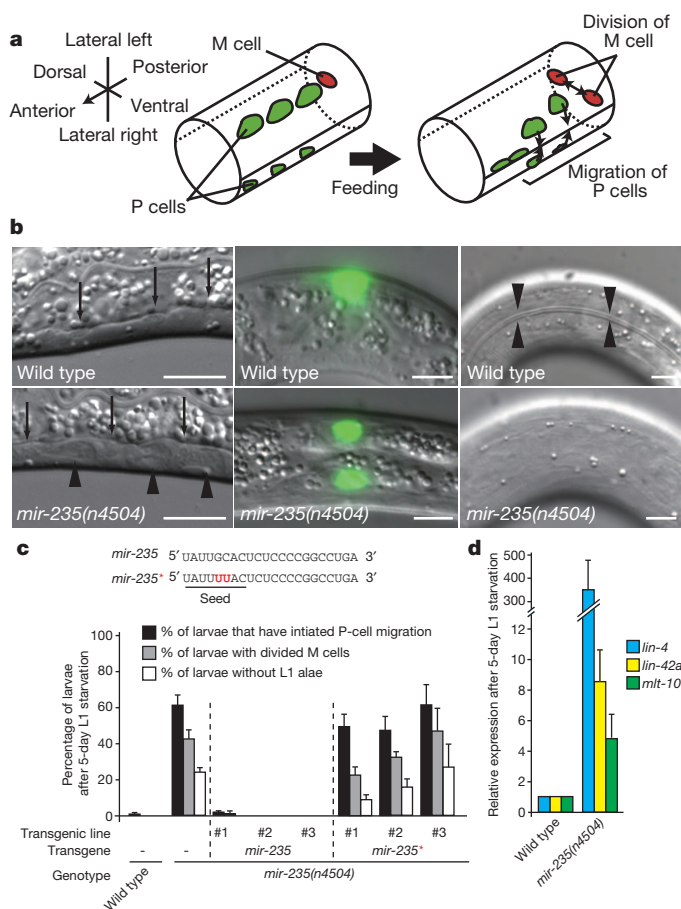

Figure 1 | *mir-235* is required for suppressing blast-cell reactivation and other postembryonic developmental events during starvation-induced L1 diapause. **a**, Early L1 development in P and M blast cells. **b**, Wild-type and *mir-235(n4504)* L1 larvae after 3-day starvation (see Methods). Left, neuronal (arrows) and migrated P cells (arrowheads) in the ventral nerve cord; middle, M cells marked by P*hlh-8::gfp* expressed by *ayIs6* (ref. 29); right, L1-specific alae (arrowheads) disappear after L1 larvae complete moulting. Scale bars, 5 μm. **c**, Rescue experiments. Mutated nucleotides of *mir-235** are shown in red. All strains carried *ayIs6* to visualize M cells. Error bars show standard error of the mean (s.e.m.). **d**, Relative abundance of *lin-4*, *lin-42a* and *mlt-10* transcripts in wild-type and *mir-235(n4504)* mutant larvae was determined by quantitative real-time polymerase chain reaction with reverse transcription (qRT–PCR). Error bars show standard deviation (s.d.). In this and subsequent figures, graphs were created using data from ≥3 independent experiments, and ≥35 animals were scored for each experiment unless otherwise indicated.

[1]Laboratory of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan.

earlier having first occurred. This may reflect concerted regulation of these events. On the other hand, both Z2/Z3 germline stem cells and Z1/Z4 somatic gonadal precursors, which normally initiate proliferation during L1 stage[20], were kept mitotically arrested in starved *mir-235(n4504)* larvae (*n* = 110). These observations suggest that *n4504* impairs a genetic pathway that arrests diverse postembryonic developmental events in a subset of somatic lineages. This contrasts with *cyclin-dependent kinase inhibitor 1 (cki-1)*, which specifically regulates cell-cycle arrest downstream of the IIS pathway during L1 diapause[4,6].

Introduction of the genomic fragment spanning the *mir-235* locus restored the developmental quiescence defects of *mir-235(n4504)* animals (Fig. 1c). By contrast, restoration of the phenotypes was completely blocked when the genomic fragment was mutated in its 'seed sequence' (referred to as *mir-235**; Fig. 1c), which is generally essential for target recognition[11]. Furthermore, *tm4127*, which deletes a significant portion of the coding region of *T09B4.7* (Supplementary Fig. 3), did not cause defects in developmental arrest like *n4504* (*n* = 155). Thus, these findings indicate that *mir-235* activity is essential for maintaining developmental arrest.

To reveal the spatiotemporal pattern of *mir-235* expression, we next generated transgenic worms that expressed green fluorescent protein (GFP) under the control of 2.3 kilobases (kb) of the 5′ upstream region of the miR-235 stem-loop (P*mir-235::gfp*). P*mir-235::gfp* is apparently expressed in the entire hypodermis from late embryogenesis to adult (Supplementary Fig. 5). In addition to the hypodermis, its expression was also seen in glia-like socket cells surrounding amphid neurons during L1 diapause (AmsoL/R; Fig. 2a, b). These observations prompted us to examine in which tissue *mir-235* functioned to maintain the developmental arrest.

Expression of *mir-235* under the control of the hypodermis-specific *dpy-7* promoter restored the defects caused by *mir-235(n4504)* in suppressing reactivation in P and M cells, and moulting (Fig. 2c). The *dpy-7* promoter is active in most hypodermal cells including P neuroblasts, raising the possibility that *mir-235* autonomously regulates P-cell migration. However, all the phenotypes were also suppressed by expressing *mir-235* in the hypodermis excluding P neuroblasts (Supplementary Fig. 6), suggesting that the hypodermal activity of *mir-235* can non-autonomously regulate P and M blast cells. Similarly, expression of *mir-235* driven by the pan-neuronal *rgef-1* promoter as well as the glial enhancer of the *ztf-16* gene, both of which are active in amphid socket cells, can also restore the defects of *mir-235(n4504)* animals (Fig. 2c and Supplementary Fig. 7). By contrast, expression of *mir-235* driven by the *pgp-1* promoter in the intestine, in which expression of P*mir-235::gfp* was not detected, did not affect the defects (Fig. 2c). Thus, these observations support the model that *mir-235* activities in the hypodermis and amphid socket glia redundantly maintain developmental arrest.

To reveal the temporal pattern of *mir-235* expression, we next detected its transcripts by northern blotting (Supplementary Fig. 8). The level of mature miR-235 progressively increased during the period of L1 diapause (Fig. 3a). Conversely, upon feeding, its expression gradually decreased and became almost undetectable after 8 to 10 h of feeding, around which time P and M cells exit from quiescence (Fig. 3b and Supplementary Fig. 9).

Previous studies showed that loss of *daf-16* (*foxo* orthologue), which is negatively regulated by the IIS pathway, results not only in defects in arresting somatic developmental events like those seen with *mir-235* (ref. 6), but also in a significant decrease in the survival rate during L1 diapause[6,21]. However, loss of *mir-235* does not affect the survival rate during diapause (Supplementary Fig. 10), suggesting that *mir-235* does not regulate IIS pathway activity. Conversely, we proposed that the IIS pathway is responsible for downregulation of *mir-235* upon feeding.
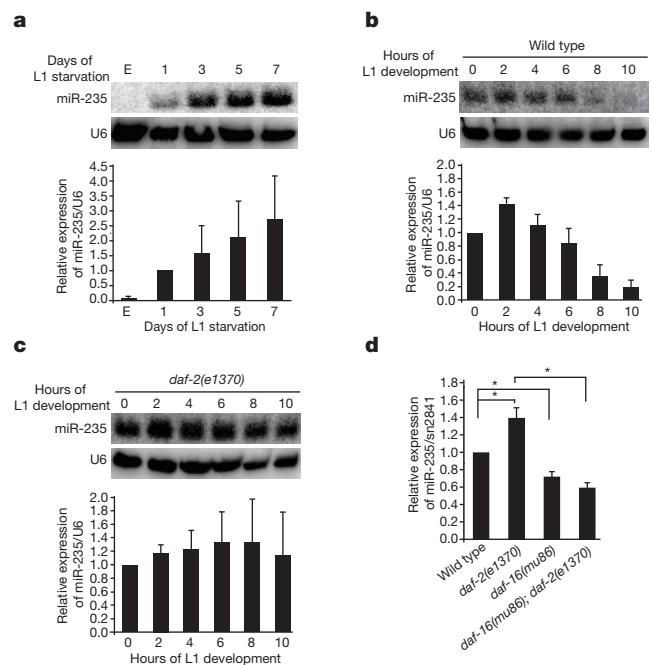

Figure 2 | *mir-235* acts in the hypodermis and glia. a, b, Expression of P*mir-235::gfp* during L1 diapause. **a**, Seam, hyp7 and P cells are the major hypodermal cells that cover most of the body, and are labelled V, hyp7 and P, respectively. **b**, The arrow indicates a cell body, and arrowheads indicate a process of an amphid socket cell (AmsoL/R). Scale bars, 10 μm. **c**, Effects of tissue-specific expression of *mir-235* on the defects of *mir-235* mutants. Data for untransformed animals are the same as Fig. 1c. Error bars show s.e.m.

Figure 3 | Expression of *mir-235* is negatively regulated by feeding via the IIS pathway. a–c, Expression of mature miR-235 in starved (**a**) and synchronously grown (**b**, **c**) L1 larvae of the indicated genotypes was analysed by northern blotting. Synchronized growth was initiated by feeding L1-arrested animals (see Methods). The average amount of *mir-235* transcript was calculated after U6 RNA normalization (*n* = 3). E, embryonic stage. **d**, Relative abundance of *mir-235* transcript in L1 larvae of the indicated genotypes after 24 h L1 starvation was analysed by TaqMan qRT–PCR (*n* = 3). *P < 0.05 (paired two-tailed Student's *t*-test). Error bars show s.d.

Consistent with this idea, reduced activity of *daf-2* (insulin receptor orthologue) impaired the feeding-induced downregulation (Fig. 3c and Supplementary Discussion). Furthermore, even during L1 diapause, the levels of mature *mir-235* transcript were further increased in *daf-2(lf)* mutants compared with wild-type animals, and this upregulation was dependent on the activity of *daf-16* (Fig. 3d). Of the 5′ upstream region of the miR-235 stem-loop, 2.3 kb does not possess any consensus DAF-16 binding sites[22,23]. Consistent with this, interference of the IIS pathway does not significantly affect levels of pri-miR-235 (Supplementary Fig. 11). Thus, these findings suggest that the IIS pathway negatively controls expression of *mir-235*, probably at the post-transcriptional level, coupling it to nutrient availability.

We next sought to identify a target of miR-235 that has a key role in the release of developmental quiescence. Among the miR-235 target genes predicted by multiple algorisms such as TargetScan and miRanda, we focused our analysis on the *nhr-91* gene[24], because it is among several genes whose mRNAs possess more than one miR-235 complementary site (Fig. 4a) and its mammalian orthologue, germ cell nuclear

factor (GCNF) has been shown to promote differentiation of neural stem cells in mice[25,26]. The level of *nhr-91* transcripts was increased in starved *mir-235(n4504)* L1 larvae, compared with wild type (Fig. 4b). In fed wild-type animals, significant upregulation of *nhr-91* transcripts temporally coincides with downregulation of mature miR-235 after 8 and 10 h of feeding (Fig. 3b and Supplementary Fig. 12a). Similarly, a reporter gene, in which the *gfp-pest* complimentary DNA was inserted between the promoter and the 3′ UTR of *nhr-91* (P*nhr-91::gfp-pest::nhr-91* 3′ UTR), also showed a burst of expression in the hypodermis, where P*mir-235::gfp* was also expressed, after 8 to 10 h of feeding (Supplementary Fig. 12b). The level of *nhr-91* transcripts was further increased in *mir-235* mutants compared with wild-type animals after 8 h of feeding (Supplementary Fig. 12a), suggesting that *mir-235* functions not only to suppress expression of *nhr-91* during L1 diapause, but also to limit the kinetics of its upregulation when feeding.

We also assessed the role of miR-235 complementary sequences in the *nhr-91* 3′ UTR, by comparing the effects of wild-type and mutated *nhr-91* 3′ UTRs on the expression of upstream-fused *gfp-pest* (Fig. 4a, c–e). These *gfp-pest*, *mir-235* and *mCherry* genes, the last of which facilitates detection of transformed animals, were simultaneously overexpressed by the hypodermal-specific *dpy-7* promoter (Fig. 4c). Transgenic animals that carried the mutated *nhr-91* 3′ UTR more frequently showed de-repression of GFP expression than those harbouring the wild-type one (Fig. 4d, e). Finally, although loss of *nhr-91* does not result in constitutive L1 arrest, delay in P-cell migration or any noticeable developmental deficiency (data not shown), it significantly suppressed the defects caused by *mir-235(n4504)* during L1 diapause (Fig. 4f, g and Supplementary Discussion). Altogether, these findings revealed that *nhr-91* is one of the key targets of miR-235 in the regulation of developmental quiescence.

We have demonstrated that *mir-235* has an essential role in arresting diverse postembryonic developmental events including blast-cell reactivation, activation of the heterochronic pathway and moulting cycle, and that its expression is downregulated by feeding via the IIS pathway. In addition to *C. elegans*, there are several examples of stem- and blast-cell behaviours being coupled to nutritional state in mammals and *Drosophila*[2,3,27]. Because orthologues of miR-235 are found among many animal species, it would be of great interest to examine whether these miRNAs are also involved in nutritional regulation of stem and blast cells.
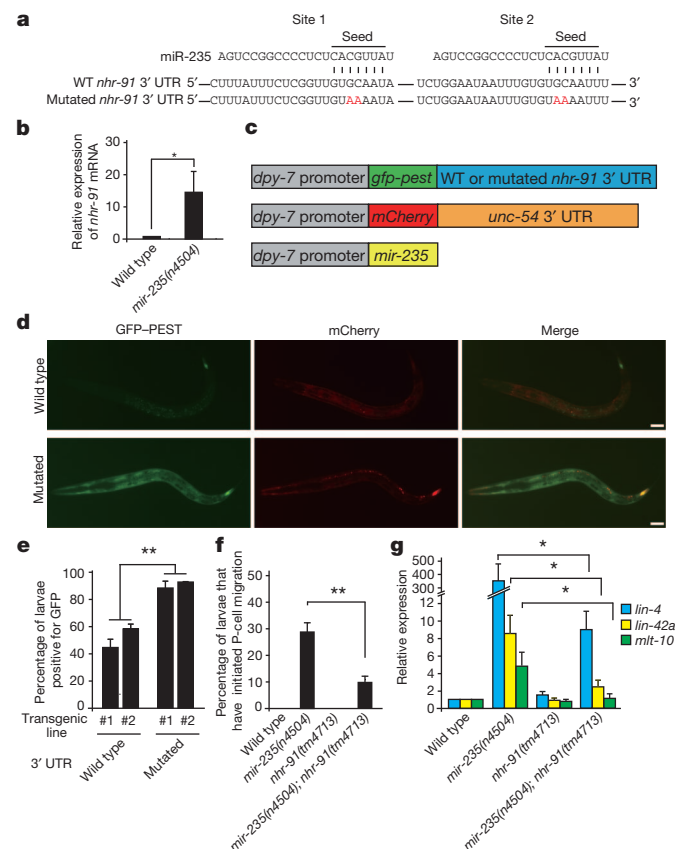

Figure 4 | *nhr-91* mRNA is a target of miR-235. **a**, Predicted miR-235 binding sites on the 3′ UTR of the *nhr-91* mRNA. Mutated nucleotides for experiments shown in **d** and **e** are indicted in red. WT, wild type. **b**, Relative abundance of *nhr-91* transcripts in 5-day L1 starved animals determined by qRT–PCR. *$P < 0.05$ (paired two-tailed Student's *t*-test; $n = 3$). **c**, Indicated constructs were transformed together as multi-copy extrachromosomal arrays for experiments shown in **d** and **e**. **d**, Expression of the reporter transgenes. Scale bars, 10 μm. **e**, Fractions of mCherry-positive animals that show detectable GFP–PEST expression. ≥75 animals were scored for each experiment. **$P < 0.001$ (Scheffé's method; $n \geq 3$). **d**, **e**, L1 larvae synchronously fed for 12 h at 25.5 °C were used. **f**, **g**, *nhr-91(tm4713)* suppresses defects of *mir-235* mutants. Data were collected after 5-day L1 starvation. **$P < 0.001$ (unpaired two-tailed Student's *t*-test; $n = 10$ for *mir-235(n4504)*, $n = 9$ for *mir-235(n4504);nhr-91(tm4713)*). **g**, Relative amounts of indicated transcripts were determined by qRT–PCR. Data for wild type and *mir-235(n4504)* in **g** are the same as Fig. 1d. *$P < 0.05$ (paired two-tailed Student's *t*-test; $n = 3$). Error bars show s.d. (**b**, **g**) and s.e.m. (**e**, **f**).

METHODS SUMMARY

***C. elegans* culture.** Sterile L1 animals were prepared for starvation by bleaching gravid adults as previously described[21]. One-hundred thousand embryos were hatched in a 15 ml polypropyrene tube containing 10 ml of complete S medium[28] (10 μl of cholesterol (5 mg ml⁻¹ in EtOH) per 10 ml of complete S medium were added before use) with rotation using a rotator, ROTAMIX (ATR), at approximately 30 r.p.m. at 20 °C, and hatched L1 larvae were continuously cultured in the same tube. L1 larvae 'after 3-day L1 starvation' (Fig. 1b) indicates larvae that were allowed to hatch and cultured in nutritionally poor, complete S medium for 3 days after the alkali/bleach treatment. To prepare synchronously grown L1 larvae (Figs 3b, c, 4d, e and Supplementary Figs 9, 12), sterile L1-arrested animals cultured in M9 buffer at 20 °C for 24 h after the bleach treatment were washed in M9 buffer three times, transferred onto 10 cm 4× peptone NGM plates seeded with *Escherichia coli* OP50 strain at a concentration of 100,000 larvae per plate (considered to be 0 hour of L1 development), allowed to grow at 25.5 °C, and subsequently harvested at the indicated times.

Other methods are described in Methods and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

1. Jasper, H. & Jones, D. L. Metabolic regulation of stem cell behavior and implications for aging. *Cell Metab.* **12,** 561–565 (2010).
2. Ables, E. T. & Drummond-Barbosa, D. Food for thought: neural stem cells on a diet. *Cell Stem Cell* **8,** 352–354 (2011).

3. Nakada, D., Levi, B. P. & Morrison, S. J. Integrating physiological regulation with stem cell and tissue homeostasis. *Neuron* **70**, 703–718 (2011).

4. Hong, Y., Roy, R. & Ambros, V. Developmental regulation of a cyclin-dependent kinase inhibitor controls postembryonic cell cycle progression in *Caenorhabditis elegans*. *Development* **125**, 3585–3597 (1998).

5. Fukuyama, M., Rougvie, A. E. & Rothman, J. H. *C. elegans* DAF-18/PTEN mediates nutrient-dependent arrest of cell cycle and growth in the germline. *Curr. Biol.* **16**, 773–779 (2006).

6. Baugh, L. R. & Sternberg, P. W. DAF-16/FOXO regulates transcription of *cki-1*/Cip/Kip and repression of *lin-4* during *C. elegans* L1 arrest. *Curr. Biol.* **16**, 780–785 (2006).

7. Gems, D. *et al.* Two pleiotropic classes of *daf-2* mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. *Genetics* **150**, 129–155 (1998).

8. Kao, G. *et al.* ASNA-1 positively regulates insulin secretion in *C. elegans* and mammalian cells. *Cell* **128**, 577–587 (2007).

9. Olive, V., Jiang, I. & He, L. *mir-17-92*, a cluster of miRNAs in the midst of the cancer network. *Int. J. Biochem. Cell Biol.* **42**, 1348–1354 (2010).

10. Ibáñez-Ventoso, C., Vora, M. & Driscoll, M. Sequence relationships among *C. elegans*, *D. melanogaster* and human microRNAs highlight the extensive conservation of microRNAs in biology. *PLoS ONE* **3**, e2818 (2008).

11. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).

12. Yi, R., Poy, M. N., Stoffel, M. & Fuchs, E. A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* **452**, 225–229 (2008).

13. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).

14. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).

15. Miska, E. A. *et al.* Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, 2395–2403 (2007).

16. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).

17. Meli, V. S., Osuna, B., Ruvkun, G. & Frand, A. R. MLT-10 defines a family of DUF644 and proline-rich repeat proteins involved in the molting cycle of *Caenorhabditis elegans*. *Mol. Biol. Cell* **21**, 1648–1661 (2010).

18. Monsalve, G. C., Van Buskirk, C. & Frand, A. R. LIN-42/PERIOD controls cyclical and developmental progression of *C. elegans* molts. *Curr. Biol.* **21**, 2033–2045 (2011).

19. Feinbaum, R. & Ambros, V. The timing of *lin-4* RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. *Dev. Biol.* **210**, 87–95 (1999).

20. Kimble, J. & Hirsh, D. The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol.* **70**, 396–417 (1979).

21. Fukuyama, M. *et al. C. elegans* AMPKs promote survival and arrest germline development during nutrient stress. *Biol. Open* **1**, 929–936 (2012).

22. Furuyama, T. *et al.* Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).

23. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).

24. Gissendanner, C. R., Crossgrove, K., Kraus, K. A., Maina, C. V. & Sluder, A. E. Expression and function of conserved nuclear receptor genes in *Caenorhabditis elegans*. *Dev. Biol.* **266**, 399–416 (2004).

25. Fuhrmann, G. *et al.* Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Dev. Cell* **1**, 377–387 (2001).

26. Akamatsu, W., DeVeale, B., Okano, H., Cooney, A. J. & van der Kooy, D. Suppression of Oct4 by germ cell nuclear factor restricts pluripotency and promotes neural stem cell development in the early neural lineage. *J. Neurosci.* **29**, 2113–2124 (2009).

27. Yilmaz, O. H. *et al.* mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).

28. Lewis, J. A. & Fleming, J. T. Basic culture methods. *Methods Cell Biol.* **48**, 3–29 (1995).

29. Harfe, B. D. *et al.* Analysis of a *Caenorhabditis elegans* Twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev.* **12**, 2623–2635 (1998).

Author Contributions H.K., M.F. and A.K. designed and performed the experiments; all authors analysed and interpreted the data; M.F., K.K. and T.K. supervised the project; H.K. and M.F. wrote the manuscript with comments from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.F. (mfukuyam@mol.f.u-tokyo.ac.jp).

METHODS

C. elegans culture. Sterile L1 animals were prepared for starvation by bleaching gravid adults as previously described[21]. One-hundred thousand embryos were hatched in a 15 ml polypropyrene tube containing 10 ml of complete S medium[28] (10 μl of cholesterol (5 mg ml^{-1} in EtOH) per 10 ml of complete S medium were added before use) with rotation using a rotator, ROTAMIX (ATR), at approximately 30 r.p.m. at 20 °C, and hatched L1 larvae were continuously cultured in the same tube. L1 larvae 'after 3-day L1 starvation' (Fig. 1b) indicates larvae that were allowed to hatch and cultured in nutritionally poor, complete S medium for 3 days after the alkali/bleach treatment. To prepare synchronously grown L1 larvae (Figs 3b, c, 4d, e and Supplementary Figs 9, 12), sterile L1-arrested animals cultured in M9 buffer at 20 °C for 24 h after the bleach treatment were washed in M9 buffer three times, transferred onto 10 cm 4× peptone NGM plates seeded with *E. coli* OP50 strain at a concentration of 100,000 larvae per plate (considered to be 0 hour of L1 development), allowed to grow at 25.5 °C, and subsequently harvested at the indicated times.

