THIS WEEK

Politics without the poison

If the US midterm elections are to kickstart progress in Congress on urgent issues such as climate and basic research, the mudslinging and mayhem have got to stop.

As the race to next month's midterm elections for the US Congress enters the home straight, scientists and non-scientists alike could be forgiven a shudder of disgust over their elected representatives. Hyper-partisan fighting in recent years, characterized by venom, invective and mindless, just-say-no obstructionism, has made progress virtually impossible on national priorities such as climate, immigration reform and crumbling infrastructure. Despite survey after survey showing that voters want to see hands across the political divide, US politicians have been unable or unwilling to make productive deals to address the nation's clear and present problems.

Asked to describe how the United States fell short of ideal at a focus group last week, one voter from Keansburg, New Jersey, described a food fight in the cafeteria of a local middle school. The students had gleefully flung chicken casserole at each other, then left it to the school staff to clean up the mess. According to a *New York Times* reporter sitting in on the focus group, other participants struck much the same note. Candidates who emerge victorious should keep that Keansburg story in mind. Voters on all sides sense that too many privileged Americans, including the politicians for whom they end up casting their ballots, are engaged in reckless behaviour that leaves a mess behind.

CHANCE FOR EARLY ACTION

The situation is not entirely irredeemable, even if, as seems likely, the Republicans take control of one or both houses of Congress and set themselves up in opposition to the administration. The new Congress has a number of ready-made opportunities for early action.

Health-care reform, for example, has been a food fight all of its own (see page 763). The complex measure was finally passed last March with no Republican support, in the face of baseless propaganda about provisions such as 'death panels', which would supposedly decide whose life was valuable enough to save. One thing that both sides should be able to agree is the need to slow the explosive growth in the cost of health care. A key way to achieve those savings is the widespread use of electronic medical records, which promise to make a national health-care system more efficient overall, and, with appropriate privacy safeguards, could provide researchers with a rich new source of public-health statistics. Another way is through funds for comparative-effectiveness research, which compiles empirical data about which treatments work best for each condition. Comparative effectiveness has been controversial, and the source of some of the death-panel rhetoric. But both approaches are essential tools in cost control. Whatever happens to the reform bill overall, the new Congress should continue to fund them vigorously.

Just as contentious are climate and energy (see page 762). The current Congress has failed to pass cap-and-trade legislation designed to limit US greenhouse-gas emissions, thanks chiefly to strong Republican opposition. Even on this poisoned political ground, everyone in the new Congress should be able to agree to a strong push on energy efficiency.

Such an effort could come through federal efficiency codes for power plants and transmission lines, or for new and retrofitted buildings. Increased efficiency, by whatever route, is the easiest and cheapest way to limit carbon emissions, and has the added appeal of decreasing the nation's dependence on imported oil. Politicians unwilling to make the first argument may well be persuaded by the second.

If Congress cannot even agree to that, when legislators are totally gridlocked on a major policy issue, often the easiest way forward is to put more money into research and development, which has long enjoyed bipartisan support on Capitol Hill. The Advanced Research Projects Agency-Energy, created by the 2009 Stimulus Act, offers Congress a ready-made mechanism to funnel money into energy research (see page 760). And the 2007 America COMPETES Act, assuming it is renewed after the elections, gives Congress a mechanism to boost physical-sciences funding. Granted, the new Congress may not be especially generous to science, given anxiety over the federal deficit. But legislators would be foolish to subject research to the deep cuts proposed in the United Kingdom and elsewhere. Quite aside from the disruption that would cause for individual science programmes, it would undermine the potential for long-term economic growth.

> *"Voters sense that too many privileged Americans, politicians included, are behaving recklessly."*

RESEARCH OFF THE RAILS

Finally, the new Congress should squarely address the question of federally funded research on human embryonic stem (ES) cells. Such studies are currently at the mercy of the courts: guidelines set out by President Barack Obama in 2009 were challenged as a violation of the 1995 Dickey–Wicker amendment, legislation that bars federal funding for any research in which human embryos are destroyed. How the amendment applies to current research is debatable, to say the least — it was written three years before human ES cells were even isolated — and the case could take years to resolve. But in the meantime, investigators in the field could be wracked by injunctions that force them to stop work any time a judge decrees it, as has already happened once.

Congress should take this issue out of the courts by clarifying exactly what it wants for human ES-cell research. The best route would be to explicitly endorse the continuation of such work under the Obama guidelines. But even if Congress instead bars federal funding for the research completely, at least that would be a decision. Scientists could plan their careers around it. To leave things as they are, with researchers jerked around and left in limbo, would be unconscionable. Congress is there to make decisions. Those decisions should be based on sound evidence and public and national interest. Its members must remember that, unlike the Keansburg school floors, there is no one available to clean up their mess. ∎

AFLO/NATUREPL.COM

One backyard at a time

Local action can curb habitat loss and counter global pessimism on biodiversity.

This month's global summit on biodiversity in Nagoya, Japan, is unlikely to draw the same level of attention as last year's climate talks in Copenhagen, but the problem it aims to address is as pressing, if not more so. From a global perspective, the news on biodiversity is bad, although at the local or national level, there is room for hope. So let's get the bad news out of the way. The Nagoya meeting, which aims to pursue the agendas of the international Convention on Biological Diversity, will document failures to show any real progress in sustainable consumption of biological resources, or to approach initial targets in conserving species and preserving habitats (see page 764). In some respects, even the basic terms of the convention have been opened up for renegotiation.

One example of why biodiversity matters is that about a dozen species make up around 80% of today's total global crop production. This is a lot of eggs in very few baskets. Maintaining biodiversity, especially among indigenous species, is crucial for the optimization of agricultural production across the range of climates and terrains. Witness the growing shift in African food production away from introduced crops such as barley and towards indigenous crops such as sorghum as countries pursue greater food security.

There are obvious parallels between the world's approaches to the biodiversity and climate problems. Both focus on global targets and the need for multilateral action to curb irreversible damage. Both are moving towards market solutions: funnelling cash from the developed world to poorer countries to preserve carbon and protect life. Biodiversity even has its own Stern Review of economic impacts. Led by banker Pavan Sukhdev, The Economics of Ecosystems and Biodiversity (TEEB) project, to be formally released in Nagoya, has valued the annual contribution to humankind of global flora and fauna at US$5 trillion. 'Ecosystem services' is a clumsy term, but one that conservation scientists must get used to. Policy-makers and economists must in turn accept that the evidence they need to put a tradable price on biodiversity is currently as rare as some of the species they aim to save.

How can science help? Satellite monitoring is improving in coverage and resolution. But, too often, restrictions are placed on access to the resulting data, and policy-makers must address this. Satellites cannot do all of the work and people are needed on the ground to survey, monitor and count, particularly in developing countries. At a Science and Technology in Society forum on sustainability and biodiversity last week in Kyoto, Japan, there was clear concern about taxonomy's declining status. Although molecular biology can sample habitats for genetic diversity, and microbiology sheds light on microbial populations in habitats ranging from soil to our own guts, scientists are steadily losing the ability to map and understand life as most people experience it, at the organism level. For natural capital to be banked, that has got to change.

Unlike climate change, which probably needs a global framework for regional regulation of emissions to work, much national legislation to tackle biodiversity loss already exists. Stronger enforcement and stricter penalties would be a relatively easy win. Those politicians who have, rightly, talked up the threat of global warming should remember that the consequences for plants and animals are one of the climate impacts they seek to avoid. On current trends, how many threatened species will survive hunting, trapping and habitat destruction only to be wiped out by rising temperatures?

Biodiversity loss is best tackled locally, and here is the good news. From London's River Thames to Lake Biwa, Japan's largest freshwater lake, good management and a whole-system approach are restoring ecosystems and bringing benefits to local people. To build on such successes, scientists and policy-makers need to engage more seriously with those in the social sciences, who themselves must develop greater ambition to make a positive difference. Whether in Eritrea or the Amazon, convincing people to behave differently demands an understanding of indigenous populations' practical, cultural and spiritual relationships with their natural environment. The global picture is gloomy. But if each country can empower the citizens and international research communities that focus on local threats, using the tools of all the sciences, there could be movement rather than paralysis. ∎

Probe retires to a place in the Sun

A fitting end for the satellite that shed light on the age, shape and fabric of the Universe.

Farewell, then, to NASA's Wilkinson Microwave Anisotropy Probe (WMAP): the little satellite that could. Last month, the spacecraft fired its boosters for the final time and fell into a graveyard orbit around the Sun, where it will remain as a permanent fixture in the Universe it helped to decode.

Launched in June 2001, the WMAP makes a strong case to be considered one of the greatest scientific experiments of all time. It turned cosmology from informed guesswork into a precision science, and brought our fuzzy understanding of the nature of the Universe into breathtaking focus. Along the way, its findings yielded some of the most highly cited papers in physics — as well as bewitching images that introduced millions of people to the truths and enduring mysteries of the cosmos.

The WMAP studied the remnants of heat that lingered after the Big Bang, a pattern frozen in time when the Universe was only about 380,000 years old and which has since stretched to microwave wavelengths. Subtle differences seen in the texture of this 'cosmic microwave background' by the WMAP have revealed the geometry, composition and age of the Universe. The probe showed that the Universe is flat and probably endless, and produced the first fine-resolution, full-sky map of the cosmic microwave background. It also revealed how this original light is polarized — the blueprint for the first galaxy formation. Physicists have yet to encounter dark matter or dark energy, but the WMAP has already audited the expected contribution of these components to the Universe: 23.3% and 72.1%, respectively. And it has determined that ordinary matter makes up just 4.6% of the Universe, to within 0.1%.

The probe's discoveries have resonated far beyond the scientific sphere. When singer–songwriter Katie Melua sang, "We are 12 billion light years from the edge/That's a guess/No one can ever say it's true" in her 2005 hit *Nine Million Bicycles*, science writer Simon Singh complained that the lyrics did not reflect current scientific knowledge. The WMAP, after all, had recently determined the age of the Universe with great accuracy. Melua agreed to re-record the song using Singh's revision: "We are 13.7 billion light years from the edge of the observable Universe/That's a good estimate with well-defined error bars/And with the available information." (See go.nature.com/ONJSQG.)

Science marches on, and a replacement for the WMAP is already in orbit: the European Space Agency's Planck satellite. Whatever this latest spacecraft may discover, the legacy of the Wilkinson Microwave Anisotropy Probe is assured. ∎

Publish your computer code: it is good enough

Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says **Nick Barnes**.

I am a professional software engineer and I want to share a trade secret with scientists: most professional computer software isn't very good. The code inside your laptop, television, phone or car is often badly documented, inconsistent and poorly tested.

Why does this matter to science? Because to turn raw data into published research papers often requires a little programming, which means that most scientists write software. And you scientists generally think the code you write is poor. It doesn't contain good comments, have sensible variable names or proper indentation. It breaks if you introduce badly formatted data, and you need to edit the output by hand to get the columns to line up. It includes a routine written by a graduate student which you never completely understood, and so on. Sound familiar? Well, those things don't matter.

That the code is a little raw is one of the main reasons scientists give for not sharing it with others. Yet, software in all trades is written to be good enough for the job intended. So if your code is good enough to do the job, then it is good enough to release — and releasing it will help your research and your field. At the Climate Code Foundation, we encourage scientists to publish their software. Our experience shows why this is important and how researchers in all fields can benefit.

Programs written by scientists may be small scripts to draw charts and calculate correlations, trends and significance, larger routines to process and filter data in more complex ways, or telemetry software to control or acquire data from lab or field equipment. Often they are an awkward mix of these different parts, glued together with piecemeal scripts. What they have in common is that, after a paper's publication, they often languish in an obscure folder or are simply deleted. Although the paper may include a brief mathematical description of the processing algorithm, it is rare for science software to be published or even reliably preserved.

Last year's global fuss over the release of climate-science e-mails from the University of East Anglia (UEA) in Norwich, UK, highlighted the issue, and led the official inquiry to call for scientists to publish code. My efforts pre-date the UEA incident and grew from work in 2008 based on software used by NASA to report global temperatures. Released on its website in 2007, the NASA code was messy and proved difficult for critics to run on their own computers. Most did not seem to try very hard, and nonsense was written about fraud and conspiracy. With other volunteers, I rewrote the software to make it easier for non-experts to understand and run. All software has bugs, and we found a number of minor problems, which had no bearing on the results. NASA fixed

> NOBODY IS ENTITLED **TO DEMAND** TECHNICAL SUPPORT FOR FREELY PROVIDED CODE: IF THE FEEDBACK IS UNHELPFUL, **IGNORE IT.**

them and now intends to replace its original software with ours.

So, openness improved both the code used by the scientists and the ability of the public to engage with their work. This is to be expected. Other scientific methods improve through peer review. The open-source movement has led to rapid improvements within the software industry. But science source code, not exposed to scrutiny, cannot benefit in this way.

NO EXCUSES

If scientists stand to gain, why do you not publish your code? I have already discussed misplaced concern about quality. Here are my responses to some other common excuses.

It is not common practice. As explained above, this must change in climate science and should do so across all fields. Some disciplines, such as bioinformatics, are already changing.

People will pick holes and demand support and bug fixes. Publishing code may see you accused of sloppiness. Not publishing can draw allegations of fraud. Which is worse? Nobody is entitled to demand technical support for freely provided code: if the feedback is unhelpful, ignore it.

The code is valuable intellectual property that belongs to my institution. Really, that little MAT-LAB routine to calculate a two-part fit is worth money? Frankly, I doubt it. Some code may have long-term commercial potential, but almost all the value lies in your expertise. My industry has a name for code not backed by skilled experts: abandonware. Institutions should support publishing; those who refuse are blocking progress.

It is too much work to polish the code. For scientists, the word publication is totemic, and signifies perfectionism. But your papers need not include meticulous pages of Fortran; the original code can be published as supplementary information, available from an institutional or journal website.

I accept that the necessary and inevitable change I call for cannot be made by scientists alone. Governments, agencies and funding bodies have all called for transparency. To make it happen, they have to be prepared to make the necessary policy changes, and to pay for training, workshops and initiatives. But the most important change must come in the attitude of scientists. If you are still hesitant about releasing your code, then ask yourself this question: does it perform the algorithm you describe in your paper? If it does, your audience will accept it, and maybe feel happier with its own efforts to write programs. If not, well, you should fix that anyway. ■ **SEE NEWS FEATURE P. 775**

Nick Barnes *is director of the Climate Code Foundation, Sheffield S17 4DL, UK.*
e-mail: nb@climatecode.org

Nitrogen radical synthesized

Nitrogen is essential to life. A crucial step in the nitrogen cycle is the conversion of the nitrite ion NO_2^- to the nitric oxide molecule NO by the enzyme copper nitrite reductase, which is found in a variety of bacteria.

Trevor Hayton and his colleagues at the University of California, Santa Barbara, have synthesized a copper-nitrosyl radical that has been proposed to play a key part in this reaction. They found that it contains a surprisingly long bond between the copper and nitrogen atoms, and suggest that this arrangement provides a model of how the nitric oxide interacts with the copper during its formation.

J. Am. Chem. Soc. doi:10.1021/ja105930b (2010)

Soccer Samaritans

Fans of different soccer teams are often sworn enemies — but would they go so far as to refuse to help a rival fan in pain?

Grit Hein at the University of Zurich in Switzerland and her colleagues used functional magnetic resonance imaging to scan the brains of fans of two rival football teams, identified by coloured wristbands (pictured). The

Colourful bacterial resistance

Parrots' brightly coloured feathers are not only useful for visual signalling, but also protect against microbial damage.

Edward Burtt Jr at Ohio Wesleyan University in Delaware and his colleagues exposed parrot feathers of varying colours and species to *Bacillus licheniformis*, a feather-degrading bacteria, over a five-day period.

They found that white feathers degraded the most rapidly. Yellow feathers, which contain low levels of psittacofulvin, a pigment that is resistant to bacterial breakdown and found only in parrots, came a close second.

Red feathers, which contain higher levels of psittacofulvins, and blue, green and black feathers, which all contain melanin, broke down much more slowly.

Biol. Lett. doi:10.1098/rsbl.2010.0716 (2010)

volunteers watched either a fellow fan or a rival receive a painful electric shock to their hand. Fans could then choose to help the other person by enduring half the pain themselves, or to watch either the person endure the pain, or a video about football.

Activity in two different brain areas predicted fans' reactions. In those that chose to help, a region associated with empathy called the anterior insula was activated. By contrast, in those that let their rival suffer, the nucleus accumbens showed activation — which was stronger if they had rated that particular foe more negatively before the test.

Neuron 68, 149–160 (2010)

Keeping the peace

Most of the proteins — known as cytokines — in the interleukin 1 family are thought to spark inflammation while defending the body. Charles Dinarello at the University of Colorado Denver in Aurora and his colleagues show that one cytokine in the group instead dampens inflammation. The finding reveals the function of IL-37, which has been mysterious since it was discovered a decade ago.

The researchers first used RNA interference to turn off production of IL-37 in blood cells. Those cells showed an increased immune response when treated with molecules common to bacterial membranes to mimic infection. When they expressed the human form of IL-37 in transgenic mice, they found that it protected the animals against organ damage caused by severe infection, probably by reducing the production of pro-inflammatory cytokines.

Nature Immunol. doi:10.1038/ni.1944 (2010)

NETWORKS

Online peer-pressure threshold

An analysis of 100 million installations of applications — or 'apps' — on the popular social-networking site Facebook demonstrates that online social influences have an on–off pattern. Social pressure pushed apps that were installed by more than about 55 users per day to heights of popularity. Below this tipping point, users seemed to install or not install apps regardless of what their friends or others in the network were doing, suggesting that there was no social influence.

Felix Reed-Tsochas and Jukka-Pekka Onnela of the University of Oxford, UK, identified the behavioural threshold, which has never been seen in the offline world, by combing through data from 2,705 apps. The researchers consider it to be an inherent property of this particular networking systems and expect it to be observed in other online networks.

Proc. Natl Acad. Sci. USA 10.1073/ pnas.0914572107 (2010)

GEOPHYSICS

Planet Earth, squeezed

Earth isn't perfectly round, and the factors underlying its shape have undergone a rethink.

The theoretical shape of an entirely fluid, rotating Earth is an ellipsoid with a difference of about 21 kilometres between the equatorial and the polar radius. In reality, however, convective processes in Earth's mantle, triggered by differences in density, account for an additional 100-metre or so difference between the radii.

Frédéric Chambat at the University of Lyon in France and his colleagues recalculated the difference in radii caused by rotation. By subtracting this from the observed value, they found that convection causes a 113-metre difference between the equatorial and the polar radius instead of the 98 metres that is currently used in calculations.

Geophys. J. Int. doi:10.1111/ j.1365-246X.2010.04771.x (2010)

PALAEONTOLOGY

Introducing Sarahsaurus

Long-necked dinosaurs called sauropodomorphs originated in the Southern Hemisphere during the mid to late Triassic period, some 230 million years ago. From there, they dispersed to all early continents, but there is some debate as to how. Some have assumed that a single taxon dispersed widely across the Pangaea supercontinent and diversified later. But Timothy Rowe at the University of Texas at Austin and his colleagues argue that this ubiquitous 'cosmopolitan' dinosaur never existed.

They recently described a new taxon named Sarahsaurus (fossil veterbrae pictured), found in Arizona, and analysed its relationships to other North American sauropodomorphs. Their analysis suggests separate arrivals for different groups. Geographical isolation events, extinctions and dispersal happened multiple times throughout the Mesozoic period, they argue.

Proc. R. Soc. B doi:10.1098/ rspb.2010.1867 (2010)

PLANT EVOLUTION

Model plant's secret past

The darling of plant research may be millions of years older than botanists thought.

Arabidopsis thaliana is a dainty weed used by plant biologists around the globe, but the details of its evolutionary history are controversial. Sarah Mathews and Mark Beilstein of the Arnold Arboretum of Harvard University in Jamaica Plain, Massachusetts, and their colleagues re-examined its fossil record, including several fossils that other studies had overlooked.

The team determined that the genus *Arabidopsis* split from the genus *Brassica*, which includes broccoli and cabbage, about 43 million years ago — two to three times earlier than previously thought. The results suggest that the *Arabidopsis* genome has been changing more slowly than expected, and that members of the order Brassicales, which includes *Arabidopsis*, coevolved with butterflies that are able to detoxify the plants' defensive chemicals.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0909766107 (2010)

NEUROBIOLOGY

Powerless against Parkinson's

Cellular power outages may contribute to Parkinson's disease, according to an analysis of gene expression in hundreds of brain tissue samples.

ECOLOGY

The tundra warms and grows

HIGHLY READ on www. esajournals.org in September

The effects of climate change on tundra in the high Arctic are chronicled in this study by James Hudson and Greg Henry of the University of British Columbia in Vancouver, Canada.

Data collected from study plots over a 13-year period and survey data covering 27 years on the tundra of Ellesmere Island in Nunavut, Canada, show an ecosystem 'in transition'. Temperatures have risen and the growing season has lengthened. The biomass of mosses has increased by 74% and that of evergreen shrubs by 60%. The total biomass of the system has increased significantly, and vegetation has grown taller. But because there was plenty of open ground at the site into which plants could expand, these changes did not result in decreases in any group. The research indicates that climate change has already begun to increase plant productivity in the high Arctic.

Ecology 90, 2657–2663 (2009)

When Clemens Scherzer at the Brigham and Women's Hospital in Cambridge, Massachusetts, and his colleagues looked for pathways that are wrongly expressed in Parkinson's disease, they found ten that had not previously been linked to the condition. All ten are important for proper functioning of the cell's energy-generating structures, the mitochondria.

In particular, genes under the control of a protein called PGC-1α are expressed at low levels in patients with Parkinson's disease. When the team overexpressed PGC-1α in rat neurons grown in culture, the cells became more resistant to chemicals that damage mitochondria and cause Parkinson's-like changes. The results suggest that PGC-1α activation could one day be used to treat the disease.

Sci. Transl. Med. 2, 52ra73 (2010). For a longer story on this research see go.nature.com/ q3orms

NEW ONLINE

Papers published this week at nature.com

SEVEN DAYS *The news in brief*

POLICY

Genetics advice

An influential panel that advised the US government on health and genetics had its final meeting on 6 October in Bethesda, Maryland. The Secretary's Advisory Committee on Genetics, Health, and Society was created in 2003 to address issues such as genetic discrimination, gene patenting and genetic testing. The panel had "fulfilled its charge" and other government bodies would carry its mantle, said John Burklow, a spokesman for the National Institutes of Health, its parent agency. But current and former committee members were puzzled by last month's decision to dissolve the panel. See go.nature.com/miz9gy for more.

Oil-spill response

A White House-appointed commission investigating the BP Deepwater Horizon oil spill in the Gulf of Mexico released its draft reports on 6 October. It criticized the US federal government for being ill-prepared to use large volumes of dispersants, and for taking too long to embrace estimates of the rate of flow from the leaking well that were higher than the official federal line. Jane Lubchenco, administrator of the National Atmospheric and Oceanic Administration, responded that low early flow-rate estimates did not affect the federal response. The commission's final report is scheduled to be released in January 2011.

Climate deadlock

The last formal round of talks before this year's climate summit in Cancún, Mexico, produced little progress towards any international climate deal. The meeting in Tianjin, China, was marked

Toxic sludge leaves lasting legacy

Scientists assessing the 4 October toxic spill from a Hungarian aluminium oxide factory say recovery could take years. "All life is dead," says János Szépvölgyi, director of the Hungarian Academy of Sciences' Institute of Materials and Environmental Chemistry in Budapest. "A 2–5-centimetre-thick layer of caustic mud is covering the soil." Up to 1 million cubic metres of alkaline sludge have contaminated thousands of hectares of land north of Lake Balaton. The red colour is from iron oxide, but scientists are more worried about dissolved heavy metals, such as arsenic and chromium, entering groundwater. As *Nature* went to press, eight people were reported to have died, and a second spill was feared. See go.nature.com/a7rm8x for more.

by disagreements between the world's largest polluters, China and the United States, over pledging and verifying emissions cuts. The UN Framework Convention on Climate Change conference in Cancún starts on 29 November.

Scientist sacked

Israel's minister of education, Gideon Sa'ar, fired his chief scientist, Gavriel Avital, on 4 October. It was not a surprise: Avital has taken vocal stands against evolution and climate change since he was appointed in December 2009. An aeronautical engineer and a politician, Avital is reported to have said that Israel's science textbooks should offer pupils alternatives to evolution, and has published articles claiming that there is no connection between human activity and climate change.

Innovation mapped

European research commissioner Máire Geoghegan-Quinn published the European Commission's long-awaited proposal for boosting innovation on 6 October. Part of the European Union's economic strategy plan ('Europe 2020'), the proposal, dubbed Innovation Union, wraps together familiar aims — such as increasing research and development expenditure and improving mobility between Europe's 27 member states — with a list of steps for removing bottlenecks that prevent ideas from reaching the market. The European Council and Parliament will discuss the proposal over the next weeks.

Fisheries fight

The process of attaining the Marine Stewardship Council's (MSC's) 'sustainable' certification for a fishery (see *Nature* **467**, 28–29; 2010) took a novel turn this week in the case of the Antarctic toothfish. After protests from some scientists (see *Nature* **467**, 15; 2010), on 11 October an independent adjudicator for the London-based MSC asked

Moody Marine — the company assessing the toothfish — to amend several of the scores in its sustainability assessment. No case has reached this stage before. Opponents of the listing could still lose their fight: revised scores may be sufficient to declare the fishery sustainable.

RESEARCH

Big Bang probe end

Nine years after launch, a spacecraft that measured the faint afterglow of the Big Bang to unprecedented accuracy has ceased operations. On 6 October, researchers at the NASA Goddard Space Flight Center in Greenbelt, Maryland, announced that the Wilkinson Microwave Anisotropy Probe (WMAP) had officially ended its mission. The probe was disposed of by sending it into orbit around the Sun on 8 September. See page 752 for more.