Strains. The strains used in this study are as follows: wild-type Bristol N2, PD4666 *ayIs6[Phlh-8::gfp]*, MT17997 *mir-235(n4504)*, CB1370 *daf-2(e1370)*, CF1038 *daf-16(mu86)*, YB2072 *nhr-91(tm4713)*, NC138 *dpy-20(e1282); wdIs3*, NC197 *wdIs4; dpy-20(e1282)*, YB2148, YB2149, YB2150 *mir-235(n4504); ayIs6; tdEx[mir-235 genomic fragment + Ppgp-1::mcherry::unc-54 3′ UTR + pRF4]*, YB2255, YB2256, YB2257 *mir-235(n4504); ayIs6; tdEx[mir-235 mutated genomic fragment + Ppgp-1::mcherry::unc-54 3′ UTR + pRF4]*, YB2055, YB2056, YB2057 *mir-235(n4504); ayIs6; tdEx[Pdpy-7::mir-235 + Ppgp-1::mcherry::unc-54 3′ UTR + pRF4]*, YB1777, YB1778, YB1782 *mir-235(n4504); ayIs6; tdEx[Prgef-1::mir-235 + Prgf-1::mcherry::unc-54 3′ UTR + pRF4]*, YB2061, YB2062 *mir-235(n4504); ayIs6; tdEx[Ppgp-1::mir-235 + Ppgp-1::mcherry::unc-54 3′ UTR + pRF4]*, YB2414 *tdIs43[Pmir-235::gfp::unc-54 3′ UTR + Ppgp-1::mcherry::unc-54 3′ UTR + pDP#MM51]*, YB2311 *tdIs38[Pnhr-91::gfp-pest::nhr-91 3′ UTR + Ppgp-1::mcherry::unc-54 3′ UTR + pDP#MM51]*, YB2242, YB2243 *unc-119(ed4); tdEx[Pdpy-7::gfp::nhr-91 3′ UTR + Pdpy-7::mcherrry::unc-54 3′ UTR + Pdpy-7::mir-235 + pDP#MM51]*, YB2247, YB2248 *unc-119(ed4); tdEx[Pdpy-7::gfp::nhr-91 mutated 3′ UTR + Pdpy-7::mcherrry::unc-54 3′ UTR + Pdpy-7::mir-235 + pDP#MM51]*, YB2474, YB2475 *mir-235(n4504); ayIs6; tdEx1651, tdEx1652[Pztf-16::mir-235 + Pztf-16::gfp::unc-54 3′ UTR + pRF4]*, YB2513, YB2514, YB2515 *mir-235(n4504); ayIs6; tdEx1640, tdEx1641, tdEx1642[Pdpy-7::QF::SL2::mcherry::unc-54 3′ UTR + Phlh-3-enhancer::delta-pes-10::QS::SL2::mcherry::unc-54 3′ UTR + QUAS::mir-235 + QUAS::gfp::unc-54 3′ UTR + pRF4]*.

Construction of plasmids is described in Supplementary Information.

Measurement of survival rate during L1 diapause. To determine survival rates during L1 diapause, 100,000 embryos were hatched and subsequently starved in 10 ml M9 buffer in a 15 ml tube at 20 °C with rotation as described above. Viability was scored as previously described[21].

Total RNA isolation. Worms were harvested and washed with M9 buffer several times to remove *E. coli*. One millilitre of RNAiso plus (TAKARA) was added and vortexed with acid washed beads (SIGMA; diameters 425–600 μm) for 15 min. After extraction using chloroform and acid phenol (pH 5.0)/chloroform (1:1) solution, total RNA was precipitated, adding an equal volume of isopropanol, and was subsequently washed with 70% of ethanol. RNA was finally dissolved in RNase-free water and the absorbance was measured to certify its quality.

Northern blotting. An appropriate amount of total RNA was mixed with an equal volume of 2× formamide loading dye (98% of deionized formamide, 10 mM of EDTA pH 8.0, 0.025% of xylene cyanol, 0.025% of bromophenol blue) and denatured by heating at 65 °C for 15 min. Total RNA was separated with 15% of denaturing polyacrylamide (urea/TBE) gels with 0.5× TBE buffer, transferred to Hybond-N+ (GE) by the semi-dry method (300 mA, 3 h), and then fixed to membranes by ultraviolet exposure (120 mJ cm^{-2}). These membranes were prehybridized by prehybridization/hybridization solution (5× SSC, 20 mM Na$_2$HPO$_4$ pH 7.0, 7% SDS, 3× Denhardt's solution, denatured salmon sperm DNA (1 μg ml^{-1}) at 42 °C for 30 min. Subsequently, probes were added and hybridized at 42 °C overnight. miR-235-specific StarFire probe (IDT) was used as the probe for mature miR-235. The probe for U6 (TGTCATCCTTGCGCAGG) was prepared by 5′-end labelling using T4PNK (TAKARA). The hybridized membrane was washed twice by 2× SSC, 0.1% SDS, and once by 0.2× SSC, 0.1% SDS at 42 °C for 15 min for each. The hybridization signals were analysed using Typhoon 9200 (GE), and their intensities were measured using ImageJ.

qRT–PCR analysis. cDNA was synthesized from 1 μg of total RNA with anchored oligonucleotide (dT) and random hexamer primers using Transcriptor High Fidelity cDNA Synthesis kit (Roche) or ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO), and RT–PCR was performed with THUNDERBIRD qPCR Mix (TOYOBO) using StepOne Plus (Applied Biosystems) according to the manufacturer's protocols. Primers used for RT–PCR: pri-miR-235, AGAGTACC CGGAGAGAAACAAATG, TGTCATCTGAAGAAAGGACACACAT; *nhr-91*, GT CAACCGGTCTACATTACGG, ACCGCTTGAAGAATCCTTTG; *lin-42a*, TGG ATATTTTAAGCTACTTGTACGATCT, GTTCTGATCCTTGAGGCTTGAGT; *mlt-10*, TACACCGGCCGAGTTTGT, AGACGAGAACAATGAAATCGC; *act-1*, TGGG ACAGAAGGACTCGTACGT, CCGTGCTCAATTGGGTACTTG. For quantification of small RNAs, TaqMan microRNA assay (Applied Biosystems) was used for miR-235 and lin-4 as described by the manufacturer. *act-1* and sn2841 were used as internal controls. Relative abundance was determined by the relative standard curve method.

LETTER

Lamin A/C and emerin regulate MKL1–SRF activity by modulating actin dynamics

Chin Yee Ho[1,2], Diana E. Jaalouk[2]†, Maria K. Vartiainen[3] & Jan Lammerding[1,2]

Laminopathies, caused by mutations in the *LMNA* gene encoding the nuclear envelope proteins lamins A and C, represent a diverse group of diseases that include Emery–Dreifuss muscular dystrophy (EDMD), dilated cardiomyopathy (DCM), limb-girdle muscular dystrophy, and Hutchison–Gilford progeria syndrome[1]. Most *LMNA* mutations affect skeletal and cardiac muscle by mechanisms that remain incompletely understood. Loss of structural function and altered interaction of mutant lamins with (tissue-specific) transcription factors have been proposed to explain the tissue-specific phenotypes[1]. Here we report in mice that lamin-A/C-deficient (*Lmna*[−/−]) and *Lmna*[N195K/N195K] mutant cells have impaired nuclear translocation and downstream signalling of the mechanosensitive transcription factor megakaryoblastic leukaemia 1 (MKL1), a myocardin family member that is pivotal in cardiac development and function[2]. Altered nucleo-cytoplasmic shuttling of MKL1 was caused by altered actin dynamics in *Lmna*[−/−] and *Lmna*[N195K/N195K] mutant cells. Ectopic expression of the nuclear envelope protein emerin, which is mislocalized in *Lmna* mutant cells and also linked to EDMD and DCM, restored MKL1 nuclear translocation and rescued actin dynamics in mutant cells. These findings present a novel mechanism that could provide insight into the disease aetiology for the cardiac phenotype in many laminopathies, whereby lamin A/C and emerin regulate gene expression through modulation of nuclear and cytoskeletal actin polymerization.

MKL1 (also known as MAL or MRTF-A) is a mechanosensitive transcription factor with important roles in the cardiovascular system[2,3]. Intracellular localization of MKL1 is regulated via changes in actin polymerization[4,5]. Normally, MKL1 is localized in the cytoplasm by binding to cytoplasmic G actin and constitutive nuclear export. Mitogenic or mechanical stimulation triggers RhoA-mediated actin polymerization, liberating MKL1 from G actin and exposing a nuclear localization sequence (NLS) within the actin-binding domain of MKL1 (refs 6, 7). Increased nuclear import, coupled with decreased export, causes accumulation of MKL1 in the nucleus, where it co-activates serum response factor (SRF) to turn on genes regulating cellular motility and contractility, including vinculin, actin and SRF itself[8]. Because cells from lamin-A/C-deficient mice have impaired activation of mechanosensitive genes *in vitro*[9] and *in vivo*[10], we investigated whether loss of lamin A/C could affect MKL1–SRF signalling. Nuclear translocation of endogenous MKL1 in response to serum stimulation was severely abrogated in *Lmna*[−/−] mouse embryonic fibroblasts (MEFs) compared to wild-type controls (Fig. 1a, c and Supplementary Fig. 1a). We confirmed these findings by time-lapse microscopy of cells expressing MKL1–green fluorescent protein (GFP) (Figs 1b and 2a) and in lamin-A/C-downregulated HeLa cells (Supplementary Fig. 2a, b), indicating that impaired MKL1 translocation is a general effect of loss of lamin A/C. To test whether similar defects could also result from lamin mutations associated with DCM, we investigated cells from the *Lmna*[N195K/N195K] mouse model (subsequently referred to as *Lmna* N195K cells), which develops severe DCM but lacks skeletal muscle involvement[11]. *Lmna* N195K MEFs

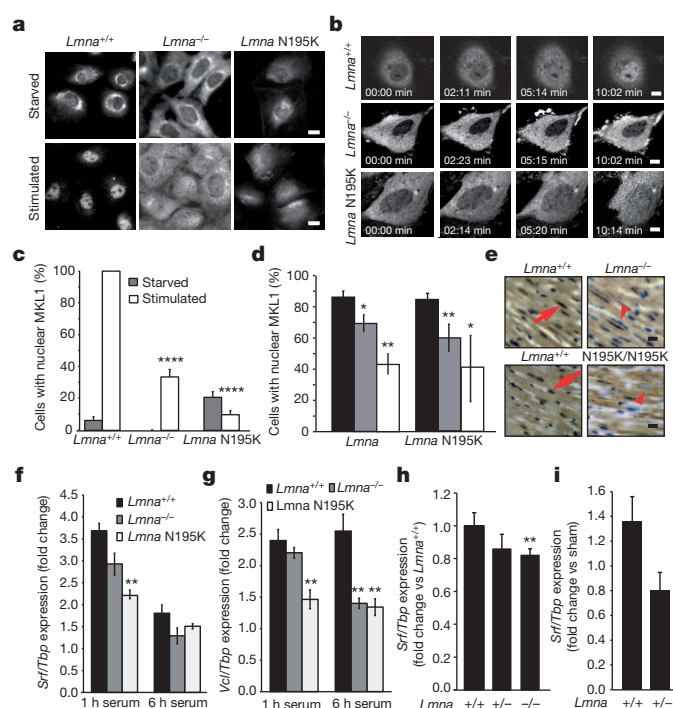**Figure 1 | Impaired nuclear translocation of MKL1 in lamin-A/C-deficient and *Lmna* N195K mutant cells. a**, *Lmna*[−/−] and *Lmna* N195K MEFs had a lower fraction of nuclear MKL1 after serum stimulation than *Lmna*[+/+] cells, on the basis of MKL1 immunofluorescence. Scale bars, 10 μm. **b**, Time-lapse sequences of cells expressing MKL1–GFP stimulated with serum (see Supplementary Videos 1–3). Scale bars, 10 μm. **c**, Quantitative analysis of MEFs with positive nuclear MKL1 staining in response to serum stimulation (*n* ≈ 50 per cell line). **d**, Quantitative analysis of myocytes with nuclear MKL1 in cardiac sections from *Lmna*[−/−] and *Lmna*[N195K/N195K] mice as well as littermate controls (*n* = 3 for each). **e**, Representative histological cardiac tissue sections from *Lmna*[−/−] and *Lmna*[N195K/N195K] mice and age-matched wild-type littermates stained for MKL1 (brown). Red arrows denote example of MKL1-positive nucleus; arrowhead denotes an MKL1-negative nucleus. Scale bars, 20 μm. **f, g**, Gene expression of serum response factor (*Srf*) and vinculin (*Vcl*) in *Lmna*[+/+], *Lmna*[−/−] and *Lmna* N195K MEFs after 1 h and 6 h of serum stimulation. Values were based on three independent experiments and were normalized to TATA binding protein (*Tbp*). **h**, Gene expression of *Srf* in *Lmna*[+/+] (*n* = 9), *Lmna*[+/−] (*n* = 11) and *Lmna*[−/−] (*n* = 10) cardiac tissue. Values were normalized to *Tbp*. **i**, Gene expression of *Srf* in *Lmna*[+/+] (*n* = 5) and *Lmna*[+/−] (*n* = 7) cardiac tissue collected 1 week after transverse aortic constriction (TAC) surgery. Values were normalized to *Tbp* and compared to those from sham mice. Statistical significance determined by Student's *t*-test, compared to *Lmna*[+/+] MEFs; *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001. Error bars indicate s.e.m.

[1]Cornell University, Weill Institute for Cell and Molecular Biology/Department of Biomedical Engineering, Ithaca, New York 14853, USA. [2]Brigham and Women's Hospital/Harvard Medical School, Department of Medicine, Boston 02115, Massachusetts, USA. [3]Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland. †Present address: American University of Beirut, Department of Biology, Beirut 1107 2020, Lebanon.
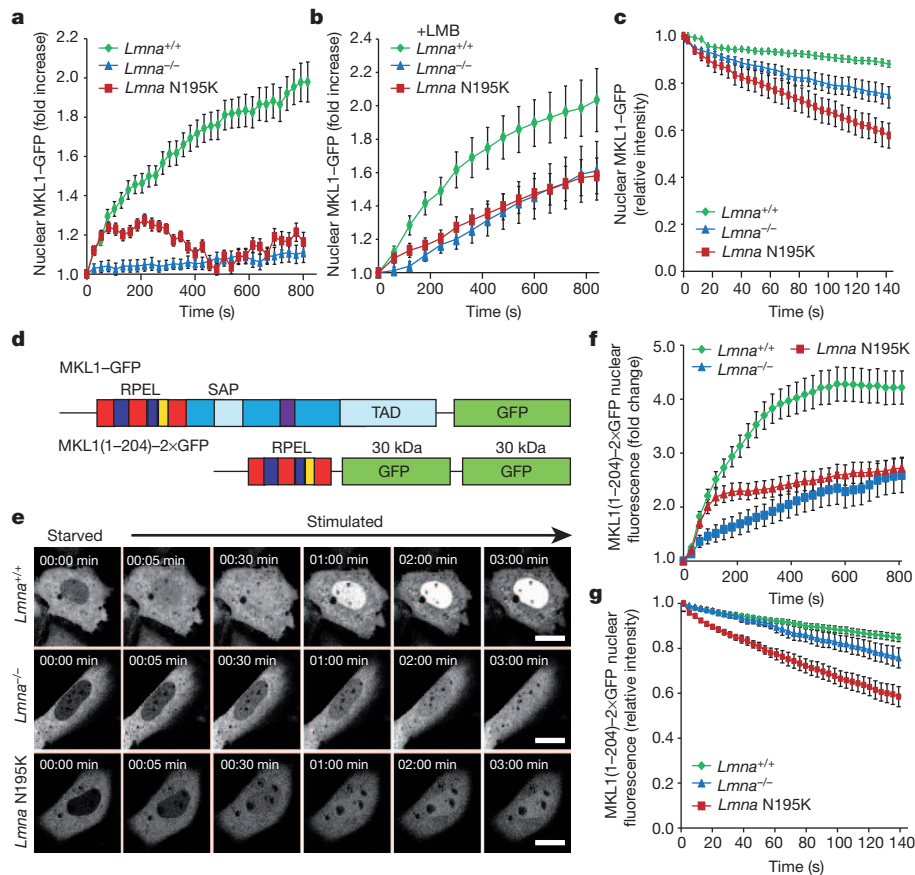
Figure 2 | Changes in nuclear import and export are specific to MKL1 and are caused by altered actin dynamics in *Lmna*−/− and *Lmna* N195K cells. **a**, **b**, Change in nuclear fluorescence intensity over time upon serum stimulation in *Lmna*+/+, *Lmna*−/− and *Lmna* N195K MEFs expressing MKL1–GFP in the absence (**a**) or presence (**b**) of leptomycin B. Values were normalized to the initial nuclear fluorescence intensity before serum addition. *n* = 20 for each cell line. **c**, Fluorescence loss in photobleaching (FLIP) experiments of MKL1–GFP to measure nuclear export. Increased loss of nuclear fluorescence indicates a higher rate of nuclear export of MKL1–GFP in lamin mutant cells. *n* = 10 for each cell line. **d**, Schematic representation (not drawn to scale) of full-length MKL1–GFP (top) and MKL1(1–204)–2×GFP (bottom), consisting of the N-terminal actin-binding domain of MKL1 fused to two GFP moieties. RPEL motifs are depicted in blue and red; NLS in yellow;

DNA-binding domain (SAP) and transcriptional activation domain (TAD) in light blue; coiled-coil domain in purple; other parts of the C terminus in dark blue. **e**, Representative frames from time-lapse series of *Lmna*+/+, *Lmna*−/− and *Lmna* N195K MEFs expressing MKL1(1–204)–2×GFP after serum stimulation (see Supplementary Videos 4–6). Scale bars, 10 μm. **f**, *Lmna*+/+ MEFs showed rapid accumulation of MKL1(1–204)–2×GFP in the nucleus upon serum stimulation, whereas nuclear accumulation was slower in *Lmna*−/− and *Lmna* N195K cells. Nuclear fluorescence intensity was normalized to the initial nuclear fluorescence before serum stimulation. *n* = 60 for each cell line. **g**, FLIP experiments in cells expressing MKL1(1–204)–2×GFP. Fluorescence intensity values were normalized to the initial nuclear fluorescence intensity before bleaching of a cytoplasmic region. *n* = 10 for each cell line. Error bars indicate s.e.m.

(Fig. 1a–c and Supplementary Fig. 1a) and bone-marrow-derived mesenchymal stem cells (Supplementary Fig. 1b) had impaired nuclear translocation of MKL1. Notably, cardiac sections from *Lmna*−/− and *Lmna*^N195K/N195K mice had significantly reduced fractions of cardiomyocytes with nuclear MKL1 (Fig. 1d, e), confirming MKL1 translocation defects *in vivo* and implicating altered MKL1 signalling in the development of cardiomyopathies in these animals.

To characterize the consequences of altered MKL1 translocation, we assessed expression of select MKL1–SRF target genes. *Lmna*−/− and *Lmna* N195K MEFs had impaired serum-induced expression of *Srf* and vinculin (Fig. 1f, g) and had fewer focal adhesions than wild-type controls (Supplementary Fig. 3c, d); expression of an SRF-dependent luciferase reporter was also significantly reduced (Supplementary Fig. 3e). Cardiac tissues from *Lmna*−/− mice had lower *Srf* and actin transcript levels than those from wild-type littermates, and activation of *Srf* expression in response to left ventricular pressure overload was impaired in *Lmna*+/− mice (Fig. 1h, i and Supplementary Fig. 3a, b), demonstrating altered MKL1–SRF mechanosignalling *in vivo*.

Experiments with an NLS–GFP–NES reporter construct consisting of GFP fused to an NLS and a nuclear export sequence (NES) revealed that general nuclear import and export were preserved in *Lmna*−/−

and *Lmna* N195K cells (Supplementary Fig. 4), as were levels and localization of the nuclear transport factor Ran and its regulator, RCC1 (Supplementary Fig. 2c–f). We then devised experiments to assess independently nuclear import and export of MKL1. Nuclear import was measured by monitoring nuclear accumulation of MKL1–GFP while blocking nuclear export with leptomycin B[8]. *Lmna*−/− and *Lmna* N195K mutant cells had significantly reduced nuclear import of MKL1 in response to serum stimulation than wild-type controls (Fig. 2b), which we confirmed with photoactivatable MKL1–PAGFP (Supplementary Fig. 5; see also Supplementary Videos 7–9). Fluorescence loss in photobleaching (FLIP) experiments revealed that lamin mutant MEFs had a significantly faster decrease in nuclear MKL1–GFP (Fig. 2c) than wild-type cells, indicating increased nuclear export of MKL1 in *Lmna*−/− and mutant MEFs.

Nuclear import and export of MKL1 are regulated by actin polymerization[8], requiring interaction between MKL1 and G actin via three amino-terminal RPEL motifs[12]. We expressed a truncated MKL1 construct, MKL1(1–204)–2×GFP, which contains the RPEL motifs but lacks the transcriptional domains (Fig. 2d) and recapitulates the actin-binding characteristics and serum-inducible translocation of full-length MKL1 (ref. 12). Nuclear accumulation of MKL1(1–204)–2×GFP was substantially

lower in $Lmna^{-/-}$ and $Lmna$ N195K cells than in wild-type cells (Fig. 2e, f), indicating that impaired nuclear translocation of MKL1 was caused by altered actin dynamics in the lamin mutant cells. As seen with full-length MKL1–GFP, FLIP studies showed that nuclear export of MKL1(1–204)–2×GFP was significantly increased in the mutant cells (Fig. 2g). In contrast, abrogation of G-actin binding by mutating all three RPEL motifs (MKL1(1–204)XXX–2×GFP) or disrupting the interaction between G actin and MKL1 with cytochalasin D resulted in nuclear accumulation of MKL1(1–204)XXX–2×GFP (Supplementary Fig. 6a) and endogenous MKL1 (Supplementary Fig. 6b), respectively, in all cell types, indicating that MKL1 can enter the nucleus of $Lmna^{-/-}$ and $Lmna$ N195K cells when decoupled from actin dynamics.

We subsequently compared actin organization between mutant and wild-type cells. Fluorescence recovery after photobleaching (FRAP) revealed that nuclear actin, which modulates nuclear export of MKL1 (ref. 8), was more mobile in $Lmna^{-/-}$ cells than in wild-type controls (Supplementary Fig. 7). $Lmna^{-/-}$ and $Lmna$ N195K MEFs also had a larger fraction of highly mobile cytoplasmic actin (Fig. 3a). Furthermore, $Lmna^{-/-}$ and $Lmna$ N195K cells were slower to reassemble stress fibres after disruption of actin filaments with cytochalasin D (Fig. 3b, c). In addition, whereas wild-type cells increased their ratio of F actin to G actin upon serum stimulation, $Lmna^{-/-}$ and $Lmna$ N195K MEFs had a consistently weaker response (Fig. 3d). These findings indicate that actin polymerization is altered in $Lmna^{-/-}$ and $Lmna$ N195K cells and that altered actin organization may be responsible for the impaired nuclear translocation of MKL1.
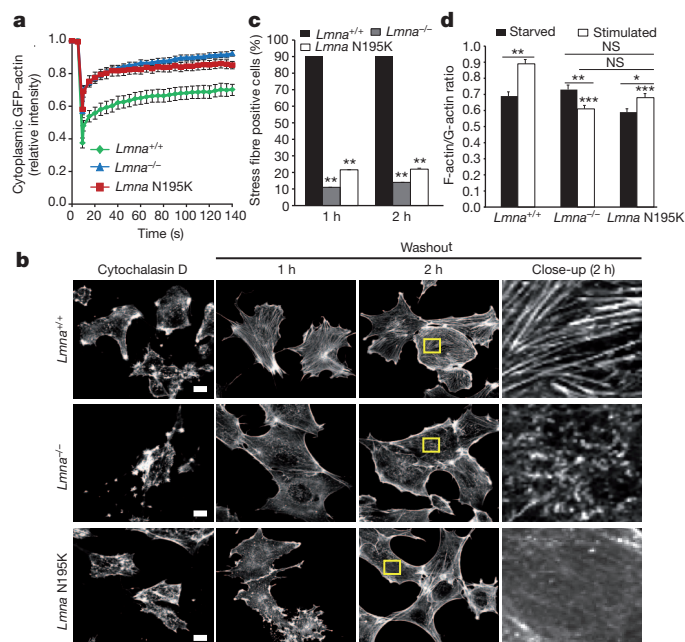
Figure 3 | $Lmna^{-/-}$ and $Lmna$ N195K cells have altered actin dynamics and polymerization kinetics. a, Fluorescence recovery after photobleaching (FRAP) studies with GFP–actin revealed increased cytoplasmic actin mobility in $Lmna^{-/-}$ and $Lmna$ N195K cells relative to $Lmna^{+/+}$ controls. $n = 20$ for each cell line. **b**, Representative images of $Lmna^{+/+}$, $Lmna^{-/-}$ and $Lmna$ N195K MEFs stained for actin stress fibres with phalloidin after cytochalasin D washout. The right column contains close-up images of the regions marked by the yellow rectangle. Scale bars, 10 μm. **c**, $Lmna^{+/+}$ MEFs had a larger fraction of cells containing stress fibres at 1 h and 2 h after cytochalasin D washout than $Lmna^{-/-}$ and $Lmna$ N195K cells. $n = 50$ for each cell line. **d**, Comparison of F-actin/G-actin ratio in starved and serum-stimulated $Lmna^{+/+}$, $Lmna^{-/-}$ and $Lmna$ N195K MEFs based on phalloidin (F actin) and DNase 1 (G actin) staining. Difference in the F-actin/G-actin ratio in serum-starved cells was not statistically significant (NS); $n = 35$ for each cell line. $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$; all comparisons relative to corresponding $Lmna^{+/+}$ cells unless indicated otherwise by horizontal bars. Error bars indicate s.e.m.

What causes disturbed actin organization in lamin mutant cells? Because lamins contribute to nucleo-cytoskeletal coupling and impaired nucleo-cytoskeletal coupling can disturb perinuclear actin organization, we tested whether disrupting nucleo-cytoskeletal coupling with dominant-negative nesprin mutants (consisting of the KASH domain of nesprin) could reproduce defects in MKL1 translocation. However, expression of dominant-negative KASH had no effect on MKL1 localization (Supplementary Fig. 8). Emerin, an inner nuclear membrane protein associated with X-linked EDMD[13], is an actin pointed-end capping protein that promotes actin polymerization *in vitro*[14] and requires lamin A/C for proper localization. In $Lmna^{-/-}$ and $Lmna$ N195K MEFs, emerin was more mobile and mislocalized from the nuclear envelope compared with wild-type cells (Fig. 4a and Supplementary Fig. 9a). Hemizygous emerin-null male mice ($Emd^{-/Y}$) MEFs displayed the same impaired nuclear translocation of MKL1 as lamin mutant cells, which could be rescued by re-introduction of exogenous emerin (Fig. 4b–d). FRAP studies in $Emd^{-/Y}$ MEFs demonstrated that exogenous emerin completely restored actin mobility to levels of wild-type cells (Fig. 4e, f). Ectopic expression of emerin also markedly improved nuclear translocation of MKL1 in $Lmna^{-/-}$ and $Lmna$ N195K cells (Fig. 4d and Supplementary Fig. 9b) by increasing the amount of emerin available at the nuclear envelope. In contrast, expression of emerin mutants unable to bind actin and to promote actin polymerization[14] failed to restore nuclear translocation of MKL1 and caused dominant-negative defects in wild-type cells (Fig. 4d). These data indicate that emerin is a crucial modulator of actin polymerization and that loss of emerin from the nuclear envelope causes disturbed actin dynamics and impaired MKL1 signalling.

Taken together, our data suggest a novel mechanism for nuclear envelope proteins to regulate MKL1–SRF signalling by modulating actin polymerization. We propose that emerin primarily affects nuclear actin polymerization, which controls nuclear export and transcriptional activity of MKL1 (refs 2, 8). Altered MKL1–SRF signalling could then further affect cytoskeletal actin, as MKL1 and SRF are master regulators for numerous cytoskeletal proteins, including actin and actin-binding proteins, consistent with the reduced cytoskeletal stiffness reported in $Lmna^{-/-}$ MEFs[9,15,16]. Given the low levels of emerin at the outer nuclear membrane[17] and the fact that the fraction of emerin at the endoplasmic reticulum increases in $Lmna^{-/-}$ and $Lmna$ N195K cells (Supplementary Fig. 10), it is likely that emerin has only a limited direct effect on cytoplasmic actin polymerization. Nonetheless, we cannot exclude the possibility that emerin (and lamins) may have additional effects on MKL1. For example, direct interaction of lamin A/C with nuclear G actin[18] could further contribute to the altered actin dynamics in $Lmna^{-/-}$ and $Lmna$ N195K cells, as lamins, together with emerin and spectrin IIa, may form a nuclear cortical actin network[14]. Furthermore, emerin can inhibit or reduce the nuclear accumulation of other transcription factors, including β-catenin, Lmo7 and phospho-ERK1/2 (ref. 19).

$Lmna^{-/-}$ and $Lmna^{N195K/N195K}$ mice develop DCM and have defects in cytoskeletal organization and focal adhesions[9,20,21], consistent with impaired MKL1–SRF signalling[22]. Underscoring the crucial role of MKL1–SRF in cardiac function, cardiac-specific deletion of SRF in adult mice results in DCM[23]. Considering the marked similarity of the cardiac phenotype observed in these mice with those in EDMD and DCM patients, we propose that impaired MKL1–SRF signalling and the resulting alterations in cytoskeletal organization may have a pivotal role in the development of cardiac defects and muscle phenotypes in various laminopathies. Surprisingly, although $Emd^{-/Y}$ cells have obvious defects in nuclear stability and mechanotransduction signalling[24,25], emerin-deficient mice—unlike human patients with emerin mutations —lack an overt muscular phenotype[26], indicating additional layers of complexity, species-specific differences, and possible compensation in the $Emd^{-/Y}$ mice. Nonetheless, $Emd^{-/Y}$ animals show delays in muscle repair[26], consistent with a role of MKL1 in satellite cells and skeletal muscle regeneration[27] and providing additional support for the involvement of impaired MKL1 signalling in nuclear envelopathies.
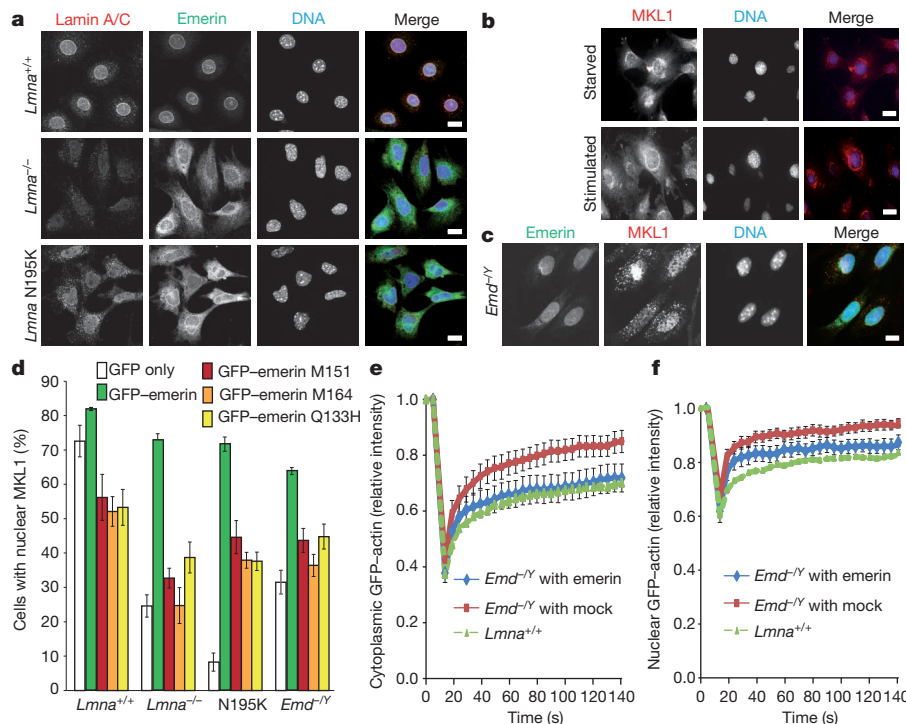
Figure 4 | Emerin expression rescues actin dynamics and restores MKL1 nuclear translocation in $Lmna^{-/-}$ and $Lmna$ N195K cells. **a**, Representative immunofluorescence images showing mislocalization of emerin from the nuclear envelope in $Lmna^{-/-}$ and $Lmna$ N195K MEFs. Scale bar, 10 μm. **b**, $Emd^{-/Y}$ MEFs had the same defects in MKL1 translocation as lamin mutant cells (compare with Fig. 1a). Scale bar, 10 μm. **c**, Stable expression of HA–emerin in $Emd^{-/Y}$ MEFs restored normal nuclear MKL1 localization ($71.1 \pm 6.02\%$) in response to serum stimulation. Scale bar, 10 μm. **d**, Quantification of nuclear MKL1 localization upon serum stimulation in $Lmna^{+/+}$, $Lmna^{-/-}$, $Lmna$ N195K and $Emd^{-/Y}$ MEFs transiently expressing GFP–emerin, emerin mutants that do not bind to actin (GFP–M151,

GFP–M164, GFP–Q133H), or GFP vector alone. Cells were categorized as either having 'nuclear' or 'diffuse/cytoplasmic' localization of MKL1. Expression of GFP–emerin restored serum-induced nuclear localization of MKL1 in $Lmna^{-/-}$, $Lmna$ N195K and $Emd^{-/Y}$ cells. $n = 50$ for each cell line. Statistical significance determined by one-way analysis of variance (ANOVA) ($P \le 0.001$) with Dunnett multiple comparison post test. Each group was compared to $Lmna^{+/+}$ expressing GFP–emerin. **e**, **f**, FRAP analysis of GFP–actin mobility in the cytoplasm (**e**) and in the nucleus (**f**) of $Emd^{-/Y}$ MEFs stably expressing either HA–emerin or a mock control. $n = 10$ for each cell line. $Lmna^{+/+}$ data reproduced from Fig. 3a for comparison. Error bars indicate s.e.m.

Our findings further illustrate the wide-ranging impacts of mutations in nuclear envelope proteins. Treatment of lamin and emerin mutant mice with MAPK inhibitors can reduce cardiac and skeletal phenotypes[28], which may be attributed at least in part to the effect on MKL1 signalling, as inhibiting ERK1/2 activity is expected to increase nuclear localization of MKL1 by reducing its nuclear export[20]. These findings encourage further approaches to correct impaired MKL1–SRF signalling to ameliorate the devastating cardiac disease associated with many laminopathies.

METHODS SUMMARY

Details for standard cell biology techniques (western blotting, immunofluorescence labelling, time-lapse microscopy and photobleaching experiments) can be found in the Methods. Mouse models have been described previously[11,26,29]. All animal work was conducted in accordance with relevant guidelines and regulations.

Full Methods and any associated references are available in the online version of the paper.

1. Ho, C. Y. & Lammerding, J. Lamins at a glance. *J. Cell Sci.* **125,** 2087–2093 (2012).
2. Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nature Rev. Mol. Cell Biol.* **11,** 353–365 (2010).
3. Parmacek, M. S. Myocardin-related transcription factors: critical coactivators regulating cardiovascular development and adaptation. *Circ. Res.* **100,** 633–644 (2007).
4. Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113,** 329–342 (2003).
5. Mouilleron, S., Guettler, S., Langer, C. A., Treisman, R. & McDonald, N. Q. Molecular basis for G-actin binding to RPEL motifs from the serum response factor coactivator MAL. *EMBO J.* **27,** 3198–3208 (2008).
6. Hirano, H. & Matsuura, Y. Sensing actin dynamics: structural basis for G-actin-sensitive nuclear import of MAL. *Biochem. Biophys. Res. Commun.* **414,** 373–378 (2011).
7. Pawłowski, R., Rajakyla, E. K., Vartiainen, M. K. & Treisman, R. An actin-regulated importin α/β-dependent extended bipartite NLS directs nuclear import of MRTF-A. *EMBO J.* **29,** 3448–3458 (2010).
8. Vartiainen, M. K., Guettler, S., Larijani, B. & Treisman, R. Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. *Science* **316,** 1749–1752 (2007).
9. Lammerding, J. *et al.* Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction. *J. Clin. Invest.* **113,** 370–378 10.1172/JCI19670 (2004).
10. Cupesi, M. *et al.* Attenuated hypertrophic response to pressure overload in a lamin A/C haploinsufficiency mouse. *J. Mol. Cell. Cardiol.* **48,** 1290–1297 (2010).
11. Mounkes, L. C., Kozlov, S. V., Rottman, J. N. & Stewart, C. L. Expression of an LMNA-N195K variant of A-type lamins results in cardiac conduction defects and death in mice. *Hum. Mol. Genet.* **14,** 2167–2180 (2005).
12. Guettler, S., Vartiainen, M. K., Miralles, F., Larijani, B. & Treisman, R. RPEL motifs link the serum response factor cofactor MAL but not myocardin to Rho signaling via actin binding. *Mol. Cell. Biol.* **28,** 732–742 (2008).
13. Fairley, E. A., Kendrick-Jones, J. & Ellis, J. A. The Emery-Dreifuss muscular dystrophy phenotype arises from aberrant targeting and binding of emerin at the inner nuclear membrane. *J. Cell Sci.* **112,** 2571–2582 (1999).
14. Holaska, J. M., Kowalski, A. K. & Wilson, K. L. Emerin caps the pointed end of actin filaments: evidence for an actin cortical network at the nuclear inner membrane. *PLoS Biol.* **2,** e231 (2004).
15. Nikolova-Krstevski, V. *et al.* Nesprin-1 and actin contribute to nuclear and cytoskeletal defects in lamin A/C-deficient cardiomyopathy. *J. Mol. Cell. Cardiol.* **50,** 479–486 (2011).
16. Hale, C. M. *et al.* Dysfunctional connections between the nucleus and the actin and microtubule networks in laminopathic models. *Biophys. J.* **95,** 5462–5475 (2008).
17. Salpingidou, G., Smertenko, A., Hausmanowa-Petrucewicz, I., Hussey, P. J. & Hutchison, C. J. A novel role for the nuclear membrane protein emerin in association of the centrosome to the outer nuclear membrane. *J. Cell Biol.* **178,** 897–904 (2007).
18. Simon, D. N., Zastrow, M. S. & Wilson, K. L. Direct actin binding to A- and B-type lamin tails and actin filament bundling by the lamin A tail. *Nucleus* **1,** 264–272 (2010).

19. Wilson, K. L. & Berk, J. M. The nuclear envelope at a glance. *J. Cell Sci.* **123,** 1973–1978 (2010).
20. Muehlich, S. *et al.* Serum-induced phosphorylation of the serum response factor coactivator MKL1 by the extracellular signal-regulated kinase 1/2 pathway inhibits its nuclear localization. *Mol. Cell. Biol.* **28,** 6302–6313 (2008).
21. Nikolova, V. *et al.* Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *J. Clin. Invest.* **113,** 357–369 (2004).
22. Morita, T., Mayanagi, T. & Sobue, K. Reorganization of the actin cytoskeleton via transcriptional regulation of cytoskeletal/focal adhesion genes by myocardin-related transcription factors (MRTFs/MAL/MKLs). *Exp. Cell Res.* **313,** 3432–3445 (2007).
23. Parlakian, A. *et al.* Temporally controlled onset of dilated cardiomyopathy through disruption of the SRF gene in adult heart. *Circulation* **112,** 2930–2939 (2005).
24. Lammerding, J. *et al.* Abnormal nuclear shape and impaired mechanotransduction in emerin-deficient cells. *J. Cell Biol.* **170,** 781–791 (2005).
25. Rowat, A. C., Lammerding, J. & Ipsen, J. H. Mechanical properties of the cell nucleus and the effect of emerin deficiency. *Biophys. J.* **91,** 4649–4664 (2006).
26. Melcon, G. *et al.* Loss of emerin at the nuclear envelope disrupts the Rb1/E2F and MyoD pathways during muscle regeneration. *Hum. Mol. Genet.* **15,** 637–651 (2006).
27. Mokalled, M. H., Johnson, A. N., Creemers, E. E. & Olson, E. N. MASTR directs MyoD-dependent satellite cell differentiation during skeletal muscle regeneration. *Genes Dev.* **26,** 190–202 (2012).
28. Muchir, A., Shan, J., Bonne, G., Lehnart, S. E. & Worman, H. J. Inhibition of extracellular signal-regulated kinase signaling to prevent cardiomyopathy caused by mutation in the gene encoding A-type lamins. *Hum. Mol. Genet.* **18,** 241–247 (2009).
29. Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* **147,** 913–920 (1999).

METHODS

Plasmids. MKL1–GFP was a gift from A. Kapus; MKL1(1–204)–2×GFP, MKL1 (1–204)XXX–2×GFP, NLS–GFP–NES and MKL1–PAGFP were obtained from M. Vartiainen. The NLS–GFP–NES reporter construct consists of GFP fused to an NLS and an importin α/β-dependent NES, using the same import/export mechanism as MKL1 (ref. 30). The MKL1(1–204)XXX–2×GFP construct contains alanine substitutions in all three of the MKL1 RPEL motifs, which abrogates binding to G actin and leads to constitutively nuclear localization independent of actin polymerization[12]. GFP–RCC1 was obtained from B. Paschal. GFP–actin was from F. Gertler and emerin–GFP was obtained from H. Worman.

Lamin and emerin mouse models. $Lmna^{-/-}$ mice along with heterozygous ($Lmna^{+/-}$) and wild-type ($Lmna^{+/+}$) littermates were obtained from crossing $Lmna^{+/-}$ mice. Similarly, $Lmna^{N195K/N195K}$ and heterozygous ($Lmna^{N195K/+}$) and wild-type ($Lmna^{+/+}$) littermates resulted from breeding $Lmna^{N195K/+}$ animals. $Lmna^{-/-}$ mice develop severe muscular dystrophy and cardiomyopathy and die at 4–8 weeks of age[29]. $Lmna^{N195K/N195K}$ mice develop severe dilated cardiomyopathy and die prematurely around 12–16 weeks of age[11]. Genotype was determined by polymerase chain reaction of genomic tail DNA, as described previously[29]. The mouse colonies were derived from breeders provided by C. Stewart[11,26,29]. All mice were maintained in the animal facility at Cornell University following protocols approved by the Cornell University Institutional Animal Care and Use Committee (IACUC).

Cell lines, transfection and drug treatment. Immortalized MEFs from $Lmna^{+/+}$, $Lmna^{-/-}$, $Lmna^{N195K/N195K}$ and $Emd^{-/Y}$ mice were a gift from C. Stewart. Cells were maintained in Dulbecco's Eagle's Modified Media (DMEM) with GlutaMax (Gibco, Invitrogen) containing 10% fetal bovine serum (FBS) (PAA), and 1% penicillin/streptomycin at 37 °C in a humidified atmosphere with 5% carbon dioxide. Serum starvation was done by withdrawing serum and incubating the cells in 0.3% FBS for 24 h. Starved cells were stimulated with DMEM with 15% FBS. Transient transfection was carried out using Lipofectamine Reagent (Invitrogen) according to manufacturer's protocol. 1 µg of plasmid DNA was used per transfection reaction. For cytochalasin D (Sigma Aldrich) treatment, a stock of 1 mM was prepared by reconstituting the drug in DMSO and a working concentration of 1 µM was used for all experiments. For leptomycin B (Sigma Aldrich) treatment, a working concentration of 5 nM was used for all experiments.

Antibodies and immunofluorescence staining. Goat anti-MKL1 (C-19 and H-180), goat anti-actin (C-11) conjugated with horseradish peroxidase (HRP) and goat anti-lamin A/C (N-18) antibodies were purchased from Santa Cruz Biotechnology. Mouse anti-Ran antibody (610341) was purchased from BD Biosciences. Mouse anti-emerin antibody (NCL-emerin) was a product from Novocastra. Mouse anti-paxillin (05-417) was purchased from Millipore. Rabbit anti-tubulin (ab6046) was purchased from Abcam. Secondary antibodies conjugated to Alexa-488 or Alexa-568 fluorophores were purchased from Molecular Probes (Invitrogen). For immunofluorescence staining, cells grown at subconfluency were collected and fixed with 4% paraformaldehyde/phosphate-buffered saline (PBS) and permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature. Cells were grown at similar subconfluency for all cell lines tested. Primary antibodies in 4% bovine serum albumin in Tris-buffered saline (TBS) with 0.05% Tween 20 were incubated for 1 h at room temperature or overnight at 4 °C. Cells were washed three times and incubated with appropriate secondary antibodies for 1 h at room temperature. Slides were mounted in Prolong Gold Anti-Fade medium with 4′,6-diamidino-2-phenylindole (DAPI) (Molecular Probes, Invitrogen). Images were collected and analysed on a Zeiss LSM 700 confocal microscope (Carl Zeiss). Images were captured using identical exposure times for each cell line.

Immunohistochemistry. Hearts from 4-week-old $Lmna^{-/-}$ mice and 8-week-old $Lmna^{N195K/N195K}$ mice along with wild-type and heterozygous littermate controls were collected and fixed in 4% paraformaldehyde in PBS before processing for paraffin embedding and cutting. The sections were deparaffinized with xylene and rehydrated. Antigen retrieval was performed by incubation with sodium citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0) (Invitrogen Inc.) at 95 °C for 20 min. The sections were then incubated with 3% hydrogen peroxide (H_2O_2) for 20 min to quench endogenous peroxidase activity. Blocking of nonspecific sites was done by incubating the slices with 10% horse serum in PBS for 20 min. Anti-MKL1 antibody (Santa Cruz Biotechnologies) was used at 1:50 dilution at 4 °C for 12–16 h to detect endogenous MKL1. A biotinylated anti-goat IgG secondary antibody (Vector Laboratories Inc.) was used at 1:400 dilution at room temperature for 1 h. The biotinylated secondary antibody was then detected using the VECTASTAIN ABC system (Vector Laboratories Inc.) which uses a preformed macromolecular complex between avidin and biotinylated horseradish peroxidase. A working solution of 3,3′-diaminobenzidine (DAB) was used as a substrate for the peroxidase. The sections were then counterstained with haematoxylin before air-drying and mounting. Images were acquired using a Zeiss Axiovert

200 inverted microscope (Carl Zeiss) equipped with a ×20 objective with an AxioCam ICc1. The number of cells positive for nuclear MKL1 was counted manually on at least five random microscopic fields per section and normalized to the total number of cardiac myocytes.

Western blotting. Cells were lysed using radioimmunoprecipitation assay (RIPA) buffer with freshly added protease inhibitor cocktail. The cells were scraped off using a cell scraper and incubated on ice for 30 min. The lysate was then cleared by centrifuging at 13,200 r.p.m. at 4 °C. Protein concentration was estimated by Bradford assay. Thirty micrograms of protein re-suspended in Laemmli sample buffer was loaded per sample. Cytoplasmic and nuclear fractions were prepared using the Pierce NE-PER Nuclear and Cytoplasmic Extraction kit according to manufacturer's instructions. Denatured proteins were resolved on 4–15% Nu-PAGE bis-tris polyacrylamide gels and blotted to a polyvinylidene fluoride (PVDF) membrane. Blocking was done with incubation in 10% non-fat dry milk in TBS with 0.1% Tween-20. The membrane was then probed with primary antibodies in 5% milk in TBST at 4 °C overnight and sequentially detected with horseradish peroxidase conjugated secondary antibodies. The signal was revealed by autoradiography using enhanced chemiluminescence (ECL) (Pierce, Thermo Fisher Scientific Inc.).

F- and G-actin assays. For the cytochalasin D washout experiment, MEFs were treated with 1 µM cytochalasin D (Sigma Aldrich) for 30 min; subsequent drug washout was performed by rinsing the cells with three changes of medium. Cells were fixed with 4% paraformaldehyde/PBS at 1 h or 2 h after washout and then permeabilized with 0.2% Triton X-100 in PBS. Stress fibres were visualized using Phalloidin-Alexa 568. For fluorescence labelling of F and G actin, starved and stimulated $Lmna^{+/+}$, $Lmna^{-/-}$ and $Lmna$ N195K MEFs were fixed and permeabilized as described above. Cells were then stained with Phalloidin-Alexa 568 to label F actin and DNase 1-Alexa 488 (Invitrogen) to label G actin[31]. Slides were mounted in Prolong Gold Anti-Fade medium with DAPI (Invitrogen). Images were collected and analysed on a Zeiss LSM 700 confocal microscope (Carl Zeiss). Images were captured using identical exposure times for each cell line.

Real-time PCR. Total RNA from cell lines was extracted using the Qiagen RNeasy kit (Qiagen) according to manufacturer's instructions. Total RNA from cryopreserved tissues was extracted using TRIzol Reagent (Life Technologies, Invitrogen Inc.) according to manufacturer's instructions. RNA was reverse-transcribed to cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Real-time PCR was carried out using SYBR-Green technology (Applied Biosystems) in a total volume of 25 µl. Gene expression for *Srf* and *Vcl* was quantified. Values were normalized to an endogenous control, TATA binding protein (*Tbp*), and compared to unstimulated samples with the $\Delta\Delta C_t$ method. Data are based on results from three independent experiments.

SRE-luciferase assay. SRF forms a complex over the SRE/CArG element upon receiving upstream signals from the MAPK pathway and/or the RhoA pathway. SRE activity was measured using the Dual-Glo SRE-Luciferase Assay (Promega) according to manufacturer's instructions. Briefly, $Lmna^{+/+}$, $Lmna^{-/-}$ and $Lmna$ N195K MEFs were transfected with either the SRE reporter construct or positive and negative controls. Dual-luciferase data from starved and serum-stimulated transfected cells were then collected on a luminometer. The fold change of SRE activity for each cell line was determined by comparing normalized luciferase activities of the reporter in stimulated versus starved samples.