Regulatory science

The US Food and Drug Administration (FDA) has developed a US$25-million-a-year programme to modernize the science behind its safety evaluations, the agency's head, Margaret Hamburg, announced on 6 October. The Advancing Regulatory Science initiative will update methods for tracking bacteria in the food supply, develop computer models for toxicology testing and improve laboratory assays to ease development of generic drugs, among other projects. The move follows a 2007 report by an FDA advisory board that said the agency was unable to fulfil its mission because its scientific base had eroded.

Asteroid dust hope

The Hayabusa space explorer might have picked up dust at the Itokawa asteroid, from which it returned in June 2010 after a seven-year mission. Researchers at the Japan Aerospace Exploration Agency (JAXA) said on 6 October that they had scraped 100 or so micrometre-sized grains from an inner container in Hayabusa's recovery capsule. The quantity of particles is encouraging, but there is no evidence yet to conclude that they are from the asteroid. Definitive tests will take several months.

FUNDING

Neglected diseases

The Global Fund to Fight AIDS, Tuberculosis and Malaria raised only US$11.7 billion for 2011–13, after a meeting of its donor countries in New York City on 4–5 October. The fund, based in Geneva, Switzerland, had hoped for $20 billion. See page 767 for more.

UK protest rally

Several hundred scientists gathered in London on 9 October to protest against upcoming cuts to government science and engineering funding (pictured). The rally organizers — pressure groups Science is Vital and the Campaign for Science and Engineering — reiterated messages that cutting science funding would harm the economy and drive British scientists abroad.

BUSINESS

Super cool deal

The world's largest order for superconducting wire was announced on 6 October. American Superconductor, based in Devens, Massachusetts, said it had received an order for 3 million metres of wire from South Korean company LS Cable, headquartered in Anyang-Si near Seoul. LS Cable will use the wire — the cost of which was not disclosed — to make superconducting cables, initially for South Korea's electricity network. The wire is made using yttrium barium copper oxide (YBCO); discovered in 1987, this ceramic compound superconducts at temperatures up to 93 kelvin (−180 °C), so can be cooled with liquid nitrogen. See go.nature.com/bok2qw for more.

AWARDS

Chemistry Nobel

Richard Heck, Ei-ichi Negishi and Akira Suzuki were awarded the 2010 Nobel Prize in Chemistry on 6 October, for their development of reactions to connect carbon-chain molecules using palladium catalysts. See page 765 for more.
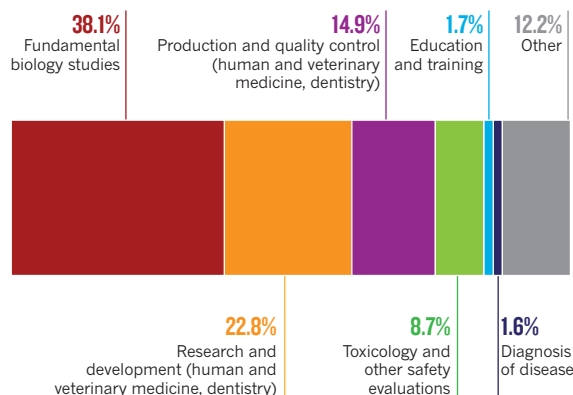
TREND WATCH

About 12 million animals were used in science in the European Union during 2008, say statistics published last week — roughly the same as in 2005. The fraction of animals used in fundamental biology research has risen by 5% since 2005, because of greater use of transgenic mice. Alternative ways of measuring toxicity slightly cut the number of rats and rabbits used in chemical-safety tests, but 54,000 more mice were needed to test batches of the wrinkle-reducing botulinum toxin than in 2005. See go.nature.com/uunigs for more.

EUROPE'S USE OF RESEARCH ANIMALS
In 2008, fundamental biology studies accounted for the largest single proportion of animal experiments in the European Union.

38.1% Fundamental biology studies
14.9% Production and quality control (human and veterinary medicine, dentistry)
1.7% Education and training
12.2% Other
22.8% Research and development (human and veterinary medicine, dentistry)
8.7% Toxicology and other safety evaluations
1.6% Diagnosis of disease

COMING UP

14 OCTOBER
NASA hosts a symposium in Arlington, Virginia, to celebrate 50 years since the agency started an exobiology programme.
go.nature.com/yfdlud

18–19 OCTOBER
Leading lights of stem-cell biology gather at the Royal Society in London to discuss future prospects for the field.
go.nature.com/lawddu

18–29 OCTOBER
A summit in Nagoya, Japan, discusses how to preserve the world's biodiversity.
go.nature.com/v2kgi7 and see page 764 for more.

20 OCTOBER
The extent of the UK government's cuts to funding for science and education will be revealed when it publishes its three-year spending plan.

NEWS IN FOCUS

B. SMIALOWSKI/GETTY

US MIDTERM ELECTIONS

Volatile forces shape US vote

Science could face budget constraints and ideological challenges in a post-midterm Congress.

BY IVAN SEMENIUK

"Want to cap Congress — not carbon?" The question appeared, slogan-style, in an online notice urging Virginia voters to attend a Tea Party convention held last week in the state's capital, Richmond. Organizers say that more than 2,800 people attended the event, the largest meeting yet for the new political movement that is injecting an unusual degree of volatility into this year's US midterm-election season.

Strong opposition to greenhouse-gas regulation is just one manifestation of a broader Tea Party theme: keeping government small. Now the Tea Party phenomenon (and the public ire it has tapped) has become the defining feature of a campaign that will determine who, on

2 November, occupies each seat in the House of Representatives and more than one-third of the seats in the Senate. With Democrats currently holding narrow majorities in both chambers, even a modest change at the polls could spell a wholesale transfer of power in Congress and a seismic shift in the US political landscape. The governorships of 37 states are also in play this year, including that of California, a research powerhouse. It is difficult to predict how all this will affect scientists and the government agencies that fund them, but with the US economy staggering under the burden of high unemployment and record-setting federal deficits, the new Congress, whatever its ideological bent, will preside over diminished resources and difficult choices.

"In the face of fiscal constraints to come, making decisions on where to cut and how that

will affect our research and innovation effort is a very serious issue," says Anne Solomon, a senior adviser on science and technology at the Center for the Study of the Presidency and Congress, a think tank based in Washington DC.

An issue paper co-authored this year by Solomon calls for a "science and technology-enhanced Congress", in which legislators are broadly knowledgeable about science and have better access to technical expertise on the complex issues they face — from energy policy, to education, to economic and security matters. In fact, the opposite is likely to be true of the next Congress.

Several senior figures who are strong on science policy are retiring, or may be ousted in a Tea Party-boosted vote swing. Solomon worries that this could mean an influx of new players on Capitol Hill who are less committed to ▶

▶ funding science research and education, and who lack "the general science and technology savvy" to make informed decisions.

The Tea Party's conservative message could also affect how science is treated in Congress. Last week's convention in Richmond featured appearances by politicians including former Pennsylvania senator Rick Santorum, who once tried to amend an education bill to promote the teaching of intelligent design, and Virginia attorney general Ken Cuccinelli, who recently renewed his efforts to investigate and discredit climate scientist Michael Mann, formerly at the University of Virginia in Charlottesville.

But the Tea Party's biggest influence stems from its small-government, cost-cutting agenda. This has been taken on board by the Republicans, who are currently leading in the polls and look set to win the House from their Democratic rivals. The Republicans' 45-page policy document, unveiled on 23 September, is telling in both its emphasis and its omissions. Although the words "tax", "taxes" and "taxpayer" appear 56 times in the document, the words "science", "research" and "education" do not figure once.

In the current Democrat-controlled Congress, science was given plenty of attention in spite of the economic crisis. The National Science Foundation and the National Institutes of Health received influxes of cash as part of President Barack Obama's economic stimulus package. Legislation such as the health-care reform bill and a bill to deal with oil spills (not yet passed by the Senate) also include provisions for funding research and development. Disputes between the parties on science spending have tended to be "differences of opinion on how much to invest in basic research, not on whether or not it was important", says Joanne Carney, director of the Center for Science, Technology and Congress at the American Association for the Advancement of Science in Washington DC.

In his inaugural address, Obama promised to "restore science to its rightful place", and he appointed prominent researchers and science-policy experts to key government roles. But as the stories on the following pages detail, science faces plenty of challenges for the post-midterm Congress to address.

And given the changing political climate, the new Congress looks likely to be even more polarized than the last. If Republicans are running the House, the Obama administration will be dealing with an adversary rather than an ally as it tries to push through changes in areas such as energy and education. As concerns about the deficit combine with an enthusiasm for scaling back government, science funding could see heavy reductions. ■ **SEE EDITORIAL P.75I**

# Deficit poses threat to science

Research programmes in the United States seem to be heading for a cliff, no matter who wins in Congress.

BY EUGENIE SAMUEL REICH

Unlike their counterparts elsewhere, US scientists have been sitting pretty in the wake of the global economic downturn. The 2009 American Recovery and Reinvestment Act, or stimulus bill, pumped an extra US$31 billion into science, and President Barack Obama's budget request for fiscal year 2011 included generous increases for several science-funding agencies.

But going into the midterm elections, a different narrative is emerging. Republicans are running on a platform to reduce the $1.4-trillion US deficit, which seems likely to entail freezes or effective cuts for at least some science programmes. If the Democrats retain control of Congress by a thin margin, policy experts say that they are likely to interpret the loss of seats as a call to rein in spending too. "Science is pretty well supported by both sides, but it's a matter of balancing investment with the deficit," says Patrick Clemins, director of the research and development budget and policy programme at the American Association for the Advancement of Science in Washington DC.

A Republican win would probably spell problems for America COMPETES, a 2007 act of Congress that set the National Science Foundation (NSF), the National Institute of Standards and Technology and the Department of Energy's Office of Science on a path to double their funding over ten years. America COMPETES originally passed with bipartisan support, and stimulus funding provided
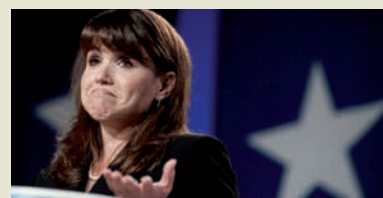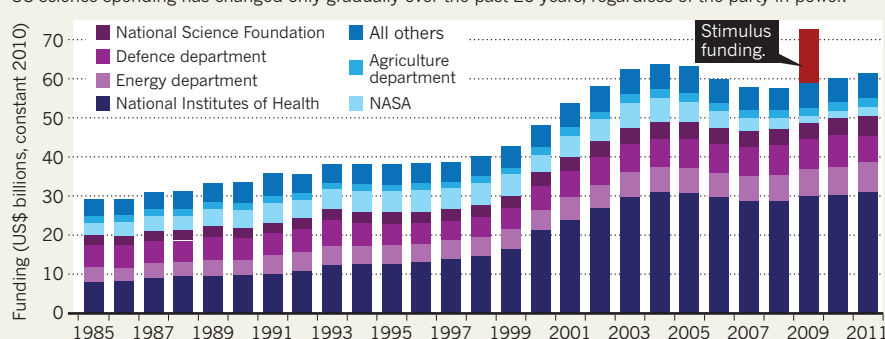
STATE WATCH: DELAWARE

Having defeated high-profile moderate Republican Michael Castle in her state's primary campaign, Tea Party-endorsed Delaware Senate candidate Christine O'Donnell (pictured) now faces scrutiny over some of her public statements, including a 2007 assertion that "American scientific companies are cross-breeding humans and animals and coming up with mice with fully functioning human brains". She has also called evolution a myth and opposes stem-cell science. O'Donnell's election race against Democrat Christopher Coons may serve as a barometer for how far US voters will tolerate views unfriendly to science from the resurgent right.

a further boost for research in 2009. But COMPETES expired at the end of September 2010 because Republicans objected to its high funding levels and slowed the passage of its reauthorization. Even if COMPETES passes

FEDERAL FUNDS FOR RESEARCH

US science spending has changed only gradually over the past 25 years, regardless of the party in power.

- National Science Foundation
- Defence department
- Energy department
- National Institutes of Health
- All others
- Agriculture department
- NASA

Stimulus funding.

in the weeks after the November elections and before new members take their seats in January, it may not receive the funding it requires in what is expected to be a tighter federal budget for fiscal year 2012.

The looming budget shortfall has led to concern about a 'cliff effect', in which projects funded by the stimulus bill find themselves without support when its funds run out, which could occur even if Democrats retain control. Paradoxically, a resounding Republican win might be better for science funding than a small Democratic majority, says Christopher Hill, a science-policy expert at George Mason University in Arlington, Virginia. Republicans are bigger spenders than their rhetoric suggests, he says.

In the past, both parties have consistently supported science funding (see 'Federal funds for research'). A doubling of the National Institutes of Health budget from 1999 to 2003 was initiated during Democratic president Bill Clinton's tenure with the approval of a Republican Congress. From 2001 to 2007, the House Committee on Science and Technology, which drafts legislation authorizing science budgets, was headed by Sherwood Boehlert, a moderate New York Republican and a strong supporter of funding the NSF and the energy department. The next chair was Tennessee Democrat Bart Gordon, who secured bipartisan support for COMPETES but plans to retire after this Congress. If control changes, Ralph Hall, a Texan Republican, is in line to take over. Last spring, Hall pushed for a three-year budget freeze for the agencies targeted for increases by COMPETES.

The potential arrival of Tea Party-backed candidates in Congress with a focus on reduced government and unfavourable attitudes towards science (see 'State watch: Delaware') is likely to be significant. Under a Republican Congress, Hill says that some symbolic projects such as FutureGen — a carbon capture and storage initiative in Meredosia, Illinois — may be axed. But others, such as the Advanced Research Projects Agency–Energy are better protected against major cuts, he says.

Barry Toiv, vice-president for public affairs at the Association of American Universities in Washington DC, is pragmatic about the prospects of working with a Republican Congress. "If the House and/or Senate does turn over, we will need to make a renewed effort to ensure the new leadership understands the value of research investments." ∎

ELECTIONS

↻ **NATURE.COM**
More on the midterm elections at:
www.nature.com/
midterm2010

NIH prepares for loss of political champions

Supporters call for Congress to pass stem–cell bill and NIH budget boost during post–election session.

BY MEREDITH WADMAN

The last time a bill supporting human embryonic stem-cell research was introduced to the US House of Representatives, it was co-authored by a seasoned Republican congressman from Delaware. Announcing the bill in March, Mike Castle hailed President Barack Obama's executive order that lifted restrictions on federal funding for the controversial research, and declared that "Congress must act to ensure that an overarching ethical framework is signed into law."

Castle, who gave up his House seat to run for the Senate, lost his party's nomination to Tea Party-backed opponent Christine O'Donnell. O'Donnell is against stem-cell science, which some equate to abortion — a deeply divisive issue in US politics. This is reflected by a state ballot that would grant human rights to embryos from the moment of their creation (see 'State watch: Colorado').

Even more worrying for researchers is a lawsuit that seeks to suspend federal funding for the research. The case could overturn guidelines implementing Obama's order as early as next month. That possibility has advocates calling for Congress to pass the Castle bill, co-sponsored by Diana DeGette (Democrat, Colorado), during its post-election session, before the new congress is seated.

"Should there be an adverse ruling blocking stem-cell research, certainly there would be much more pressure on Congress to act," says Tony Mazzaschi, the senior director of scientific affairs at the Association of American Medical Colleges (AAMC) in Washington DC. Earlier this month, the AAMC asked its members to contact their representatives, urging them to use the session to enact the bill.

But with a long list of priorities demanding time during the dying days of the current Congress, lawmakers may not manage to deal with the stem-cell bill. And should the House or Senate swing to the right in the election, stem-cell legislation may falter in the next Congress — despite having passed twice in the past, only to be vetoed by then-president George W. Bush.

Castle's is not the only departure that will be keenly felt by the National Institutes of Health (NIH), by far the largest funder of US biomedical research. David Obey (Democrat, Wisconsin), chairman of the powerful House Committee on Appropriations that allots money for the NIH, is retiring. In the Senate, Arlen Specter (Democrat, Pennsylvania), who last year won an additional $10.4-billion allocation for the NIH in return for supporting Obama's stimulus bill and later switched parties to join the Democrats, lost his new party's nomination.

Even with Specter and Obey still in place, the current Democrat-controlled Congress is struggling to deliver the 3.2% boost to the agency in 2011 that Obama requested last winter. A Senate committee passed the $1-billion increase in July, but the bill has not come to a vote in either the House or the full Senate. When lawmakers reconvene after the elections, NIH advocates may not be able to withstand relentless pressure to curb non-mandatory spending. "We are going to have to work hard to make the case that NIH deserves an increase while other things are being cut," says Jennifer Zeitzer, a lobbyist for the Federation of American Societies for Experimental Biology in Bethesda, Maryland. ∎

P. GOETGHELUCK/SPL
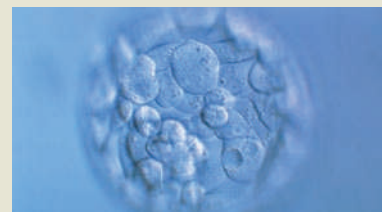
STATE WATCH: COLORADO

In Colorado, voters must decide whether to support a constitutional amendment that would extend human rights and equal protection under the law to "every human being from the beginning of … biological development". Such a vote would open human embryonic stem-cell research in the state to legal challenge, along with abortion. Although not expected to pass (a similar amendment was defeated by nearly 3 to 1 in Colorado in 2008), the vote is a reminder that a key question at the heart of the human embryonic stem-cell debate — when life begins — remains unresolved in US politics. The issue could affect US stem-cell science for another generation.

The Sun may have already set on any short-term prospects for climate legislation in the US Congress.

L. NICHOLSON/REUTERS

US MIDTERM ELECTIONS

A chilly season for climate crusaders

Open scepticism of global warming could rule next Congress.

BY JEFF TOLLEFSON

Nineteen of the 21 serious Republican challengers for seats in the US Senate believe that climate science is either "inconclusive" or "incorrect", according to an analysis by Washington DC's non-partisan *National Journal*. A more comprehensive list compiled by the left-leaning Wonk Room website suggests that 31 out of 37 Republican Senate candidates — including nine out of ten sitting senators — have recently disputed the science. Five of the remaining six actively oppose existing climate bills.

It is not clear exactly how concerns about climate-change regulations will affect the US midterm elections next month. Battles about political ideology and the state of the US economy are more pressing. But one thing is certain: scepticism about climate science has become one of the many litmus tests for candidates backed by the surging right. Even Senator John McCain of Arizona, who once championed climate legislation, has said that the world needs to know whether the scientific community's conclusions about global warming "were flawed by outside influences". In trendsetting California, where the science of climate change

is not at issue between the leading gubernatorial candidates, concern over the economy could still lead to a deferral of greenhouse-gas emission cuts (see 'State watch: California').

If Republicans take the House or Senate, US climate scientists could be targeted for investigations that challenge findings related to global warming. In the House, Darrell Issa of California, the ranking Republican on the House Committee on Oversight and Government Reform, has promised to give climate science "a careful relook". In the Senate, long-time climate sceptic James Inhofe (Republican, Oklahoma) would relish the opportunity to subpoena climate scientists for hearings before the Committee on Environment and Public Works, which he is likely to chair if Republicans take control.

Democratic leaders pushed many members to vote for a comprehensive climate bill in the House in 2009, only to see the issue fizzle out in the Senate. Republican candidates are now using that vote to campaign against Democrats such as Zack Space, who has been accused by his Republican opponent in Ohio, state senator Bob Gibbs, of voting for a "cap-and-trade energy tax that will kill over 100,000 Ohio jobs".

The use of climate science as a weapon to skewer political opponents does not bode well for bipartisan progress on climate after the election. "If the message is that climate legislation is political poison, then that will make it harder to bring it up in the future," says David Goldston, director of government affairs at the Natural Resources Defense Council in Washington DC.

Prospects for the kind of emissions-trading programme that allows polluters to buy and sell permits on a fluid carbon market have already faded, says Frank Maisano, an energy specialist with the lobby firm Bracewell & Giuliani in Washington DC. Maisano notes that most of the sitting lawmakers who are likely to lose in November — moderate Democrats and Republicans — did not support aggressive action on climate science in any case. "This is regional politics, not partisan politics."

A more divided Congress could take up smaller initiatives targeting energy efficiency, renewable energy and long-term investments in clean-energy. But the Environmental Protection Agency (EPA) is poised to act on a 2007 court ruling giving it authority to regulate carbon-dioxide emissions. Some observers suggest that the fear of direct EPA regulation could help to spur a legislative solution among moderates of both parties.

That kind of political compromise might yet be possible if the climate rhetoric tones down after the elections. "Climate-science denial is a by-product of extreme partisanship and a kind of reactionary mode among conservatives, and I expect that this will wane," says Paul Bledsoe, a senior adviser to the Bipartisan Policy Center, a centrist think tank based in Washington DC. He says that most Republicans in the current Congress accept the science even if they disagree over what to do about it. "But if large parts of the Republican Party begin to deny consensus science," he adds, "then the climate community will have to confront them about it." ∎

K. DJANSEZIAN/GETTY

STATE WATCH: CALIFORNIA

Voters in California must decide whether to effectively repeal the global-warming law AB 32, which requires the state to cut greenhouse-gas emissions to 1990 levels by 2020, a 25% decrease. 'Proposition 23' would repeal the law until state unemployment — currently 12.4% — drops to 5.5% or less for a full year. Critics say that has happened only three times in the past three decades. Both major candidates for governor, Democrat Jerry Brown and Republican Meg Whitman (pictured), oppose the proposition, although Whitman initially supported the move and has criticized the climate law. She says she will implement a one-year moratorium on AB 32 if elected.

ELECTIONS

The Constellation programme: gone in name only.

US MIDTERM ELECTIONS

Policy row launches NASA into limbo

Funding dispute leaves space programme on autopilot.

BY ADAM MANN

NASA is caught between President Barack Obama and Congress, and Republican gains in the midterm elections could prolong the agony. In February, Obama called for an end to the Constellation programme to take astronauts to the Moon and Mars after a review called it underfunded and overambitious. The administration wants to invest in new technology and private spaceflight.

Congress baulked, and on 29 September passed an act requesting funds for projects initiated under Constellation and granting less than half of the administration's request for private spaceflight. With Congress now in recess and unable to allocate funds, NASA is funded at current levels and cannot change course.

"It has to keep spending on programmes that it's going to kill, and can't start new programmes that it wants to," says Keith Cowing, editor-in-chief of the website NASA Watch.

If the Democrats retain control of the Senate and House of Representatives, they will probably quickly pass an appropriations bill to allocate the money, says John Logsdon, former director of the Space Policy Institute at the George Washington University in Washington DC. But if the Republicans prevail, they may defer negotiations until January, when they would take over the appropriations subcommittee.

Appropriators must also find US$500 million for a key project in the act: an extra flight for the retiring space shuttle. The money could come from NASA's science allocation, but the administration would object to that, says Logsdon.

For Cowing, competing interests could spell a troubled future for the space agency. "How can NASA progress if it's pulled in ten different directions?" he asks. ∎

US MIDTERM ELECTIONS

Opponents battle health–care research

Treatment–outcome studies and key health reforms face Republican antipathy and legal challenges.

BY HEIDI LEDFORD

When President Barack Obama signed the US health-care reform act into law on 23 March, he seemed to have emerged victorious from a bitter and protracted political battle. But the fight over the legislation promises to continue beyond the midterm elections next month. If the Republicans gain a significant lead in the House of Representatives, "there will be political trench warfare", says Henry Aaron, a senior fellow at the Brookings Institution, a think tank based in Washington DC.

Health insurance in the United States is provided mainly by the private sector, and 16.7% of Americans don't have any. The reform act aims to cut the number of uninsured and to stem soaring costs — two factors that threaten to cripple the US health-care system.

In addition to requiring that everyone take out health insurance, the act gives the makers of complex biologic drugs 12 years free from generic competition, and offers a tax credit to young biotechnology firms. It also sets up the Patient-Centered Outcomes Research Institute (PCORI) to coordinate comparative-effectiveness research, which compares treatment options to see which is the most effective.

But comparative-effectiveness research has proved controversial. Proponents say it will cut wasteful spending, but critics claim that it means health-care rationing and even, in an erroneous assertion, the formation of 'death panels' that limit access to expensive but life-saving treatments. Debate over the research, which has largely died down while Republicans focus on attacking insurance provisions in the health-care act, could resurface in the spring, says Dan Holler, spokesman for the conservative advocacy group Heritage Action for America, based in Washington DC.

Some Republicans boast that they will try to repeal the health-care act — a bold strategy that is unlikely to succeed. Any such attempt would presumably provoke a veto from Obama, and it is doubtful whether the Republicans could amass the two-thirds majority needed to override a presidential veto.

Another approach is to underfund specific programmes. Representative Michael Burgess (Republican, Texas), the ranking Republican on the Subcommittee on Oversight and Investigations of the House Committee on Energy and Commerce, has already pledged to take this approach. PCORI is one provision that could be vulnerable. Funding has been set aside for it to 2012; thereafter, its money will come from sources including $150 million in appropriations and a $2 fee tacked onto insurance plans. But Republicans could attack PCORI by restricting funds to the agency responsible for collecting the $2 fee. "You certainly could see a significant spending fight on comparative effectiveness," says Holler.

Meanwhile, voters in Arizona (see 'State watch: Arizona'), Colorado and Oklahoma will decide whether to amend their state constitutions to prohibit any requirement that citizens take out health insurance. In August, Missouri voters overwhelmingly passed similar legislation. Although national law would trump these state mandates, making them largely symbolic, 20 states have launched a lawsuit asserting that the health-care act violates the US constitution. ∎

STATE WATCH: ARIZONA

Arizona is ahead of the pack in the fight against health-care reform. In a narrowly defeated initiative, the state tried to ban mandatory health insurance two years ago. Now it is trying again, this time with support from Republican Governor Jan Brewer (pictured), who has also joined 19 other states in suing the federal government over the legislation. Brewer is up for election and she has used her tough stance on the reforms to fight her opponent, Democrat Terry Goddard, who supports some aspects of the legislation.

The golden poison frog (*Phyllobates terribilis*) is one of many endangered amphibian species.

CONSERVATION

Biodiversity hope faces extinction

Upcoming meeting will set out global conservation targets.