Time-lapse microscopy and photobleaching experiments. For live cell imaging, a Zeiss LSM 700 confocal microscope (Carl Zeiss) equipped with a ×63 oil immersion objective (Carl Zeiss) was used. Cells were maintained at 37 °C in HEPES-buffered DMEM for the duration of the time-lapse acquisition. Images were recorded at 30 s or 1 min intervals and analysed using the Zen software (Carl Zeiss). For photobleaching experiments, cells were plated on a coverslip and mounted onto a glass slide with a depression containing culture media. Fluorescence loss in photobleaching (FLIP) experiments were performed on a Zeiss LSM 700 confocal microscope (Carl Zeiss) using the 488 nm laser line. Cells had been serum-stimulated for 30 min before the experiments. Relative loss in nuclear fluorescence during continuous photobleaching of cytoplasmic MKL1–GFP was computed by normalizing nuclear fluorescence intensity to pre-bleach values ($t = 0$). Increased loss of nuclear fluorescence indicates a higher rate of nuclear export of MKL1–GFP. For FLIP experiments, two single scans were acquired, followed by repeated photobleaching using a single bleach pulse at intervals of 1 s for 200 iterations in defined regions of approximately 30 µm² in the cytoplasm. Single section images were then collected at 1 s intervals. For imaging, the laser power was attenuated to 2% of the bleach intensity. The relative fluorescence intensity in a region of interest was determined by normalizing fluorescence intensity in the region to the total fluorescence in the same region during prebleach. This method provides a means of quantifying nuclear export as the cytoplasmic pool of fluorescent protein is rapidly bleached and subsequent loss of fluorescence signal from the nucleus reflects nuclear export. For FRAP experiments, the cells were scanned

two times before photobleaching by scanning the region of interest 80 times at 100% laser intensity of a 488-nm laser line. Single section images were then collected at 5-s intervals with laser power attenuated to 2% of the bleach intensity. The fluorescence intensity at the region of interest at each time point was normalized to the change in total fluorescence due to bleaching and imaging, as described previously[32]. For actin FRAP experiments, cells were photobleached in defined ~2 μm diameter nuclear and cytoplasmic regions with 2 μm thickness[33]. Values were normalized to the whole-cell fluorescence or nuclear fluorescence at each time point, for cytoplasmic and nuclear actin FRAP, respectively. For photoactivation experiments, cells expressing MKL1–PAGFP were plated on glass-bottom dishes and starved for 24 h before stimulation with 15% FBS in phenol-red free DMEM for 30 min. Imaging was performed at 37 °C using an LSM 700 confocal microscope (Carl Zeiss). Photoactivation of cytoplasmic MKL1–PAGFP was carried out using the 405-nm laser at 50% laser power for 30 iterations. Sequential imaging after photoactivation was performed using a ×63 oil immersion objective using excitation from a 488-nm laser line with 2% laser intensity. The increase

of fluorescence was normalized to the initial fluorescence of the cytoplasmic activation area.

Statistical analysis. Statistical analysis was performed using GraphPad Prism (GraphPad Software Inc.). Data are presented as mean ± s.e.m. unless stated otherwise. Two-tailed unpaired *t*-test and one-way ANOVA were used as detailed in respective figure legends. Statistical significance was defined as $P < 0.05$. All results are derived from three independent experiments.

30. Kudo, N. *et al.* Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. *Exp. Cell Res.* **242,** 540–547 (1998).
31. Knowles, G. C. & McCulloch, C. A. Simultaneous localization and quantification of relative G and F actin content: optimization of fluorescence labeling methods. *J. Histochem. Cytochem.* **40,** 1605–1612 (1992).
32. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404,** 604–609 (2000).
33. McDonald, D., Carrero, G., Andrin, C., de Vries, G. & Hendzel, M. J. Nucleoplasmic beta-actin exists in a dynamic equilibrium between low-mobility polymeric species and rapidly diffusing populations. *J. Cell Biol.* **172,** 541–552 (2006).

LETTER

doi:10.1038/nature12123

Reconfiguration of the proteasome during chaperone–mediated assembly

Soyeon Park[1,2]*, Xueming Li[3]*, Ho Min Kim[3]†*, Chingakham Ranjit Singh[4], Geng Tian[1], Martin A. Hoyt[5], Scott Lovell[6], Kevin P. Battaile[7], Michal Zolkiewski[8], Philip Coffino[5], Jeroen Roelofs[4], Yifan Cheng[3] & Daniel Finley[1]

The proteasomal ATPase ring, comprising Rpt1–Rpt6, associates with the heptameric α-ring of the proteasome core particle (CP) in the mature proteasome, with the Rpt carboxy-terminal tails inserting into pockets of the α-ring[1–4]. Rpt ring assembly is mediated by four chaperones, each binding a distinct Rpt subunit[5–10]. Here we report that the base subassembly of the *Saccharomyces cerevisiae* proteasome, which includes the Rpt ring, forms a high-affinity complex with the CP. This complex is subject to active dissociation by the chaperones Hsm3, Nas6 and Rpn14. Chaperone-mediated dissociation was abrogated by a non-hydrolysable ATP analogue, indicating that chaperone action is coupled to nucleotide hydrolysis by the Rpt ring. Unexpectedly, synthetic Rpt tail peptides bound α-pockets with poor specificity, except for Rpt6, which uniquely bound the α2/α3-pocket. Although the Rpt6 tail is not visualized within an α-pocket in mature proteasomes[2–4], it inserts into the α2/α3-pocket in the base–CP complex and is important for complex formation. Thus, the Rpt–CP interface is reconfigured when the lid complex joins the nascent proteasome to form the mature holoenzyme.

The proteasome mediates selective protein degradation in eukaryotes[1]. It is composed of a 19-subunit regulatory particle (RP; also known as PA700 or the 19S complex) and a 28-subunit proteolytic CP (also known as the 20S complex) composed of four stacked heptameric rings. Ubiquitinated proteins are recognized by the RP, translocated into the CP via a channel in its outer (α)-ring, and degraded. The RP comprises a 10-subunit base and 9-subunit lid[1–4]. Central to the base is the Rpt ring, a heterohexameric ATPase complex that abuts the α-ring, with flexible C-terminal tails of multiple Rpt subunits inserting into 'α-pockets' of the CP[11–14].

The Rpt ring is formed from three modules[1]: Rpt3–Rpt6; Rpt4–Rpt5; and Rpt1–Rpt2. Within these modules there are four assembly chaperones. Although unrelated phylogenetically, these 'RP chaperones' each bind a CP-proximal C domain within a specific Rpt subunit[5,9], as follows: Rpt1–Hsm3; Rpt3–Nas6; Rpt5–Nas2; and Rpt6–Rpn14.

A debated aspect of RP assembly is whether the CP facilitates the process[15,16], although CP-dependent and CP-independent assembly pathways are not mutually exclusive. We have proposed that the insertion of Rpt tails into CP α-pockets is important for RP assembly in yeast, and that RP chaperones antagonize Rpt tail insertion into CP α-pockets by steric hindrance, thus promoting temporal order in assembly[5,6] (Supplementary Fig. 1).

The base is considered a key assembly intermediate of yeast proteasomes[5,6,9,10]. To investigate the effects of RP chaperones on the RP–CP interface, and to model the behaviour of early assembly intermediates, we developed a reconstitution assay for the base–CP complex. When purified base and CP were mixed, they formed complexes

(base₁–CP and base₂–CP; collectively termed base–CP) with an apparent dissociation constant (K_d) of ∼3 nM (Supplementary Fig. 2). Base–CP was visualized on native polyacrylamide gel electrophoresis (PAGE) by using an in-gel assay for hydrolysis of the fluorogenic peptide LLVY-AMC (Fig. 1a). When chaperones Hsm3, Nas6 and Rpn14 were added in excess at time zero, the chaperone trio inhibited complex formation beyond the detection limit. Because tail–pocket contacts mediate base–CP association[12], this experiment satisfies a key prediction of the model[5,6]—that RP chaperones antagonize insertion of Rpt tails into CP α-pockets. A fourth RP chaperone, Nas2, dissociates before base assembly[17] and is accordingly inactive in these assays (data not shown).

Antagonism of base–CP association, assessed as described earlier with chaperones in excess, remained strong when chaperones were added at 1:1 stoichiometry versus base, indicating potent interference (Supplementary Fig. 3). Base–CP association can be quantified by real-time fluorometric assays that track LLVY-AMC hydrolysis. Free CP hydrolyses LLVY-AMC slowly, owing to closure of its gated channel[1]; binding of base opens the channel via Rpt tail–α-pocket interactions[12,13]. Suppression of base–CP assembly by the chaperone trio is readily observed with the LLVY-AMC assay (Fig. 1b) at low-nanomolar levels, which are comparable to or below their estimated intracellular abundance (ref. 18 and data not shown). Each of these chaperones was found to antagonize the base–CP association individually, although with different potencies (Fig. 1c and Supplementary Fig. 4). In summary, Hsm3, Nas6 and Rpn14 act coordinately, and through a common mechanism, to antagonize the Rpt ring–CP association.

To assess further how RP chaperones regulate proteasomes, we determined the crystal structures of Hsm3 and of Hsm3 complexed with the Rpt1 C domain[19,20] (Protein Data Bank accessions 4FP7 and 4JPO; Supplementary Figs 5–9), and modelled this complex into the cryo-electron microscopy (cryoEM) structure of the yeast proteasome holoenzyme[2]. The Nas6–Rpt3 co-complex[21] was similarly modelled into the holoenzyme. The results suggest physical clashing between CP and chaperones in the holoenzyme (Fig. 1d), consistent with the steric interference hypothesis[5,6]. Although modelling is not completely predictive, these data agree with previous attempts to model chaperones into the holoenzyme[5,20]. The Rpn14–Rpt6 structure is unsolved and thus the relevance of steric interference to Rpn14 remains conjectural. The RP–CP interface is likely to be dynamic, owing to conformational changes in the Rpt ring during cycles of ATP hydrolysis (see later). Consequently, steric interference may apply to a subset of conformational states.

In the base–CP experimental model, RP chaperones may act by binding to free base, thus preventing association of base with CP. Alternatively, or in addition, chaperones might interact transiently with base–CP to promote dissociation actively. To assess these models,
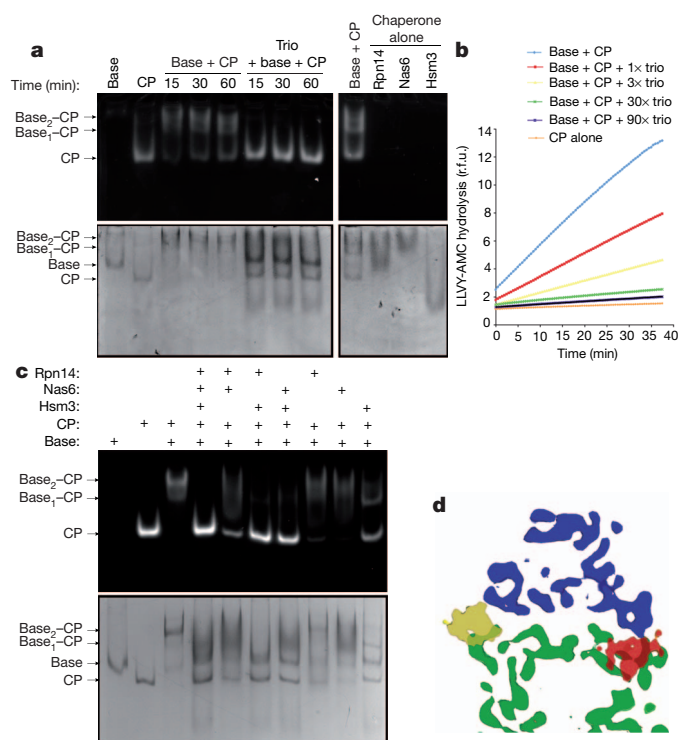
Figure 1 | Chaperones inhibit base–CP assembly. a, Purified base (160 nM) and CP (80 nM) were incubated with or without Rpn14, Nas6 and Hsm3 (trio, 1.6 μM each), and resolved by native PAGE. Top, in-gel peptidase assay (0.02% SDS); bottom, Coomassie stain. For input protein see Supplementary Fig. 2. **b**, Base (5 nM) and CP (2 nM) were challenged with chaperone trio (amounts in molar excess of base; ATP at 2 mM). In this and all real-time experiments, LLVY-AMC hydrolysis is expressed as relative fluorescence units (r.f.u.) and experiments were performed in triplicate with traces combined for presentation. **c**, Native gel analysis of base–CP formation as in **a**, after addition of chaperones to base (160 nM) singly or in combination at tenfold molar excess of base. **d**, A yeast Rpt hexamer model was built, using the hexameric P97 D1 domain structure as a template (see Supplementary Methods). This model was fit into the EM map[2] of yeast Rpt hexamer. Relative positions of Hsm3 (red) and Nas6 (yellow) on the Rpt ring (blue) were assessed by superimposing Hsm3–Rpt1C and Nas6–Rpt3C structures onto the Rpt ring model that had been fit into the EM map. A clipped view of the Rpt ring with bound chaperones and CP (green) is presented. Areas of overlap highlight steric clashes between chaperones and CP.

we assayed the time course of base–CP dissociation after chaperone addition. As a control, we examined spontaneous dissociation of base–CP using a 'CP trap' that is inactivated by the proteasome inhibitor epoxomicin. The trap captures base that had dissociated from base–CP (Supplementary Fig. 10), thus suppressing LLVY-AMC hydrolysis. Trap addition resulted in slow loss of LLVY-AMC hydrolytic activity over more than 15 min, whereas chaperone addition led to an immediate reduction in hydrolytic activity, as expected from active dissociation (Fig. 2a, left). On chaperone addition, a new, stable steady-state hydrolytic rate was established within approximately 7–8 min (Fig. 2a, right). Chaperone addition to holoenzyme had only weak dissociative effects (Supplementary Fig. 11a). Thus, RP chaperones may work preferentially on assembly intermediates.

To assess the possible relevance of Rpt ring conformational dynamics to base–CP dissociation, we compared the effects of ATP and non-hydrolysable ATPγS in the base–CP dissociation assay. Base–CP assembled normally with ATPγS (Supplementary Fig. 11b), but subsequent chaperone addition had no detectable effect on LLVY-AMC hydrolysis, indicating failure to dissociate the complex (Fig. 2b). Antagonism of base–CP association by chaperones may therefore be finely tuned to the conformational state of the Rpt ring. The simplest interpretation is that ATPγS mimics the ATP-bound state of the Rpt, with a chaperone inhibiting base–CP association when its cognate Rpt is
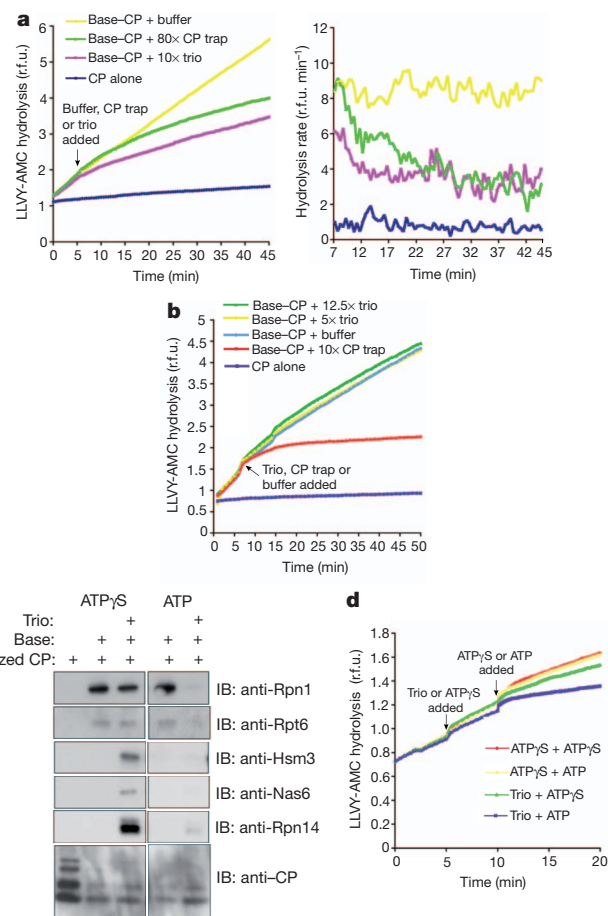

Figure 2 | Base–CP association is nucleotide dependent. a, CP (2 nM) activity stimulated by base (5 nM) was monitored over time (2 mM ATP, 50 mM KCl). At 5.5 min, chaperone trio or CP trap was added in molar excess of base or active CP, respectively. CP trap inhibits re-association of base with active CP. Right, hydrolysis rate (r.f.u. min^{-1}) over time. **b**, Purified base (5 nM) and CP (2 nM) were assembled in the presence of ATPγS (0.1 mM throughout). At 6 min, chaperone trio or CP trap were added in molar excess. **c**, CP (15 nM) was immobilized on IgG resin via ProA tag, and incubated with base (~80 nM) and chaperone trio (160 nM) in the presence of 2 mM ATP or 0.5 mM ATPγS. CP-bound proteins were washed with buffer (50 mM KCl), then eluted with TEV protease while maintaining nucleotide concentration. Immunoblots (IBs) were probed with indicated antibodies. Images are from the same gel and exposure. **d**, CP (2 nM) activity (LLVY-AMC hydrolysis; r.f.u.) was monitored in the presence of base (5 nM) and 0.1 mM ATPγS for 5 min. Chaperone trio (50 nM) or buffer alone containing 0.1 mM ATPγS was then added. At 10 min, buffer containing either ATPγS or ATP plus ATPγS was added. Final nucleotide concentrations were either 0.1 mM ATPγS or 10 mM ATP plus 0.1 mM ATPγS. See also Supplementary Fig. 12.

bound to ADP (or is free of nucleotide), but not to ATP. With ATPγS, RP chaperones could fail to antagonize base–CP simply because they cannot bind Rpt proteins under these conditions. However, chaperone–base complexes form comparably with ATP, ATPγS or ADP (Supplementary Fig. 11c).

The results described earlier suggest that a stable base–chaperone–CP co-complex may form in the presence of ATPγS. To test this, we immobilized CP to a resin, added base and chaperones, and assayed resin-bound components after washing. When chaperones were added in the presence of ATP, base dissociated from the complex (Fig. 2c), whereas, with ATPγS, base remained bound to both CP and chaperones. Thus, the chaperones' lack of effect on LLVY-AMC hydrolysis in the presence of ATPγS reflects their failure to dissociate base–CP, and the chaperones' capacity to compete with CP for occupancy of the base may be dependent on the Rpt nucleotide hydrolytic cycle.

To test whether preformed base–chaperone–CP complex is primed for dissociation, the complex was formed in the presence of ATPγS, and dissociation was then monitored after ATP addition. ATP produced rapid base–CP dissociation; LLVY-AMC hydrolysing activity decayed with a half-life of approximately 1 min (Fig. 2d and Supplementary Fig. 12). By contrast, ATP and ATPγS produced indistinguishable hydrolytic profiles in the absence of chaperones (Fig. 2d). Thus, RP chaperones actively dissociate base–CP.

To understand chaperone action within the base–CP complex better, we studied the specificity of the insertion of Rpt tails into α-pockets, interactions proposed to be under chaperone control[5,6]. We previously determined by single-particle cryoEM that the C termini of the homohexameric archaeal PAN ATPases bind the α-pockets of the homoheptameric α-ring of archaeal CP[13]. This approach is used here to assign distinct α-pocket binding preferences to each yeast Rpt tail. We determined seven subnanometre-resolution three-dimensional reconstructions of the yeast CP, one from CP alone (Supplementary Figs 13 and 14), and the others from CP incubated individually with six different Rpt peptides. Each peptide comprised 8 amino acids from the C terminus of an Rpt.

The pseudo seven-fold symmetry of the heteroheptameric α- and β-rings poses a challenge to single-particle cryoEM. To break this pseudo-symmetry, we fused a glutathione S-transferase (GST) tag to the C terminus of subunit β2 (Supplementary Fig. 13a, e), and verified that the tag does not alter CP function. Differences between maps of peptide–CP and CP alone were calculated as described[13] (Supplementary Fig. 15). At proper thresholds, difference densities correspond to peptides bound to α-pockets. Because the pocket between subunits α7 and α1 lacks the Lys residue that is required for binding the C terminus of an Rpt[11], no specific binding of any peptide to this pocket is expected. Thresholds were therefore set to show no difference density in the α7/α1-pocket, which in fact always had the lowest difference density corresponding to tail peptide (Fig. 3). As controls, difference maps between two independent three-dimensional reconstructions from two separately collected data sets of the same sample showed no significant difference density, indicating that difference densities assigned to each peptide were not generated by image misalignment or random noise (Supplementary Methods).

Figure 3a compares binding specificities of the six Rpt peptides for the seven α-pockets of the CP. In an individual difference map, the sizes of the densities in different pockets correlate with peptide-binding affinities for these pockets. Figure 3b summarizes these data. Figure 3b and c also represent previous mapping of Rpt tails to α-pockets in the holoenzyme[2–4,22]. Our data reveal an unexpected lack of specificity in the Rpt tail–α-pocket interaction. We therefore suggest that the specificity of tail–pocket interactions within the mature complex is largely guided by constraints on possible tail–pocket interactions that arise from the defined subunit arrangements of the apposed Rpt and α-rings. For example, within the holoenzyme, the Rpt2 and Rpt3 C termini insert into α3/α4- and α1/α2-pockets, respectively[2–4,22]. However, free forms of Rpt2 and Rpt3 C termini show a marked preference for noncognate pockets (Fig. 3b).

Our results raise the question of how the register of the RP–CP interface is determined. Only the Rpt6 peptide showed high specificity of binding to an established[22] cognate pocket (Fig. 3b, c). Thus, the Rpt6 tail has the specificity to serve as an anchor point for either the mature proteasome or an assembly intermediate.

The unique binding specificity of the free Rpt6 tail peptide led us to examine its physiological significance genetically. We generated substitution mutations in Rpt6 tail residues, and assessed effects on proteasome assembly using assays of cell growth (Fig. 4a) and native PAGE (Fig. 4b). Several substitutions resulted in proteasome defects, notably a block substitution of alanines for the terminal LFK sequence (Fig. 4a, b and Supplementary Fig. 16). Deletion of one residue from the C terminus had a similar effect (Fig. 4a, b and Supplementary Figs 16–18).
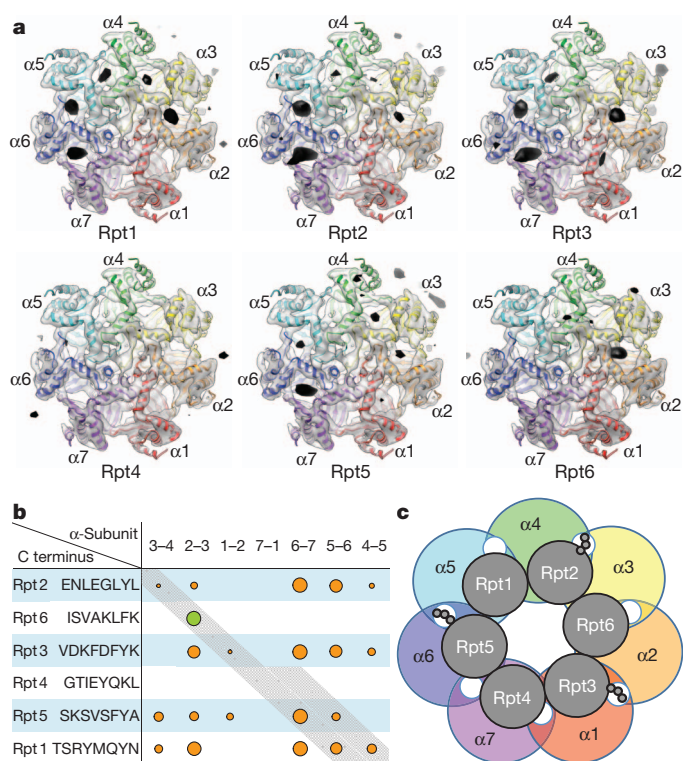

Figure 3 | Difference maps reveal binding sites of Rpt C-terminal peptides to CP α-pockets. a, Top views of three-dimensional density maps of CP superimposed with difference densities corresponding to C-terminal peptides of each Rpt. Peptides were present at 0.5 mM, CP at 1.6 μM. The amount of each peptide bound is reflected by the size of black densities within each pocket. **b,** Summary of Rpt tail peptide-binding sites and relative intensities. Sizes of circles represent the volume of difference densities generated by peptides. Grey diagonals denote Rpt tail–α-pocket mapping of intact proteasomes by crosslinking[22]. Rpt4, Rpt5 and Rpt1 each crosslink to two α-pockets, suggesting an ambiguous register. **c,** Predominant tail–pocket interactions in yeast holoenzymes as determined by cryoEM[2–4].

The exceptional specificity of the free Rpt6 tail for the α2/α3-pocket, together with its role in proteasome assembly, seem to be inconsistent with recent cryoEM studies of the yeast proteasome, which visualize the Rpt2, Rpt3 and Rpt5 tails within α-pockets, with Rpt6 apparently not fixed within an α-pocket, presumably being too flexible to be visualized[2–4]. To resolve this paradox, we first used the base–CP assembly assay to test whether the Rpt6 tail helps to stabilize the base–CP interface. Base complexes were purified from *rpt6-Δ1* and wild-type cells, then mixed with CP; the activation of LLVY-AMC hydrolysis was then assayed. At a ratio of 16:1, the *rpt6-Δ1* base remained deficient in comparison to a twofold excess of wild-type base over CP (Fig. 4c). The defect in activation can be attributed largely to reduced base–CP association, as shown by native PAGE followed by Coomassie staining (Supplementary Fig. 19).

One scenario to explain the phenotype of *rpt6* tail mutants is that their reduced proteasome levels simply reflect a lower affinity between CP and RP. To test this, we purified *rpt6-Δ1* proteasome holoenzymes and incubated them in the presence or absence of CP trap to follow their dissociation over time. Mutant and wild-type proteasomes were comparable in stability (Fig. 4d). Thus, the Rpt6 tail influences biosynthetic proteasome assembly rather than holoenzyme stability.

To investigate the base–CP interface more generally, we reconstituted this complex using base from a wild-type (*RPT6*) strain of yeast, and subjected it to cryoEM (Supplementary Fig. 20). In the three-dimensional reconstruction of base–CP (Fig. 5a), the rotational register between the Rpt ring and the CP, as well as the axial tilt of the Rpt ring from that of CP, are comparable to those in the holoenzyme. Individual
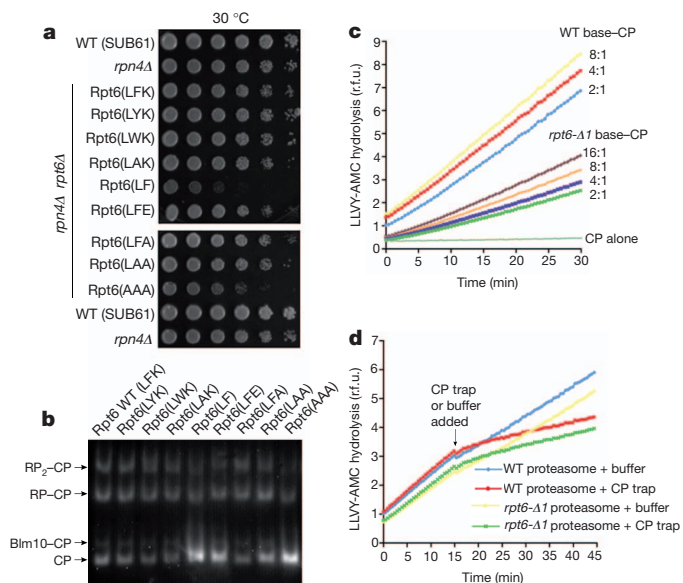
Figure 4 | Rpt6 C-terminal tail promotes formation of base–CP complex.
a, Growth defects of C-terminal *rpt6* mutants. Strains were spotted onto plates containing rich media (YPD) in fourfold serial dilutions. Plates were incubated at 30 °C for 2 days. WT, wild type Capital letters indicate the last three amino acids of Rpt6 (LFK) or mutants thereof. **b,** Whole-cell extracts (100 μg) from *rpt6* mutants as in **a** were resolved by native PAGE and subjected to LLVY-AMC assay in 0.02% SDS. **c,** Role of Rpt6 tail in base–CP association. Assembly kinetics of wild-type or *rpt6-Δ1* base with CP was measured using LLVY-AMC hydrolysis. Purified CP (2 nM) was mixed with the indicated fold excess of base (2 mM ATP). LLVY-AMC hydrolysis is indicated in r.f.u. **d,** Stability of proteasome holoenzyme (2 nM) from wild-type or *rpt6-Δ1* mutants was assessed in 2 mM ATP by adding 50-fold molar excess of CP trap or buffer alone at 15 min.

Rpt subunits were modelled into the structure as shown, based on previous characterization of the holoenzyme[2] (Supplementary Fig. 20c, d).

A key feature of the base–CP complex is the existence of a strong contact between the Rpt6 tail and the α2/α3-pocket (Fig. 5), which is inherently specific for Rpt6 tail peptide. As the holoenzyme does not show a prominent Rpt6 tail contact[2-4], the structural and functional data both suggest a transient role for the Rpt6 tail in assembly. Distinct contacts were also observed between Rpt1 and the α4/α5-pocket, as well as Rpt2 and the α3/α4-pocket (Fig. 5 and Supplementary Fig. 21), although the Rpt1 contact is relatively superficial. On the basis of the free tail peptide experiments of Fig. 3, the intrinsic specificity of Rpt2 and Rpt1 cannot explain the tail–pocket register of base–CP. This suggests that Rpt6 is a key determinant of the tail–pocket register in base–CP.

The determination of the tail–pocket register is inherently problematic because of the symmetry mismatch between the hexameric Rpt ring and heptameric α-ring, which may underlie a tendency for ambiguous register at this interface[3,4,22]. Proper register may be achieved through global optimization of tail–pocket interactions, allowing for a subset of incorrect tail–pocket alignments to be rejected despite their being stronger than the correct alignment. However, our data suggest that the tail–pocket register might be largely defined by dominant interactions of high specificity. The Rpt6 tail may perform such a function at an early stage of proteasome maturation, when the RP–CP interface is defined exclusively by Rpt tail–α-pocket contacts. This role of Rpt6 is apparently not sustained in the holoenzyme, perhaps because register is enforced by an alternative mechanism once the lid is incorporated into the complex. Subunit Rpn6 of the lid extends directly past Rpt6 to contact the CP[2,23], and may substitute for the anchoring role of Rpt6.

Although the tail–pocket register of base–CP is consistent with that of holoenzyme, the dominant tail–pocket interactions are quite different. For the holoenzyme[2-4], these are thought to be Rpt3, Rpt2 and Rpt5. These tails alternate across the ring in the holoenzyme (Fig. 3c), whereas in base–CP the dominant tail contacts appear to be collected on one side of the ring (Fig. 5), in an arrangement resembling that of the archaeal PAN complex[14].

An interesting feature of base–CP is that the neighbouring[17] Rpt6 and Rpt2 tails show strong pocket interactions. Because the tail of Rpt2 has little inherent specificity, its insertion into the α3/α4-pocket may be facilitated by Rpt6. It is consistent with the symmetry mismatch between the Rpt and α-rings that Rpt6 should preferentially promote tail insertion of its nearest neighbour, as more distant tails would fall out of phase with the CP pockets.
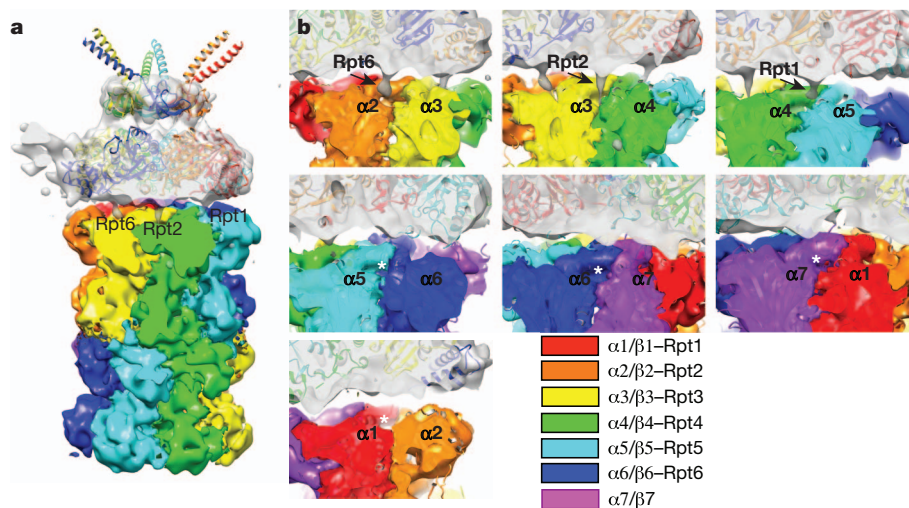

Figure 5 | Three-dimensional reconstruction of base–CP complex reveals an asymmetric interaction between the Rpt ring and the α-ring of the CP.
a, Three-dimensional reconstruction of the singly capped base–CP complex was determined by single-particle cryoEM to a resolution of ~10 Å. CP subunits are rendered in different colours as indicated. A difference map was calculated between the original three-dimensional reconstruction and one rotated 180° around the two-fold CP symmetry axis. The positive difference density (grey) corresponds to base bound to CP. It shows prominent densities from C termini of Rpt6, Rpt2 and Rpt1, which are clustered on one side of the Rpt ring, bound to specific α-pockets. **b,** Each panel shows an α-pocket. Arrows indicate densities corresponding to Rpt C termini extending towards the pockets. Asterisks indicate pockets without detectable density from Rpt C termini. Thresholds of CP and base densities are set separately but are identical in all panels. C termini of Rpt6, Rpt2 and Rpt1 are seen to insert into α-pockets.

Negative regulation of tail–pocket interactions by chaperones may potentially help in temporal ordering of the assembly pathway, in suppressing out-of-register tail–pocket interactions, and in maintaining proteasome assembly intermediates in a highly dynamic state. Our data suggest that chaperone action may be coupled to the ATPase cycle of the Rpt ring, with nucleotide controlling the competition between chaperone and CP for base interaction. The mechanism may involve changes in positioning of the Rpt C domain, which has an integrative role in that it positions the Rpt C-terminal tail, while at the same time binding chaperone on its outer face and contacting nucleotide on its inner face.

The major forms of mature proteasome differ from the ATPγS–base–CP complex in that they are associated with a mixture of ATP and ADP[24], whereas early intermediates in RP assembly are reported to have no detectable ATPase activity[25], suggesting that an ADP-free species resembling the ATPγS–base–CP complex could potentially function as a transient assembly intermediate. The fate of this complex may be to undergo chaperone-dependent dissociation upon ATP hydrolysis, or alternatively the lid may join the complex before the first round of ATP hydrolysis, to impose new modes of CP binding and suppress complex dissociation.

METHODS SUMMARY

The GST-tagged CP used for cryoEM analysis was purified using a 3× Flag tag appended to the Pre1 C terminus (β4). For the structure of CP complexed with peptide, ~0.5 mM peptide was incubated with 1.6 μM GST-tagged CP for 1 h at 37 °C directly before grid vitrification. Recombinant chaperones were purified from *Escherichia coli* using a GST tag, which was removed with Prescission protease before biochemical assays. Single-particle cryoEM studies were carried out as described[13], with details given in Supplementary Information.

1. Kish-Trier, E. & Hill, C. P. Structural biology of the proteasome. *Ann. Rev. Biophys* http://dx.doi.org/10.1146/annurev-biophys-083012-130417 (2013).
2. Lander, G. C. *et al.* Complete subunit architecture of the proteasome regulatory particle. *Nature* **482,** 186–191 (2012).
3. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl Acad. Sci. USA* **109,** 1380–1387 (2012).
4. Beck, F. *et al.* Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl Acad. Sci. USA* **109,** 14870–14875 (2012).
5. Roelofs, J. *et al.* Chaperone-mediated pathway of proteasome regulatory particle assembly. *Nature* **459,** 861–865 (2009).
6. Park, S. *et al.* Hexameric assembly of the proteasomal ATPases is templated through their C termini. *Nature* **459,** 866–870 (2009).
7. Le Tallec, B., Barrault, M. B., Guerois, R., Carre, T. & Peyroche, A. Hsm3/S5b participates in the assembly pathway of the 19S regulatory particle of the proteasome. *Mol. Cell* **33,** 389–399 (2009).
8. Kaneko, T. *et al.* Assembly pathway of the mammalian proteasome base subcomplex is mediated by multiple specific chaperones. *Cell* **137,** 914–925 (2009).
9. Saeki, Y., Toh, E. A., Kudo, T., Kawamura, H. & Tanaka, K. Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. *Cell* **137,** 900–913 (2009).
10. Funakoshi, M., Tomko, R. J. Jr & Hochstrasser, M. Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* **137,** 887–899 (2009).
11. Forster, A., Masters, E. I., Whitby, F. G., Robinson, H. & Hill, C. P. The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol. Cell* **18,** 589–599 (2005).
12. Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's α ring opens the gate for substrate entry. *Mol. Cell* **27,** 731–744 (2007).
13. Rabl, J. *et al.* Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* **30,** 360–368 (2008).
14. Smith, D. M. *et al.* ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. *Mol. Cell* **20,** 687–698 (2005).
15. Park, S., Kim, W., Tian, G., Gygi, S. P. & Finley, D. Structural defects in the RP-CP interface induce a novel proteasome stress response. *J. Biol. Chem.* **286,** 36652–36666 (2011).
16. Kusmierczyk, A. R., Kunjappu, M. J., Funakoshi, M. & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nature Struct. Mol. Biol.* **15,** 237–244 (2008).
17. Tomko, R. J. Jr, Funakoshi, M., Schneider, K., Wang, J. & Hochstrasser, M. Heterohexameric ring arrangement of the eukaryotic proteasomal ATPases: implications for proteasome structure and assembly. *Mol. Cell* **38,** 393–403 (2010).
18. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).
19. Takagi, K. *et al.* Structural basis for specific recognition of Rpt1, an ATPase subunit of 26S proteasome, by proteasome-dedicated chaperone Hsm3p. *J. Biol. Chem.* **287,** 12172–12182 (2012).
20. Barrault, M. B. *et al.* Dual functions of the Hsm3 protein in chaperoning and scaffolding regulatory particle subunits during the proteasome assembly. *Proc. Natl Acad. Sci. USA* **109,** E1001–E1010 (2012).
21. Nakamura, Y. *et al.* Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. *Biochem. Biophys. Res. Commun.* **359,** 503–509 (2007).
22. Tian, G. *et al.* An asymmetric interface between the regulatory particle and core particle of the proteasome. *Nature Struct. Mol. Biol.* **18,** 1259–1267 (2011).
23. Pathare, G. R. *et al.* The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc. Natl Acad. Sci. USA* **109,** 149–154 (2012).
24. Smith, D. M., Fraga, H., Reis, C., Kafri, G. & Goldberg, A. L. ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. *Cell* **144,** 526–538 (2011).
25. Thompson, D., Hakala, K. & DeMartino, G. N. Subcomplexes of PA700, the 19S regulator of the 26S proteasome, reveal relative roles of AAA subunits in 26S proteasome assembly and activation and ATPase activity. *J. Biol. Chem.* **284,** 24891–24903 (2009).

Author Contributions S.P. performed reconstitution of the base–CP complex and holoenzyme stability. X.L. performed all cryoEM experiments and analysis. H.M.K. and C.R.S. generated yeast strains. H.M.K. purified GST-fused CP, and participated in cryoEM experiments and analysis. C.R.S. performed purifications, and M.Z. performed ultracentrifugation. K.P.B. and S.L. determined crystal structures. J.R. and G.T. performed structural analysis and modelling. M.A.H., H.M.K. and P.C. performed phenotypic and native gel analysis of Rpt6 mutations. J.R. wrote the supplement with contributions from all authors. The manuscript was drafted by D.F. and Y.C., and modified by all authors.

Author Information Data have been deposited in the Electron Microscopy Data Bank under the following accession numbers: free CP, EMD-5593; Rpt1–CP, EMD-5611; Rpt2–CP, EMD-5612; Rpt3–CP, EMD-5613; Rpt4–CP, EMD-5614; Rpt5–CP, EMD-5615; Rpt6–CP, EMD-5616; and base₁–CP: EMD-5617. For the crystal structures, data have been deposited in the Protein Data Bank under accessions 4FP7 (Hsm3) and 4JPO (Hsm3–Rpt1 C domain). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R. (jroelofs@ksu.edu), Y.C. (ycheng@ucsf.edu) or D.F. (daniel_finley@hms.harvard.edu).

LETTER

Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits

Unnur Styrkarsdottir[1], Gudmar Thorleifsson[1], Patrick Sulem[1], Daniel F. Gudbjartsson[1], Asgeir Sigurdsson[1], Aslaug Jonasdottir[1], Adalbjorg Jonasdottir[1], Asmundur Oddsson[1,2], Agnar Helgason[1], Olafur T. Magnusson[1], G. Bragi Walters[1], Michael L. Frigge[1], Hafdis T. Helgadottir[1], Hrefna Johannsdottir[1], Kristin Bergsteinsdottir[3], Margret H. Ogmundsdottir[3], Jacqueline R. Center[4,5,6], Tuan V. Nguyen[4,6], John A. Eisman[4,5,6,7], Claus Christiansen[8], Erikur Steingrimsson[3], Jon G. Jonasson[9,10,11], Laufey Tryggvadottir[10,11], Gudmundur I. Eyjolfsson[12], Asgeir Theodors[13], Thorvaldur Jonsson[10,14], Thorvaldur Ingvarsson[15,16], Isleifur Olafsson[17], Thorunn Rafnar[1], Augustine Kong[1], Gunnar Sigurdsson[10,18], Gisli Masson[1], Unnur Thorsteinsdottir[1,10] & Kari Stefansson[1,10]

Low bone mineral density (BMD) is used as a parameter of osteoporosis. Genome-wide association studies of BMD have hitherto focused on BMD as a quantitative trait, yielding common variants of small effects that contribute to the population diversity in BMD[1–7]. Here we use BMD as a dichotomous trait, searching for variants that may have a direct effect on the risk of pathologically low BMD rather than on the regulation of BMD in the healthy population. Through whole-genome sequencing of Icelandic individuals, we found a rare nonsense mutation within the leucine-rich-repeat-containing G-protein-coupled receptor 4 (*LGR4*) gene (c.376C>T) that is strongly associated with low BMD, and with osteoporotic fractures. This mutation leads to termination of LGR4 at position 126 and fully disrupts its function. The c.376C>T mutation is also associated with electrolyte imbalance, late onset of menarche and reduced testosterone levels, as well as an increased risk of squamous cell carcinoma of the skin and biliary tract cancer. Interestingly, the phenotype of carriers of the c.376C>T mutation overlaps that of *Lgr4* mutant mice.

We selected individuals with standardized BMDs below minus one standard deviation (−1 s.d.) for inclusion in the low BMD group. As controls we combined those who had a measured BMD above −1 s.d. and, for increased power, individuals who had not had their BMD measured. The low BMD group included 4,931 individuals and the control group comprised 69,034 individuals. We then tested for associations between 34.2 million sequence variants and low BMD. The variants were identified through whole-genome sequencing of 2,230 Icelanders and then imputed into the entire sample set; 95,085 Icelanders genotyped with single nucleotide polymorphism (SNP) chips and 296,526 close relatives of those who were chip genotyped through familial imputation (Supplementary Fig. 1 and Supplementary Methods).

The two most significant associations with low BMD ($P < 5 \times 10^{-8}$) were with common variants at 13q14 that have previously been reported to be associated with BMD, represented by rs8001611 (odds ratio (OR) = 1.21; $P = 1.6 \times 10^{-12}$)[1,6], and a novel association with a group of correlated rare variants at 11p14 (Supplementary Fig. 2 and Supplementary Table 1). This group includes 16 variants with $P < 10^{-9}$, all with population frequencies of between 0.14% and 0.18%. Among the rare variants that associate with BMD at 11p14 is a SNP (University of California Santa Cruz (UCSC) Genome Browser position hg18_chr11:27369242_A) that introduces a nonsense codon

(c.376C>T; in the Refseq NM_018490 transcript) into exon 4 of the *LGR4* gene. None of the other rare variants at 11p14 were significant after accounting for the effect of c.376C>T (Supplementary Table 1), nor did any other coding variant in *LGR4* associate with BMD (Supplementary Table 2). We validated, and improved, the imputation of c.376C>T by direct genotyping (Supplementary Information); this resulted in a slightly stronger association with low BMD (OR = 4.30 and $P = 1.3 \times 10^{-10}$) (Table 1).

A common variant at this same locus, rs10835187 (minor allele frequency (MAF) 45%), was previously reported to associate with spine BMD ($P = 4.9 \times 10^{-8}$) but not with hip BMD ($P = 0.03$) in a large meta-analysis of 80,000 individuals[7]. rs10835187 is not correlated with c.376C>T ($r^2 < 0.001$) and its effect on low BMD is much weaker (OR = 1.06, $P = 0.031$). Testing the association of c.376C>T with BMD-related traits conditional on the effect of rs10835187 has very little impact on the results, and the association of rs10835187 with BMD is not changed if we adjust for c.376C>T (Supplementary Table 3). Thus, the two variants represent two independent BMD association signals in the 11p14 region. Because the association of rs10835187 with spine BMD is weak in our data set ($\beta = -0.036$ and $P = 0.0042$) we are not well powered to refine the rs10835187 association signal. However, we notice two variants in the 3′ untranslated region (UTR) region of lin-7 homologue C (*LIN7C*) at 11p14, rs3209593 and rs1140711, that are highly correlated with rs10835187 ($r^2 > 0.9$) and whose association with spine BMD is indistinguishable from that of rs10835187. The expression of *LIN7C* in blood and adipose tissue has previously been shown to be highly correlated with rs10835187. Another variant, rs61888800 ($r^2 = 0.31$ with rs10835187), located in the 5′ UTR region of the brain-derived neurotrophic factor (*BDNF*) gene could also explain the association of rs10835187 with BMD. No variant in *LGR4* can explain the rs10835187 association (Supplementary Table 4). Our data, although not conclusive, indicate that rs10835187 acts through *LIN7C* or *BDNF* rather than *LGR4* (Supplementary Information).

We examined the association of the *LGR4* c.376C>T variant with low BMD measures at separate sites, and with other osteoporosis-related traits (Table 1). All three low BMD measures associate with c.376C>T; the strongest site-specific association is with low whole-body values, with OR = 6.45 ($P = 5.2 \times 10^{-8}$). An association is also observed with risk of osteoporotic fractures (OR = 3.12, $P = 0.00013$) and with the BMD definition of osteoporosis[8] (OR = 3.27,

[1]deCODE Genetics/Amgen, 101 Reykjavik, Iceland. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden. [3]Department of Biochemistry and Molecular Biology, University of Iceland, Biomedical Center, Faculty of Medicine, 101 Reykjavik, Iceland. [4]Garvan Institute of Medical Research, 2010 Sydney, Australia. [5]St Vincent's Hospital, 2010 Sydney, Australia. [6]University of New South Wales (UNSW), 2010 Sydney, Australia. [7]University of Notre Dame Australia (UNDA), 2010 Sydney, Australia. [8]Center for Clinical and Basic Research (CCBR), 2750 Ballerup, Denmark. [9]Department of Pathology, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland. [10]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. [11]Icelandic Cancer Registry, 105 Reykjavik, Iceland. [12]Laboratory in Mjodd (RAM), Icelandic Medical Center (Laeknasetrid), 109 Reykjavik, Iceland. [13]Department of Gastroenterology, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland. [14]Department of Surgery, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland. [15]Department of Orthopedic Surgery, Akureyri Hospital, 600 Akureyri, Iceland. [16]Institution of Health Science, University of Akureyri, 600 Akureyri, Iceland. [17]Department of Clinical Biochemistry, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland. [18]Department of Endocrinology and Metabolism, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland.

Table 1 | Associations of the LGR4 c.376C>T nonsense mutation with osteoporotic traits

Phenotype	OR (95% CI)	P value	N (N_{wg})* Cases	Controls
Low BMD†				
Hip, spine or whole body	4.30 (2.76–6.71)	1.3×10^{-10}	4,931 (4,188)	69,034 (58,632)
Whole body	6.45 (3.30–12.62)	5.2×10^{-8}	1,206 (1,159)	73,566 (70,699)
Hip	3.30 (1.89–5.76)	2.6×10^{-5}	2,804 (2,349)	70,100 (58,725)
Spine	3.81 (2.25–6.46)	6.9×10^{-7}	3,015 (2,538)	69,345 (58,374)
Osteoporosis BMD‡	3.27 (1.84–5.79)	5.0×10^{-5}	2,615 (2,162)	65,375 (54,050)
Any osteoporotic fractures	3.12 (1.74–5.59)	0.00013	2,668 (2,532)	72,036 (68,364)

Phenotype	β§	P value§	N (N_{wg})*
BMD			
Whole body	−0.75 (0.16)	1.6×10^{-6}	7,359 (6,896)
Hip	−0.49 (0.12)	6.2×10^{-5}	21,024 (17,171)
Spine	−0.52 (0.15)	3.8×10^{-5}	21,001 (17,193)

The results shown are for the A allele of the reference plus strand (hg18_chr11:27369242_A), presenting the nonsense T mutation of *LGR4* at codon 126. CI, confidence interval.
* N is the total number of individuals included in the analysis, and N_{wg} is the number of included individuals who have been chip typed.
†The low BMD phenotypes are defined as those BMD values that are below −1 s.d. from the mean.
‡World Health Organization definition of osteoporosis BMD; <−2.5 s.d. in young women at the spine or hip, uncorrected for age or weight[8].
§The effect β (the regression coefficient), and the standard error (s.e.) and P values from a linear regression of the age, sex and weight-adjusted BMD values on the mutation status. All P values are corrected for relatedness using the method of genomic controls (see Supplementary Information).

$P = 0.000050$). Furthermore, the association becomes stronger if a more extreme cut-off is used to define low BMD (Supplementary Table 5). Comparison of carriers and non-carriers of c.376C>T shows that the mean BMD value is 0.49–0.75 s.d. lower in carriers than in non-carriers, depending on site (Table 1 and Supplementary Table 6).

Analysis of 5,835 individuals with repeated BMD measurements did not show an association of *LGR4* c.376C>T with age-related changes in BMD (Supplementary Table 5), suggesting that the effect of the mutation is to reduce peak bone mass rather than increasing the rate of age-related bone loss. This is further supported by our observation that the strongest site-specific association is with low whole-body BMD, which is the site with the lowest remodelling rate. However, our data do not indicate that the mutation primarily associates with cortical rather than trabecular bone (Supplementary Table 5).