BY NATASHA GILBERT

The future of the world's biodiversity hangs in the balance as countries convene for crucial negotiations next week in Nagoya, Japan. The 193 signatory nations to the Convention on Biological Diversity (CBD) pledged eight years ago to cut species loss "significantly" by this year. But studies show that the health of global biodiversity is reaching a crisis point, with extinctions of mammal and amphibian species continuing to rise (see *Nature* **462,** 263; 2009).

Conservation leaders hope that the meeting of CBD nations on 18–29 October will give the green light to a new set of international agreements to help protect biodiversity. These include 20 new ambitions to be met by 2020, such as halving the loss and degradation of forests and other natural habitats, and ensuring that agriculture and aquaculture are sustainably managed (see *Nature* doi:10.1038/news.2010.31; 2010).

Agreement on the general ambitions looks likely, but large international rifts over conservation funding and access to the genetic resources of ecologically rich nations are threatening to derail the negotiations over how to achieve them. Failure to reach agreement could be a severe setback for international efforts to save biodiversity, says Jane Smart, director of biodiversity conservation at the International Union for Conservation of Nature, headquartered in Gland, Switzerland. "It all comes down to what happens at Nagoya,"

she says. "If it does not succeed, rebuilding morale and momentum will be such a big mountain to climb."

A lack of money for conservation projects is a key reason why governments have failed to meet their 2010 targets, says a spokesman for the CBD. For example, the 33 member nations of the Organisation for Economic Co-operation and Development donated US$8.5 billion for climate-change mitigation projects in 2008, but only $3 billion annually for biodiversity.

More than 17,000 species of fauna and flora were at RISK OF EXTINCTION in 2009.

Several proposals for mustering funds are up for discussion next week. The billion-dollar budget for the Global Environment Facility — an international partnership used by wealthy countries, businesses and other stakeholders to fund biodiversity projects — could be doubled, for example. Japan is expected to use the meeting to announce a trust fund for biodiversity, says Ahmed Djoghlaf, executive secretary of the CBD. "I hope Europe will do the same," he says. But Djoghlaf declined to say how much money he would like donors to pledge. "It's not a figure we are looking for, but an approach."

One fund-raising mechanism gaining support is the phasing out of subsidies that are

harmful to biodiversity. For example, global subsidies to marine fisheries are estimated at around $15 billion–$35 billion annually, according to 2008 figures from the United Nations Environment Programme. The Food and Agriculture Organization of the United Nations says that fish stocks in almost one-third of these fisheries are close to collapse or have already done so.

"If we cut subsidies, this will free up a lot of funds that can be put back into protecting fisheries," says Lina Barrera, director of biodiversity and ecosystem services at Conservation International, a non-profit green group based in Washington DC. Barrera is optimistic that the proposal to cut subsidies will be approved, because it has come through pre-meeting negotiations "relatively unscathed".

But there is likely to be tense debate on a proposed legally binding agreement to tighten rules for access to nations' genetic resources, such as plants that potentially produce pharmaceuticals. Biodiversity-rich countries, which are typically in the developing world, are concerned that the access and benefits scheme that has been proposed is not strong

Over half of all MEDICINAL PLANTS in Africa face EXTINCTION.

enough to guarantee them financial compensation if their resources are developed into commercial products by foreign companies, which are often based in developed nations. "The wording is very ambiguous. Some countries feel they could be making resources available on terms that are not good for them, and that they could be giving up their profits," says Geoff Burton, an expert on biodiversity law at the United Nations University Institute of Advanced Studies in Yokohama, Japan.

Developing countries want the rules to cover products that have been generated in the past, but companies and developed countries say that only products worked on after the regulations are legally implemented should be affected. However, the scheme could give basic researchers easier access to genetic resources: earlier this year, negotiators reached a broad agreement that non-commercial research should not be inhibited by any future rules (see page 779).

Burton is concerned that failure to reach agreement on the more contentious aspects of the rules could derail the non-commercial exemption before it is signed by all the parties. But Djoghlaf is optimistic that countries will agree to a set of core principles, and then set a timetable for hammering out the finer details by 2012. "I can't imagine closing the meeting without agreement on the access and benefit sharing scheme," he says. ■ SEE EDITORIAL P.752

↪ NATURE.COM
For breaking news from the Nagoya summit, see:
nature.com/news

AWARDS

Chemistry Nobel won by … chemists

Synthesis method scoops prize for first time in five years.

BY RICHARD VAN NOORDEN

For three of the past four years, the committee that awards the Nobel Prize in Chemistry has scattered its medals in biological fields (see 'Biology's decade'), prompting inevitable and often fruitless introspection from chemists about their field's identity. So this year's award has been welcomed as a return to what many see as chemistry's core, almost monastic, discipline: synthesizing molecules.

Richard Heck, Ei-ichi Negishi and Akira Suzuki won for their development of reactions that stitch together molecules by cleanly connecting chains of carbon atoms, using small amounts of catalysts based on the metal palladium. The processes they uncovered for forming carbon–carbon bonds in the 1960s and 1970s are now a staple of chemistry textbooks, and are used every day in industry and academia.

Chemists hoping to connect carbon chains once faced an arduous task. Reaction mixtures often produced a mess of by-products that would form a sludge at the bottom of a chemist's reaction flask. But by 1968, Heck — then at Hercules, a chemical company in Wilmington, Delaware, and later at the University of Delaware in Newark — had found that palladium combines with organic molecules to produce a catalyst that dramatically improves such tricky couplings. The palladium atom inserts itself between atoms in the carbon–carbon chains as the reaction proceeds, effectively guiding molecules towards the desired product with few messy side reactions. In the late 1970s, Negishi, of Purdue University in West Lafayette, Indiana, and Suzuki, at Hokkaido University in Sapporo, Japan, published their own variations on this theme, each using different palladium-based catalysts.
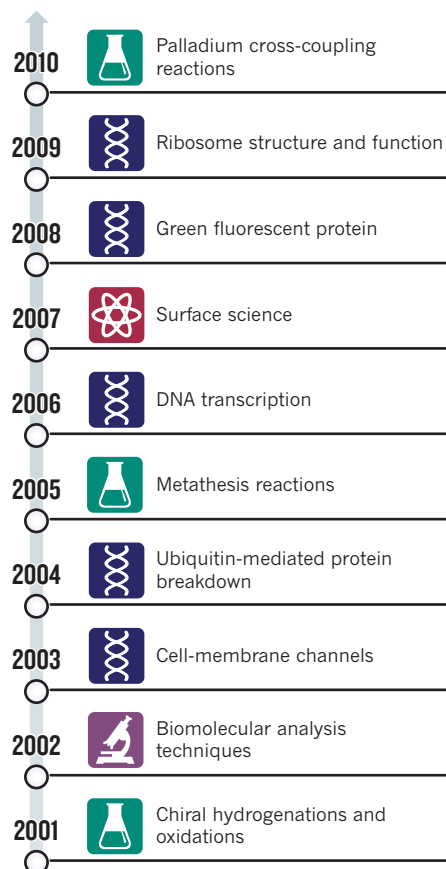
The palladium couplings didn't just ease the process of synthesizing compounds; they prompted molecule-makers to experiment with new classes of drugs, and with materials that were previously too hard to make. It took a decade before the processes caught on with industrial chemists, says Andrew Wells, a chemist at multinational drug giant Astra-Zeneca in Loughborough, UK, but about 15–30% of compounds in development at his company now rely on palladium-catalysed cross-coupling reactions. The anti-cancer drug Gleevec (imatinib) and the blood-pressure medication Diovan (valsartan), as well as the fungicide boscalid, are just some of the fruits of palladium chemistry.

Many chemists say they are delighted by the recognition of work at the heart of their discipline. "There's an extent to which chemistry is buggered at both ends," comments Eric Scerri, a philosopher of chemistry at the University of California, Los Angeles. "Physicists say it's nothing but physics. At the other end it's being encroached on by biologists who see it as just a tool for biology."

Roald Hoffmann of Cornell University in Ithaca, New York, who won the 1981 Nobel Prize in Chemistry, sees no point in complaining about the Nobel committee's biological bent. "The Nobel committee has decided that molecular biology and biochemistry are chemistry, no matter what chemists think," he says. "I don't think it's quite right, but I don't feel too bad about it." ∎

BIOLOGY'S DECADE

Since 2001, the chemistry Nobel has been awarded in biological areas more often than for synthesis, physical or analytical work

2010 Palladium cross-coupling reactions

2009 Ribosome structure and function

2008 Green fluorescent protein

2007 Surface science

2006 DNA transcription

2005 Metathesis reactions

2004 Ubiquitin-mediated protein breakdown

2003 Cell-membrane channels

2002 Biomolecular analysis techniques

2001 Chiral hydrogenations and oxidations

GENETICS

Cancer–gene testing ramps up

Thousands to get personalized medicine in Britain's National Health Service.

BY EWEN CALLAWAY

In an approach that many doctors and scientists hope will form the medical care of the future, Massachusetts General Hospital in Boston has for the past year and a half been offering people with cancer a novel diagnostic test. Instead of assessing tumours for a single mutation that will indicate whether a drug is likely to work or not, the hospital tests patients for some 150 mutations in more than a dozen cancer-causing genes, with the results being used to guide novel treatments, clinical trials and basic research. This form of personalized medicine tailors treatments on the basis of the molecular and genetic characteristics of a patient's cancer cells, potentially improving the treatment's outcome.

Now Britain is set to test whether an entire health-care system is ready for the approach. Plans were unveiled this week to deploy broad genetic testing for selected cancer patients in Britain's government-run health-care provider,

the National Health Service (NHS). This form of 'stratified medicine' uses genetic information to group patients according to their likely response to a particular treatment.

"The United Kingdom is really the ideal place to do this," says James Peach, who heads the programme for Cancer Research UK, the charity that is leading the effort. As the NHS treats millions of people each year, unprecedented numbers of suitable patients could be enrolled in the genetic-profiling programme. "The idea is to scale this up to every patient in the NHS," says Peach. In its first phase, the programme will be rolled out to as many as 12,000 NHS cancer patients over two years, beginning in early 2011. By contrast, Massachusetts General has tested about 1,600 patients, and other hospitals' efforts each number in the hundreds.

The tests, which will look for several dozen mutations in about a dozen genes linked to cancer, will be carried out on people with lung, breast, colorectal, prostate or ovarian cancers,

or metastatic melanoma, who are being treated at six NHS hospitals. Therapies that target specific tumour-causing mutations have already been approved, or are on the verge of approval, for most of these conditions, says Peach.

Testing a clinical sample for so many mutations at once is a challenge in itself. Because most existing clinical tests probe individual genes, the NHS programme is working with the Technology Strategy Board, a government agency that supports technology development, and several companies to design a customized test that detects all of these mutations in one go. The partnership, which includes the pharmaceutical multinationals Pfizer and AstraZeneca, will also design software to make the results useful to researchers and clinicians. By genotyping patients for a broad array of cancer-causing mutations, the new tests will make it easier to assign subjects to clinical trials, Peach says.

That is already happening at Massachusetts General, where the test is helping to establish

clinical trials that wouldn't otherwise have happened, says Leif Ellisen, a geneticist who helps lead the hospital's cancer testing programme. For example, its broad genetic test detects a mutation in a gene called *BRAF* that is already known to be commonly mutated in metastatic melanoma. Finding such mutations in people with lung and colon cancer made it possible to put them in a trial of an experimental treatment targeting that gene, Ellisen explains.

Basic research should also benefit from the NHS programme, says Peach. Researchers will have access to consenting patients' genetic data

as well as to medical records of the outcomes of the treatment. These data could reveal how drugs targeting one molecular pathway are affected by mutations in another gene, says Andy Futreal, a cancer geneticist at the Wellcome Trust Sanger Institute in Hinxton, UK, and an adviser to the programme.

Peach hopes that the first phase of the cancer programme will pave the way for expanding genetic testing to more patients and other conditions, such as diabetes, AIDS and even psychiatric disorders. Cancer offers a good testing ground for personalized medicine,

because numerous targeted therapies already exist, but "there's no reason why this should be restricted to cancer", says Peach.

Fabrice André, who runs a similar cancer diagnostic programme that has so far been offered to about 100 patients at the Gustave Roussy Institute in Villejuif, France, says the NHS programme could point the way to implementing personalized medicine across an entire population. "It can really change the landscape of how molecular testing is being done for cancer," he says. "If they succeed, then it's going to be a major step forward." ∎

MEDICINE

Global disease fund hit by cash crisis

Pledges fall short for fight against AIDS, TB and malaria.

BY DECLAN BUTLER

Going into a meeting of donors in New York City on 5 October, the Global Fund to Fight AIDS, Tuberculosis and Malaria was aiming high. Its representatives hoped to raise US$20 billion, enough to scale up the campaign against these diseases. But countries pledged just $11.7 billion for 2011–13, although the fund still hopes to raise the minimum $13 billion needed to maintain its current programmes.

The fund accounts for a quarter of all international financing to fight AIDS, two-thirds of that for tuberculosis, and three-quarters of that for malaria. By its own estimates, it has saved 4.9 million lives since its creation in 2002.

"It's a classic good news, bad news story,"

says Andrew Hurst, a spokesman for the fund, which is based in Geneva, Switzerland. The pledge marks a 20% increase on the $9.7 billion raised for the period 2008–10 at its last fundraising meeting in Berlin in 2007. Given today's economic climate, the outcome could have been far worse, he says.

The bad news is that the fund has not raised enough cash to reduce disease burdens among the poor by 2015, to levels to which countries have committed in the UN Millennium Development Goals. In the run-up to the meeting, the fund had estimated that it needed $20 billion to start making serious progress on this task, whereas $13 billion would buy only minimal burden reductions (see 'The human cost').

The poor funding outcome starkly highlights a chronic structural weakness of the

Global Fund: its narrow base of donors. Just a couple of dozen countries donated, and a mere seven countries gave more than $100 million — the United States, France, Germany, Japan, Canada, Norway and Australia. Three other countries traditionally supportive of the fund — the United Kingdom, the Netherlands and Sweden — were unable to officially pledge at the meeting because of the timing of national budgets, but are likely to commit sums that will allow the Global Fund to reach its $13-billion lower target. Yet dozens of rich countries, including Italy, Spain and Saudi Arabia, gave nothing. The fund is seeking to persuade more nations to pledge, says Hurst, but admits that "we still have some way to go".

For the first time in its history, the fund will now be forced to reject otherwise-solid new proposals from recipient countries, and trim others. "Today marks a sad turning point in the fight against AIDS, TB and malaria, as world leaders have officially underfinanced the Global Fund," says Jennifer Cohn, HIV/AIDS policy adviser at Médecins Sans Frontières in Geneva. "This decision will result in the deaths of millions of people from otherwise treatable diseases." ∎

⟳ NATURE.COM
Who pledged?
How much?
go.nature.com/ekhcxd

THE HUMAN COST

Estimating the effect of different funding scenarios for 2011–13 shows that although US$20 billion would allow a significant scale-up of activities, $13 billion yields far less progress towards internationally agreed targets for 2015.
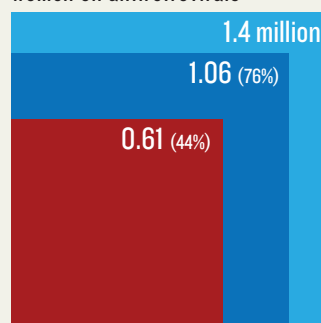
- Global target
- $20-billion fund
- $13-billion fund

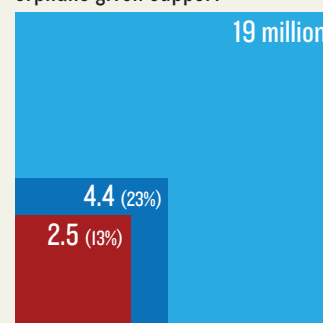**Number of people to get antiretrovirals for HIV/AIDS**
21.9 million
7.5 (34%)
4.4 (20%)
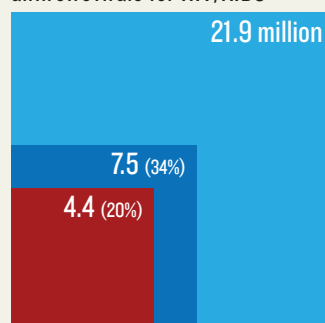
Number of insecticidal bednets to tackle malaria
264 million
190 (72%)
110 (42%)
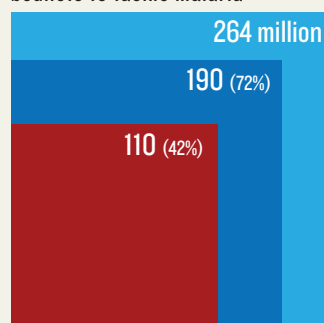
HIV-positive pregnant women on antiretrovirals
1.4 million
1.06 (76%)
0.61 (44%)

Number of AIDS orphans given support
19 million
4.4 (23%)
2.5 (13%)

The end of the cold war left uranium fuel stranded at research reactors around the world, including the Maria reactor in Poland.

A NUCLEAR EXCHANGE

*More than 100 cold-war era research reactors run on uranium
pure enough to be used in a nuclear weapon.
But switching to safer fuel isn't easy.*

BY GEOFF BRUMFIEL

I t is a warm day in early autumn, and speckled sunlight shines
through spruce and linden trees onto a crumbling road. Ahead
stands a dilapidated brick building encircled by a chain-link fence.
It seems peaceful enough. But attached to the fence is a sign that
says *Teren Nadzorowany* — 'supervised area' — topped by the faded
remains of a red trefoil: the symbol used world-over to denote a radia-
tion hazard. The sign is the only visible hint of what once lay inside the
building: enough highly enriched uranium reactor fuel to make more
than 18 nuclear bombs.

A group of visiting journalists and US officials files into the building
and looks down into pools of ultrapure water in which the fuel once lay.
The water is still and clear, the fuel basins empty. Here at the Polish Insti-
tute of Atomic Energy in Otwock-Świerk, 30 kilometres southeast of
Warsaw, fuel-removal operations have been under way for a year. They
end today: the last shipment has been pulled out of the pools and packed
away in seven bright-blue shipping containers, which are now sitting on
trucks back at the gate. Soon, the containers will begin a high-security
journey to a reprocessing plant in Russia, east of the Ural Mountains,

where the fuel's uranium will be converted into a safer form for use in power reactors.

The trip will be a homecoming of sorts: this fuel originally came to Poland from Russia more than 30 years ago. But officials of the US National Nuclear Security Administration (NNSA), sensitized by the attacks of 11 September 2001, feared that this kind of highly enriched fuel posed too tempting a target for terrorists seeking to build a nuclear weapon. So a year ago, after much negotiation with the Polish government, they began clearing out the dangerous uranium and converting the reactor to run on safer fuel. For this last day of removal, *Nature* was granted a rare chance to watch the American programme in action.

The reactor at the Polish institute is hardly unique: globally, roughly 130 research reactors in state laboratories and university campuses continue to run on bomb-grade fuel (see map). US President Barack Obama did not explicitly mention these reactors in April 2009, when he announced a substantial acceleration of established non-proliferation efforts and made a commitment to ensure the security of all vulnerable nuclear material within just four years. But many of the facilities clearly do fall into that 'vulnerable' category.

Research-reactor operators are not always enthusiastic about cooperating with the American initiative. Safer alternative fuels are available, but converting the reactors to use them is difficult, time-consuming and costly. "We are not so happy to convert our reactor," says Grzegorz Krzysztoszek, the director of the Polish facility. But the NNSA is winning the operators over by offering to remove old uranium, help with the conversion process and pay for new fuel — a US$70-million proposition in this case. The US agency's Global Threat Reduction Initiative has already helped to convert or shut down 72 vulnerable research reactors worldwide. Working on Obama's accelerated timetable, the NNSA is now hoping to spend nearly US$3 billion over the next four years to secure or convert as many reactors as possible and clear out uranium from ten countries. "It's an aggressive schedule," says Andrew Bieniawski, the head of the operation. "But I do think we can make it."

LEFT BEHIND

Poland's research reactor was built during the cold war, and is one of many that the United States and the former Soviet Union helped to construct during that period. Named Maria after the Polish-born scientist Marie Curie, the 30-megawatt behemoth was actually one of two large research reactors that the Soviets helped to build at the institute starting in the 1950s. An important part of the deal was that the Soviets would also keep the reactors supplied with highly enriched uranium fuel.

But the fall of the Soviet Union in 1991 led to hard times. Plans for a Polish commercial power reactor in the north of the country fell

GLOBAL REACTION

Around the world, some 130 nuclear research reactors run on bomb-grade highly enriched uranium (HEU); others have been shut down or converted to safer fuel, or are being converted.

RESEARCH REACTORS

■ HEU, targeted for conversion   ■ Converted   ■ Partially converted

SOURCE: NNSA
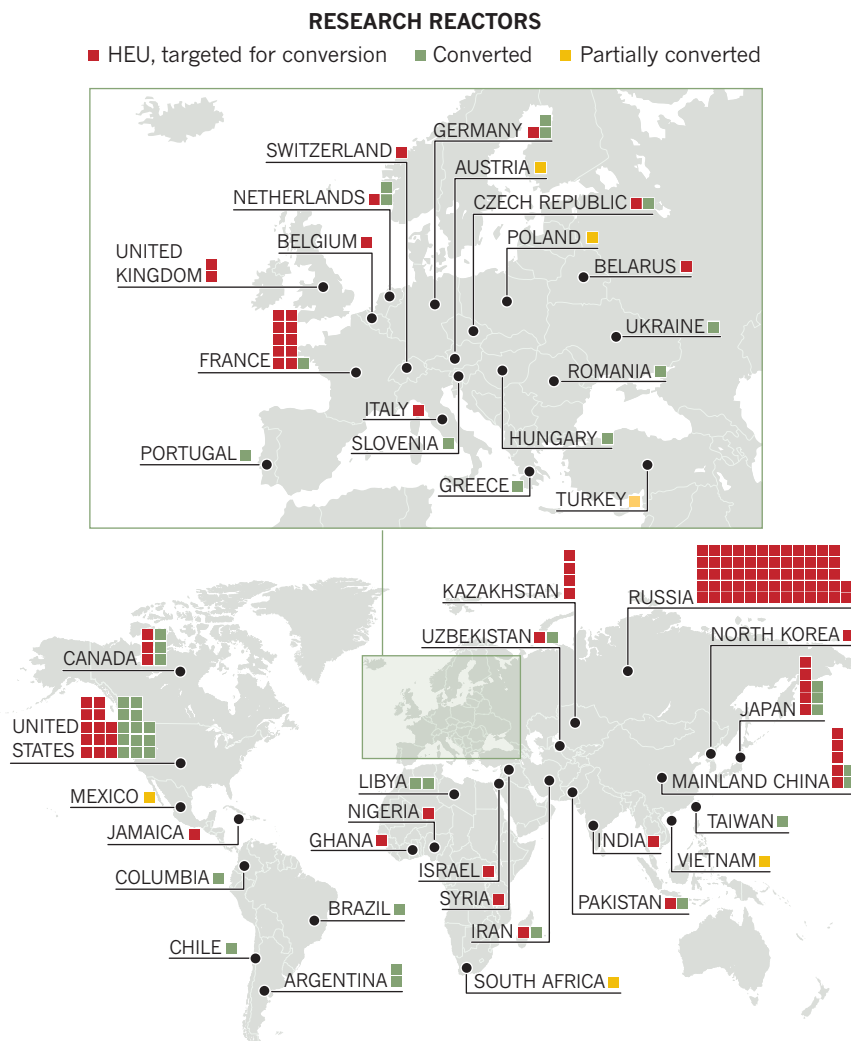
through, and the older of the institute's reactors was shut down in 1995. That left only Maria — which has become something of a relic. Trail-blazing nuclear physics is now done at particle accelerators, not research reactors.

Yet Maria hums on, Poland's last remaining tie to the nuclear age. Even if Maria is no longer at the cutting edge, Krzysztoszek says that the reactor has found its niche in more utilitarian research — primarily the use of neutrons from the core for imaging, and the production of medical isotopes. About 60% of Maria's US$5-million annual budget comes from industrial clients, including a US company that hopes to use it to produce molybdenum-99, a short-lived isotope used in medical imaging. Outages in reactors in Canada and the Netherlands have created a dire worldwide shortage of this isotope (see *Nature* **460,** 312–313; 2009), and the Polish institute hopes to profit from it. Domestically, the reactor also features in plans

for a new nuclear age of energy independence. The Polish government wants to move away from using dirty coal or Russian-supplied gas, says Krzysztoszek. By 2021–22, it hopes to have two nuclear power plants in operation. But that will require training Polish scientists and engineers in the nuclear arts, and this reactor is the only domestic facility available.

Similar stories can be heard at other research facilities around the world: they still do worthwhile work. "We're not running a reactor just to run a reactor. We do things with it," says David Moncton, director of the research reactor at the Massachusetts Institute of Technology (MIT) in Cambridge, one of a handful on US soil that continues to operate with bomb-grade fuel. The MIT reactor is used to test new nuclear fuels and components, but it also serves less-obvious functions: for example, doping silicon semiconductors by transmuting some of the silicon into phosphorus. The semiconductor wafers

are used as high-performance electronic components in hybrid vehicles.

Whether a reactor is useful or not, the fuel is still a security nightmare. The concern is that a terrorist group might somehow get hold of it and extract enough of the fissile isotope uranium-235 to construct a crude bomb. When atoms of uranium-235 split, they produce energy and neutrons. The neutrons strike nearby uranium-235 nuclei, causing them to split and creating a self-perpetuating chain reaction, which, if uncontrolled, leads to an explosion. To prevent just anyone from gaining a nuclear weapon, fuel for commercial power reactors is diluted with copious amounts of uranium-238 — a more common isotope that does not split. Because the two isotopes are chemically identical, they are impossible to separate without highly specialized equipment.

Most research reactors, however, run on up to 93% pure uranium-235, mixed with elements such as aluminium. This lets the reactors operate for longer and produce more neutrons with less fuel. But the uranium can easily be separated from other elements in the fuel, so extracting enough for a nuclear bomb is comparatively easy — and within the means of a sophisticated terrorist group.

VULNERABLE TO ATTACK

Compounding the problem of fuel security, research reactors are often housed in dilapidated state laboratories or the basements of university research buildings. "Many of these places have very minimal security," says Matthew Bunn, an expert in nuclear terrorism at Harvard University in Cambridge, Massachusetts. The threat is more than academic: in 2007, two teams of armed men assaulted the Pelindaba reactor near Pretoria, South Africa. While one team engaged the site's security forces, who fled, the other men penetrated an electrified fence and made their way to an emergency control centre inside the facility. There, they shot a worker in the chest before fleeing. They were never apprehended. No uranium-235 fuel was reported missing — indeed, no one knows what the assailants were after — but the incident underscores the vulnerability of civilian nuclear facilities.

Security was much on the mind of the NNSA negotiators as they worked out a deal with the Polish institute, where 454.9 kilograms of fresh and used highly enriched uranium fuel were left stranded by the collapse of the Soviet Union. And security remains very much on the minds of the US and Polish team members today.

By mid-afternoon, police cars have gathered at a gate near the reactor building. Officers in black T-shirts cluster in small groups near an adjacent yard where seven trucks are parked, each carrying one of the blue containers of the

Grzegorz Krzysztoszek is overseeing the reactor's conversion.

last remaining spent fuel. A military helicopter suddenly roars over the treetops and begins to circle overhead; the airborne escort has arrived, and the convoy quickly lurches into action. One by one, the trucks file out of the gate towards the railway yard — the first stop on the long journey to Russia.

With their departure, the only highly enriched uranium remaining on site lies inside the reactor itself. As in other research reactors covered by the NNSA programme, that fuel will systematically be replaced with a safer one.

This is not as easy as it sounds. Researchers in the United States and elsewhere had to go to great lengths to develop safe blends of around 20% uranium-235 that will work in the cores of research reactors but can't be fashioned into a bomb. Because the reactors are designed for highly enriched uranium, they can run only on a very high density of uranium-235 atoms, which can make the fuel less stable. So to keep it from breaking down in the intense environment at the reactor's core, researchers have had to come up with exotic mixtures of uranium and elements such as silicon and molybdenum.

That development effort was a success, and the fuel is now available for many research reactors, including Maria. But most institutions with research reactors don't have the millions of dollars needed to buy the new fuel — or to convert the core to use it. "It's a hell of a lot of work," says Michael Corradini, a mechanical and nuclear engineer at the University of Wisconsin–Madison, which last year made the switch to the safer fuel.

At Maria, even the most mundane-seeming issues can be a challenge. For example, the cooling water that helps to regulate the fission reaction doesn't circulate as well through the new blended fuel, Krzysztoszek says, so

the pumps must be upgraded. Moreover, the core will have to be converted little by little in a laborious operation that will take nearly two years.

FINAL DESTINATION

By dusk, the convoy has reached a railway yard on the outskirts of Warsaw. Shipping containers are stacked around the train, part of an effort to shield the shipment from view during loading. The watchful helicopter peels off as a heavy loader lifts the containers onto flatbed cars. As darkness falls, the train begins to roll towards the Baltic port of Gdynia, some 400 kilometres to the north.

In the end, the promise of ridding the facility of the old fuel helped to convince the Poles to convert Maria to the blended fuel. Like most countries, Poland has no long-term facility for storing spent nuclear material, and the prospect of US aid to take it back to Russia was simply too tempting to pass up. It is hoped that the expansion of the US repatriation programme, and the development of new fuels for research reactors, will convince many countries to surrender their uranium. Over the next four years, the NNSA hopes to make a similar switch

"MANY OF THESE PLACES HAVE VERY MINIMAL SECURITY."

over in about ten countries around the world, and to complete most of its work worldwide by 2020.

Right now, the US teams are stretched thin, frantically trying to accelerate the repatriation of fuel. American-supplied fuel will go to the United States, and Soviet material will go to Russia. Meanwhile, scientists at the US Department of Energy's laboratories are close to certifying another kind of even-higher density fuel that will work in reactors that cannot take the current mixtures available.

The biggest problem, says Bunn, is that Russia itself continues to operate 60-odd reactors on bomb-grade material. "They've been very resistant" to changing, he says. The NNSA still hopes to win the Russians over, and says that it already has some six reactors in the country considering conversion.

The next morning finds Bieniawski and the rest of the American team standing on the deck of a Russian cargo ship that has been specially outfitted to carry the fuel. Bieniawski watches the containers being lowered slowly into the hold. "This is the last we'll physically see of it," he says. He and his team are already mustering for a trip to another reactor site, which holds more fuel that could be fashioned into a weapon. When asked where it is, he demurs: "Let's talk about that later." ∎

Geoff Brumfiel is a senior reporter for Nature in London.

C:\lab>
f77 -o
data.exe
>
>

...ERROR

...why scientific programming does not compute

BY ZEEYA MERALI

hen hackers leaked thousands of e-mails from the Climatic Research Unit (CRU) at the University of East Anglia in Norwich, UK, last year, global-warming sceptics pored over the documents for signs that researchers had manipulated data. No such evidence emerged, but the e-mails did reveal another problem — one described by a CRU employee named "Harry", who often wrote of his wrestling matches with wonky computer software.

"Yup, my awful programming strikes again," Harry lamented in one of his notes, as he attempted to correct a code analysing weather-station data from Mexico.

Although Harry's frustrations did not ultimately compromise CRU's work, his difficulties will strike a chord with scientists in a wide range of disciplines who do a large amount of coding. Researchers are spending more and more time writing computer software to model biological structures, simulate the early evolution of the Universe and analyse past climate data, among other topics. But programming experts have little faith that most scientists are up to the task.

A quarter of a century ago, most of the computing work done by scientists was relatively straightforward. But as computers and programming tools have grown more complex, scientists have hit a "steep learning curve", says James Hack, director of the US National Center for Computational Sciences at Oak Ridge National Laboratory in Tennessee. "The level of effort and skills needed to keep up aren't in the wheelhouse of the average scientist."

As a general rule, researchers do not test or document their programs rigorously, and they rarely release their codes, making it almost impossible to reproduce and verify published results generated by scientific software, say computer scientists. At best, poorly written

...SCIENTISTS AND THEIR SOFTWARE

A survey of nearly 2,000 researchers showed how coding has become an important part of the research toolkit, but it also revealed some potential problems.

> **45%** said scientists spend more time today developing software than five years ago."

> **38%** of scientists spend at least one fifth of their time developing software.

> Only **47%** of scientists have a good understanding of software testing.

> Only **34%** of scientists think that formal training in developing software is important.

> programs cause researchers such as Harry to waste valuable time and energy. But the coding problems can sometimes cause substantial harm, and have forced some scientists to retract papers.

As recognition of these issues has grown, software experts and scientists have started exploring ways to improve the codes used in science. Some efforts teach researchers important programming skills, whereas others encourage collaboration between scientists and software engineers, and teach researchers to be more open about their code.

A PROPER EDUCATION

Greg Wilson, a computer scientist in Toronto, Canada, who heads Software Carpentry — an online course aimed at improving the computing skills of scientists — says that he woke up to the problem in the 1980s, when he was working at a physics supercomputing facility at the University of Edinburgh, UK. After a series of small mishaps, he realized that, without formal training in programming, it was easy for scientists trying to address some of the Universe's biggest questions to inadvertently introduce errors into their codes, potentially "doing more harm than good".

After decades griping about the poor coding skills of scientists he knew, Wilson decided to

see how widespread the problem was. In 2008, he and his colleagues conducted an online survey of almost 2,000 researchers, from students to senior academics, who were working with computers in a range of sciences. What he found was worse than he had anticipated[1] (see 'Scientists and their software'). "There are terrifying statistics showing that almost all of what scientists know about coding is self-taught," says Wilson. "They just don't know how bad they are."

As a result, codes may be riddled with tiny errors that do not cause the program to break down, but may drastically change the scientific results that it spits out. One such error tripped up a structural-biology group led by Geoffrey Chang of the Scripps Research Institute in La Jolla, California. In 2006, the team realized that a computer program supplied by another lab had flipped a minus sign, which in turn reversed two columns of input data, causing protein crystal structures that the group had derived to be inverted. Chang says that the other lab provided the code with the best intentions, and "you just trust the code to do the right job". His group was forced to retract five papers published in *Science*, the *Journal of Molecular Biology* and *Proceedings of the National Academy of Sciences,* and now triple checks everything, he says.

"How many fields have been held back, and how many people have had their careers disrupted, because of a buggy program?" asks Wilson.

More-rigorous testing could help. Diane Kelly, a computer scientist at the Royal Military College of Canada in Kingston, Ontario, says the problem is that scientists rely on "validation testing" — looking to see whether the answer that the code produces roughly matches what the scientists expect — and this can miss important errors[2]. The software industry relies on a different approach: breaking codes into manageable chunks and testing each piece individually, then visually inspecting the lines of code that stitch these chunks together (see 'Practicing safe software').

Many programmers in industry are also trained to annotate their code clearly, so that others can understand its function and easily build on it. But scientists often lack these communication and documentation skills. Even if researchers lift a whole working code

and reuse it, rather than writing their own, they can apply the program incorrectly if it lacks clear documentation. Aaron Darling, a computational biologist at the University of California, Davis, unwittingly caused such a mistake with his own computer code for comparing genomes to reconstruct evolutionary relationships. He had designed the program to work only with closely related organisms, but discovered that an independent group had used it to look at sequences far outside the code's working range.

"It was lucky that I came across it, because their published results were totally wrong, but they couldn't know that because I hadn't clearly documented how my code worked," says Darling. "It's not something that I am proud of, but I am careful to be more clear now."

SLAYING THE MONSTER

Problems created by bad documentation are further amplified when successful codes are modified by others to fit new purposes. The result is the bane of many a graduate student or postdoc's life: the 'monster code'. Sometimes decades old, these codes are notoriously messy and become progressively more nightmarish to handle, say computer scientists.

"You do have some successes, but you also end up with a huge stinking heap of software that doesn't work very well," says Darling.

The mangled coding of these monsters can sometimes make it difficult to check for errors. One example is a piece of code written to analyse the products of high-energy collisions at the Large Hadron Collider particle accelerator at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. The code had been developed over more than a decade by 600 people, "some of whom are excellent programmers and others who do not really know how to code very well", says David Rousseau, software coordinator for the ATLAS experiment at CERN. Wilson and his students tried to test the program, but they could not get very far: the code would not even run on their machines.

Rousseau says that the ATLAS group can test the software only on the Linux operating system at the moment, but is striving to make the code compatible with Mac computers. This is important, he says, "because different platforms expose different types of errors that may otherwise be overlooked".

Some software developers have found ways to combat the growth of monster code. One example is the Visualization Toolkit, an open-source, freely available software system for three-dimensional

<Could your code stand up to attack?>

```
C:\lab>
fil = imos
 end if

>
```

```
...PRACTICING SAFE SOFTWARE
> Five tips to make scientific code
more robust.
```

→ Use a version-control system:
Put source code, raw data files, parameters and other primary material into it to record what you did, and when.

▲ Track your materials:
Know the source of your software. Keep a record of what raw data were processed to produce a particular result, what tools were used to do the processing, and how the tools were set up.

✚ Write testable software:
Build large codes from smaller, easily testable chunks.

← Test the software:
And get somebody else to read it and look for bugs.

↕ Encourage sharing of software:
Make the code that you use in research freely available, when possible.

computer graphics. People can modify the software as they wish, and it is rerun each night on every computing platform that supports it, with the results published on the web. The process ensures that the software will work the same way on different systems.

That kind of openness has yet to infiltrate the scientific research world, where many leading science journals, including *Nature*, *Science* and *Proceedings of the National Academy of Sciences*, do not insist that authors make their code available. Rather, they require that authors provide enough information for results to be reproduced.

THE SEARCH FOR SOLUTIONS

In November 2009, a group of scientists, lawyers, journal editors, and funding representatives gathered for the Yale Law School Data and Code Sharing Roundtable in New Haven, Connecticut, where they recommended that scientists go further by providing links to the source-code and the data used to generate results when publishing. Although a step in the right direction, such requirements don't always solve the problem. Since 1996, The *Journal of Money, Credit and Banking* has required researchers to upload their codes and data to an archive. But a 2006 study revealed that of 150 papers submitted to the journal over the preceding decade that fell under this requirement, results could be independently replicated with the materials provided for fewer than 15 (ref. 3).

Proponents of openness argue that researchers seeking to replicate published results need access to the original software, but others say that more transparency may not help much. Martin Rees, president of the Royal Society in London, says it would be too much to ask reviewers to check code line by line. And in his own field of astrophysics, results can really be trusted only in cases in which a number of different groups have written independent codes to perform the same task and found similar results. Still, he acknowledges that "how to trust unique codes remains an issue".

There are signs that scientific leaders are now taking notice of these concerns. In 2009, the UK Engineering and Physical Sciences Research Council put out a call for help for scientists trying to create usable software, which led to the formation of the Software Sustainability Institute (SSI)

at the University of Edinburgh. The SSI unites trained software developers with scientists to help them add new lines to existing codes, allowing them to tackle extra tasks without the programs turning into monsters. They also try to share their products across disciplines, says Neil Chue Hong, the SSI's director. For instance, they recently helped build a code to query clinical records and help monitor the spread of disease. They are now sharing the structure of that code with researchers who are trying to use police records to identify crime hot spots. "It stops researchers wasting time reinventing the wheel for each new application," says Chue Hong.

Another solution is to bring trained computer scientists into research groups, either permanently or as part of temporary alliances. Software developer Nick Barnes has set up the Climate Code Foundation, based in Sheffield, UK, to help climate researchers. He was motivated by problems with NASA's Surface Temperature Analysis software, which was released to the public in 2007. Critics complained that the program, written in the scientific programming language Fortran, would not work on their machines and they could therefore not trust what it said about global warming. In consultation with NASA researchers, Barnes rewrote the code in a newer, more transparent

programming language — Python — reducing its length and making it easier for people who aren't software experts to understand how it functions. "Because of the immense public interest and the important policy issues at stake, it was worth taking the time to do that," says Barnes. His new code shows the same general warming trend as the original program.

In the long term, though, Barnes says that there needs to be a change in the way that science students are trained. He cites Wilson's online Software Carpentry course as a good model for how this can be done, to equip students with coding skills. Wilson developed the week-long course to introduce science graduate students to tools that have been software-industry standards for 30 years — such as 'version control', which allows multiple programmers to make changes to the same code, while keeping track of all changes.

Science administrators also need to value programming skills more highly, says David Gavaghan, a computational biologist at the University of Oxford, UK. "There needs to be a real shift in mindset away from worrying about how to get published in *Nature* and towards thinking about how to reward work that will be useful to the wider community."

Gavaghan now uses the software industry's 'master–apprentice' approach to train graduate students in his lab. New software projects are split up into bite-sized chunks, with each segment assigned to a pair of programmers — one experienced and one novice — who work together on it. "It forces students to become consistent code-builders," says Gavaghan.

Bringing industrial software-development practices into the lab cannot come too soon, says Wilson. The CRU e-mail affair was a warning to scientists to get their houses in order, he says. "To all scientists out there, ask yourselves what you would do if, tomorrow, some Republican senator trains the spotlight on you and decides to turn you into a political football. Could your code stand up to attack?" ■ **SEE WORLD VIEW, P.753**

Zeeya Merali is a freelance writer in London.

1. Hannay, J. E. *et al. Proc. 2nd Int. Workshop on Software Engineering for Computational Science and Engineering* (2009).
2. Kelly, D. *IEEE Software* **24,** 119–120 (2007).
3. McCullough, B. D., McGeary, K. A. & Harrison, T. D, *J. Money Credit Bank.* **38,** 1093–1107 (2006).

COMMENT

A researcher prepares to analyse plant samples from the biodiversity-rich regions around Hanoi.

P. GINTER/SCIENCE FACTION/CORBIS

Biology without borders

Fundamental research must not be hampered by an international agreement on sharing the benefits from national biodiversity, says **David Schindel**.

The supreme decision-making body of the United Nations Convention on Biological Diversity (CBD) meets in Nagoya, Japan, on 18–29 October 2010 for its tenth biennial conference. One of the most important items on the agenda is a new protocol which, if enacted, would specify how countries that are parties to the convention control access to their 'genetic resources' (including whole organisms, tissue samples and DNA extracts) and what benefits they can expect from sharing them. The negotiators' focus on genetic resources used to develop commercial products[1] has left non-commercial academic research in a perilous position[2]. One-size-fits-all legislation could have devastating effects on research conducted by foreign and local investigators, and even on the technological growth and economies of developing countries.

According to the CBD, countries can control access to their own species and set the terms for sharing any benefits resulting from their use by foreigners. Since 1993, only 15 of the 193 countries that have ratified the convention have passed legislation and created regulations to control access (another 58 have either legislation or regulations in place; see 'Where countries stand'). Most are hoping for a long-awaited international agreement to set global standards. The tenth Conference of the Parties (COP-10) could provide this.

Just last month, a CBD working group agreed on a new section to the draft protocol, proposed by the European Union and Japan. This directs CBD countries to encourage research that contributes to the conservation and sustainable use of biodiversity. More specifically, it directs them to create simplified access procedures for non-commercial research, with the understanding that mechanisms for handling unanticipated commercial applications may have to be developed. It is crucial that the parties to the convention approve the global access and sharing agreement only if this amendment is included.

SOVEREIGN RIGHTS

The rich biological diversity of many developing countries has long attracted biologists interested in evolution and ecology, as well as researchers looking for compounds that could be developed into products such as drugs and cosmetics. Before the creation of the CBD, most government ministries didn't pay much attention to the collecting ▶

activities of biologists. Then, beginning in the 1950s, 'bioprospecting' led to the development of highly profitable commercial products. The pharmaceutical company Eli Lilly, for instance, created the anti-cancer drugs vincristine and vinblastine from *Catharanthus roseus,* the Madagascar periwinkle plant — and it is still not clear which country (or countries), if any, deserve a share of its profits. Environment ministries began to see wild species as valuable natural resources alongside timber and fisheries. Later, the CBD, which came into force at the end of 1993, affirmed that living species were no longer the common heritage of mankind by declaring that "States have sovereign rights over their own biological resources".

One of the objectives of the convention, in addition to conserving species and encouraging sustainable development, is to ensure "the fair and equitable sharing of the benefits arising out of the utilization of genetic resources". To this end, a working group of CBD country representatives (mainly from environment ministries) has been operating since 2004 with the goal of creating an International Regime on Access and Benefit Sharing (IR-ABS) by the time of COP-10. With the deadline looming, those involved have ramped up their efforts to create a workable agreement. Only in the past six months, however, have the negotiators agreed that the regime should be a legally binding treaty, not a voluntary measure. Also, many crucial issues remain unresolved. It's unclear, for instance, whether — to facilitate urgent research on pandemics — disease-related samples should be exempt, or whether the agreement should cover only material expatriated since the convention came into force.

CAUGHT IN THE CROSSFIRE

So far, negotiations have been adversarial between at least five stakeholder groups — of which only the CBD nations can raise formal objections to the text. First, leaders of biodiversity-rich countries such as Brazil, Indonesia and South Africa want payback for the decades to centuries during which they did not share in the proceeds of species exported for cultivation and commercial exploitation. Second, the heads of industrialized countries such as Canada and members of the European Union want to ensure that their nations' biotechnology sectors can continue to do business without falling into legal quagmires. Third, several non-governmental organizations, such as the ETC Group in Ottawa, Canada, have pilloried virtually all international movement of biological specimens as 'biopiracy'[3]. Fourth, indigenous communities often consider local species to be their own domain, rather than that of their government, and they seek recognition and reward for their traditional knowledge and conservation efforts.

WHERE COUNTRIES STAND

The varying extent to which countries have brought in legislation and regulation to control foreign access to genetic resources in accordance with the Convention on Biological Diversity.

3 Non signatories (United States, Andorra, Holy See)

58 Parties with legislation or regulation

Total parties 193

117 Parties with no legislation or regulation

15 Parties with both; includes Australia, Brazil

SOURCE: CBD
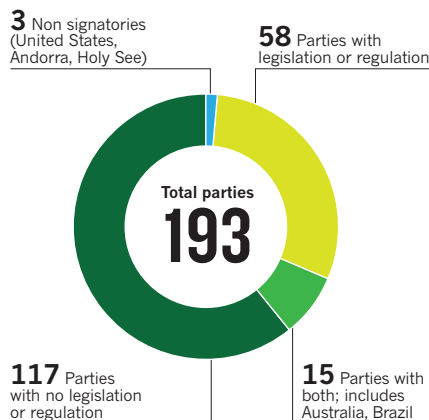
Caught in the cross-fire is a fifth group: taxonomists, ecologists and other non-commercial researchers.

In many countries, the decision on whether and when foreign researchers can collect and remove samples is now partly in the hands of provinces and local communities. A taxonomist from the Museum of Natural History in Paris, for example, recently spent two years negotiating with local officials in the Philippines before obtaining a permit to collect species of marine invertebrate. In other countries, such as Indonesia, India or Colombia, waiting periods of a year are common, and many credible projects never gain access. Indeed, separate jurisdictions (for example, from agriculture and environment ministries, or from state as well as local authorities) make it hard even for local researchers to do field-based research[4].

A global agreement — done correctly — could remove much of the confusion, save researchers' time and effort, encourage basic research, reduce governmental costs and ensure a fairer distribution of any benefits.

Admittedly, it can be hard to separate commercial from non-commercial research. Neither can be defined according to who does the work or where a study is conducted. Most big universities have technology-transfer offices devoted to commercializing basic research. A University of Florida researcher invented the sports drink Gatorade, for example. Also, big companies sometimes conduct 'blue sky' research — seven Nobel Prizes have been awarded for work conducted at Bell Labs in Murray Hill, New Jersey.

However, access agreements are generally made one project at a time. So the country providing the biological material could require that each researcher: makes his or her findings public; foregoes any intellectual property rights and activities that might lead to the commercialization of the work; applies for a new agreement should he or she come across something of potential commercial value. Alternatively, agreements could include profit-sharing terms that would come into play if researchers developed commercial intent while carrying out the work. If biologists want to move specimens or samples to a foreign museum, the agreement could specify what downstream uses of the specimens are prohibited — and that any loan transactions be disclosed so that the provider country can track the movement of the samples.

Australia, one of the CBD countries with access regulation and legislation in place, already has streamlined, standardized agreements for foreign researchers wanting to pursue non-commercial research[5] — where the application process typically takes a few days. In the past 5 years, about 450 permits have been granted, only 10 of which have involved potential commercial applications. Several taxonomists and ecologists studying organisms in marine or hot-spring environments in three of the non-commercial projects recently discovered compounds that may have biomedical or industrial applications. In each case, the researchers have been able to follow a clear and predictable legal process. New commercial access and sharing

C. CASE/ALAMY

Ecuadorian farmers can better manage land thanks to non-commercial research on local ecology.

agreements have been negotiated with the environment ministry, allowing them to continue with their research[6].

Some industrialized countries that want to defend their business sectors may resist demands for greater transparency. In doing so, they will undermine their non-commercial research sectors, which rely on open, trusting relationships with provider countries. On the other side of the fence, some biodiversity-rich developing countries may be tempted to demand a defensive, protectionist regime that assumes all research projects are equally likely to yield big commercial benefits. Yet excessive red tape could force non-commercial researchers to conduct their work elsewhere.

If this happens, developing countries will lose on two fronts. The public knowledge created by pure research is essential for countries to support any claims that a commercial product came from samples taken from their territories. Biodiversity inventories and databases, museum collections and DNA bar-coding (www.barcodeoflife.org) — identifying species using short DNA sequences — provide openly accessible documentation of which species occur in which countries. Additionally, international collaborations offer the opportunity to develop universities, museums, scientific expertise and economic growth. Non-commercial research on the ecology of native grasses, trees and shrubs, and invasive plants in Ecuador, for example, is helping local farmers to better manage their land.

With just a few days before COP-10, consensus on the IR-ABS still seems a long way off. All sides should back the text on non-commercial research that was recently added to the draft protocol, regardless of where they stand on other issues relating to the treaty. If not, the CBD's two other long-term goals will be jeopardized: conserving biological diversity and promoting its sustainable use. ∎

David E. Schindel *is executive secretary of the Consortium for the Barcode of Life, a global initiative supported by the Alfred P. Sloan Foundation and hosted by the National Museum of Natural History, Smithsonian Institution in Washington DC.*
e-mail: schindeld@si.edu

1. Secretariat of the Convention on Biological Diversity *Access and Benefit-Sharing in Practice: Trends in Partnerships Across Sectors, Technical Series No. 38* (2008).
2. Secretariat of the Convention on Biological Diversity *Appendix II of the Bonn Guidelines on Access to Genetic Resources and Fair and Equitable Sharing of the Benefits Arising out of their Utilization* (2002); available at http://go.nature.com/a8jqao
3. http://go.nature.com/3atqx5
4. Grajal, A. *Conserv. Biol.* **13**, 6–10 (1999).
5. Australian Natural Resource Management Ministerial Council *Nationally Consistent Approach for Access to and the Utilisation of Australia's Native Genetic and Biochemical Resources* (2002); available at http://go.nature.com/he7ikv
6. Personal communication from Australia's designated competent national authority for genetic resources.

President Barack Obama touring an oncology lab in Bethesda, Maryland, in September 2009.

A two–year plan for US science

Daniel S. **Greenberg** sets out five things that the White House and Capitol Hill can and should accomplish between now and the 2012 election.

The widely expected rightward shift in the coming US midterm election reflects hostility to the current government's spending and programmes. Will that be bad for science, with its lifelines to the US treasury? Maybe not. With some exceptions, the scientific enterprise is out of the line of political fire. Both parties trumpet the importance of science. And there's a lot they can do for the enterprise, even with low-growth or stagnant budgets. In most instances, the opening move belongs to the president, who has demonstrated solid support for science.

What should be on Barack Obama's to-do list for science in the next two years? First, assuming little or no increase in total federal research and development (R&D) spending, the president should seek to reverse the growing emphasis on defence-related R&D relative to support for civilian programmes. In the 1970s, at the height of the cold war, government spending on civilian and military research was about the same. The civilian R&D budget was even slightly ahead in 1979, at US$51.1 billion, compared to $48.9 billion for the Pentagon. Today, defence R&D accounts for about 59% of the federal R&D budget — $85.3 billion for the Pentagon, compared with $60.2 billion for civilian R&D. There ▶

ELECTIONS

is little sense in that, given the winding down of the war in Iraq, plans to get out of Afghanistan and the growing scientific and technological competence of the nation's industrial competitors. Even as China builds up its military strength, the United States remains overwhelmingly ahead in military R&D.