The *LGR4* gene encodes a cell-surface receptor that directly binds R-spondins[9–11], secreted agonists of the Wnt pathway[12]. The Wnt signalling pathway has been shown to regulate bone mass[13,14] and to be essential for differentiation of osteoblasts both in skeletogenesis and in the maintenance of the adult skeleton[15]. Furthermore, mutations in the *LRP5* and *SOST* genes, which take part in the Wnt pathway, have previously been identified in families with rare bone disorders[13], and common variants with small effects on BMD at these loci have been identified in genome-wide association studies[2,3,7]. On the basis of LGR4 potentiation of Wnt signalling, the lack of LGR4 should be reflected in less Wnt signalling in osteoblasts. Indeed, $Lgr4^{-/-}$ mice have delayed osteoblast differentiation and mineralization during embryonic bone formation. Postnatal bone remodelling (bone density, bone formation rate, osteoid formation) is also significantly impaired in these mice[16]. The *LGR4* c.376C>T mutation is predicted to truncate the LGR4 protein, normally composed of 951 amino acids, at position 126, thus fully disrupting its function (Supplementary Figs 3 and 4). We detected reduced levels of mutated *LGR4* messenger RNA isolated from white blood cells and adipose tissue of heterozygous c.376C>T carriers (Fig. 1a), consistent with a nonsense-mediated decay of transcripts containing nonsense mutations[17] (Fig. 1b and Supplementary Fig. 5).

To investigate whether the *LGR4* c.376C>T nonsense mutation is present in other populations, we screened the public Exome Variant Server database[18] (EVS), and directly genotyped the variant in two sample sets of Northern European descent; the PERF study of 3,032 Danish postmenopausal women[19] and 1,393 individuals in the Australian DOES study[20]. The mutation was neither present in the EVS database nor in the Australian samples. One carrier of the nonsense mutation was found in the Danish sample, who was subsequently shown to be of Icelandic origin (Supplementary Information). Investigation of the size of the haplotype background shared by the carriers indicates that the mutation was introduced into the Icelandic gene pool about

400 years ago (Supplementary Information). Although the c.376C>T mutation seems to be specific to the Icelandic population, two truncating (frameshift) mutations in *LGR4* that could mediate similar effects to c.376C>T were found in the EVS database. The combined frequency of these two truncating mutations in the EVS is similar to that of the c.376C>T mutation in Iceland. We genotyped these two truncating mutations in the PERF and DOES samples by Sanger sequencing. Neither mutation was found to be present in these samples.

Various abnormalities other than those related to bone formation and remodelling have been reported in *Lgr4* mutant mice[10,21–28]. To get a more detailed picture of the overall effect of the c.376C>T mutation, we investigated its effect on a variety of human diseases and other traits for which we have genotype and phenotype information (Supplementary Information). This revealed a complex pattern of similarities between phenotypes displayed by *Lgr4* mutant mice and the carriers of the c.376C>T mutation (Supplementary Table 7): reduced birth size (weight $P = 0.00026$, length $P = 0.0055$); lower weight ($P = 0.0064$); electrolyte disturbances (with reduced plasma sodium ($P = 0.00014$) and bicarbonate levels ($P = 0.0017$)); elevated potassium levels ($P = 0.00063$) (Table 2); and normal serum levels of calcium ($P = 0.25$) and phosphate ($P = 0.90$). Furthermore, c.376C>T associates with reduced testosterone levels ($P = 0.00022$) and late onset of menarche ($P = 0.0018$),
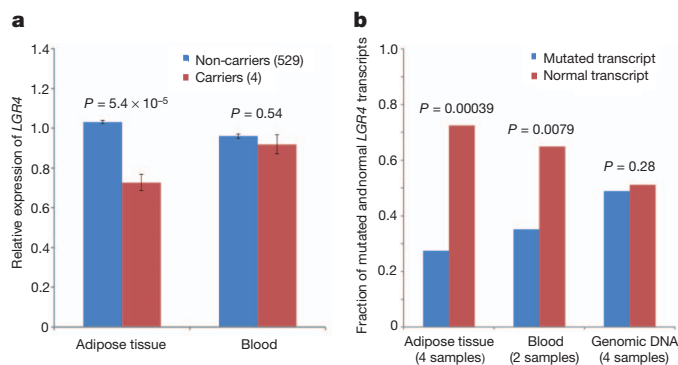

Figure 1 | Reduced expression of the *LGR4* transcript containing the c.376C>T mutation. a, Expression of *LGR4* in adipose tissue and blood in c.376C>T carriers and non-carriers, shown as 10 to the power of average mean log expression ratio (MLR). P values are from regression of the MLR on the carrier status, adjusting for age and sex, and differential cell counts for blood. Error bars show standard error of the mean. **b**, Proportion of mutated and normal *LGR4* transcripts in adipose tissue and white blood cells from heterozygous carriers of c.376C>T based on complementary DNA sequencing. Proportion of the two alleles determined by sequencing genomic DNA from heterozygous carrier is also shown. The P values are from a one-sample *t*-test.

Table 2 | The *LGR4* c.376C>T mutation is associated with a wide range of phenotypes

Phenotype	β (s.e.)	P value	Effect*	N (N_{wg})
Birth length	−0.799 (0.29)	0.0055	−0.03 cm	12,661 (1,485)
Birth weight	−1.046 (0.29)	0.00026	−0.46 kg	12,678 (1,485)
Weight	−0.237 (0.09)	0.0064	−3.0 kg	76,499 (65,461)
Sodium	−0.18 (0.05)	0.00014	−0.71 mmol l^{-1}	69,597 (63,194)
Potassium	0.16 (0.05)	0.00063	0.061 mmol l^{-1}	70,799 (64,152)
Bicarbonate	−0.34 (0.11)	0.0017	−1.2 mmol l^{-1}	14,563 (13,783)
Testosterone	−1.17 (0.32)	0.00022	−4.0 pg ml^{-1}	2,781 (2,556)
Age at menarche	0.38 (0.12)	0.0018	0.48 years	39,574 (35,757)

Phenotype	OR (95% CI)	P value	N (N_{wg}) Cases	N (N_{wg}) Controls
Gallbladder and biliary duct cancer	9.85 (2.61–37.2)	0.00074	320 (99)	39,680 (12,276)
SCC	2.81 (1.29–6.12)	0.0092	1,508 (958)	144,768 (91,968)

Association of the c.376C>T mutation with phenotypes in deCODE's database. Included is, for each trait, the effect estimate β and s.e. from linear regression of the trait on the number of copies of the mutation, or the OR and the 95% CI from logistic regression for a case–control analysis, and the corresponding P values. All P values have been adjusted for relatedness using the method of genomic control. For detailed comparison, see Supplementary Table 7.
*The estimated difference in age- and gender-adjusted trait values between carriers and non-carries of the mutation.

possibly reflecting defects in the development of the male reproductive tract[21] and developmental delays[26], respectively, as has been described for *Lgr4* mutant mice. Many other abnormalities have been described for *Lgr4* mutant mice[10,22,24,27,28], warranting further investigation of the *LGR4* c.376C>T carriers.

In addition to the phenotypes previously described for *Lgr4* mutant mice, we found the *LGR4* c.376C>T nonsense mutation to associate with an increased risk of squamous cell carcinoma of the skin (SCC) (OR = 2.81, P = 0.0092), and cancers of the gallbladder and biliary tract (OR = 9.85, P = 0.00074) (Table 2). The loss-of-function nature of the c.376C>T mutation suggests a tumour suppressor role for *LGR4* in these cancers. This is further supported by a predisposition to SCC in homozygous carriers of a loss-of-function mutation in the LGR4 ligand, R-spondin 1 (ref. 29). *Lgr4$^{-/-}$* mice lack gallbladders[23], demonstrating an essential role for *LGR4* in the development of this organ. Classical tumour suppressor genes are characterized by loss of the wild-type allele in tumours of heterozygous carriers (loss of heterozygosity (LOH)). We tested for LOH at the *LGR4* locus in DNA isolated from biliary tract tumours of four carriers of the c.376C>T mutation (Supplementary Information and Supplementary Fig. 6). The results indicate LOH in three out of the four carriers, suggesting that *LGR4* acts as a tumour suppressor gene in biliary tract cancers. Increased cancer risk has not been described in *Lgr4$^{-/-}$* mice; however, these mice have a short lifespan and the oldest mice analysed so far have not been older than 24 weeks, which may not be sufficient time for cancer development.

Through whole-genome sequencing of Icelanders, we have found a rare nonsense mutation, c.376C>T, within the *LGR4* gene that leads to termination of the LGR4 protein at position 126. Carriers of this mutation are at increased risk of various conditions, similar to the phenotypes that are observed in *Lgr4* mutant mice (Table 3). Our results highlight the value of human genome sequence information in the context of rich phenotypic information from which the effects of rare deleterious mutations can be directly assessed in humans, creating a human model of physiological disturbance or disease.

METHODS SUMMARY

The Icelandic, Danish and Australian BMD samples were described previously[3,19,20]. In the discovery analysis, cases were defined as those with adjusted BMD levels less that −1 s.d. from the mean at the hip (total hip), lumbar spine (L2–L4) or whole body, whereas the control group comprised individuals with BMDs above −1 s.d. and those with no BMD information available. Serum electrolytes and all other trait measurements were normalized to a standard normal distribution and adjusted for sex, year of birth and age at measurement. SCC and biliary tract cancer cases were identified in the nationwide Icelandic Cancer Registry, and self-reported information on age at menarche and menopause was retrieved from the Icelandic Cancer Registry.

About 34.2 million sequence variants (30.6 million SNPs and 3.6 million indels) identified through whole-genome sequencing (to an average sequencing depth of >10×) of 2,230 Icelanders using Illumina GAIIx and HiSeq2000 instruments, were imputed into 95,085 Icelanders genotyped with Illumina SNP chips and phased using long-range phasing. Familial imputation methods were used to impute the variants into 296,526 ungenotyped relatives of the 95,085 individuals using the nationwide Icelandic genealogy database. Association testing for case–control analysis was performed using logistic regression, matching controls to cases based on how informative the imputed genotypes were, whereas quantitative traits were tested using a generalized form of linear regression. Single SNP genotyping was carried out by the Centaurus (Nanogen) platform and deletions identified from the EVS were genotyped by Sanger sequencing.

Differences in expression of *LGR4* mRNA in blood and adipose tissue between carriers and non-carriers of c.376C>T were tested using multiple regression, adjusting for age and gender, and for differential cell counts in blood. Differences in the normalized amount of mutated versus wild-type *LGR4* transcripts in carriers of c.376C>T were tested using one-sample *t*-test.

Table 3 | Phenotypic similarities between human carriers of c.376C>T and *Lgr4* mutant mice

Human c.376C>T carriers	*Lgr4* mutant mice
Low BMD	Low BMD[16]
Osteoporotic fractures	Defective bone formation and remodelling[16]
Reduced birth size	Reduced birth size[16,26]
Reduced weight	Reduced weight[16,26]
Elevated serum potassium	Elevated serum potassium[25]
Reduced serum sodium	Reduced serum sodium[25]
Reduced serum bicarbonate	Reduced serum bicarbonate[25]
Serum calcium unchanged	Serum calcium unchanged[16]
Serum phosphate unchanged	Serum phosphate unchanged[16]
Late age at menarche	Developmental delay[16,25,26]
Reduced free testosterone	Defective male reproductive tract[21]
Gallbladder and biliary duct cancers	Lack of gallbladder and cystic duct[23]

For a more detailed comparison between the phenotypes shown by *Lgr4* mutant mice and carriers of c.376C>T, see Supplementary Table 7.

1. Styrkarsdottir, U. *et al.* Multiple genetic loci for bone mineral density and fractures. *N. Engl. J. Med.* **358,** 2355–2365 (2008).
2. Richards, J. B. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371,** 1505–1512 (2008).
3. Styrkarsdottir, U. *et al.* New sequence variants associated with bone mineral density. *Nature Genet.* **41,** 15–17 (2009).
4. Kung, A. W. *et al.* Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am. J. Hum. Genet.* **86,** 229–239 (2010).
5. Duncan, E. L. *et al.* Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet.* **7,** e1001372 (2011).
6. Rivadeneira, F. *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nature Genet.* **41,** 1199–1206 (2009).
7. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44,** 491 (2012).
8. Kanis, J. A. *et al.* A reference standard for the description of osteoporosis. *Bone* **42,** 467–475 (2008).

9. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proc. Natl Acad. Sci. USA* **108,** 11452–11457 (2011).
10. de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476,** 293–297 (2011).
11. Glinka, A. *et al.* LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling. *EMBO Rep.* **12,** 1055–1061 (2012).
12. Yoon, J. K. & Lee, J. S. Cellular signaling and biological functions of R-spondins. *Cell. Signal.* **24,** 369–377 (2012).
13. Ralston, S. H. & Uitterlinden, A. G. Genetics of osteoporosis. *Endocr. Rev.* **31,** 629–662 (2010).
14. Richards, J. B., Zheng, H. F. & Spector, T. D. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nature Rev. Genet.* **13,** 576–588 (2012).
15. Long, F. Building strong bones: molecular regulation of the osteoblast lineage. *Nature Rev. Mol. Cell Biol.* **13,** 27 (2012).
16. Luo, J. *et al.* Regulation of bone formation and remodeling by G-protein-coupled receptor 48. *Development* **136,** 2747–2756 (2009).
17. Balasubramanian, S. *et al.* Gene inactivation and its implications for annotation in the era of personal genomics. *Genes Dev.* **25,** 1–10 (2011).
18. NHLBI. Exome Sequencing Project. Exome Variant Server v.0.0.17 (NHLBI, Seattle, Washington, 2012).
19. Bagger, Y. Z. *et al.* Links between cardiovascular disease and osteoporosis in postmenopausal women: serum lipids or atherosclerosis per se? *Osteoporos. Int.* **18,** 505–512 (2007).
20. Nguyen, T. V., Sambrook, P. N. & Eisman, J. A. Sources of variability in bone mineral density measurements: implications for study design and analysis of bone loss. *J. Bone Miner. Res.* **12,** 124–135 (1997).
21. Mendive, F. *et al.* Defective postnatal development of the male reproductive tract in LGR4 knockout mice. *Dev. Biol.* **290,** 421–434 (2006).
22. Kato, S. *et al.* Leucine-rich repeat-containing G protein-coupled receptor-4 (LGR4, Gpr48) is essential for renal development in mice. *Nephron, Exp. Nephrol.* **104,** e63–e75 (2006).
23. Yamashita, R. *et al.* Defective development of the gall bladder and cystic duct in Lgr4- hypomorphic mice. *Dev. Dyn.* **238,** 993–1000 (2009).
24. Oyama, K., Mohri, Y., Sone, M., Nawa, A. & Nishimori, K. Conditional knockout of Lgr4 leads to impaired ductal elongation and branching morphogenesis in mouse mammary glands. *Sex Dev.* **5,** 205–212 (2011).
25. Wang, J. *et al.* GPR48 increases mineralocorticoid receptor gene expression. *J. Am. Soc. Nephrol.* **23,** 281–293 (2012).
26. Mazerbourg, S. *et al.* Leucine-rich repeat-containing, G protein-coupled receptor 4 null mice exhibit intrauterine growth retardation associated with embryonic and perinatal lethality. *Mol. Endocrinol.* **18,** 2241–2254 (2004).
27. Jin, C. *et al.* GPR48 regulates epithelial cell proliferation and migration by activating EGFR during eyelid development. *Invest. Ophthalmol. Vis. Sci.* **49,** 4245–4253 (2008).
28. Weng, J. *et al.* Deletion of G protein-coupled receptor 48 leads to ocular anterior segment dysgenesis (ASD) through down-regulation of Pitx2. *Proc. Natl Acad. Sci. USA* **105,** 6081–6086 (2008).
29. Parma, P. *et al.* R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nature Genet.* **38,** 1304 (2006).

LETTER

Crystal structure of the integral membrane diacylglycerol kinase

Dianfan Li[1], Joseph A. Lyons[1], Valerie E. Pye[1]†, Lutz Vogeley[1], David Aragão[1]†, Colin P. Kenyon[2], Syed T. A. Shah[1], Christine Doherty[1]†, Margaret Aherne[1] & Martin Caffrey[1]

Diacylglycerol kinase catalyses the ATP-dependent phosphorylation of diacylglycerol to phosphatidic acid for use in shuttling water-soluble components to membrane-derived oligosaccharide and lipopolysaccharide in the cell envelope of Gram-negative bacteria[1]. For half a century, this 121-residue kinase has served as a model for investigating membrane protein enzymology[1–6], folding[7,8], assembly[9–12] and stability[1,13]. Here we present crystal structures for three functional forms of this unique and paradigmatic kinase, one of which is wild type. These reveal a homo-trimeric enzyme with three transmembrane helices and an amino-terminal amphiphilic helix per monomer. Bound lipid substrate and docked ATP identify the putative active site that is of the composite, shared site type. The crystal structures rationalize extensive biochemical and biophysical data on the enzyme. They are, however, at variance with a published solution NMR model[14] in that domain swapping, a key feature of the solution form, is not observed in the crystal structures.

Diacylglycerol kinase (DgkA) is a unique enzyme, structurally and functionally. A solution NMR model of wild-type DgkA revealed a compact, domain-swapped trimer[14]. This, along with extensive biochemical and biophysical data, was used to rationalize its stability, folding, active site architecture, catalytic mechanism and substrate selectivity. Because it was a backbone-only model, with motional disorder over its first 25 residues, there was a need for a high-resolution

crystal structure. Traditional crystallization approaches had failed to produce quality crystals[15]. Given the success of the lipidic cubic phase (*in meso*) method with similarly sized proteins[16], it was considered a worthwhile crystallization strategy.

Initial trials yielded minute crystals with monoolein (9.9 MAG) as host lipid at 20 °C. Rounds of optimization with shorter chain mono-acylglycerols (MAGs)[17,18] using rationally designed mutant forms of the kinase provided quality crystals at 4 °C. Molecular replacement with the published model[14] as a template failed to provide a solution. To phase the X-ray data, seleno-methionine (Se-Met), as well as heavy-atom pre-labelling, co-crystallization and soaking, and engineered single-cysteine mutants for mercury labelling, were tested. Phases were obtained using Se-Met single-wavelength anomalous diffraction (SAD) from crystals of a stable active mutant, $\Delta 7$ (seven changes relative to wild type; Supplementary Figs 1 and 2). Structures for the wild type and a thermo-stabilized $\Delta 4$ (four changes relative to wild type)[13] form of DgkA were obtained by molecular replacement using the $\Delta 7$ structure (Methods). Here, we report on structures of the wild-type, $\Delta 4$ and $\Delta 7$ constructs obtained with 7.8 MAG. The highest resolution was for $\Delta 7$ DgkA at 2.05 Å (Supplementary Table 1).

The crystal structure of all three constructs reveals a trimer with layered packing[16] (Fig. 1 and Supplementary Fig. 3). Across constructs, subunits and folds are alike (Supplementary Figs 4–6). This, coupled
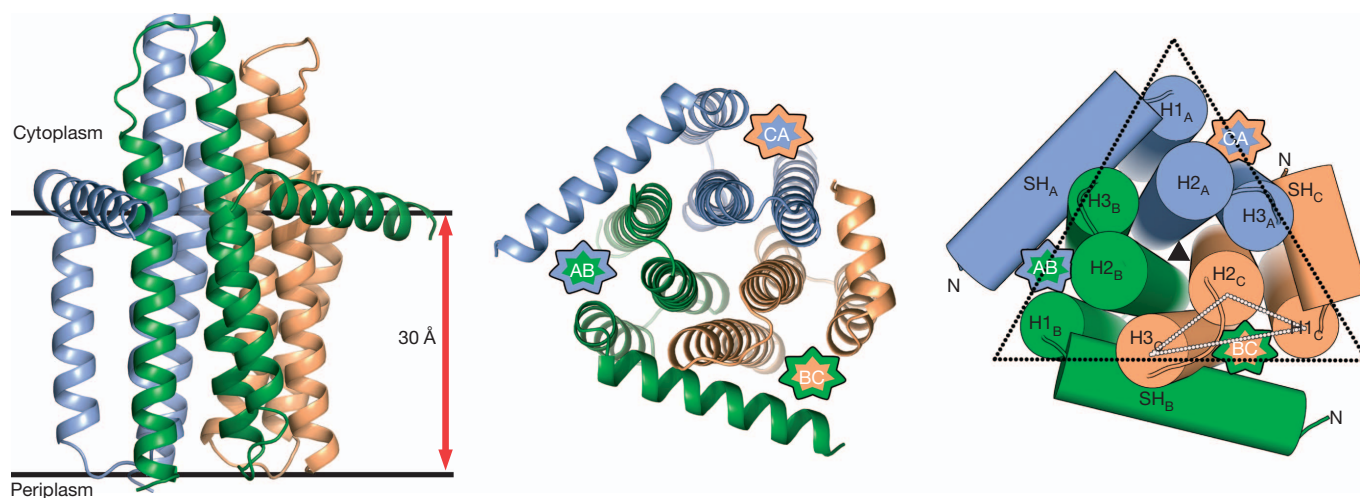

Figure 1 | Crystal structure of Δ4 DgkA. a, Structure (ribbon model) viewed from the membrane plane. Individual subunits are coloured blue, green and orange. Membrane boundaries are based on hydrophobic thickness calculations from the PPM server[28]. **b**, As in **a**, viewed from the cytoplasm. Putative active sites are designated on the basis of the identities of the subunits (A, B and C) contributing to each site and demarked with a seven-pointed star. **c**, As in **b**, with helices shown as cylinders and the structure notation

introduced. The solid black triangle marks the axis of an approximate three-fold symmetry. Dotted triangles define the transmembrane helical core of the trimer (black) and the arrangement of transmembrane helices (H1–H3) in individual subunits (white). The N terminus in the model of each subunit is identified. It leads directly into an amphiphilic surface helix (SH), the length of which varies between subunits.

[1]School of Biochemistry and Immunology & School of Medicine, Trinity College Dublin, Dublin 2, Ireland. [2]CSIR, Biosciences, Meiring-Naude Road, Pretoria 0184, Gauteng, South Africa. †Present addresses: Clare Hall Laboratories, Cancer Research UK, Blanche Lane, South Mimms, Hertfordshire EN6 3LD, UK (V.E.P.); Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria VIC 3168, Australia (D.A.); The Matrix Biology Group, Kennedy Institute of Rheumatology, University of Oxford, 65 Aspenlea Road, Hammersmith, London W6 8LH, UK (C.D.).
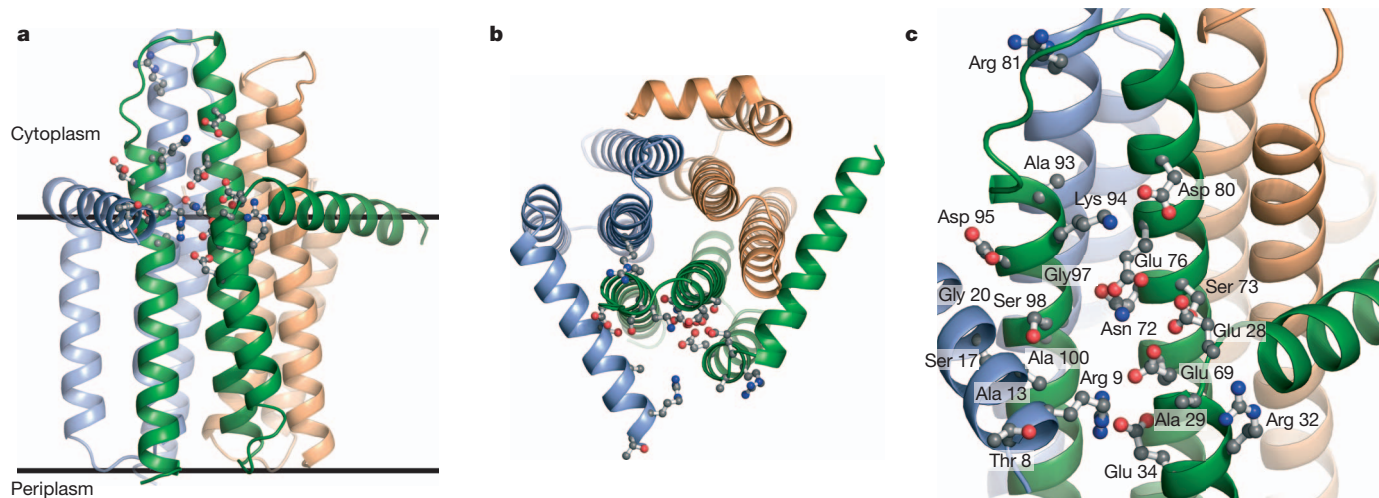
Figure 2 | Rationalizing functional biochemistry with the crystal structure of DgkA. Residues in DgkA, in which mutations to cysteine reduce kinase activity to 7% or less that of wild-type activity[14], are mapped onto the crystal structure of Δ4 DgkA. **a**, View in the membrane plane. The protein (ribbon model) is displayed with chains A, B and C coloured blue, green and orange, respectively. Residues are shown in ball and stick and colour-coded according to atom type (nitrogen, blue; oxygen, red; carbon, grey). **b**, View from the cytoplasm. **c**, An expanded view of the putative active site (asAB) of the enzyme, in which crucial kinase activity residues[14] are clustered.

with the fact that the three constructs are active (Supplementary Fig. 2), indicates that all three are structurally and functionally similar. Δ4 provided the most complete model and is used for the structure description that follows.