Better to shift some of those defence R&D dollars to academic and commercial research. That may seem hard to do, given the traditionally strong Republican backing for military funding. But prominent Republicans, with former house speaker Newt Gingrich in their lead, are among the most vociferous exponents of science as the key to economic strength. Furthermore, because the military increasingly dips into the civilian economy for dual-use technologies, the shift may actually be beneficial for the armed services.

AGE BIAS

With defence R&D money diverted to civilian accounts, real remedies should be directed at the neglect of youth in the allocation of research grants. The worst offender is the National Institutes of Health (NIH), where the average age for receiving a first grant has risen to 42 from about 36 in 1981 (see chart). In 2006, the NIH moved to reverse this age trend with grants for young researchers and further moves are afoot this year (see *Nature* **467**, 635; 2010). But the geriatric tilt persists. If budget growth is unavailable for a properly funded youth programme, then a reallocation from the old to the young must be made, painful as that would be. Given Congress's tradition of strong support for the NIH, even while other science budgets languish, a strong presidential

⟳ **NATURE.COM**
See midterm election special at
go.nature.com/dparwu

pitch for additional funds to support young health researchers is likely to be favourably received. The extraordinary doubling

THE GERIATRIC TILT OF US HEALTH RESEARCH

The age at which National Institutes of Health (NIH) grant receipients win their first grant (R01 or equivalent) has been increasing for decades, despite moves to reverse the trend.

All NIH grant recipients under 35 years old
13% — 1980s
2% — 2000s

Should military spending account for 58% of the US research and development budget?
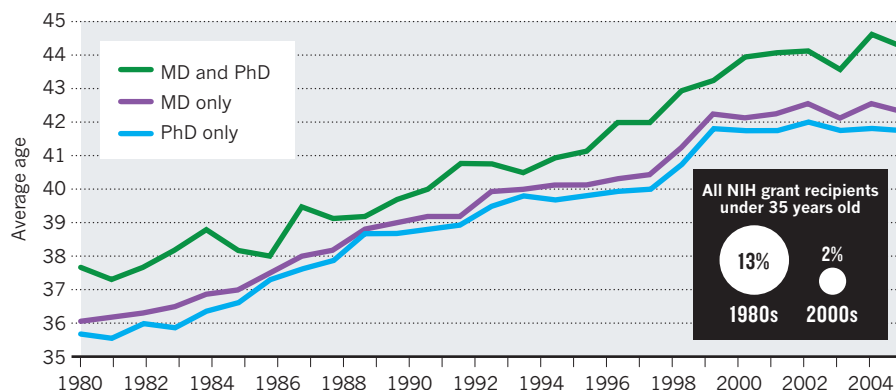
of the NIH budget between 1998 and 2003, for example, was initiated by congressional Republicans.

Tax credits should be used to prod industry into providing greater support for the university-based science programmes that underpin much of America's high-tech success. Industrial money accounts for a mere 5–7% of all the money spent on science-related activities in academia. When industrial firms do pitch in, it is usually to exploit the specialized skills of university science and engineering departments — such as when a group of energy companies including Exxon-Mobil provided funding for a Global Climate and Energy Project at Stanford University in California in 2002. Sadly, there's no reason to expect that industry, with minor exceptions, will recognize its own self interest and provide string-free money for the education system on which it depends for trained personnel and innovative research.

At virtually no cost, the White House can bolster public confidence in science by cleaning up the conflict-of-interest mess that repeatedly makes headlines, particularly concerning financial links between academia and the pharmaceutical industry. New regulations drafted by the NIH are a step in the right direction, requiring disclosure of all outside income above $5,000, down from $10,000 under the old rules. But the new regulations fall short by relying on university administrators to monitor their employees' outside financial dealings — a task they don't like and that many, in effect, ignore. Meanwhile, the NIH itself has followed a policy of 'trust but don't verify'. Spot checks by NIH staff should be part of any reform effort. And penalties should be strengthened — currently they are ridiculously mild, commonly involving two or three years of ineligibility to apply for the next grant.

The financial future for science is complicated by the injection last year of more than $18 billion in stimulus money for federal research agencies. Directed at boosting

employment, the money came with orders to spend it all within two years. This predictably set off an avalanche of research proposals. But the money will soon be gone, and will not be replaced under the normal budget process. Given the disruptions caused by abrupt financial ups and downs in the conduct of research, this is an area that merits some relief from the budget restraints that lie ahead.

Should all of these steps be taken, it won't boost or protect all areas of science — particularly not the ceaseless battleground of reproductive biology. If Congress moves to the right with more Republican seats, the stem-cell issue will flare even higher than it has in recent months, with renewed attempts to prohibit all research based on cells derived from human embryos. Don't expect comity on that topic.

Obama has repeatedly demonstrated respect and support for science in public statements and policy-making. His appointees for heading federal research agencies have all received strong approval in the scientific community. His embrace of science is widely emulated in Congress. Even in the stormy politics anticipated for the next two years, the steps cited above are achievable, and they are important because they can ensure the continued excellence of American science. ■ **SEE EDITORIAL P.751**

Daniel S. Greenberg *is a journalist based in Washington DC, and author of several books on science policy and politics. His debut novel, recently published, is* Tech Transfer: Science, Money, Love, and the Ivory Tower.
e-mail: danielg523@aol.com

Rare but catastrophic climate events, such as drought, can have unfathomable costs.

CLIMATE CHANGE

Insurance for a warming planet

Climate policy should be viewed as protection against uncertain future risks, says **Martin L. Weitzman**.

Bjørn Lomborg has been a lightning rod for controversy since he published *The Skeptical Environmentalist* in 2001. Yet in the time between his first book and this third edited volume, there has been a sea change in his attitude to climate change. Lomborg, director of the Copenhagen Consensus Center think tank, now characterizes the fundamental question as "not *if* we should do something about global warming, but rather *how best* to go about it".

Smart Solutions to Climate Change presents economic analyses of eight proposed solutions to climate-related problems: climate engineering; mitigation of carbon dioxide, methane and black carbon;

carbon sequestration by forests; market and policy-driven adaptation to climate change; technology-led climate policy; and technology transfer. Each proposal is set out, critiqued from two alternative perspectives and summarized by an expert panel of five economists. It is a constructive book that focuses seriously on finding effective ways to combat global warming, and the differences of opinions it expresses are stimulating and enlightening. But the book falls short in its treatment of risk.

To help prioritize the proposals, each analysis calculates a cost/benefit ratio. However, the estimates used are of uneven quality. Some solutions, such as technology-led policy, are too vague for a

meaningful value to be assigned. And cost alone is not the best way to choose between options — geoengineering, for example, is expensive in terms of risk but may be necessary if we are faced with a disaster scenario such as runaway temperatures. It makes more sense to think of the solutions as making up a portfolio of options, including others such as nuclear power, guided by risk analysis.

All of the cost/benefit estimates in *Smart Solutions* are based on deterministic models — uncertainty doesn't figure much in this book. The assessments rely on joint computer modelling of economic growth and climate change to examine the tradeoffs: whether or not we incur the costs of mitigation now to benefit from less-severe climate change in the future. Key parameters are approximated by firm values, such as the median or mean, rather than a probability distribution. The modelling thus becomes a knob-twiddling exercise in optimizing outcomes,

Smart Solutions to Climate Change: Comparing Costs and Benefits
BJØRN LOMBORG
Cambridge University Press: 2010. 432 pp. $29.99

where it is easy to flirt with high carbon dioxide concentrations.

Such modelling breeds complacency — temperature targets can be hit exactly, economic and ecological damages from high temperatures are low to begin with, and the pain of action now is greater than the pain of damages in a century or two when discounted at current interest rates. But in reality, there is no such thing as hitting a target of 2 °C, 4 °C or any other temperature change. Everything is probabilistic.

The economics of climate change is mainly about decision-making under extreme uncertainty. Climate-change analysis is hampered by many unknowns in the science combined with an inability to evaluate meaningfully the welfare losses from increased global temperatures. The values of key future parameters — global and regional average temperatures, damages to the world's economy and ecology, welfare, costs of unproven technologies and so forth — cannot be known now. Instead, they must be treated as random variables, yet to be drawn from some probability distribution that itself is uncertain.

A striking feature of the economics of climate change is that rare but catastrophic events may have unfathomable costs. Deep uncertainty about the unknown unknowns of what might go wrong is therefore coupled with essentially unlimited liability. The resulting battle between declining probabilities and increasing damages is difficult to resolve. Alas, this uncertainty can figure prominently in evaluations of climate-change policies. Its absence in a book dealing with economic comparisons of smart solutions is a serious omission.

When confronted with the possibility of extreme damages at low probabilities, most people do not look to averages. Instead they think about how much insurance they need, and can afford to buy, to survive those events. Climate policy is better viewed as buying insurance for the planet against extreme outcomes than as the solution to a multivariate problem over which we have control. To analyse policies in terms of deterministic cost/benefit ratios is to marginalize the very possibilities that make climate change so grave.

Lomborg concludes that, if we value our planet's future, we must "start seriously focusing, right now, on the most effective ways to fix global warming". Despite its limitations, *Smart Solutions* marks symbolically the end of one stage of thinking about climate change and the beginning of another. ∎

Martin L. Weitzman *is professor of economics in the Department of Economics, Harvard University, Cambridge, Massachusetts 02138, USA.*

Books in brief
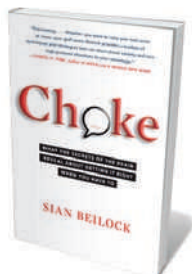
Choke: What the Secrets of the Brain Reveal About Getting It Right When You Have To
Sian Beilock FREE PRESS *304 pp. $26 (2010)*
When the pressure's on, we've all 'choked' — hit the wrong note, flunked an exam or messed up an interview. Cognitive psychologist Sian Beilock explains why. Describing how memory works, she shows that experts whose minds brim with facts are more likely to freeze than novices. Social stereotyping also leads us to underperform. Beilock's solutions for big occasions are simple: reaffirm your self-worth, write away your worries and keep practising. If the worst happens, pause and refocus.
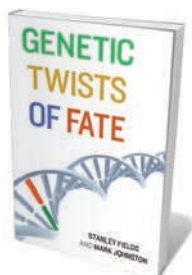
Genetic Twists of Fate
*Stanley Fields and Mark Johnston* MIT PRESS *240 pp. $24.95 (2010)*
Minuscule inherited changes in our DNA can have major effects on our lives. Geneticists Stanley Fields and Mark Johnston explain how genes affect our health, from conditions such as Alzheimer's disease, cancer, diabetes and depression to rare genetic disorders. Giving the science a personal twist, they relate how a mother was wrongly accused of killing her son when the cause of death was in fact a rare inherited condition, and how former US president Dwight Eisenhower's hereditary heart disease was treated with an anticoagulant derived from rat poison.
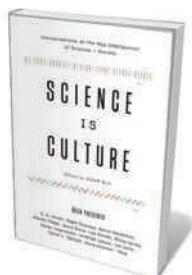
Science is Culture: Conversations at the New Intersection of Science and Society
*Edited by Adam Bly* HARPER PERENNIAL *368 pp. $15.99 (2010)*
Science is often divorced from mainstream culture. This collection of conversations between 44 top scientists and thinkers from the humanities, first published in *Seed* magazine, aims to blur the boundaries. Entomologist E. O. Wilson discusses evolution with philosopher Daniel Dennett; linguist Noam Chomsky and sociobiologist Robert Trivers debate war and deceit; and astrobiologist Jill Tarter muses on alien life and reality with Will Wright, designer of the computer game *Spore*.
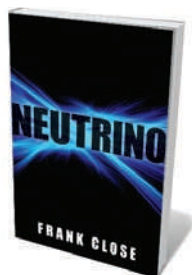
Neutrino
Frank Close OXFORD UNIVERSITY PRESS *176 pp. £9.99 (2010)*
By the time you have read this paragraph, some 50 trillion neutrinos will have passed through your body. Formed in stars and through radioactivity, these enigmatic particles rarely interact, travelling through matter almost unseen. Particle physicist Frank Close explains how it took 26 years for the neutrino — 'little neutron' in Italian, as named by Enrico Fermi — to be detected in the lab after its prediction by Wolfgang Pauli in 1930. Close describes ongoing attempts to capture neutrinos, to determine their mass and to understand their significance in the Universe.
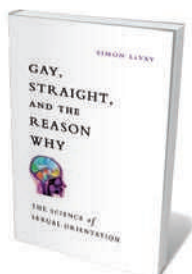
Gay, Straight, and the Reason Why: The Science of Sexual Orientation
Simon LeVay OXFORD UNIVERSITY PRESS *432 pp. $27.95 (2010)*
The theory that sexual orientation has a biological basis receives support in neuroscientist Simon LeVay's book. Relating evidence from genetics, neuroscience and developmental biology, he suggests that prenatal interactions between hormones and the developing brain influence adult sexuality. LeVay, who published a 1991 *Science* paper on brain differences in gay and straight men, believes we should accept that homosexuality in humans is biologically hardwired, as it may be in other species such as geese.

DRUG DEVELOPMENT

The invisible front line

The paid participants in phase I clinical trials need more protection, finds **Meredith Wadman**.

An essential step in bringing a new drug to market is the first testing of the product in humans. Phase I clinical trials are safety studies, typically conducted in fewer than 100 healthy subjects. The new drug's concentration and duration in the blood are tracked at various doses, and any toxic effects are monitored. Volunteers are paid for their participation.

Since the US Food and Drug Administration barred prisoners — until then the main source of paid subjects — from participating in clinical trials in 1980, a new class of 'professional human guinea pig' has stepped forward. Roberto Abadie, a visiting ethnographer at the City University of New York, provides a disturbing account of volunteers' experiences in *The Professional Guinea Pig*.

Phase I trials can be a risky business. In 2001, Ellen Roche, a previously healthy 24-year-old, died during a study of an asthma drug at Johns Hopkins University in Baltimore, Maryland. In 2006, at London's Northwick Park Hospital, six male volunteers were left in a critical condition after they were injected with an experimental treatment for rheumatoid arthritis and leukaemia (see *Nature* **440**, 388–389; 2006).

With so much at stake for trial subjects, public protection and drug companies alike, the phase I system is worth examining. Abadie surmises that paid volunteers are here to stay, but that the commercial incentives that attract them put them in peril. More should be done to protect them, he says, including removing the veil of secrecy that currently makes it impossible for the public to know who these volunteers are, how many trials they have participated in, and what, if any, side effects they experience in the long term.

Abadie studied a group of 18 professional human guinea pigs who live from week to week on the money they are paid for intermittent drug trials. His interviewees were mostly white males living in communal houses, who supplemented their unpredictable income through temporary jobs. They came to his attention through *Guinea Pig Zero*, a small-

The Professional Guinea Pig: Big Pharma and the Risky World of Human Subjects
ROBERTO ABADIE
Duke University Press: 2010. 216 pp. $22.95

circulation publication charting the experiences of professional volunteers edited by Robert Helms, one of Abadie's interviewees. Helms was involved in the first known 'strike' during a phase I clinical trial. In late December 2002, subjects at Jefferson Hospital in Philadelphia, Pennsylvania, succeeded in wringing a substantial pay increase from a major drug company by threatening to walk out halfway through tests of an anti-anxiety drug.

Abadie concedes that his group is not representative. The bulk of people who regularly enrol in phase I trials in Philadelphia — a hot spot for clinical studies because of the number of drug firms in the area — are African American or Hispanic, poor and relatively uneducated. They are also difficult to track down. Abadie points out that no demographic data are available on the whole phase I trial population. Companies disclose neither the number of trials they are conducting nor the number of subjects they enrol.

Money is the sole incentive for Abadie's professional guinea pigs. At the time of his research in 2003 and 2004, payments ranged from US$200–400 a day for trials lasting three or four days, to $5,000 for studies lasting up to four weeks. To fight attrition, the money is doled out in strategic increments, with the largest sum paid at the end. "You became addicted to the trials, to the easy money," one subject observes. But it is a coercive relationship, another admits. They sense that they are bodies for hire, with echoes of prostitution: Shon (a pseudonym) says that he experiences being stripped of his clothing and attached to a heart monitor as dehumanizing. "When they ask you to pee, you pee," says another.

The financial motive is in stark contrast to the altruistic or personal drivers of volunteers for phase II and III trials, who usually have the relevant disease for the drug being

tested. Abadie interviewed for comparison a group involved in late tests of HIV drugs. These people took part because they wished to improve treatments generally, to benefit from them therapeutically if possible, and to gain a sense of empowerment in the fight against their disease.

The phase I subjects accept the short-term risks involved in trial participation — they perceive them as similar to, for instance, working in a restaurant kitchen. If they experience uncomfortable symptoms, they are reluctant to drop out because of the loss of pay. Between trials, they flush their bodies with diets of apples, cranberry juice and marigold flowers.

Most avoid thinking about the long-term risks. Abadie worries this is a consequence of the increasing sums being paid for phase I trials. Current ethical frameworks, he argues, inadequately address the effects of these incentives. Informed consent forms cover the trial at hand, but not the risks of serial trial participation. Abadie notes that the subjects lack the workforce stability and permanent community that support others in hazardous occupations, such as coal miners.

Conceding that payments for phase I trials will endure — a switch to an unpaid volunteer base would reduce numbers and almost halt drug development — Abadie recommends that greater occupational protections, such as the right to form or join labour organizations, be extended to subjects. However, he argues, the industry has tried to forestall such protections by calling subjects 'paid volunteers', placing them outside the umbrella of labour law in the United States.

"Subjects sense that they are bodies for hire, with echoes of prostitution."

Abadie also proposes the setting up of a centralized registry to identify all phase I participants, list the trials that each undertakes and document their short- and long-term side effects. To deter companies from performing unnecessary phase I trials, he suggests imposing a tax on those firms that test medicines that merely add to a class of therapies already on the market, favouring profit over scientific advancement.

Although the voice of the drug companies is missing from Abadie's ethnographic study, any improvement in the situation of phase I volunteers will need the industry's engagement. It is hard to imagine that the pharmaceutical industry will welcome his suggestions with open arms. But *The Professional Guinea Pig* raises important questions. It should not take another Ellen Roche or another Northwick Park for us to make these volunteers safer, and to dignify their important role. ∎

Meredith Wadman *is a biomedical reporter for* Nature *in Washington DC.*

D. HATFULL/THE OXFORD SCIENCE PARK

Conrad Shawcross's *Fraction (9:8)*: inspired by nineteenth-century machines that visualized sound.

CULTURE

Metallic music

Sculptor Conrad Shawcross captures the fundamentals of physics and music, finds **Martin Kemp**.

The physics of music was apparently born in a Greek blacksmith's shop in the sixth century BC. The philosopher Pythagoras, who believed that everything was governed by numbers, heard the hammers beating out a piece of iron on the anvil. According to his later biographer Iamblichus, he recognized in the sounds "the concord of the octave, the fifth, and the fourth". Pythagoras' intervals were encapsulated in equations in the centuries that followed — from the 'music of the spheres' that explained the heavenly orbits, to the proportionality of the human body, as believed by Leonardo da Vinci.

During the eighteenth and nineteenth centuries, ingenious devices translated music into visual form. Among them were the famous plates of physicist and musician Ernst Chladni from 1787, which recorded in sand the nodal patterns of vibrations produced by drawing a bow across the edge of pieces of metal. The harmonograph, perfected by Scottish mathematician Hugh Blackburn, was more theoretically driven. It consists of two adjustable pendulums swinging at right angles to one another, one holding a pen, the other a drawing surface. By varying the relative frequency and phases of the pendulums — which can be adjusted to musical intervals — a range of ellipses, spirals and complex loops can be created. These include the figure-of-eight curves named after Jules Antoine Lissajous, the French

physicist who used tuning forks and beams of light to see sound waves.

The harmonograph has inspired a notable new sculpture by Conrad Shawcross, a young British artist who is becoming famed for his invention of wondrous scientific and philosophical machines that are usually on a large scale (see go.nature.com/o3x8ts). Shawcross's devices often exploit natural ratios and movement in many planes. His recent works include *The Nervous System* (2003), a loom spinning hundreds of metres of helical rope per day; and *Continuum* (2004), a wooden spring wound into a toroidal form to signify the cycles of time. Shawcross's sculptures reflect his well-informed thoughts about the theories of science, uncertainty and the limits of knowledge. He champions theoretical physicist Werner Heisenberg's dictum that "what we observe is not nature itself, but nature exposed to our method of questioning".

Shawcross's hanging sculpture, *Fraction (9:8)*, installed last month in the foyer of the Sadler Building at the heart of the Oxford Science Park, UK, is both typical and atypical of his work. True to his canon, Shawcross devised a machine to sketch the sculpture's form, adapted from the harmonograph. He set the pendulums to achieve a ratio of 9:8 in frequency — the musical interval of a major second, corresponding to the note D in the Pythagorean scale. Rather than using a fixed drawing surface, he pulled a roll of paper across the swinging plate to create a long oscillation that tapers away into silence. The trace, he confesses, involved "a lot of human intervention".

The figure was translated by computer into three dimensions, and fabricated by automotive-industry robots into the aluminium vortex, 14 metres long, which hangs from wires attached at critical balance points. Its machine finish is a departure for Shawcross, whose sculptures usually retain signs of the 'visual noise' that arose during their making. However, although it represents a Platonic, archetypal note, the sculpture is not invariable. As the spectator walks by, different visual rhythms appear together with varied degrees of order and disorder. Only from the bridge crossing at first-floor level do its symmetries emerge clearly.

Music in metal is an appropriately Pythagorean idea. It recalls the name given to the final movement in Handel's harpsichord *Suite No. 5 in E major*, the 'Harmonious Blacksmith'. Like Pythagoras' story, its alternative name came later and belongs to legend. But legends have their own kind of truth, as does capturing the essence of what science and art find it productive to investigate. ■

Martin Kemp *is emeritus professor of art history at the University of Oxford, UK.*

CORRESPONDENCE

Taxonomy: add a human touch too

Norman MacLeod and colleagues' call to develop automated species-identification systems is laudable (*Nature* **467**, 154–155; 2010), but let's not forget a core feature of taxonomic work that depends on a scholar's input — the discovery of new characters.

Unexpected evolutionary novelties in morphology and physiology, for example, are what make taxonomic exploration rewarding. Taxonomists set out to discover and track such novelties and their evolutionary history. It is in this sense that taxonomy provides the empirical basis for understanding speciation and phylogeny.

There is a place for automated pattern detection, but it would not work with the 5,000 species of *Drosophila*, say, which are identified by their many different structures. Taxonomy can independently test and verify identifications without relying on patterns of single characters, as the long list of synonymies in any biological group testifies. The practice of taxonomic revision and publishing detailed monographs ensures that character distributions, species status and phylogenetic relationships are subject to repeated and critical testing.

We should beware the trend to confuse automatic identification tools with those that are useful for discovering new species. The emerging field of cybertaxonomy is an advance only if it is understood as enhancing and enabling theory-rich descriptive taxonomy, not replacing it.

As in many other modern scientific fields, including diagnostic medicine and molecular genetics, a final step involving a human expert is essential.
Antonio G. Valdecasas *Museo*

Nacional de Ciencias Naturales, CSIC, Spain,valdeca@gmail.com
Quentin D. Wheeler *Arizona State University, USA*

Taxonomy: include social networking

Help with the shortage of professional taxonomists needed to identify organisms (*Nature* **467**, 154–155; 2010) may also come from an unexpected source — social networking on the Internet.

Through social networking, the identification process can be made more efficient while simultaneously spreading real taxonomic knowledge. The facility is available to anyone, unlike other technologies that require specialized equipment.

In its first year of operation, the website iSpot (http://ispot.org.uk) has helped 6,000 users to identify 25,000 sightings of some 2,500 species, from lichens to birds. The website works by linking experts (including amateur experts) with beginners through a sophisticated reputation system that encourages users to help and learn from each other.

Eventually, DNA bar-code matching and image recognition might be added to the tools available. But these will be aids, not replacements, for people learning how to identify species.
Jonathan Silvertown *The Open*
*University, UK,
j.silvertown@open.ac.uk*

Call to reshape university curricula

Universities worldwide are undergoing structural transformation, but reform of curricula is also key to their renewal. We invite scholars, educators and administrators to collaborate with us to reshape university teaching and learning.

Universities are experiencing a crisis of purpose, focus and content, rooted in a fundamental confusion about all three. The crisis is all the more visible as their pace of social, intellectual and technological change falls increasingly out of step with that outside. Furthermore, universities are largely reactive where they should be visionary and critical.

Curricula are mostly separated from research, and subjects are taught in disciplinary isolation. Knowledge is conflated with information and is too often presented as static rather than dynamic.

There can be no standard formula to rectify these problems, given the diversity of institutional structures and cultural differences among universities. However, our working group of scholars has made a start. To inform international dialogue and guide the experimental process of redesigning curricula, we have just published a set of 11 principles for rethinking undergraduate courses worldwide (see www.curriculumreform.org), and we welcome any suggestions.
Yehuda Elkana
Wissenschaftskolleg zu Berlin, Germany
Manfred D. Laubichler *Arizona State University, USA,
manfred.laubichler@asu.edu*
Adam S. Wilkins *Centre de Recherches Interdisciplinaires, France*

Science as a threat to far-right fringe

The scorn of the US far-right 'Tea Party' fringe for science, particularly relating to sustainability, climate change and biodiversity, stems from a perceived threat to its idealized views of how the world should be (*Nature* **467**, 133; 2010).

The great (and undervalued) psychologist George Kelly described threat as "the awareness of an imminent comprehensive change in one's core structures". Science opens up our culture's core structures to criticism and calls their importance into question — one example being non-sustainable approaches to managing the planet.

The far-right's response to the threat posed by modern environmental science goes roughly like this. First, discredit a piece of scientific data — the 'hockey-stick' graph, for example. When that fails to remove the threat of climate-change science, discredit the scientists (as happened in 'Climategate'). When that doesn't work either, discredit science as a method for understanding the world.
Richard Kool *Royal Roads University, British Columbia, Canada,
rick.kool@royalroads.ca*

Alternative view of Serengeti road

Andrew Dobson and colleagues overstate the argument against a road across the Serengeti National Park in Tanzania (*Nature* **467**, 272–273; 2010). Migrant wildebeest already regularly cross transportation arteries in the Serengeti and in adjacent linked ecosystems. Established evidence of impacts on declines in migratory wildlife populations is centred not on roads, but on fencing — a separate and more remote possibility.

Conservation policy in the Serengeti has too often been based on idealized notions of what landscapes should look like. The Serengeti ecosystem has been shaped by thousands of years of human use, and is now groomed by park managers rather than by local inhabitants.

The ecosystem is supported and enhanced by pastoralist land use in surrounding buffer zones.

Dobson *et al.* do not consider the likely importance of the road for the land use, mobility, livelihoods and welfare of people living near the park.

Wildlife tourism is a major contributor to Tanzania's economy. But local people, whose land-use practices have created and now maintain the ecosystems valued by Western conservationists, are being dispossessed by state and international elites who capture much of the tourism revenue and reinvest it in conservation-incompatible land use.

Wherever the road goes, it will have an environmental impact as well as socio-economic costs and benefits that require rigorous analysis of good data.
Katherine Homewood
University College London, UK,
k.homewood@ucl.ac.uk
Daniel Brockington *University of Manchester, UK*
Sian Sullivan *Birkbeck, University of London, UK*

International, not 'campus', please

Your News story on China's science journals (*Nature* **467,** 261; 2010) inaccurately described the *Journal of Zhejiang University-Science* (*JZUS*) as a 'campus' journal. In fact, it is an international publication with a pool of some 7,600 referees from more than 67 countries (go.nature.com/pknY2R). On average, 64.4% of its contributions come from outside Zhejiang University, Hangzhou, of which 50% are from more than 46 countries and regions. The editorial board is international too.

There are roughly 5,000 science, technology and medical periodicals in China, including more than 200 published in English. Around 1,000 are university ('campus') journals that mostly publish papers from a specific institution, many of which were launched during the country's early reform and opening-up stage from 1978 into the 1980s. This also happened in Japan — the only difference being that most university journals in China are multidisciplinary in

content, whereas those in Japan usually focus on one subject or subject group.
Yuehong (Helen) Zhang *Journal of Zhejiang University-Science, China, jzus@zju.edu.cn*

Reward quality not quantity

China is not the only country facing a glut of low-impact papers in scientific journals (*Nature* **467,** 252; 2010). Shutting down journals will not help (*Nature* **467,** 261; 2010), as the papers will inevitably resurface elsewhere. As many have pointed out, it is the incentives responsible for the glut that need to change.

What is required is a new academic culture that rewards quality over quantity, an idea whose time has finally come. For example, Germany's main research foundation, the DFG, recently ruled that a maximum of only five papers should be listed in any grant application; compare this with the 10 papers allowed in applications to the US National Science Foundation.

More universities and colleges could limit the number of publications permitted when applying for tenure or promotion. Researchers would then be able to focus their efforts on publishing articles only when there is something substantial to report.
Stan W. Trimble *on behalf of 7 co-authors*, University of California, Los Angeles, USA, trimble@geog.ucla.edu*
**A full list of signatories is available online at http://dx.doi.org/10.1038/467789b*

Basic science in Russia under threat

The president and prime minister of Russia proclaim their support for science and many ambitious projects (*Nature* **467,** 251; 2010). However, the actual support for basic research in Russia will decrease drastically this year.

The main source of peer-reviewed funding is the

Russian Foundation for Basic Research (RFBR), a non-profit organization that is largely backed by the state, as well as by voluntary contributions from enterprises, organizations and individuals. Its budget will be cut in 2011 to 6 billion roubles (US$200 million), half of what it was in 2009.

The budgets managed directly by the Russian ministry of research decreased 4.5% in 2010 compared with 2009, but are expected to rise in 2011. Although this does not sound that bad, the funds distributed by the ministry seem to be based on non-transparent schemes and arbitrary decisions, and are often awarded without proper peer review.

It is essential to expand state support for the RFBR and save it from closure. The government should recognize that investing in internationally recognized peer-reviewed research is better than funding ambitious and costly projects that have an uncertain outcome.
Yegor Vassetzky *Institut Gustave Roussy, France, vassetzky@igr.fr*

Better governance to save rainforests

Urgent measures are needed to help Papua New Guinea to develop sustainably. The country is beset by corporate misdealing and weak governance, and consequently its biologically and culturally rich forests are imperiled.

Attempts by the World Bank and other organizations to improve forest governance have been largely frustrated. As a result, overcutting of forests continues apace, with most accessible forests likely to be logged or to disappear in just one to two decades.

Foreign corporations are driving much of this forest exploitation. Trees are mostly cut by Malaysian logging companies and then shipped as raw logs to Chinese manufacturers of wood products for export.

Papua New Guinea's raw-log exports yield only meagre income and employment for local communities, and proceeds are often concentrated in the

hands of political elites.

Environmental prospects in Papua New Guinea might actually be worsening. In May, the country's parliament stripped indigenous groups of important land-protection rights, making it difficult for them to sue offending corporations for environmental damage. The few conservation areas in the country are mostly small and poorly protected, and the government has frozen proposals for 120 new conservation areas to avoid conflicts with loggers. A government pledge to phase out raw-log exports by 2010 has also gone unfulfilled.

Papua New Guinea should immediately reduce its raw-log exports, which will promote local wood-processing industries, training and employment. It should also reinstate the rights of traditional communities to legally challenge foreign corporations for environmental damage. And it should use international climate-change mitigation funds — which could eventually exceed US$500 million a year — to promote new conservation areas and improve forest management and governance.
William Laurance *on behalf of 6 co-authors*, James Cook University, Australia, bill.laurance@jcu.edu.au*
**A full list of signatories is available online at http://dx.doi.org/10.1038/467789d*

Left-handed DNA can be right

I must protest against the knowing groans expended every time a scientist is personally affronted by the image of a left-handed DNA helix (*Nature* **467,** 401; 2010). I would remind the groaners that DNA handedness depends on its molecular sequence — witness a seminal paper you published more than 30 years ago: 'Molecular structure of a left-handed double-helical DNA fragment at atomic resolution' (A. H.-J. Wang *et al. Nature* **282,** 680–686; 1979).
Albert Erives *Dartmouth College, USA, aerives@dartmouth.edu*

George C. Williams
(1926–2010)

Incisive thinker who influenced a generation of evolutionary biologists.

In 1978, at the age of 52, the great evolutionary theorist George C. Williams began to chronicle his own senescence, recording once a year how long it took to run 1,700 metres round a track in Stony Brook, New York. Williams presented the graph of his 12 years of slowing speed at his acceptance speech for the Crafoord Prize in Bioscience that he shared with Ernst Mayr and John Maynard Smith in 1999. He later published it in *The Quarterly Review of Biology*, with which he was involved for 32 years. The plot encapsulated his lifelong fascination: why do we decline with age?

Williams died on 8 September, aged 84. Little known to the public, this tall, reserved man with an Abraham Lincoln beard will be remembered by evolutionary biologists as one of the most incisive thinkers of the twentieth century. His major contribution, the theory of gene-level natural selection, left a profound and enduring stamp on fields from sociobiology and evolutionary psychology to behavioural ecology. He spoke slowly and little, but when he spoke, you listened: his words were full of insight and flashes of dry wit.

GENE-LEVEL SELECTION

After a stint in the US Army, working on a water purification plant in Italy during the Second World War, Williams finished his BA in zoology at the University of California, Berkeley in 1949. He got his PhD from the University of California, Los Angeles in 1955 for work on the ecology of the blenny — a type of fish. There followed a postdoc at the University of Chicago and an assistant professorship at Michigan State University. In 1960, Williams moved to the State University of New York at Stony Brook. He later became one of the first professors in its newly formed Department of Ecology and Evolution. There he remained until his retirement in 1990 — the year I arrived as assistant professor and inherited his freezer of Icelandic eel samples. He had spent two sabbaticals in Iceland, was fluent in Icelandic and published on the European and American species of eels and their potential hybrids on Iceland.

In 1957, he published his seminal paper 'Pleiotropy, Natural Selection, and the Evolution of Senescence' in the journal *Evolution*. He argued that genes that enhance fitness early in life but have detrimental effects later in life — genes with 'antagonistic pleiotropic effects' — would
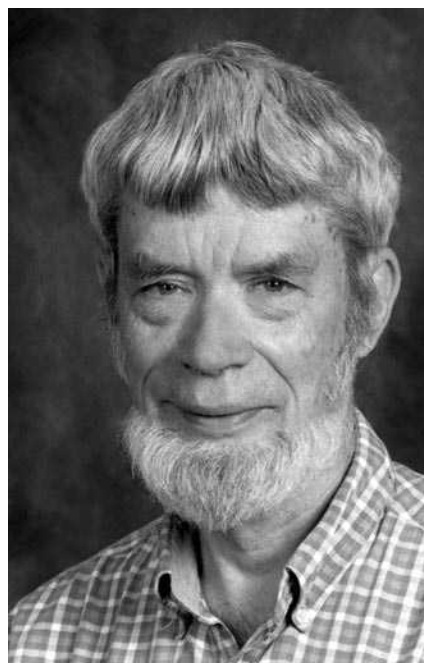
be expected to persist and even increase in abundance as long as, on balance, they boost an individual's fitness. He also pointed out that selection should be weaker in older age because fewer individuals are alive to be subject to it — an idea for which Williams shares credit with Peter Medawar.

The dominant narrative of early 1960s evolutionary biology was that natural selection acts at the level of the group or even for 'the good of the species'. Even death was explained in a group-selectionist light — as creating space for the next generation. Williams skewered this thinking, which he felt was "sloppy" and "anti-Darwinian", in his most influential book, *Adaptation and Natural Selection* (1966). In it he proposed that natural selection almost always acts more directly, swiftly and strongly at the level of the gene or the individual than at the level of the group or even species. He also railed against 'pan-adaptationism' — the idea that every feature is adaptive: he showed that adaptations have to have fitness-enhancing effects at the level of the individual rather than at the level of the species.

Adaptation and Natural Selection was way ahead of its time; its impact was felt for decades. Following Williams, E. O. Wilson extended gene-level and individual selection in *Sociobiology: The New Synthesis* — his controversial 1975 book on the role of

genetics in social behaviour, even of humans. And Richard Dawkins's 1976 book *The Selfish Gene* popularized some of Williams's ideas. That said, gene-level selection and inclusive fitness were not universally accepted then, and still meet with occasional criticism — notably from researchers trying to explain altruism and eusociality, for example. These ideas remain, nonetheless, cornerstones of modern biological theory.

COMPETITION NOT COOPERATION

Williams made further influential contributions. With his 1975 book *Sex and Evolution*, he was among the first to offer explanations for the puzzling prevalence of sexual reproduction. He pointed out that it is yet another example of competition, not cooperation, being the dominant force in evolution — with genes from each parent battling for influence within the same genome. (He saw a bright future for the fields of genetic imprinting and epigenetics.)

Williams went even further with his reductionist view of natural selection in *Natural Selection: Domains, Levels and Challenges*, his 1992 book about information and matter. He pointed out that what is of importance in evolution is the information that is contained in genes, genotypes and gene pools, not the physical objects — a position reminiscent of Dawkins's 'meme' concept.

Williams returned late in life to his abiding concern — ageing. In 1994 he wrote the book *Why We Get Sick: the New Science of Darwinian Medicine* with the physician Randolph Nesse. Williams and Nesse proposed that disease symptoms should be understood, and treatment informed, by the long evolutionary history that shaped immune responses. Their work has spawned a new field of study, evolutionary medicine.

It is a cruel irony that this brilliant man who first explained senescence died of Alzheimer's disease. ∎

> *"He spoke slowly and little, but when he spoke, you listened: his words were full of insight and flashes of dry wit."*

Axel Meyer *is professor of zoology and evolutionary biology in the Department of Biology at the University of Konstanz, D78457 Konstanz, Germany.*
e-mail: axel.meyer@uni-konstanz.de

MEDIA SERVICES, STONY BROOK UNIV.

NEWS & VIEWS

Accidental investigation

Astronomers would be expected to recognize comets easily when they see them, not least because of the objects' bright tails. But when planetary accidents try to fool them, their job is no longer that simple. SEE LETTERS P.814 & P.817

DAVID NESVORNÝ

Suppose you point a telescope up at the sky one dark night and, to your surprise, spot a diffuse object with a bright tail. You track down the object the next night, and the following nights, and see it slowly drifting among the stars. A cross-check against catalogues of astronomical objects tells you that no known body in the Solar System should have these sky coordinates and drift. You might think, then, that you have discovered a comet, right? Wrong. In this issue, Snodgrass et al.[1] (page 814) and Jewitt et al.[2] (page 817) report a startling discovery which shows that what looks like a comet can be something else altogether if you look closer (Fig. 1).

The object studied by the authors[1,2] — P/2010 A2 as it is designated — was found by the LINEAR robotic survey telescope on 6 January 2010. Soon after, astronomer Peter Birtwhistle, while observing from the Great Shefford Observatory in Berkshire, UK, reported it as having the appearance of a "headless comet with a straight tail". Something was wrong, however, because the orbit of P/2010 A2 was located in the innermost part of the asteroid belt, a swarm of rocky bodies between the orbits of Mars and Jupiter. Comets do not have such orbits.

To resolve the problem, Jewitt et al.[2] took several high-resolution images of P/2010 A2 with the Hubble Space Telescope. They found a speckle of light right at the place where the tail starts, slightly offset from it but otherwise following the tail's apparent motion on the sky. (This point-like body was first spotted in previous observations made by Jewitt et al.[3] and Licandro et al.[4] using ground-based telescopes.) This could not be a comet, they reasoned, because the point-like object lacked signs of a dusty envelope, known as a halo, which is produced by sublimating ices found around active cometary nuclei. Instead, Jewitt et al.[2] suggest, along with Snodgrass et al.[1], that P/2010 A2 is the debris of a collision between two small, rocky asteroids.

As Japanese astronomer Kiyotsugu Hirayama proposed in his groundbreaking work[5] in 1918, asteroids can hit each other, and when they do, they leave behind clumps of fragments known as asteroid families. Consider the orbital speed

Figure 1 | Comet? Not really. Snodgrass et al.[1] and Jewitt et al.[2] argue that P/2010 A2, an object that was originally identified as a comet, is actually the debris of a collision between two small, rocky asteroids. This image was taken with the Spacewatch 1.8-metre telescope on 8 January 2010 soon after the object's discovery, showing it under-resolved and with a comet-like appearance. The other objects are stars smeared out by the motion of the telescope while it was tracking the moving P/2010 A2. The horizontal band in the lower part of the image is an artefact of the detector.

of P/2010 A2, and that of other asteroids on similar orbits, which is about 20 kilometres per second. This value exceeds — in roughly the ratio of the speed of a Formula One racing car to that of a casual walk — the speed with which fragments can be ejected from an asteroid collision. The fragments will, at least initially, have similar orbital speeds, and will be identified on the generic asteroid background as a dense clump. But their orbits will later start to vary, owing to gravitational interactions with planets. Hirayama's ingenious mathematical method allows us to recognize the fragments, even if their orbits later become dispersed by gravity, by recovering the initial, dense-clump state.

More than 50 asteroid families have been identified through Hirayama's method in modern catalogues containing orbital information for some 350,000 asteroids[6]. Most of the known asteroid families correspond to disruptive collisions that occurred hundreds to thousands of million years ago. But in 2002, by propagating the fragments' orbits into the past, astronomers

were able to precisely date an asteroid family that was only 5.8 million years old[7] — a mere blink of an eye by astronomical standards. That was a sign of things to come.

To determine the time of the collision that produced P/2010 A2, Snodgrass et al.[1] observed the object on 16 March 2010 with a camera on board the Rosetta spacecraft, which was approaching the asteroid belt for its fly-by of asteroid 21 Lutetia. The spacecraft, located at the time some 10° above the orbital plane of P/2010 A2, offered a better vantage point than Earth for observations. The Rosetta images revealed that the tail of debris particles is straight, and tilted at about 30° relative to the orbital direction of P/2010 A2. What does this tell us about the time of collision?

Small particles, upon release from their parent object, follow different orbits around the Sun, because the force exerted on them by solar photons is stronger for smaller particles than for larger ones. After some time, particles of different sizes, but produced at the same time, will appear spread along a line known as

a synchrone. Because the solar-photon force does not change the orbital inclination, synchrones corresponding to different collision times will form a fan-like structure in the parent object's orbital plane.

In their study, Snodgrass et al.[1] show, using a computer model of particle orbits, that the P/2010 A2 tail can be fitted by a single synchrone corresponding to 10 February 2009, plus or minus a week. They thus find that the collision happened a mere 11 months before P/2010 A2 was discovered by the LINEAR telescope. This result is solid. So, unbelievable as it seems, astronomers are now able to monitor collisions between asteroids in real time.

The early detection of debris produced by asteroid collisions can be helpful in many ways. For example, the present length of the

P/2010 A2 tail — about 200,000 km — indicates that the particles in the tail are millimetres in size and larger. Where, then, are the really small, submillimetre particles that asteroid break-ups are known to shed into interplanetary space[8]?

There seem to be at least two options. Either the impact did not have enough power to pulverize the original rock into very small grains — which may hint at the structural properties of the asteroid interior — or the tail's relatively large particles predate the impact. The latter possibility would imply that small asteroids such as P/2010 A2 are, at least in large part, accumulations of gravel. Further investigation should be carried out to distinguish between the two options. After all, asteroids can bump into

all sorts of things, and before they do, we had better know what they are made of. ∎

David Nesvorný *is at the Southwest Research Institute, Department of Space Studies, Boulder, Colorado 80302, USA.*
e-mail: davidn@boulder.swri.edu

1. Snodgrass, C. *et al. Nature* **467,** 814–816 (2010).
2. Jewitt, D. *et al. Nature* **467,** 817–819 (2010).
3. Jewitt, D., Annis, J. & Soares-Santos, M. *IAU Circ.* **9109,** 3 (2010).
4. Licandro, J., Tozzi, G. P., Liimets, T., Haver, R. & Buzzi, L. *IAU Circ.* **9109,** 4 (2010).
5. Hirayama, K. *Astron. J.* **31,** 185–188 (1918).
6. www.minorplanetcenter.org/iau/MPCORB.html
7. Nesvorný, D., Bottke, W. F. Jr, Dones, L. & Levison, H. F. *Nature* **417,** 720–722 (2002).
8. Farley, K. A., Vokrouhlický, D., Bottke, W. F. & Nesvorný, D. *Nature* **439,** 295–297 (2006).

IMMUNOLOGY

Egocentric pre–T–cell receptors

The T–cell receptor on the surface of T cells requires antigen recognition to function. Structural studies reveal that its predecessor, the pre–T–cell receptor, is much more independent. SEE LETTER P.844

BERNARD MALISSEN & HERVÉ LUCHE

T cells are one of the main players in adaptive immunity. They express T-cell receptors (TCRs) on their surface, through which they detect antigens. Slight variations in the antigen-binding site of TCRs allow T cells to detect a large variety of antigens. On page 844 of this issue, Pang et al.[1] reveal the structure of the pre-TCR — a receptor that is expressed early in T-cell development and is subsequently replaced by a TCR — providing insight into its dimerization.

TCRs comprise an α-chain and a β-chain, each of which consists of a variable (V) domain and a constant (C) immunoglobulin-like domain (Fig. 1a). The genes encoding the V domains are unusual in that they are assembled, during T-cell development, from smaller DNA building blocks called gene segments. Consequently, a minimum amount of genetic information is combined — that is, the gene segments undergo cleavage and nucleotide deletion and addition before being re-joined — to generate a large collection of TCRα and TCRβ chains that differ in their antigen preference.

The pre-TCR, by contrast, consists of a pTα chain and a TCRβ chain. It lacks a TCRα chain because TCRβ gene rearrangements precede those of the TCRα gene. Instead, its pTα gene, which does not undergo rearrangements, encodes an invariant transmembrane

glycoprotein with a single, extracellular C-type immunoglobulin domain and a cytoplasmic tail[2]. Immature T cells constitutively express pTα and the signalling components required for pre-TCR function — execution of a molecular programme that leads to commitment of

developing T cells to the αβ-T-cell lineage. Successful expression of TCRβ chains is therefore the rate-limiting step in the assembly of the pre-TCR.

On recognizing antigenic ligands on the surface of other cells, the αβTCRs of mature T cells can form dimers that drive T-cell activation. Pre-TCRs, however, seem capable of dimerization and signalling independently of a ligand[3]. Instead, synthesis of TCRβ chains and their subsequent incorporation into pre-TCRs suffice to trigger the signals that allow immature T cells to escape programmed cell death and to develop further. (Developing T cells are programmed to die unless 'rescued' by the pre-TCR signals.)

Pang and colleagues[1] resolve the crystal structure of the pre-TCR's extracellular domain, shedding light on this receptor's functioning. The structure reveals that pTα contains a typical C-type immunoglobulin domain, and
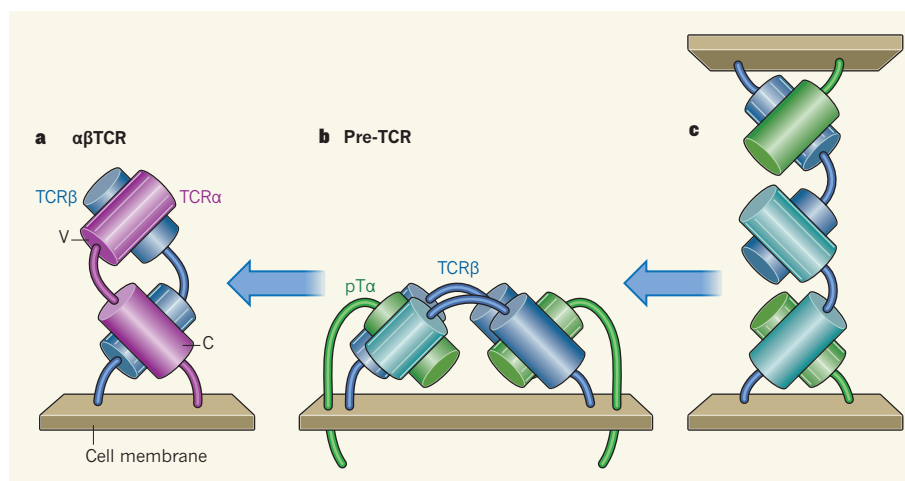

Figure 1 | Receptor evolution. a, The αβTCR, which consists of a TCRα chain and a TCRβ chain, each containing a variable (V) domain and a constant (C) immunoglobulin-like domain, lies perpendicular to the cell membrane. **b**, The pre-TCR consists of a pTα chain and a TCRβ chain. Pang et al.[1] find that two pre-TCR molecules form a functional dimer independent of ligand stimulation. This dimer lies almost parallel to the plane of the membrane. **c**, The evolutionary forerunners of the present-day pTα chain might have enabled interactions between TCRβ-like chains expressed on opposite cells[10]. By acquiring the capacity to interact with the hydrophobic face of the Vβ-like domain, such proto-pTα chains might have led to the emergence of the pre-TCR.

that this domain interacts with the C domain of the associated TCRβ chain. Within the crystal lattice, two pre-TCR molecules adopt a head-to-tail dimeric arrangement in which each pTα C domain is sandwiched between the C domain of its partner TCRβ chain and the V domain of the TCRβ chain of the other pre-TCR molecule (Fig. 1b). This quaternary arrangement, in which a large, surface-exposed hydrophobic patch on the V domain of a TCRβ is shielded by the C domain of pTα, provides an immediate explanation for the ability of the pre-TCR to dimerize autonomously.

Pang and colleagues' ligand-independent dimerization model also implies that the segments that connect the extracellular domains of pre-TCR molecules to the cell membrane have sufficient rotational freedom to allow interactions between the pTα chain of one pre-TCR with the V domain of the TCRβ chain of another pre-TCR on the surface of the same cell. This surmise is based primarily on the authors' analysis of the crystal lattice, and on their biophysical analysis of soluble pre-TCR. To validate it, Pang et al. tested the effect of mutations in amino-acid residues predicted to affect pre-TCR dimerization.

They found that these residues are essential for dimerization of pre-TCR in solution and for its punctate distribution in cells that were not in contact with each other; this suggests that the pre-TCRs dimerize and probably function even in the absence of pre-TCR expressed on the surface of another cell. Moreover, given that the residues within the V domain of the TCRβ chain that contact pTα are evolutionarily conserved in many Vβ genes, Pang and colleagues' model of pre-TCR assembly also explains the ability of pTα to pair with any Vβ domain[4].

On the basis of these data[1], the role of TCRβ in the pre-TCR complex does not involve its antigen-binding ability and differs from its contribution to a second checkpoint that occurs later in the life of a developing T cell — the TCRαβ selection. (At this point, pTα chains are replaced by TCRα chains and the resulting αβTCRs are checked for their antigenic specificity.) Instead, the partnership between pTα and TCRβ at the level of the pre-TCR may have evolved through co-option of a proto-pTα chain, the primordial function of which was limited to allowing interaction between TCRβ-like chains expressed on opposite cells as hinted by a crystal structure[5] (Fig. 1c).

But proteins related to pTα are found only in mammals, suggesting that αβTCR development in other vertebrates does not rely on pre-TCR signalling[6]. So what bonus was conferred by the emergence of a pTα chain? It has been argued[7] that, if rearrangements of the TCRα and the TCRβ genes were to happen concomitantly, and if their outcome were to be checked in a single step, developing T cells would not meet the stringent specificity requirements that characterize TCRαβ selection in mammals. Therefore,

it seems likely that the pre-TCR maximizes the generation of a large collection of T-cell clones, each expressing a TCR endowed with a unique antigenic specificity. This allows mammals to successfully cope with any antigen.