Each subunit (identified A–C) within a trimer forms a bundle of three transmembrane helices, H1–H3 (Fig. 1 and Supplementary Fig. 7). When viewed from the cytosol, the centres of mass of the three helices within a subunit roughly describe an isosceles triangle (Fig. 1c). H1 is the shortest at 19 residues. It is preceded by an N-terminal amphiphilic surface helix (Supplementary Fig. 6b), which is expected to reside on the cytosolic side of the membrane. H2 and H3 extend into the cell by about ten residues and are connected by a ∼five-residue cytosolic loop, likely to be quite mobile (Supplementary Fig. 5). On the other side of the membrane, H1 and H2 are connected by a four-residue periplasmic loop.

The three subunits are arranged around an approximate three-fold symmetry axis that passes through the centre of the trimer normal to the membrane plane (Fig. 1c). The core of the trimer is made up of H2 from each subunit forming a parallel three-helix bundle. Extending away from the core, H1 and H3 from each subunit form the sides of an equilateral triangle that inscribes the trimer with H1 occupying apex positions. Viewed from the cytosol, the surface helix, which is N-terminal to H1, angles away from the trimer core contacting H3 from an adjacent subunit. Together, the surface helix N terminus of one subunit and H1–H3 of an adjacent subunit create the putative active site (Fig. 1). Thus, for example, active site AB (asAB) is formed by surface helix (SH)$_A$, H1$_B$, H2$_B$ and H3$_B$, consistent with the composite, shared site model[4].

Separate studies have shown that the DgkA trimer has three active sites[4], with moderate heterollostery[1,3] (Supplementary Discussion). The enzyme has one amphiphilic substrate, diacylglycerol (DAG), and a water-soluble substrate, Mg^{2+}-ATP. Thus, DgkA catalyses phosphoryl transfer expected to take place at a polar/apolar interface. Given that the reaction involves direct phosphorylation[3] via a pentavalent transition state, the two substrates must come into close proximity with the γ-phosphate of ATP next to the primary hydroxyl of DAG. There is only one location (per active site) on the kinase where this can happen. That is in the pocket created by the surface helix of one subunit and the polar/apolar regions of H1–H3 of an adjacent subunit (Fig. 1). This prediction matches well the active site residues mapped in previous studies[1,14] (Fig. 2).

DgkA is a promiscuous enzyme. It can work with different hydroxy lipids as substrates, including the MAGs used to create the mesophase for crystallization[19–21]. One of the putative active sites in Δ7 includes density well described by MAG (Fig. 3 and Supplementary Fig. 8). The lipid head-group resides at the predicted membrane polar/apolar interface. Its acyl chain extends away from the interface into the hydrophobic membrane core next to H3. This tentatively identifies the lipid substrate half of a DgkA active site.

DgkA is a unique kinase with a distinctive active site. It has no recognizable nucleotide sequence or structural binding motifs[22,23]. With a view to identifying the ATP-binding site, crystals of Δ7 DgkA were soaked with the ATP analogue, adenylylmethylenediphosphonate (AMPPCP). This caused the crystals to 'dissolve' (Supplementary Fig. 9). Additional soaking with ATP, ADP, AMP, ATPγS, AMPPNP and dATP, but not with GTP, CTP, UTP or TTP, led to crystal dissolution (Supplementary Fig. 9). Thus, although a liganded structure did not materialize, the result is consistent with an adenine nucleotide-induced conformational and/or packing change
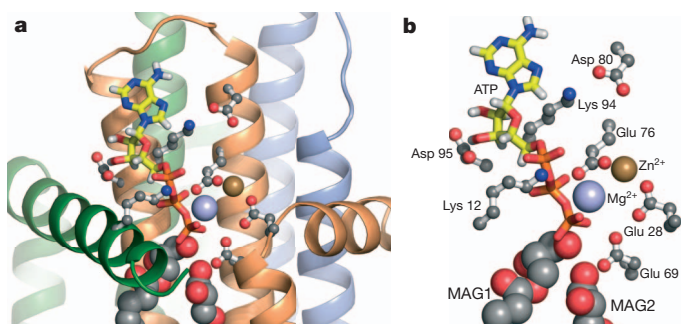

Figure 3 | Putative active site of DgkA complete with lipid and Mg^{2+}-ATP substrates and activating zinc. a, A putative active site of Δ4 DgkA is shown superposed with 7.8 MAG lipid substrate, the zinc ion from Δ7 DgkA, and Mg^{2+}-ATP from the docking calculation. The ATP (thick sticks, atoms coloured by type) extends ∼20 Å from the cytoplasmic loop between H2 and H3 to the polar headgroup of the lipid substrate (large spheres, atoms coloured by type) at the polar/apolar interface of the membrane. Residues in the vicinity of the substrates that are known to be crucial for kinase activity[14] are shown in ball and stick, as in Fig. 2. Zinc and magnesium are shown as brown and blue spheres, respectively. **b**, Expanded view of **a**, with the ribbon model removed for clarity.

in the crystalline kinase. These findings support the view that the kinase adopts a functional form in the crystal.

In the absence of a nucleotide-bound structure for DgkA, ATP has been docked into the Δ7 model with the adenyl moiety interacting with the cytosolic loop (Fig. 3, Supplementary Discussion and Supplementary Fig. 10). The triphosphate extends in the direction of the lipid-binding region of the putative active site, occupied by a lipid molecule. The observation of distinct binding sites, one (modelled) for ATP and one for the lipid substrate, is consistent with the finding that the mechanism of phosphoryl transfer is of the random-equilibrium type, in which the two substrates bind independently of one another[3,24].

For full activation, DgkA requires free divalent cation (magnesium, manganese, cadmium, zinc or cobalt)[25]. Here, zinc was found coordinated by the carboxyls of Glu 28 and Glu 76 (Supplementary Fig. 11), which are both vital to catalytic activity[14]. Superposing this onto the docked ATP-Δ7 DgkA structure positions the zinc 4.9 Å from the γ-phosphate (Fig. 3), where it may have a role in substrate placement for reaction.

The reaction product, phosphatidic acid, with its relatively bulky and charged phosphate, is expected to leave the active site for the surrounding membrane via the opening between the surface helix and H1 (Fig. 1). This may be facilitated electrostatically by Glu 69 and Glu 76 in H2 creating a push, and by Arg 9 and Lys 12 in the surface helix creating a pull on the anionic product (Supplementary Fig. 12). All four residues are important for activity[14].

A previous study[4], using subunit mixing/complementarity experiments, provided evidence for shared sites in DgkA. Ala 14 and Glu 76 were identified as active site residues contributed from one subunit, with Glu 69, Asn 72 and Lys 94 deriving from another.

The current structures show that the active sites are indeed composed of residues from two subunits consistent with the shared sites model (Fig. 1). For purposes of discussion, we focus on just one active site (asBC) (Supplementary Fig. 13). Highlighting the shared sites residues identified previously[4], in this active site it is clear that Glu 69 and Asn 72 on H2$_C$ and Lys 94 on H3$_C$ are on one half-site, whereas Ala 14 on SH$_B$ is on the other. Glu 76, however, is on H2. If it is to be on the same half-site as Ala 14 then it must be on H2$_B$, which is in asAB, not asBC. The current structure is therefore consistent with Ala 14 residing on one of the shared half-sites, and Glu 69, Asn 72 and Lys 94 on the other. It is for these that maximum complementation was observed. For Glu 76, the experimental data were less convincing[4].

In the previous study[4], Asp 95 was shown to influence at least two active sites. For purposes of rationalizing this observation we focus on asBC and asAB (Supplementary Fig. 13). Asp 95 in H3$_C$ contacts SH$_B$, which bridges two active sites. Its N terminus contributes to asBC, whereas its carboxy terminus, which extends into H1$_B$, contributes to asAB. Thus, modifying Asp 95 is likely to affect SH$_B$ which, in turn, will influence both asBC and asAB, consistent with the experimental observation[4].

The only other model for DgkA with which to compare our crystal structures was determined using solution NMR[14] and the two are different (Fig. 4). The biggest disparity relates to domain swapping, which is present in the solution but not in the crystal structure. In the former, H3 from one subunit contacts H1 and H2 from an adjacent subunit in classic domain swapped fashion (Fig. 4a). The active site, in turn, is mapped to the space created by H1 and H3 from one subunit and H2 from another. As a result of these disparate quaternary structures, the architecture and chemistry of the active sites in the two models differ markedly (Fig. 4 and Supplementary Discussion). Notably, both are consistent with the shared sites model but for different reasons.

From the quality of the diffraction data and the corresponding refinement statistics (Supplementary Table 1) it is evident that the structures reported here represent what exists in the crystal. We must consider then if the crystal structure described is physiological, because it is possible for crystal lattice restraints to impose unusual and perhaps even unnatural conformations. However, we report here two different
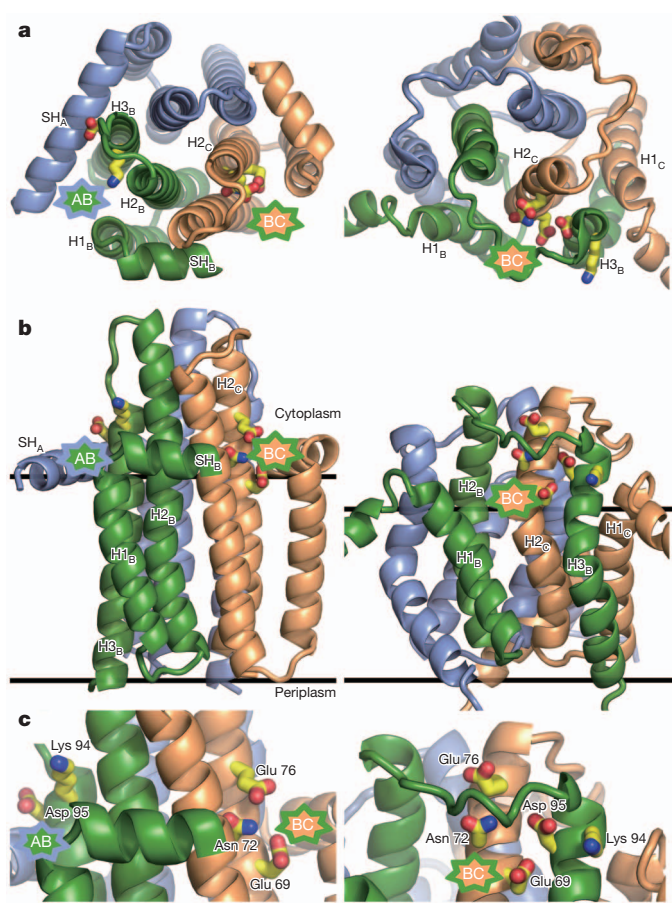

Figure 4 | Comparison of the crystal and solution NMR structures of DgkA. a, b, A view of the crystal (left) and solution structures (right) from the cytoplasm (**a**) and the membrane (**b**) plane. **c,** An expanded view of the putative active site region. The five most highly conserved residues and how they contribute to a single active site (asBC) in the solution structure are shown (right panels), as described previously[14]. In the crystal structure, three of these five residues (Glu 69, Asn 72 and Glu 76) contribute to one active site, asBC; two (Lys 94 and Asp 95) are in asAB. The surface helix of the solution model, which is ill-defined, has been truncated for space reasons. SH$_B$ of the crystal structure has been trimmed at residue 18 to reveal the active site (asBC). Membrane boundaries are on the basis of hydrophobic thickness calculations from the PPM server[28].

crystal-packing arrangements and the overall fold of the protein is the same in both. Furthermore, we have obtained a crystal structure of the enzyme with lipid-substrate-binding sites occupied, modelled ATP into the binding site, found that the enzyme is active under crystallization conditions and after crystal dissolution (Supplementary Fig. 14), and have shown that nucleotide and ATP analogue soaking cause crystals to dissolve, suggesting an adenine-specific, ligand-induced conformational change. All point to a protein with a functional form in the crystal. The solution structure, by contrast, is a low-resolution, composite model that traces only the protein backbone. It was arrived at using connectivity and distance spectroscopic and cross-linking measurements, the tolerances of which can be considerable[14] (Supplementary Discussion, Supplementary Table 2 and Supplementary Fig. 15). Presumably, long-range distance restraints, as defined by residual dipolar couplings and paramagnetic resonance enhancements that have an inherent ambiguity, were used to assist in positioning swapped domains in the micellarized trimer. Because subunit and atom ambiguity in symmetric homo-oligomeric assemblies are complications inherent to structure determination by NMR[26,27], the possibility that the solution structure of DgkA is so affected cannot be ruled out.

DgkA is a unique kinase of which a detailed catalytic mechanism has not been established. With the crystal structures reported here, a rational testing of hypotheses relating to mechanism can now be undertaken. DgkA is probably activated *in vivo* under hypo-osmotic conditions[3]. Activation contributes to enhanced production of the osmoregulant, membrane-derived oligosaccharide. The physico-chemical properties of membranes change in response to osmotic stress. We speculate that the surface helix connected to H1 by a short hinge (Fig. 1) acts as a sensor of osmotic stress by responding to bilayer thickness and lateral pressure. Given that the surface helix is a part of the putative active site, adjustments in its conformation and position relative to H1, and to the rest of the trimer as the membrane responds to hypo-osmotic stress, will alter active site architecture and lead to enhanced kinase activity.

METHODS SUMMARY

Procedures for the expression, isolation, purification and functional assay of wild-type, Δ4 and Δ7 DgkA are described in the Methods. Crystals were grown in glass sandwich plates at 4 °C by the *in meso* method using 7.8 MAG[17,18] as the host lipid. Crystals were harvested and snap-cooled directly in liquid nitrogen.

X-ray diffraction data were collected at beamline 23-ID-B (GM/CA-CAT) of the Advanced Photon Source, and the microfocus beamline I24 of the Diamond Light Source. Full details of diffraction data collection and processing, and structure solution and refinement are given in Supplementary Table 1 and are described in the Methods.

Full Methods and any associated references are available in the online version of the paper.

1. Van Horn, W. D. & Sanders, C. R. Prokaryotic diacylglycerol kinase and undecaprenol kinase. *Annu. Rev. Biophys.* **41,** 81–101 (2012).
2. Schneider, E. G. & Kennedy, E. P. Phosphorylation of ceramide by diglyceride kinase preparations from *Escherichia coli. J. Biol. Chem.* **248,** 3739–3741 (1973).
3. Badola, P. & Sanders, C. R. *Escherichia coli* diacylglycerol kinase is an evolutionarily optimized membrane enzyme and catalyzes direct phosphoryl transfer. *J. Biol. Chem.* **272,** 24176–24182 (1997).
4. Lau, F. W., Chen, X. & Bowie, J. U. Active sites of diacylglycerol kinase from *Escherichia coli* are shared between subunits. *Biochemistry* **38,** 5521–5527 (1999).
5. Pilot, J. D., East, J. M. & Lee, A. G. Effects of bilayer thickness on the activity of diacylglycerol kinase of *Escherichia coli. Biochemistry* **40,** 8188–8195 (2001).
6. Lahiri, S., Brehs, M., Olschewski, D. & Becker, C. F. Total chemical synthesis of an integral membrane enzyme: diacylglycerol kinase from *Escherichia coli. Angew. Chem. Int. Edn Engl.* **50,** 3988–3992 (2011).
7. Sanders, C. R. *et al. Escherichia coli* diacylglycerol kinase is an alpha-helical polytopic membrane protein and can spontaneously insert into preformed lipid vesicles. *Biochemistry* **35,** 8610–8618 (1996).
8. Gorzelle, B. M. *et al.* Reconstitutive refolding of diacylglycerol kinase, an integral membrane protein. *Biochemistry* **38,** 16373–16382 (1999).
9. Vinogradova, O., Badola, P., Czerski, L., Sonnichsen, F. D. & Sanders, C. R. *Escherichia coli* diacylglycerol kinase: a case study in the application of solution NMR methods to an integral membrane protein. *Biophys. J.* **72,** 2688–2701 (1997).
10. Nagy, J. K., Lau, F. W., Bowie, J. U. & Sanders, C. R. Mapping the oligomeric interface of diacylglycerol kinase by engineered thiol cross-linking: homologous sites in the transmembrane domain. *Biochemistry* **39,** 4154–4164 (2000).
11. Wen, J., Chen, X. & Bowie, J. U. Exploring the allowed sequence space of a membrane protein. *Nature Struct. Biol.* **3,** 141–148 (1996).
12. Smith, R. L., O'Toole, J. F., Maguire, M. E. & Sanders, C. R. Membrane topology of *Escherichia coli* diacylglycerol kinase. *J. Bacteriol.* **176,** 5459–5465 (1994).
13. Zhou, Y. & Bowie, J. U. Building a thermostable membrane protein. *J. Biol. Chem.* **275,** 6975–6979 (2000).
14. Van Horn, W. D. *et al.* Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* **324,** 1726–1729 (2009).
15. Lorch, M. *et al.* How to prepare membrane proteins for solid-state NMR: A case study on the α-helical integral membrane protein diacylglycerol kinase from *E. coli. ChemBioChem* **6,** 1693–1700 (2005).
16. Caffrey, M., Li, D. & Dukkipati, A. Membrane protein structure determination using crystallography and lipidic mesophases: recent advances and successes. *Biochemistry* **51,** 6266–6288 (2012).
17. Caffrey, M., Lyons, J., Smyth, T. & Hart, D. J. Monoacylglycerols: The workhorse lipids for crystallizing membrane proteins in mesophases. *Curr. Top. Membr.* **63,** 83–108 (2009).
18. Li, D., Shah, S. T. A. & Caffrey, M. Host lipid and temperature as important screening variables for crystallizing integral membrane proteins in lipidic mesophases. Trials with diacylglycerol kinase. *Cryst. Growth Des.* http://dx.doi.org/10.1021/cg400254v (2013).
19. Walsh, J. P., Fahrner, L. & Bell, R. M. *sn*-1,2-diacylglycerol kinase of *Escherichia coli*. Diacylglycerol analogues define specificity and mechanism. *J. Biol. Chem.* **265,** 4374–4381 (1990).
20. Bohnenberger, E. & Sandermann, H. Jr. Lipid dependence of diacylglycerol kinase from *Escherichia coli. Eur. J. Biochem.* **132,** 645–650 (1983).
21. Li, D. & Caffrey, M. Lipid cubic phase as a membrane mimetic for integral membrane protein enzymes. *Proc. Natl Acad. Sci. USA* **108,** 8639–8644 (2011).
22. Matte, A. & Delbaere, L. T. J. ATP-binding motifs. *eLS* http://dx.doi.org/10.1002/9780470015902.a0003050.pub2 (2010).
23. Krissinel, E. On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* **23,** 717–723 (2007).
24. Ullrich, S. J., Hellmich, U. A., Ullrich, S. & Glaubitz, C. Interfacial enzyme kinetics of a membrane bound kinase analyzed by real-time MAS-NMR. *Nature Chem. Biol.* **7,** 263–270 (2011).
25. Walsh, J. P. & Bell, R. M. *sn*-1,2-diacylglycerol kinase of *Escherichia coli*. Mixed micellar analysis of the phospholipid cofactor requirement and divalent cation dependence. *J. Biol. Chem.* **261,** 6239–6247 (1986).
26. Martin, J. W., Yan, A. K., Bailey-Kellogg, C., Zhou, P. & Donald, B. R. A geometric arrangement algorithm for structure determination of symmetric protein homo-oligomers from NOEs and RDCs. *J. Comput. Biol.* **18,** 1507–1523 (2011).
27. Martin, J. W., Yan, A. K., Bailey-Kellogg, C., Zhou, P. & Donald, B. R. A graphical method for analyzing distance restraints using residual dipolar couplings for structure determination of symmetric protein homo-oligomers. *Protein Sci.* **20,** 970–985 (2011).
28. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40,** 370–376 (2012).

METHODS

Protein. The *dgkA* wild-type gene and a *dgkA* mutant gene encoding for a thermostable quadruple mutant Δ4 DgkA (Ile53Cys, Ile70Leu, Met96Leu and Val107Asp)[13] were synthesized (Genescript) and cloned into pTrcHisB using NcoI and EcoRI restriction enzyme sites. A septuple variant, Δ7 DgkA (Ala41Cys, Cys46Ala, Ile53Val, Ile70Leu, Met96Leu, Val107Asp, Cys113Ala) was generated using PCR-based site-directed mutagenesis and verified by DNA sequencing.

DgkA native proteins were produced as described[21] with an additional size-exclusion chromatography step. Se-Met labelling of Δ7 DgkA was performed following an established protocol using the methionine auxotroph B893(DE3) strain in M9 minimal media supplemented with Se-Met[29]. Selenium incorporation into DgkA proteins was verified by mass spectrometry.

Enzyme assays. For the stability assay of DgkA proteins, a coupled assay was used to monitor kinase activity. The consumption of ATP in the phosphorylation of monoolein was coupled through the sequential action of pyruvate kinase and lactate dehydrogenase to the oxidation of NADH monitored by a change in absorbance (A) at 340 nm, as described[3,21]. The assay mix contained 21 mM decyl-maltoside, 7.2 mM monoolein (as substrate), 1.7 mM cardiolipin, 0.1 mM EGTA, 0.1 mM EDTA, 1 mM phosphenolpyruvate, 3 mM ATP, 15 mM magnesium acetate, 50 mM LiCl, 0.2 mM dithiothreitol (DTT), 0.3 mM NADH, 20 U ml^{-1} each of pyruvate kinase and lactate dehydrogenase, 75 mM PIPES, pH 6.9. The assays were run at 0.25 μg DgkA protein per millilitre in 384-well plates.

To assay Δ7 DgkA activity under crystallization conditions, 20 μl mesophase, formed by mixing 10 μl DgkA solution (12 mg ml^{-1}), and 10 mg 7.8 MAG, was bathed in 1 ml crystallization precipitant (4.5% (v/v) 2-methyl-2,4-pentanediol (MPD), 0.1 M sodium chloride, 60 mM magnesium acetate, 50 mM sodium citrate, pH 5.6) supplemented with 10 mM ATP. After incubating at 4 °C for 0, 0.5, 1, 2, 3, 5 and 10 h, 3 μl of the bathing solution was removed and transferred to 60 μl ADP assay mix (0.2 mM NADH, 1 mM phosphenolpyruvate, 15 mM magnesium chloride, 20 U ml^{-1} each of pyruvate kinase and lactate dehydrogenase, 75 mM PIPES, pH 6.9) in a 384-well plate. After incubation for 10 min at 30 °C, the drop in $A_{340\,nm}$ of the well contents was used to calculate activity. Control measurements were made using protein-free buffer. Assays were run in triplicate.

Crystallization. *In meso* crystallization trials began with the reconstitution of the protein into the bilayer of the lipidic cubic mesophase. This was done following a standard protocol[30]. The protein solution at 12 mg ml^{-1} was homogenized with 7.8 MAG in equal parts by volume using a coupled syringe device at room temperature (20–22 °C). Crystallization trials were set up by transferring 50 nl of the protein/lipid cubic mesophase onto a silanized 96-well glass crystallization plate, which was subsequently covered with 800 nl precipitant solution using *in meso* robots[31]. Wells were sealed with a glass coverslide. The glass sandwich plates were stored in a walk-in refrigerator at 4–6 °C for crystal growth. Crystallization progress was monitored using normal and cross-polarized light microscopy (Eclipse E 400 Pol, Nikon) at 4–6 °C. Precipitant solutions consisted of 4–6% (v/v) MPD, 0.1 M sodium chloride, 0.1 M lithium nitrate, 60 mM magnesium acetate, and 50 mM sodium citrate, pH 5.6 for wild-type and Δ4 DgkA. Lithium nitrate was not used for Δ7 DgkA crystallization. Bipyramid shaped crystals grew to a maximum size of $50 \times 50 \times 50$ μm^3 and $50 \times 75 \times 100$ μm^3 in 45–60 days for wild-type and Δ4 DgkA, respectively. For Δ7 DgkA, rectangular crystals grew to a maximum size of $10 \times 50 \times 150$ μm^3 in 45–60 days (Supplementary Fig. 9). Wells were opened with a tungsten-carbide glass cutter (Silverline) and the crystals were harvested using 30–100 μm micromounts (MiTeGen) at 4–6 °C. Crystals were snap-cooled directly in liquid nitrogen[32].