Pang and co-workers' structure[1] is of significance not just because it elucidates the unique mechanism by which assembly of the pre-TCR is directly coupled to its signalling role. The ligand-independent dimerization that the authors describe for pre-TCR might occur more generally during the development of adaptive immunity. Developing B cells also use a pre-B-cell receptor that, although built up differently to pre-TCR[8], monitors successful rearrangement at the level of the gene encoding the immunoglobulin heavy chains. Whether, akin to the pre-TCR, the pre-B-cell receptor also functions through ligand-independent dimerization remains to be seen. ∎

Bernard Malissen and **Hervé Luche** are at the Centre d'Immunologie de Marseille Luminy, INSERM, CNRS, Université de la Méditerranée, Parc Scientifique de Marseille Luminy, Case 906, 13288 Marseille Cédex 9, France.
e-mail: bernardm@ciml.univ-mrs.fr

1. Pang, S. S. et al. Nature 467, 844–848 (2010).
2. Saint-Ruf, C. et al. Science 266, 1208–1212 (1994).
3. Yamasaki, S. et al. Nature Immunol. 7, 67–75 (2005).
4. Wilson, A., Maréchal, C. & MacDonald, H. R. J. Immunol. 166, 51–57 (2001).
5. Bentley, G. A., Boulot, G., Karjalainen, K. & Mariuzza, R. A. Science 267, 1984–1987 (1995).
6. Kreslavsky, T., Gleimer, M. & von Boehmer, H. Curr. Opin. Immunol. 22, 185–192 (2010).
7. Malissen, B., Ardouin, L., Lin, S. Y., Gillet, A. & Malissen, M. Adv. Immunol. 72, 103–148 (1999).
8. Bankovich, A. J. et al. Science 316, 291–294 (2007).

GLACIOLOGY

Ice–sheet advance in Antarctica

Reliable forecasts of sea–level rise depend on accurately modelling the dynamics of polar ice sheets. A numerical framework that better reflects ice–sheet basal drag adds greater realism to such models.

FABIEN GILLET-CHAULET & GAËL DURAND

Sea levels are rising, with about one-third of the increase being attributable to the release of ice from coastal glaciers in Greenland and Antarctica. A decade ago came news that these glaciers were accelerating, and they have continued to do so. But our poor understanding of the processes that are driving ice discharge casts doubt on the reliability of sea-level forecasts for the coming centuries[1]. In a paper in Geophysical Research Letters, Morlighem et al.[2] provide a leap forward in the modelling of ice-sheet dynamics that promises to improve matters considerably.

Confident predictions of the evolution of a system obviously require a proper description of its current state. For polar ice sheets, the advent of remote-sensing techniques has been a great help, and it was with such data that their significant mass loss was diagnosed. A particular example of rapid change is that two-thirds of the mass loss from the West Antarctic Ice Sheet can be attributed to the Pine Island Bay sector, where ice release has doubled over the past decade[3]. From a broader perspective, we know the surface elevation and surface velocities of a fair portion of both the Greenland and Antarctic ice sheets, and are starting to have a reasonable view of the changes that

have occurred over the past ten years. But other crucial factors operating at the base of the ice sheets, such as the drag exerted by the underlying bedrock on the ice, are much less well known.

Although description of the ice sheets' current state is reaching maturity, forecasts require a numerical model that adequately describes the thermo-mechanical processes involved. Given the continental scale of the ice masses concerned, and the limited computing resources, most ice-sheet models use approximations of the ice-flow equations[4]. These approximations are fully justified when applied to the part of an ice sheet that is grounded (resting on bedrock), or the floating extensions (the ice shelves). But they fail to properly describe the flow at the transition zone in between, called the grounding line[5] (Fig. 1). We are therefore faced with deficient models for the very location at which most changes are now occurring. Moreover, ice-sheet models were mainly designed to study the long-term variations of ice masses, and so were focused on palaeoclimatic simulations and reconstruction of the area covered by ice in the past. So modelling of the configuration of present-day ice sheets arose more as a by-product — a final step — of such simulations, rather than as an objective in itself. These
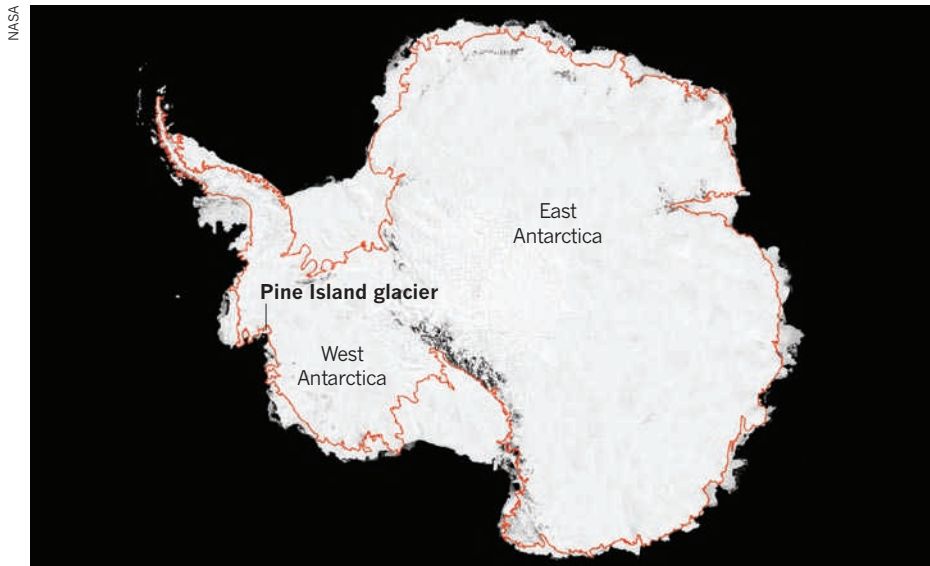
Figure 1 | The Antarctic grounding line. The grounding line is the transition point of a glacier from flow on bedrock to floating ice shelf, and is a crucial location for estimating ice loss to the ocean. Morlighem and colleagues' numerical treatment[2], as applied to the Pine Island glacier, will improve modelling of ice flow in this zone. Identifying the location of the grounding line around Antarctica is another task, and the red line on this map, generated from satellite data, shows just such a depiction. The map was presented earlier this year by a team led by researchers at NASA's Goddard Space Flight Center (see ref. 12).

models, then, only partly represent current ice-sheet geometry and flow.

Help is now at hand through the application of inverse methods. Such methods involve extracting model parameters from available data, and have shown their worth in improving model forecasts in meteorology and oceanography. Although they were introduced in glaciology almost 20 years ago[6], they have remained largely unexploited. But new initiatives have recently emerged[7,8], among which the work of Morlighem *et al.*[2] offers the most accomplished attempt to tackle the weaknesses in existing ice-sheet models.

First, using the finite-element method, Morlighem *et al.* solved the full set of flow equations (the Stokes equations) and two approximations of those equations commonly used in ice-sheet models. Second, they generalized a particular inverse method, namely the control method[9], to plausibly infer the basal drag, exerted by the bedrock on the overlying glacier, from measured surface velocities. Finally, they applied their methodology over the Pine Island drainage basin, successfully reducing the gap between surface observations and model results. The comparison of the basal-drag patterns, obtained by the inverse method for each flow approximation, confirms the need for careful flow resolution in the vicinity of the grounding line.

Challenges remain, however, in terms of both method and computation. Robust simulations will require computation of grounding-line motion, which is at the heart of a 40-year debate. Recent progress has been made in this respect, from both theoretical[10] and numerical perspectives[11], showing that a significant grid refinement (to the subkilometre scale)

is required in this transition zone between grounded and floating ice. Such a refinement, together with the extension of Morlighem and colleagues' methodology from a drainage basin to a whole ice sheet, will require massive parallel computation facilities. Nonetheless, we have reached the stage at which more robust numerical tools will soon allow model simulations to be run forward in time, and proposal of prognostic simulations of the behaviour of polar glaciers. ∎

Fabien Gillet-Chaulet and **Gaël Durand** *are in the Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), CNRS/UJF-Grenoble I, 38400 St Martin d'Hères, France.* e-mails: fabien.gillet-chaulet@lgge.obs.ujf-grenoble.fr; durand@lgge.obs.ujf-grenoble.fr

1. Solomon, S. D. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
2. Morlighem, M. *et al. Geophys. Res. Lett.* **37,** doi:10.1029/2010GL043853 (2010).
3. Rignot, E. *et al. Nature Geosci.* **1,** 106–110 (2008).
4. Ritz, C., Rommelaere, V. & Dumas, C. *J. Geophys. Res.* **106,** doi:10.1029/2001JD900232 (2001).
5. Vieli, A. & Payne, A. J. *J. Geophys. Res.* doi:10.1029/2004JF000202 (2005).
6. MacAyeal, D. R. *J. Glaciol.* **39,** 91–98 (1993).
7. Heimbach, P. & Bugnion, V. *Ann. Glaciol.* **50,** 67–80 (2009).
8. Arthern, R. J. & Gudmundsson, G. H. *J. Glaciol.* **56,** 527–533 (2010).
9. Vieli, A. & Payne, A. J. *Ann. Glaciol.* **36,** 197–204 (2003).
10. Schoof, C. *J. Geophys. Res.* doi:10.1029/2006JF000664 (2007).
11. Durand, G., Gagliardini, O., de Fleurian, B., Zwinger, T. & Le Meur, E. *J. Geophys. Res.* doi:10.1029/2008JF001170 (2009).
12. www.nasa.gov/topics/earth/features/antarctica-outline.html

50 YEARS AGO

In a recent address given in the University of Nottingham, Sir William Slater, secretary of the Agricultural Research Council, made a strong plea that universities should lead rather than follow contemporary thought, and in particular that they should be sufficiently independent financially to be able to allocate sufficient funds for research on lines of work entirely of their own choosing. Sir William was discussing the relative values of such research and of sponsored research, undertaken with the assistance of outside bodies ... The problem of course arises out of the extent to which the universities are now dependent on public funds, and how far, even with the University Grants Committee, it is possible for their independence from outside pressure to be fully maintained. **From** *Nature* **15 October 1960.**

100 YEARS AGO

The second congress on radiology and electricity was held at Brussels on September 13–15 ... [The] president called upon Madame Curie to give an account of the recent experiments made in Paris to isolate metallic radium. It will be remembered that this metal has hitherto not been separated from its salts, although a radium amalgam was obtained some years ago by Coehn. The beautiful experiments described by Madame Curie, resulting in the isolation of metallic radium, must be regarded as a triumph in chemical manipulation when it is remembered that ... the operations had to be carried out with minute quantities of material in such a way as to avoid loss of the precious substance during the process. These experiments should remove all possible doubt that radium is, in fact, an element belonging to the same group of metals as barium. **From** *Nature* **13 October 1910.**

CANCER

Oncogenes in context

In certain types of gastrointestinal cell, mutations in the protein KIT give rise to gastrointestinal stromal tumours. Why are other cell types that express KIT not affected? The answer lies with a second protein. SEE LETTER P.849

MICHAEL C. HEINRICH & CHRISTOPHER L. CORLESS

Among sarcomas — cancers of connective tissues — the gastrointestinal stromal tumour (GIST) is the most common in humans. In 80–85% of cases, GISTs are associated with gain-of-function mutations in the receptor tyrosine kinase KIT. On page 849 of this issue, Chi et al.[1] show that KIT has a partner in crime in the form of the transcription factor ETV1.

Most GISTs harbour non-germline mutations in KIT. In the rare familial GISTs, an activating KIT mutation is inherited. People carrying this mutation show diffuse, abnormal proliferation (hyperplasia) of a gastrointestinal cell type called the interstitial cell of Cajal (ICC), and are at high risk of developing GISTs throughout life[2]. ICCs are located in the muscular wall of the gut and function as pacemaker cells to regulate the process of peristalsis, whereby muscle contractions propel food through the digestive tract. GISTs are believed to originate from ICC stem/progenitor cells or from mesenchymal stem cells that give rise to ICC progenitors[3].

In the gut wall, ICCs occur in several distinct anatomical locations, but GISTs arise from only two specific populations: myenteric ICCs and intramuscular ICCs. Notably, the hyperplasia seen in humans and mice with germline activating KIT mutations is also confined to these two ICC populations[1,2,4]. What's more, these humans and mice are strongly predisposed to developing GISTs, but do not commonly develop cancers in other tissues with high KIT expression — such as those containing melanocytes, germ cells or haematopoietic cells[2,4].

So why is the pathogenicity of mutant KIT restricted to specific cell types? A tissue-restricted effect of oncogene activation — or of loss of tumour-suppressor genes — is a feature of many cancers. For example, the fused gene *BCR–ABL* functions as an oncogene in haematopoietic stem cells (most commonly causing chronic myeloid leukaemia) but not in other cell types[5]. In addition, loss of the *APC* tumour-suppressor gene is an initiating event in the formation of colonic polyp/adenocarcinoma but not other common solid tumours[6]. Like KIT, therefore, these genes act as 'gatekeepers' in specific cellular contexts, but what constrains their effects remains poorly understood.

Chi et al.[1] show that, where GISTs are concerned, ETV1 cooperates with mutated KIT in a cell-specific way to cause tumours. ETV1 is a member of the ETS family of transcription factors, which has at least 29 members. All ETS transcription factors have an evolutionarily conserved 80-amino-acid-long DNA-binding domain, which binds to the consensus DNA sequence GGA(A/T). Individual ETS proteins regulate the activity of gene promoters directly through their intrinsic activation or repression domains, and/or by interaction with other proteins[7].

Examining existing gene-expression profiles of GISTs and non-GIST sarcomas, Chi et al. show that ETV1 is part of the GIST expression signature. They report high expression of this protein, and of its messenger RNA, in all GIST samples tested, and show that reducing ETV1 expression decreases GIST-cell proliferation, as well as the tumorigenicity of GIST cells when these are injected into immunodeficient mice. The authors do not, however, find any evidence of genomic alteration affecting the *ETV1* gene.

Chi and co-workers also show that, in mice carrying germline KIT mutations, ETV1 expression in the gut is restricted to only those subsets of ICCs that give rise to GISTs (myenteric and intramuscular ICCs). In mice lacking ETV1, these specific subsets occur in lower numbers than in wild-type mice, indicating a requirement for ETV1 in these cells' development. Thus, ETV1 functions as a lineage-specific survival factor for the cells that give rise to GIST.

The authors' data also elucidate the signalling cascade that links KIT to ETV1. Chemically inhibiting KIT-dependent activation of the MAP-kinase enzyme resulted in a rapid loss of ETV1, suggesting that this transcription factor functions downstream of KIT. The ability of KIT or MEK inhibitors to decrease ETV1 levels was abrogated by inhibition of the protein-degrading machinery, the proteasome (Fig. 1).

Genomic alteration of ETS genes has been reported in several human cancers. For instance, translocation mutations affect ETS family members in Ewing's sarcoma (most commonly *EWS–FLI1*, but also *ERG*, *ETV1*, *ETV4* and *FEV*) and in prostate cancer (most commonly *TMPRSS2–ERG*, but also *ETV1*, *ETV4* and *ETV5*)[7–9]. Furthermore, mutations in the ETS protein SPI1 are associated with some cases of acute myeloid leukaemia, and high-level amplification of ETV1 has been reported in a significant minority of melanoma cancers[7,10]. Intriguingly, melanoma cell lines depend on ETV1 not just for proliferation, but also for growth independent of tissue anchorage. What's more, the ability of ETV1 to make melanocytes cancerous depends on constitutive activation of the MAP-kinase signalling pathway by gain-of-function mutations in the BRAF or NRAS proteins[10].

What's surprising about Chi and colleagues' results is that they do not show ETS-dependent
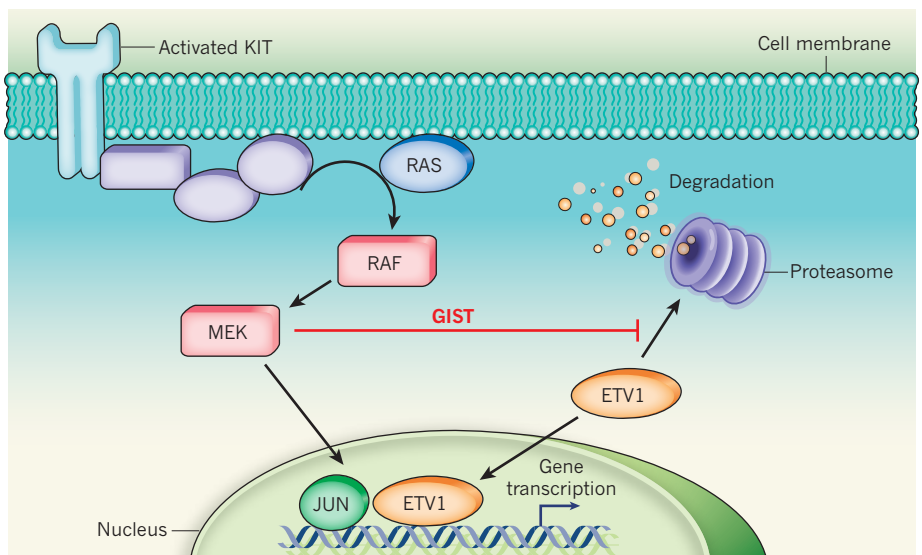
Figure 1 | Gastrointestinal stromal tumour at a cellular level. Activation of the receptor tyrosine kinase KIT triggers the RAS/RAF/MEK pathway. This pathway ultimately results in changes in gene transcription mediated by activating specific transcription factors such as JUN. The transcription factor ETV1 acts downstream of the RAS/RAF/MEK pathway and can directly regulate gene expression. Chi et al.[1] find that ETV1 is essential for KIT-mediated development of gastrointestinal stromal tumours (GISTs). Normally, proteasomal degradation of ETV1 leads to a decrease in ETV1-dependent gene transcription. In GIST, signalling from mutant KIT leads to increased activity of MEK, which then blocks ETV1 degradation. The combination of MEK activation and developmental expression of ETV1 produces the gene expression profile that is characteristic of GIST.

oncogenesis related to a genomic alteration. Instead, the increased expression of ETV1 in GISTs seems to be a consequence of KIT signalling and the underlying gene-expression profile inherent to specific subsets of ICCs. In other words, to induce GISTs, KIT mutations act within the cellular context provided by ETV1 expression.

On the basis of this evidence[1], ETV1 could not only be a useful diagnostic marker for GISTs, but also a drug target. The use of the anticancer drug imatinib and other KIT-kinase inhibitors has revolutionized the treatment of metastatic GISTs and increased the median survival of patients by 3–4 years. Imatinib, however, must be given continuously, because interrupted treatment with this drug is associated with rapid tumour re-growth[11]. Moreover, recent work indicates that, although differentiated GIST cells are KIT-dependent and imatinib-sensitive, the underlying GIST stem/progenitor pool is KIT-independent and imatinib-resistant[3]. Whether the survival of GIST stem/progenitor cells depends on ETV1 is therefore a pertinent question. If so, targeting ETV1 in GISTs may produce the cure that currently eludes KIT inhibitors such as imatinib.

More broadly, Chi and colleagues' data[1] suggest that tissue-specific expression of ETS family members may provide an oncogenic context for other human cancers. Determining which ETS family members provide an oncogenic context for other cancers requires further research. ∎

Michael C. Heinrich *is in the Department of Medicine and* **Christopher L. Corless** *is in the Department of Pathology, the Portland VA Medical Center and OHSU Knight Cancer Institute, Portland, Oregon 97239-3098, USA.* e-mail: heinrich@ohsu.edu

1. Chi, P. *et al. Nature* **467,** 849–853 (2010).
2. Kleinbaum, E. P. *et al. Int. J. Cancer* **122,** 711–718 (2008).
3. Bardsley, M. R. *et al. Gastroenterology* **139,** 942–952 (2010).
4. Sommer, G. *et al. Proc. Natl Acad. Sci. USA* **100,** 6706–6711 (2003).
5. Honda, H. *et al. Blood* **85,** 2853–2861 (1995).
6. Fodde, R., Smits, R., Hofland, N., Kielman, M. & Meera Khan, P. *Cytogenet. Cell Genet.* **86,** 105–111 (1999).
7. Seth, A. & Watson, D. K. *Eur. J. Cancer* **41,** 2462–2478 (2005).
8. Clark, J. P. & Cooper, C. S. *Nature Rev. Urol.* **6,** 429–439 (2009).
9. Ordóñez, J. L., Osuna, D., Herrero, D., de Álava, E. & Madoz-Gúrpide, J. *Cancer Res.* **69,** 7140–7150 (2009).
10. Jané-Valbuena, J. *et al. Cancer Res.* **70,** 2075–2084 (2010).
11. Le Cesne, A. *et al. Lancet Oncol.* **11,** 942–949 (2010).

Competing interests The authors declare competing financial interests. See online article for details.

Phase matters for aerosols

Organic aerosol particles are ubiquitous in the atmosphere. In forests, such particles can occur in solid form — a finding that will lead to a re-evaluation of how they are formed, and their properties and effects. SEE LETTER P.824

PAUL J. ZIEMANN

Most schoolchildren, mischievous or not, know that a thrown egg sticks to a wall because it is predominantly liquid, but that a rock bounces because it is solid. Adults know this too, and, reporting on page 824 of this issue, Virtanen *et al.*[1] cleverly use this simple principle, along with electron microscopy, to determine the phase of much smaller airborne objects — atmospheric aerosol particles. Their measurements show that, contrary to conventional wisdom, the abundant microscopic organic particles that form in boreal forests by condensation of photo-oxidized tree emissions can be solid. This result has implications for understanding the formation and properties of these and other organic particles that are thought to affect both climate and human health.

The atmospheric aerosol is a complex mixture of primary and secondary material[2]. Primary aerosol is directly emitted in particulate form and consists of soil dust, sea spray, plant material and soot. Secondary aerosol consists of organic compounds and ammonium nitrates and sulphates — which form by condensation and particle-phase reactions of products of the atmospheric oxidation of volatile organic compounds, sulphur oxides and nitrogen oxides — and ammonia. Because of their mechanisms of formation, secondary and combustion-derived particles are typically less than about one micrometre in diameter, whereas most other particles are larger. Submicrometre-size particles can efficiently scatter and absorb radiation and serve as cloud condensation nuclei (CCN), thereby affecting climate, and can also be readily inhaled and thus adversely affect human health.

An important property of particles is their phase (liquid or solid). This affects gas–particle partitioning and reactions of organic compounds, oxidants, water vapour and other species, thereby influencing particle composition, optical properties and CCN activity. The phase behaviour of systems composed of simple inorganic salts and water can be predicted reasonably well by thermodynamics. But this is not yet possible for the complex organic and organic–inorganic–water mixtures present in the atmosphere, which laboratory studies indicate can form amorphous glasses, rubbers, gels and (viscous) liquids[3,4].

Under these circumstances, measurements of the phase of authentic atmospheric organic particles become essential, and they are provided here by Virtanen and colleagues[1]. In studies carried out in a boreal forest in Finland and in a plant chamber, they determined the phase of freshly formed, submicrometre-size particles composed almost entirely of secondary organic aerosol (SOA) formed by the reactions of hydroxyl (OH) radicals and ozone (O_3) with terpenes emitted from coniferous trees.

Particle phase was determined using an electrical low-pressure impactor (ELPI). In this device, aerosol particles carried in a stream of air are charged and then sequentially separated on the basis of size-dependent aerodynamic properties. Particles in different size ranges then hit separate plates in order of decreasing size, and the associated electrical current, which is proportional to concentration, is measured. To determine particle phase, Virtanen et al.[1] took advantage of a normally undesirable property

of inertial impactors. Although liquid particles stick to plates, solid particles can bounce off, so instead of being collected on the appropriate plate for their size they are deposited on a plate intended to collect smaller particles. This behaviour skews the measured particle-size distribution to smaller sizes.

By comparing the size distribution measured with an ELPI to one calculated using an ELPI model and the true size distribution measured with an electrical mobility instrument (which operates on a different principle and gives reliable size distributions), Virtanen et al.[1] were able to quantify the extent to which bounce occurred. From comparisons of the bounce of SOA particles with that of standard crystalline, liquid and solid glassy particles, they concluded that the SOA particles were solid. Analysis of collected SOA particles by scanning electron microscopy gave results that supported the ELPI measurements, and also showed that the particles were not crystalline but amorphous, indicating that they may exist in a solid, glassy state.

The results of this study have consequences for understanding and modelling atmospheric organic aerosol and its environmental effects. In continental regions, the main component of submicrometre-size particles is usually organic matter, which in urban, rural and remote areas is often dominated by SOA[5]. In most models this material is treated as an absorbing phase, into which organic compounds can freely partition and subsequently react without diffusion limitations[6]. When the particles are glassy solids, however, this is not the case. New models

will need to be developed for such systems.

Perhaps small modifications will suffice — such as treating the organic component of particles as a two-phase system, with a liquid organic layer in which partitioning and condensed-phase reactions occur freely, similar to most current models, and a solid, glassy phase with which this layer exchanges material over longer timescales. However, it is more likely that models will need to include organic and inorganic components, and an aqueous phase[3,4]. For example, Virtanen et al.[1] note that boreal forest particles deliquesce (take up water and become liquid) at a relative humidity somewhere between 30% and 70%, and thus are not always solid. Ideally, laboratory studies will also incorporate this additional chemical complexity, so that models can be rigorously tested, and field studies will be designed to determine the applicability of the models to the real atmosphere. ∎

Paul J. Ziemann *is at the Air Pollution Research Center and the Department of Environmental Sciences, University of California, Riverside, California 92521, USA. e-mail: paul.ziemann@ucr.edu*

1. Virtanen, A. *et al. Nature* **467,** 824–827 (2010).
2. Pöschl, U. *Angew. Chem. Int. Edn* **44,** 7520–7540 (2005).
3. Marcolli, C., Luo, B. & Peter, T. *J. Phys. Chem. A* **108,** 2216–2224 (2004).
4. Mikhailov, E., Vlasenko, S., Martin, S. T., Koop, T. & Pöschl, U. *Atmos. Chem. Phys.* **9,** 9491–9522 (2009).
5. Zhang, Q. *et al. Geophys. Res. Lett.* doi:10.1029/2007GL029979 (2007).
6. Kroll, J. H. & Seinfeld, J. H. *Atmos. Environ.* **42,** 3593–3624 (2008).