Crystal soaking *in situ* was carried out by cutting a window in the coverglass of the *in meso* crystallization sandwich plate and injecting through it a solution containing nucleotide, nucleotide analogue or glutaraldehyde into the precipitant solution bathing the mesophase. When a second soaking was required excess bathing solution was wicked away with tissue paper[32], fresh solution with or without additive was added through the window, and the well was sealed with ClearSeal film (Hampton). Crystals were harvested as described above.

Data collection and processing. X-ray diffraction data were collected on the 23-ID-B beamline of the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) at the Advanced Photon Source (APS), Argonne, Illinois, and the I24 beamline at the Diamond Light Source (DLS), Didcot, Oxford. Data were collected using a 10-μm collimated minibeam at the GM/CA-CAT[33], whereas a 10-μm microfocus beam was used at the DLS[34]. Diffraction images taken with a tenfold attenuated beam were used to locate and centre on highly ordered regions of the crystal[35]. For the high-resolution data set collected on crystals of Δ7 DgkA, the data reduction strategy involved combining a complete low-resolution (3.0 Å) data set recorded from a single crystal with 18 high-resolution 10° wedges of data collected from multiple crystals.

Data were reduced with xia2 (ref. 36) using XDS[37], XSCALE and SCALA[38]. Optimum data wedges were identified by data quality and isomorphic unit cell

parameters. These data were scaled in XSCALE and merged in SCALA (Supplementary Table 1). Similarly, a 200-fold redundant anomalous Se-Met data set to 3.0 Å resolution was obtained by merging 19 data sets. Complete data sets for Δ4 DgkA and wild type were obtained from single crystals that were reduced with xia2 as outlined above.

Structure solution and refinement. Molecular replacement with the published NMR model[14] as a template failed to provide a solution. Wide-search molecular replacement[39] against every known protein fold, and BALBES[40], also failed to provide a solution. The structure of Δ7 DgkA was determined by SAD phasing using the anomalous contributions of the Se-Met derivative[41]. Seventeen potential selenium sites were identified and refined with SHELX CDE[42], yielding a pseudo-free correlation coefficient of 65%. Phases were extended to the native 2.05 Å data set using Phenix[43]. The initial model was built with Phenix.autobuild and yielded an initial model of 440 residues that gave an R_{free} value of 0.35. It was evident that the asymmetric unit contained two trimers of DgkA. The final structure was attained by iterative model building in Coot[44] and refinement with Phenix.refine[43]. Optimal weights between X-ray target and stereochemistry were assigned automatically in Phenix.refine. Density modification and solvent flattening were recalculated on a regular basis until the maps showed minimal improvements. Se-Met phases were included throughout most of the refinement but were then excluded for the final rounds of refinement owing to marginal anisotropy between the Se-Met and native data. Non-crystallographic symmetry (NCS) restraints were not used as the polypeptide chains varied to a reasonable extent within the asymmetric unit. TLS (translation/libration/screw) groups were used in the latter stages of refinement, two groups per chain, as identified using the TLSMD server[45].

Initial phases for Δ4 DgkA were obtained by molecular replacement as performed by the program Phaser[46] using the more complete trimer of the Δ7 DgkA model (chains A–C). In subsequent cycles of iterative model building and refinement the programs Coot and O (ref. 47) were used for model building, whereas the programs Buster[48] and Phenix were used for refinement in the initial and later stages, respectively. Refinement in Buster was performed with automatic NCS determination, whereas NCS was not used for refinement in Phenix. Four TLS groups per chain were generated using the TLSMD web server and used by both refinement programs.

Initial phases for wild-type DgkA were obtained as above for Δ4 DgkA using the Δ4 DgkA model. The program Coot was used for model building, and the program Phenix was used for refinement. NCS was used throughout the refinement. TLS groups (nine for chain A, four for chain B, and two for chain C) were generated using Phenix and were used throughout the refinement.

Lipid, crystallization reagents and unassigned density. Electron density for a total of 13 lipid molecules (both enantiomers of 7.8 1-MAG) were identified in a composite omit $2F_o - F_c$ map of Δ7 DgkA. Acetate and citrate molecules were similarly identified and modelled in the electron density map. Tentatively modelled lipid and solvent molecules were removed if no clear hydrogen bonding was present and/or the model was not justified by the electron density.

Zinc. The location of a single metal ion in the asymmetric unit was determined in a $F_o - F_c$ map of Δ7 DgkA. The ion identity was confirmed by X-ray fluorescence spectroscopy of crystals and an anomalous difference map generated from X-ray diffraction data collected at the zinc edge (9.6767 keV; Supplementary Table 1 and Supplementary Fig. 11).

Water. In the Δ7 DgkA model, water molecules were assigned based on sigma-A-weighted $2F_o - F_c$ electron density maps contoured at 1σ using standard geometrical and chemical restraints, for a total of 188 water molecules per asymmetric unit.

Electrostatics. The electrostatic surface potentials for DgkA were calculated using the Advanced Poisson-Boltzmann Solver (APBS) plugin, as implemented in PyMol, using default parameters[49].

Docking. ATP is ~20 Å long, equalling the distance between the cytosolic loop and the primary hydroxyl of the putative lipid substrate in Δ7 DgkA. ATP with varying α/β/γ-phosphate protonation was docked into energy minimized Δ7 DgkA (Lys 12 unprotonated, Mg^{2+} coordinated to Asn 72 and Glu 76) using LibDock[50]. Poses with the highest LibDock scores were energy minimized to zero energy over 2,000 steps using the Smart minimizer with a Generalized Born with Molecular Volume implicit solvent model (Discovery Studio 3.5, Accelrys). The optimum pose was selected based on coordination with the cytosolic loop, lipid substrate, Mg^{2+}, and putative active site residues, Lys 12 and Lys 94.

Protein figures were generated and rendered in PyMOL.

29. Doublié, S. Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol.* **363,** 91–108 (2007).
30. Caffrey, M. & Porter, C. Crystallizing membrane proteins for structure determination using lipidic mesophases. *J. Vis. Exp.* **45,** e1712 (2010).
31. Li, D., Boland, C., Walsh, K. & Caffrey, M. Use of a robot for high-throughput crystallization of membrane proteins in lipidic mesophases. *J. Vis. Exp.* **67,** e4000 (2012).

32. Li, D., Boland, C., Aragao, D., Walsh, K. & Caffrey, M. Harvesting and cryo-cooling crystals of membrane proteins grown in lipidic mesophases for structure determination by macromolecular crystallography. *J. Vis. Exp.* **67,** e4001 (2012).

33. Fischetti, R. F. *et al.* Mini-beam collimator enables microcrystallography experiments on standard beamlines. *J. Synchrotron Radiat.* **16,** 217–225 (2009).

34. Evans, G., Axford, D. & Owen, R. L. The design of macromolecular crystallography diffraction experiments. *Acta Crystallogr. D* **67,** 261–270 (2011).

35. Cherezov, V. *et al.* Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm size X-ray synchrotron beam. *J. R. Soc. Interface* **6** (suppl. 5), S587–S597 (2009).

36. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).

37. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).

38. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).

39. Stokes-Rees, I. & Sliz, P. Protein structure determination by exhaustive search of Protein Data Bank derived databases. *Proc. Natl. Acad. Sci. USA* **107,** 21476–21481 (2010).

40. Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D* **64,** 125–132 (2008).

41. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).

42. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66,** 479–485 (2010).

43. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).

44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).

45. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39,** 109–111 (2006).

46. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

47. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).

48. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60,** 2210–2221 (2004).

49. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).

50. Diller, D. J. & Merz, K. M. Jr. High throughput docking for library design and library prioritization. *Proteins* **43,** 113–124 (2001).

CAREERS

VISUAL MOZART/IMAGEZOO/CORBIS

TRAVELLING

Here, there and everywhere

Scientists often criss-cross the globe in pursuit of their careers. The practice can be rewarding — and taxing.

BY NEIL SAVAGE

Ed Kearns did not travel much when he was younger. He grew up in Syracuse, New York, and in 1978 moved to Cambridge to attend the Massachusetts Institute of Technology (MIT), then Harvard University. By then, he had been to California twice and Mexico once, and pretty much nowhere else. But as a graduate student studying physics, he began work on a project at Fermilab near Batavia, Illinois, which required stays of six to nine months in the Chicago area. As a postdoc, he joined the Monopole Astrophysics and Cosmic Ray Observatory at Gran Sasso National Laboratory in Abruzzo, Italy, and had to go there five or six times a year. "That sort of got me geared up for international travel being part of my work," he says.

In 1996, Kearns joined the Super-Kamiokande experiment, in which equipment in an old zinc and lead mine under Mount Kamioka near Hida, Japan, searches for the by-products of proton decay. That meant regular trips between Boston, Abruzzo and Hida. Now, Kearns is a professor at Boston University, studying neutrino physics and particle astrophysics. The trips to Italy have tapered off, but he continues to travel to Japan five times a year, usually for about ten days.

Travel is a given in his field. Many particle physicists teach at universities, but lots of the big research projects are international collaborations built in tunnels and abandoned mines, which require huge financial investments and so cannot be done at researchers' home institutions. "I can count on my hand the number of places in the world where this sort of science can be done," says Kearns. "You sort of have to go where the action is."

Scientists in both academia and business find themselves hopping between time zones and learning to feel at home on two or more continents. They are driven by research interests, personal desire and the demands of an increasingly interconnected world. Navigating these posts requires deft multitasking, a love of travel and careful coordination with colleagues and family.

Some tackle different aspects of their research in different locations. Johannes Lercher works on the fundamental understanding of catalysis as a chemist at the Technical University of Munich (TUM) in Germany. But he is also director of the Institute for Integrated Catalysis at the Pacific Northwest ▶

National Laboratory (PNNL) in Richland, Washington, where he explores catalysis as a means of turning biomass into fuel. The TUM group grounds him in basic theory, whereas the PNNL team specializes in an application and is well stocked with expensive imaging equipment. "It broadens your horizons. You can do things you cannot do in a single location," says Lercher. "The difference in attention between an educational, academic institution and a national lab is very stimulating."

FAR AND AWAY

Software developer Julian Cerruti says that having bases in two countries gives him a global perspective. "It keeps what I do on an international level," says Cerruti, who moved from Argentina to San Francisco, California, in 2004 to take a job with IBM Research before returning to his native country a few years later. For the next five years, he travelled regularly between the two; last autumn he started his own software development firm, Creativa 77 in Buenos Aires, but he still makes trips to San Francisco to work with a client. An international perspective is important, he says, when the microchips that will run his software might be designed in Japan, and the computer-memory technology might come from Germany. He also thinks that it distinguishes his business from others in Argentina. "A vast majority of the work that is done here is for local consumption," he says.

Cerruti has turned his experience in remote working to his advantage. He says that it is less common to move to a new city for work in Argentina than in the United States. By allowing his employees to work remotely, he has attracted a wide pool of talented individuals who can stay in their home towns. And trekking to California has its perks when it comes to scheduling meetings and making contacts. "I think the rule of thumb is that the person who lives farthest has the highest priority in the calendar," he explains. He has started an online forum (http://remotework.forumatic.com) for people to discuss the ins and outs of living in one place and working elsewhere.

Globetrotters must be easily accessible to colleagues and family members. There is a growing range of online tools to keep people connected: Dropbox or Basecamp for sharing files; Skype, Google Hangouts, FaceTime and other video-conferencing tools for talking; and e-mail. "If Skype didn't exist, we'd all be dead or in debt," jokes Gabriella Sciolla, a physicist at Brandeis University in Waltham, Massachusetts, who works on the ATLAS detector

"If Skype didn't exist, we'd all be dead or in debt."
Gabriella Sciolla

at CERN, Europe's particle-physics lab near Geneva, Switzerland.

But in conventional working, much collaboration comes during impromptu conversations in hallways, or from free-ranging discussions over coffee. The structured and time-limited format of conference calls is not really conducive to creativity, says Carlo Ratti, who runs the SENSEable City Laboratory at MIT, which studies how digital technology affects urban centres around the world. And others agree that social interactions are key for creating bonds between collaborators. So despite the many ways of communicating virtually, travel continues to be important.

Being physically present makes people feel that they are actual contributors to a project. And sometimes it is easier to explain things when you can pick up an object or make a quick scribble on a napkin, says Cerruti. "Having the high bandwidth you can have when you're in person is very useful."

Even the team behind Skype does not attempt to do all its work over the Internet. "There's still no substitute for having dinner with somebody and actually shaking hands instead of waving at them over a video link," says Eric Brown, general manager for integrated communications at Skype. He travels between his main office in Palo Alto, California, and the headquarters of company owner Microsoft in Redmond, Washington; he also goes regularly to London, where the product development team is based, and to New York to meet some of his team members. He says that experience of using Skype while travelling helps developers to improve their product.

HOMEWARD BOUND

The itinerant lifestyle intensifies the work–family balancing act that many researchers already face. Lercher says that he would not have done it before his children had left home. If his wife decides that she is tired of his travelling, he will stop. "Family is a very important factor in this," he says. Cerruti returned to Argentina so that his children could grow up near his then-wife's family, but travelled to California for work. "I originally started doing that as a matter of policy, to make sure that they didn't forget about me," he says.

Accommodating family can require major coordination. Sciolla and her husband, Harvard physicist Masahiro Morii, are both involved in ATLAS. They are supposed to attend the same meetings, but have to take turns so that one can stay home and take care of their children. Sciolla got her PhD at the University of Turin in Italy; Morii earned his at the University of Tokyo. The two met as postdocs at the SLAC National Accelerator Laboratory in Menlo Park, California, and they are now settled in Boston, with regular trips to Geneva.

To balance her travel with her teaching, Sciolla mainly goes abroad during school breaks. For a non-American, the Thanksgiving

"Working abroad keeps what I do on an international level."
Julian Cerruti

holiday is an ideal time to go to Europe, she says. She will also travel during exams: someone else can supervise them, and she marks them when she returns. Kearns, too, says that it is very important to him not to miss classes; instead, he misses out on a lot of weekends. Both Sciolla and Kearns sometimes double the number of classes that they teach in one semester, in exchange for having another semester free to travel. And Kearns is willing to cut his trips short to attend graduation at Boston, where strong faculty attendance is expected. Keeping various colleagues happy requires a lot of organization, the travellers say.

That goes for mentoring as well, although it depends on the protégé. A fair bit of managing or supervising can be done remotely, as long as everyone keeps in frequent contact. "As each student is a person with their own strengths and weaknesses, there is no single recipe," says Kearns. "Some do quite well being left alone, others require frequent course correction."

Lercher handles his two workloads — "the joke is I have two full-time jobs but I only get half pay at both" — by giving up administrative duties such as committee work. When Sciolla is on campus, she volunteers for the committees that require lots of work, to compensate for the time that she is away.

They all experience the vagaries of travel. Brown complains of the poor quality of aeroplane and hotel Internet connections, but Ratti finds it helpful to be cut off from e-mail so that he can concentrate on other tasks. Cerutti knows which security gate at the Dallas/Fort Worth airport in Texas is the quickest (it's the one on the left), and that if the flight from Brazil arrives early, he will be stuck in customs for an hour and a half. At CERN, "you catch every single bug from everywhere in the world", says Sciolla. "They all come to Geneva."

The rewards go beyond professional stimulation and networking. Kearns has learned to appreciate the small differences between places — the music that is played in Hida at 7 p.m. every night, and the time of evening when everyone in Gran Sasso walks up and down the street conversing with each other. He wishes that more scientists would work in far-away places to reap the benefits. "The opportunity to do research has not only expanded their horizons in sciences," says Kearns, "but also expanded their horizons as citizens of the world." ∎

Neil Savage *is a freelance writer based in Lowell, Massachusetts.*

TURNING POINT
Joanna Wysocka

A Polish-born stem-cell biologist at Stanford University in California, Joanna Wysocka won the Vilcek Prize for Creative Promise in Biomedical Science on 1 April. The prize recognizes immigrant scientists who have made a contribution to US society.

Why did you come to the United States?
When I was still at university, I got a volunteer position working for a summer at Cold Spring Harbor Laboratory in New York while visiting my then boyfriend (now husband). I worked with different researchers, and Winship Herr, a molecular biologist working on gene regulation, invited me to come back for a PhD.

What was the defining moment of your PhD?
My research led me to work on chromatin, the DNA and proteins that make up chromosomes. It was an exciting time because related enzymes were being purified and published every few weeks. I decided to do a postdoc with David Allis, a biologist at the Rockefeller University in New York, after seeing him give an exciting talk on chromatin.

How did that affect your career?
David has never lost his childlike drive for doing science. His career path reinforced my idea that you should do what is interesting to you, no matter what others think. He was working on modifiers of the chromatin component histone at a time when most people would have considered them boring structural proteins. But his discoveries opened up the field of chromatin regulation.

How did you make a name for yourself?
As a postdoc, I discovered that a structural fold found in human proteins, called a PHD finger, can interpret modifications of histones. Since then, hundreds of papers have been published on PHD fingers that recognize different histone modifications — shedding light on everything from basic biology to diseases.

Why did you decide to work on stem cells?
When I started my assistant professorship at Stanford in 2006, a lot of my colleagues anticipated that I would continue my postdoc work. But I wanted to understand how functional complexity arises from a single set of instructions: the genome. That means understanding development, and stem-cell models offered a way to do that.

Was that a daunting move?
It was risky. It took three years to publish the first papers, and my start-up money was almost spent. But having those start-up funds made it a perfect time to try high-risk projects. If I had followed up on my postdoc work, I probably would have published faster, but I have no regrets. I'm glad I had that freedom.

What motivates you?
I'm addicted to discovery. There is this moment when we get a glimpse of an interesting result, when we don't understand what is going on. I live for those moments.

Do you do anything to hone your creativity?
Yes, it is important to me. I make a conscious effort to do the most interesting next experiment, not the obvious one. I also make an effort to hire people with backgrounds including cell biology, development, biochemistry, physics and structural biology to foster a creative environment.

What is the most important lesson you have learned about running a lab?
When I started, I didn't know how to maintain an appropriate distance from my students to be an effective leader. But I realized that students need guidance more than a friend.

What is it like to get the Vilcek award?
It's fantastic because the prize is for immigrants. I received it around the time that I was awarded tenure, so it has been a time of reflection. We are always caught up in the next paper or grant, but the award made me step back and think that I came here 15 years ago with a single suitcase, got tenure at Stanford and made a career in my adopted country, which has been very generous to me. ∎

INTERVIEW BY VIRGINIA GEWIN

Science advocate post

Two US bills introduced on 8 May, one in the Senate and one in the House of Representatives, would create a national post to promote science education and scientific achievement. The Science Laureate would be a working researcher who would talk to citizens about science's crucial role in the economy and the need for financial support, says Nathan Click, a spokesman for Senator Mazie Hirono (Democrat, Hawaii), who is a co-sponsor of the Senate bill. "If the public begins to value basic research, that percolates up to lawmakers," says Click. The laureate would also present highlights from his or her own work to inspire young people to pursue science degrees. "We want to get students and communities energized about cutting-edge research and give them role models," says Click.

Network for success

Early-career researchers and entrepreneurs benefit from strong online social networks, finds a study published earlier this month. Peter Gloor, who studies innovation at the Massachusetts Institute of Technology in Cambridge, and his colleagues analysed the LinkedIn, Facebook and e-mail networks of thousands of US and Swiss researchers and entrepreneurs (P. A. Gloor *et al. Int. J. Organ. Des. Eng.* **3,** 67–85; 2013). Those with extended networks that included contacts who were eminent in their fields were most successful according to measures such as publications, grants and venture funding. "Networking is crucial," concludes Gloor.

Credit check

Women who collaborate with men on managerial activities undervalue their own roles, says a study released on 7 May (M. C. Haynes and M. E. Heilman *Pers. Soc. Psych. Bull.* http://doi.org/mh8; 2013). Researchers got 330 students to perform tasks such as making staffing decisions with a simulated partner. Women generally gave male but not female partners more credit than they gave themselves — unless they were told that their own work was crucial or well executed. Women who devalue their contributions could be unlikely to seek out roles such as lead investigator on a large collaboration, says co-author Michelle Haynes, a psychologist at the University of Massachusetts Lowell.

AN ALIEN NAMED TIM

The enemy within.

BY MICHAEL HAYNES

23:00 ship's time

"Mwahahahaha!" The stowaway alien — Tim, as he frequently reminded us — let out an evil laugh. An odd action, considering he was bound to a table with the ship's supply of nanorope.

I glanced at the captain, Bob. He glanced back at me. We both glanced at Tim.

"What?" Tim asked. "Is this not the victory laugh of the evil genius?"

"Yes," Bob said.

"Loosed when the evil genius has his enemies right where he wants them?"

"Quite," I said.

"Mwahahahaha!"

"Shouldn't that be our line?" Bob asked.

30 minutes earlier

The cargo-hold door was rattling. As this, generally, was not a good thing, Bob sent me to check it out.

We weren't armed; the trade routes had been peaceful for decades. The biggest risk to cargo haulers was catching something from one of the alien women at the waystations. A raygun wasn't much good against that sort of trouble.

I grabbed the longest screwdriver from the repair kit and went back towards the hold. Maybe there was just a screw loose.

I keyed in the door combination. It swung open and an alien sprang into the cabin.

"Freedom!" it yelled.

"What are you doing here?"

The alien shrugged. "I'm going to steal your ship, hijack your cargo, take it to my homeworld, and thereby use it to launch an invasion of your worlds."

"You're going to invade our worlds?"

"Uh-huh!"

"With processed soycheese and Antarean toe-fungus remover?"

The furry little thing just smiled at me.

"Seriously?"

"Have you ever mixed them together and let them ferment for one of your weeks?"

I hadn't. "Okay, fine. Let's say you could do that." I made a mental note to suggest to Interstellar that we not haul these products together again. "You said you were going to steal our ship. You and what army?"

"No army. Just Tim."

"Tim?"

The alien smiled again and pointed at himself.

"Everything okay back there?" Bob yelled.

"Stowaway," I replied. Bob groaned. I knew why; he was thinking of all the paperwork.

"Is it a cute girl?" Bob asked.

"No. It's an ugly —"

"Hey!"

"— little alien. Green, furry. A bit smelly too."

"You try hanging out in a cargo bay with Antarean toe-fungus remover and tell me how you smell!"

"Fine," Bob said. "Throw it out the airlock."

I made a grab for Tim, who dodged my reach.

"I wouldn't advise this," Tim said.

"Why not?" I asked, lunging ineffectively.

"I've told you my intentions. That makes me an enemy combatant, so I'm entitled to a hearing by the Tribunal on Interspecies Conflict."

My third try at snagging Tim was a success.

"That's a good point," I said. "Except for one thing." I marched him towards the airlock. "Who's going to tell anyone that you didn't get your hearing?"

Tim squirmed, making it hard to key in the airlock code.

"My superiors know," Tim said as the door slid open and I tossed him into the airlock.

"I'm sure they'll be anxious to talk about their failed attack."

I began to enter the digits to close the door and cycle the airlock.

"Wait!"

I held my finger over the last button. "Yes?"

Tim stood, head hung. "What if Tim gave you the method for creating the toxic chemical from soycheese and fungus remover?"

I hesitated. This was out of my usual bailiwick. I also didn't want to learn some time that millions had died from a soycheese–antifungal attack that I could've helped prevent.

"You'll give that to me?"

Tim shook his head. I reached for the number pad again. "I'll give

it to your bosses," he yelled, "when we get to your destination."

I retrieved him from the airlock and he went without a fight. I wouldn't fight, either, if someone was taking me out of an airlock rather than stuffing me into one.

I explained the situation to Bob.

"What'll we do with him until we get to Tyrena?" Bob asked.

"You could put me in the hold," Tim said.

"With the raw materials?" Bob said. "I don't think so."

"You have nanorope?"

"Naturally," I said. We used it to stow anything that needed to be tied down. The built-in nanobots made it indestructible unless cut with the matching knife. Bob nodded and soon I had Tim bound to our dining table.

Everything was quiet for a few minutes until Tim started his evil laugh.

23:15 ship's time

"I think our friend has a screw loose," Bob muttered. Tim was still cackling from the table; if his plan was to annoy us to death, he had a good start.

I went to grab a snack. As I got halfway across the cabin, the ship shuddered, knocking me down. Alarms wailed.

"What the hell?" Bob growled.

"Mwahahahaha!"

I looked at Tim. He was sitting up, bits of nanorope strewn around him.

"Sorry. Tim lied. Sort of."

The ship lurched again and I could feel it accelerating.

"What are you talking about?" I asked.

"Tim said he had no army. It was true then. But not now."

I looked again at the bits of nanorope — well, just bits of *rope*, now. The ship's engines whined, straining at the demand placed on them by the nanobots now under Tim's command. Bob stood beside me, fists clenched.

"Give us back control of our ship!"

Tim shook his head and laughed again. He brushed aside what was left of the rope and stood up.

"I would throw you out the airlock," he said, "but you're too big for me and way too big for my nano-minions. But my people will give you a warm welcome on our homeworld."

He strolled past us, sat in the captain's chair and grinned at the now-useless control panel.

"Mwahahahaha!" ∎

Michael Haynes *is a database administrator, writer, reader, movie watcher and hockey fan.*