Political evolution

Phylogenetic methods of evolutionary biology can be used to study socio-political variation mapped onto linguistic trees. The range of political complexities in Austronesian societies offers a good test case. SEE ARTICLE P.801

JARED DIAMOND

Today, the whole world is divided into states whose political organization consists of multi-level bureaucracies frustratingly familiar to all of us whenever we try to deal with government agencies. But as Europeans expanded around the globe in recent centuries, they encountered non-European peoples living under simpler political organizations, variously termed chiefdoms, tribes or bands, of which the simplest was egalitarian and leaderless[1]. The archaeological record suggests that all societies were politically simple until around 10,000 years ago, and that the first states didn't arise until about 3400 BC. Through what stages has political

complexity changed with time? On page 801 of this issue, Currie et al.[2] describe a big advance in understanding this central problem of political science. They have achieved it by applying the quantitative methods of phylogenetic-tree analysis now routine in evolutionary biology.

Currie et al. took advantage of a particularly suitable database: the territorial expansion and political evolution of speakers of Austronesian languages. From that family's differentiation on the island of Taiwan around 3200 BC, Austronesian speakers spread through the Philippines into Indonesia, west to coastal southeast Asia and Madagascar, and east through the Pacific Ocean to evolve into Polynesians and colonize every habitable Pacific island from Hawaii and Easter Island to New Zealand[3,4].

One advantage of the Austronesian expansion for scholars is that the Austronesian language tree is densely filled out, with 1,200 extant languages, and without the language extinctions that drastically pruned the Indo-European language tree[4]. A second advantage is that, over most of the Polynesian realm, Austronesian people were the first human colonists, thereby decreasing complexities associated with population mixing[3]. Finally, ancestral Austronesian societies underwent spectacular political differentiation[5] to give rise to examples of the entire range of political organization, from small egalitarian societies such as Borneo's Iban, through the simple chiefdoms of Easter Island and New Zealand and the complex chiefdoms of Tahiti and Sumatra, to the Javan and Malagasy states (Fig. 1).

Evolutionary biologists have for years been using phylogenetic trees, based on genetic data, to evaluate hypotheses about the evolution of species morphologies, physiologies and behaviours. From those genetically constructed trees, and from (for example) morphologies of the species now constituting the tips of the trees' branches, biologists reconstruct the most probable ancestral states and paths of morphological evolution[6]. Similarly, Currie et al.[2]

Figure 1 | Extremes of Austronesian political organization. The temple of Borobudur, on the Indonesian island of Java, was the product of a highly organized, ancient Javanese state. By contrast, tribesmen of the Iban group of Sarawak, Borneo, epitomize the simplest type of egalitarian, leaderless society.

construct the Austronesian phylogenetic tree on the basis of linguistic data (vocabulary lists in Austronesian languages), mark the political organization of recent Austronesian societies constituting the tips of the tree's branches, and thereby reconstruct ancestral states and evolutionary paths of Austronesian political organization. Their database consists of 84 Austronesian societies for which adequate vocabulary lists and recent political status are both known.

Sociologists, political scientists, archaeologists and anthropologists have debated the paths of political evolution. Does political complexity increase only in small steps, or also in big jumps? Does complexity increase only unidirectionally, or can complex societies revert to simpler ones? If they can revert, is the reversion only in small steps, or can it also occur in big jumps? Currie *et al.* use these various possible paths to construct six competing models of political evolution, which they test against their trees.

The results are clear. First, political evolution increases only in small steps: states and complex chiefdoms don't form directly from leaderless societies. This conclusion fits historical observations of the formation of complex societies (for instance, the Malagasy, Cherokee and Zulu states), when one unit at the next-lower level succeeded in conquering or incorporating its neighbours. Second, political complexity can decrease as well as increase, in agreement with abundant evidence of the disintegration of states and chiefdoms. Finally, unlike increases of complexity, declines can plunge a society politically several stages backwards. This can happen if a small group breaks off from a large society to form a small new society (as in the colonization of the Chatham Islands from New Zealand), or if political institutions disintegrate (as on Mangareva in Polynesia).

Knowledgeable readers of the paper by Currie *et al.*, or of this account, will raise a host of detailed objections. But but but … why do they call Hawaii a complex chiefdom rather than a proto-state? Why did they characterize political complexity coarsely by those four levels, rather than by some finer scheme? A short response is: read the paper's extensive online supplementary material.

The quantitative modelling techniques described there had previously received little application to the problem of the evolution of political complexity. Almost all of us in this field (myself included) have instead proceeded usually by narrative accounts of individual cases, less often by narrative comparisons of selected cases, and infrequently by comprehensive narrative surveys. My first reaction to Currie and colleagues' paper was one of surprise: why hadn't we used their method before, because it is so obviously superior?

One answer is that attributed to Christopher Columbus, after no one in his audience could solve his challenge to make an egg stand up without falling over: they (and we) never thought of Columbus's solution. (It was to stand the egg on end in a small pile of salt.) The other answer is that implementation of Currie and colleagues' simple idea required huge quantities of data and technical knowledge: lists of 210 vocabulary terms for more than 500 Austronesian languages, ethnographic information on the political complexity of 113 Austronesian societies, and mastery of the notorious difficulties of phylogenetic-tree construction and model evaluation.

What future extensions of this analysis might there be? Other aspects of Austronesian societies, such as canoe design, lend themselves to a similar approach[7]. Other political radiations besides that of Austronesian societies may similarly be mapped onto language-based phylogenetic trees. A prime target is the African societies speaking Niger–Congo languages, whose political systems were mostly chiefdoms but included a few states and possibly some simpler societies. The Niger–Congo language family shares the advantages of language diversity, tree density and time depth with the Austronesian language family, but reconstructions of its family tree are less advanced[8]. Finally, there's a grand challenge: can this approach ever succeed with the drastically pruned Indo-European language tree and the apparently less varied Indo-European political organization? ∎

Jared Diamond *is in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA. e-mail: jdiamond@geog.ucla.edu*

1. Johnson, A. W. & Earle, P. *The Evolution of Human Societies* (Stanford Univ. Press, 2000).
2. Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T. & Mace, R. *Nature* **467,** 801–804 (2010).
3. Kirch, P. V. *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact* (Univ. California Press, 2000).
4. Gray, R. D., Drummond, A. J. & Greenhill, S. J. *Science* **323,** 479–483 (2009).
5. Kirch, P. V. *The Evolution of the Polynesian Chiefdoms* (Cambridge Univ. Press, 1984).
6. Pagel, M. *Nature* **401,** 877–884 (1999).
7. Rogers, D. S. & Ehrlich, P. R. *Proc. Natl Acad. Sci. USA* **105,** 3416–3420 (2008).
8. Holden, C. & Gray, R. in *Phylogenetic Methods and the Prehistory of Languages* (eds Forster, P. & Renfrew, C.) 19–31 (McDonald Inst. Archaeol. Res., 2006).

CORRECTION

In the News & Views article "Nanoscience: Dark-hot resonances" by Mark I. Stockman (*Nature* **467,** 541–542; 2010), the 130-nm and 20-nm scale bars in Figure 2 should have indicated a distance between the particle rims, not their centres.

ARTICLE

Rise and fall of political complexity in island South–East Asia and the Pacific

Thomas E. Currie[1,2], Simon J. Greenhill[3,4], Russell D. Gray[3], Toshikazu Hasegawa[1] & Ruth Mace[2]

There is disagreement about whether human political evolution has proceeded through a sequence of incremental increases in complexity, or whether larger, non-sequential increases have occurred. The extent to which societies have decreased in complexity is also unclear. These debates have continued largely in the absence of rigorous, quantitative tests. We evaluated six competing models of political evolution in Austronesian-speaking societies using phylogenetic methods. Here we show that in the best-fitting model political complexity rises and falls in a sequence of small steps. This is closely followed by another model in which increases are sequential but decreases can be either sequential or in bigger drops. The results indicate that large, non-sequential jumps in political complexity have not occurred during the evolutionary history of these societies. This suggests that, despite the numerous contingent pathways of human history, there are regularities in cultural evolution that can be detected using computational phylogenetic methods.

The emergence of large-scale, complex societies from small groups of closely related individuals since the end of the last ice age has been described as "history's broadest pattern"[1]. On the basis of comparative ethnographic and archaeological data, anthropologists have classified human political organization into a limited number of forms. These classifications have frequently been organized into evolutionary sequences[1–6]. For example, under one scheme[3] small 'bands' of family groups evolve into larger kinship-based groups, or 'tribes'. Bands and tribes are generally egalitarian with only informal and ephemeral leadership roles. Tribes evolve into chiefdoms, where political leadership over a collection of local groups is centralized in a hereditary office. Chiefdoms in turn evolve into states, characterized by a centralized political bureaucracy that contains a number of more specialized administrative offices[4,6].

These classifications have been criticized for not capturing enough of the variation in political organization, or the alternative historical pathways that may have led to this variation[7–9]. For example, it has been argued that societies labelled as tribes, chiefdoms or states do not evolve from one into another but are instead the outcomes of alternative evolutionary trajectories[10,11]. Traditional social evolutionary theories have also tended to focus on changes towards increased hierarchical organization, yet there are numerous examples of societies that have decreased in complexity[10–12].

A phylogenetic approach

These ideas have been debated largely in the absence of rigorous, quantitative tests. Although archaeology has a central role in answering such questions[6,13], the archaeological record may not contain sufficient information to distinguish between competing hypotheses. Therefore, it is important to supplement archaeological investigations with other techniques[6,14–16]. Just as evolutionary biologists routinely use phylogenetic trees constructed using genetic data to test evolutionary hypotheses about such things as morphology, physiology, or behaviour[17], anthropologists have recently begun to use cultural phylogenies to test hypotheses about human social and cultural evolution[18].

Trees built using the similarities between basic vocabularies in different languages have been shown to be good indicators of population history[19]. By mapping data about the characteristics of societies onto the tips of these trees (Fig. 1), we can use phylogenetic comparative methods to make inferences about what societies were like in the past and how they have changed over time (Supplementary Fig. 2). In this study we statistically evaluate competing models of political evolution using data from Austronesian-speaking societies of the Pacific and island South-East Asia.

The ancestors of modern-day Austronesians arose in Taiwan around 5,200 years ago, and rapidly spread through island South-East Asia into Oceania, eventually reaching as far afield as Madagascar, Hawaii, Easter Island and New Zealand[19,20]. Recently, phylogenetic trees showing the historical relationships between these societies have been inferred using basic vocabulary data, by applying the same techniques that biologists use with genetic data to infer the shared ancestry of species[19]. The range of forms of political organization in Austronesian societies in conjunction with this knowledge of their ancestral relationships enables us to test between competing hypotheses of the evolution of political organization. We define levels of political complexity by the number of hierarchical decision-making levels[4,21,22], which accords well with traditional social evolutionary classifications: societies lacking permanent leadership beyond the local community are labelled 'acephalous'; those with a single level beyond the local community are labelled 'simple chiefdoms'; those with two levels represent 'complex chiefdoms'; and societies with more than two levels are 'states'. For example, in the complex chiefdoms of traditional Hawaiian society local communities (villages) were organized into districts under the control of lesser chiefs, with several districts ruled over by a paramount chief[20]. The geographic and phylogenetic distributions of the 84 Austronesian societies for which ethnographic and linguistic data were available are shown in Fig. 1.

We evaluate six different models for the evolution of political organization derived from discussions in the literature (Fig. 2). Three of these models reflect a defining feature of traditional social evolutionary theories, that is, that political evolution has proceeded

[1]Evolutionary Cognitive Science Research Center, Graduate School of Arts and Sciences, University of Tokyo 153-8902, Japan. [2]Human Evolutionary Ecology Group, Department of Anthropology, University College, London WC1H 0BW, United Kingdom. [3]Department of Psychology, University of Auckland, Auckland 1142, New Zealand. [4]Computational Evolution Group, University of Auckland, Auckland 1142, New Zealand.
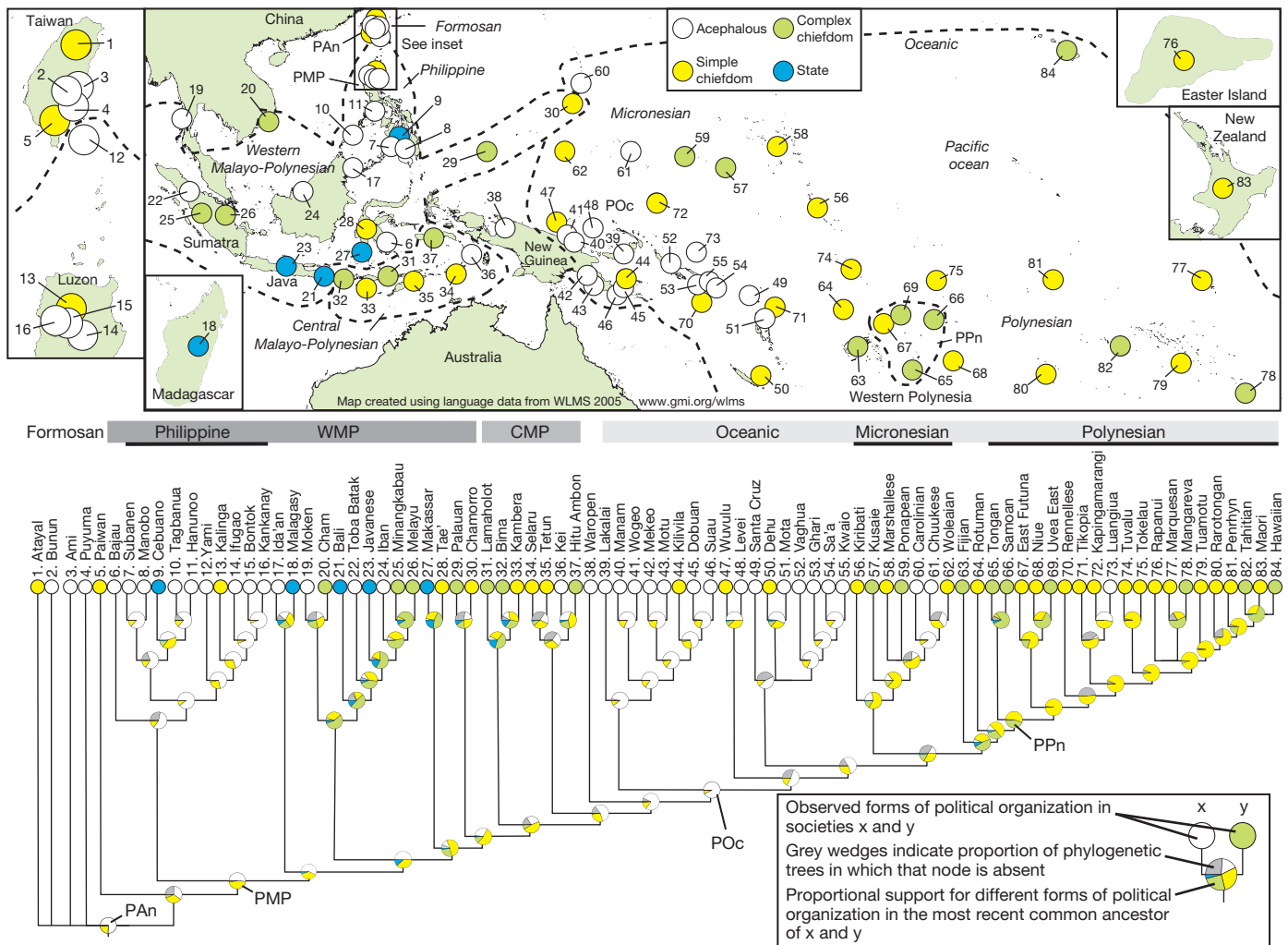
Figure 1 | Phylogenetic relationships and geographical location of 84 Austronesian societies. Majority-rule consensus tree summarizing the 1,000 Austronesian basic vocabulary trees (major linguistic sub-groupings indicated) (see also Supplementary Fig. 1). Circles at the nodes in the consensus tree indicate proportional levels of support for different forms of political organization in ancestral societies under the best-fitting model of evolution (unilinear). Nodes relating to often discussed ancestral Austronesian communities are highlighted: proto-Austronesian (PAn; ancestor to all Austronesian societies), proto-Malayo-Polynesian (PMP), proto-Oceanic (POc) and proto-Polynesian (PPn). Approximate locations of these ancestral communities are indicated on the map. Branch lengths in this figure are arbitrary (that is, not proportional to the degree of linguistic change or time). WMP, Western Malayo-Polynesian; CMP, Central Malayo-Polynesian.
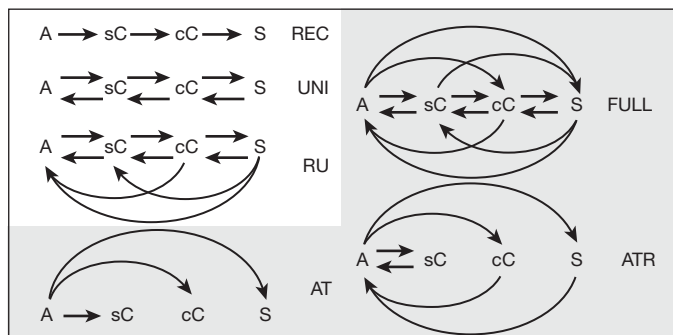
Figure 2 | Models of political evolution tested in this study. The three models on the white background (UNI, unilinear; RU, relaxed unilinear; REC, rectilinear) represent the idea that political evolution follows a sequence of incremental steps in the direction of increasing complexity, but differ as to whether and how decreases can occur. The three models on the grey background (AT, alternative trajectories; ATR, alternative trajectories (reversible); FULL) allow non-sequential increases in political complexity. Forms of political organization: A, acephalous society; sC, simple chiefdom; cC, complex chiefdom; S, state.

through a sequence of incremental increases in hierarchical complexity[1–6]. In these models it is possible for societies to change from an acephalous form of organization to a simple chiefdom, from a simple chiefdom to a complex chiefdom, or from a complex chiefdom to a state. However, the direct transitions from acephalous to complex chiefdom, acephalous to state, and simple chiefdom to state cannot occur. The rectilinear model reflects the extreme view that societies pass through the forms of increasing political complexity sequentially and that decreases in hierarchical organization cannot occur. This view is often associated with the classical evolutionists such as Spencer and Morgan[5]. We specified two further sequential models in which decreases are also possible; the unilinear model, which allows decreases to adjacent less hierarchical forms of organization, and the relaxed unilinear model, which permits decreases to any less hierarchical form. In contrast, the other three models represent the idea that the evolution of increasing hierarchical political organization has not been sequential. It has been proposed that different forms of political organization represent the outcomes of separate evolutionary pathways, evolving directly from an acephalous form of organization[10,11]. We therefore specified two models: alternative trajectories, where only non-sequential increases are possible, and alternative trajectories (reversible), where both increases and decreases are non-sequential. Finally, under a further model—termed the full

model—any change in political organization is possible, reflecting the idea that political evolution has been completely unconstrained.

We tested support for these models using two phylogenetic comparative techniques—maximum likelihood (ML) and Bayesian reversible-jump Markov chain Monte Carlo (RJMCMC)[23]—using the data on political organization described earlier and a sample of 1,000 Austronesian basic vocabulary phylogenetic trees (see Methods and Supplementary Methods). Both techniques estimate the probable rates of change between different forms of organization over the phylogenetic trees given the distribution of data at the tips of the trees. Performing the analyses over a sample of trees allows us to take into account the uncertainty about the historical relationships between Austronesian societies.

Results and discussion

In both the ML and RJMCMC analyses the unilinear model is the best supported, closely followed by the relaxed unilinear model (Table 1). The full model is ranked third and there is evidence against it from the RJMCMC analysis. The alternative trajectories (reversible) model and the two models that do not allow decreases in hierarchical organization (rectilinear and alternative trajectories) are poor fits to the data and receive no support from the RJMCMC analyses. The results of this analysis can be visualized by examining the estimated rates of change between categories of political organization (Fig. 3 and Supplementary Fig. 3). The rate parameters relating to non-sequential increases (acephalous to complex chiefdoms or states, and from simple chiefdoms to states) are often estimated as zero (see also Supplementary Methods and Supplementary Table 4). This indicates that increasing political complexity evolves incrementally whereas non-sequential increases do not occur. Both sequential and non-sequential transitions from more to less hierarchical forms of organization are consistently estimated as being above zero. This demonstrates that decreases in complexity do occur, which is not consistent with the rectilinear model.

We can also use these phylogenetic comparative methods to examine the probable forms of political organization in ancestral Austronesian societies. Figure 1 indicates the inferred probabilities of particular forms of political organization at the nodes in the consensus tree under the best-fitting model of evolution (unilinear). These ancestral states indicate that proto-Austronesian society was politically acephalous. There is evidence that political organization became more hierarchical as the Austronesians moved through island South-East Asia, with some societies, such as in Java and Bali, eventually evolving into states with strong influences from Indian societies. The results also indicate that early Austronesian societies became less hierarchical after entering the region around New Guinea. This potential decrease in hierarchical organization could perhaps be due to founder effects of small populations, contact with indigenous, non-Austronesian-speaking populations, or other ecological factors[24]. Our results indicate that leadership in proto-Oceanic (Fig. 1) society did not extend beyond local communities, consistent with archaeological evidence that settlements were generally small-scale[20]. Archaeological and linguistic evidence
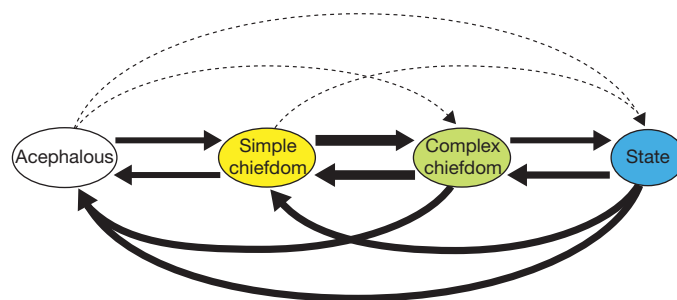
Figure 3 | Estimated rate parameters for transitions between different forms of political organization under the RJMCMC analysis. Arrow thickness is proportional to the mean estimated value of the associated rate parameter. Transitions from acephalous society to state, acephalous society to complex chiefdom and simple chiefdom to state are represented by a dotted line as they are often estimated as zero (Supplementary Methods, Supplementary Table 4 and Supplementary Fig. 3). These results are consistent with the idea that political evolution has followed a sequence of incremental increases in hierarchical organization.

also indicates that some form of social ranking may have been present in these societies[16,20,24], which later became elaborated in simple chiefdoms such as the Trobriand Islanders (Kilivila), and in the more complex societies that developed relatively recently in Polynesia and Micronesia[20] (see Supplementary Discussion). The estimated ancestral state at the proto-Polynesian node is consistent with the inference from prehistorians that hereditary chiefs were an important aspect of proto-Polynesian node society[14]. The ancestral states in Oceania illustrate the sequential, incremental nature of the evolution of increasing hierarchical political organization, starting with an ancestral acephalous form of organization in proto-Oceanic society, followed by a transition to simple chiefdoms in societies ancestral to Micronesians and Polynesians, a form of organization that was continued in many societies such as the Tikopia and the Marshallese. Simple chiefdoms subsequently developed into more complex chiefdoms such as Tonga, Tahiti, Hawaii and Ponape. However, the occurrence of several societies that seem to have decreased in complexity, such as the Polynesian society on Luanguia atoll, highlights that political evolution has not been a unidirectional process and that reductions in hierarchical organization have also featured in Austronesian history.

In the present day, larger scale and more hierarchical forms of political organization hold sway[1]. It has been argued that this is as a result of the advantage in direct or indirect competition between groups that such forms of organization can have, owing to improvements in the coordination, cooperation and division of labour between individuals within the groups[21,22,25]. Despite this potential advantage, our results indicate that political evolution is constrained to follow only incremental increases in complexity. This could be due to such factors as an evolved social psychology adapted to life in small-scale groups, the difficulty in reorganizing existing institutions that rely on the coordination of large numbers of individuals, or the requirement for the development of other institutions before more hierarchical organization is stable[25]. For example, Spencer[6] argues that the unification of the Hawaiian islands was only made possible by the development of administrative institutions such as local land managers, which enabled centralized control to be extended over a wider area. Our results are also consistent with the view that the evolution of more hierarchical forms of political organization results from the welding together of the smaller pre-existing units[1,2]. This may involve several 'cycles' of the aggregation and break-up of units until appropriate institutions evolve that enable stable forms of increased hierarchical organization[6,13,26]. In Austronesian societies, this joining together of units has been witnessed historically in Madagascar in the late 1700s and Hawaii in the early 1800s, where states were formed by the unification through conquest of one complex chiefdom over other neighbouring chiefdoms[6,26].

Table 1 | Relative rankings of models of political evolution

Model	Rank	Mean AIC	Bayes factors
Unilinear	1	201.70	157.71
Relaxed unilinear	2	202.97	114.62
Full	3	208.67	0.03
Alternative trajectories (reversible)	4	208.71	0
Rectilinear	5	238.69	0
Alternative trajectories	6	273.33	0

Akaike information criterion (AIC) values calculated using the ML of the data given each model of evolution. Lower AIC values indicate greater support. Bayes factors are calculated based on the number of times the RJMCMC visits a particular model of evolution in comparison to the expected number of times (Supplementary Methods). A Bayes factor <1 is evidence against that model. A Bayes factor >100 is considered decisive evidence in favour of a hypothesis. The alternative trajectories (reversible), relaxed unilinear and alternative trajectories models were not visited at all by the RJMCMC and therefore receive no support from these analyses.

Interestingly, although the unilinear model was the best supported, our results show that the relaxed unilinear model was also well-supported. It is therefore possible that political organization may fall back to less hierarchical forms non-sequentially. This could occur if small, peripheral groups break away from the control of a centralized state or complex chiefdom to found new societies with fewer levels of political organization, or it could occur as part of a rapid, more widespread societal collapse and the breakdown of political institutions leaving smaller, less politically complex groups in some regions[12]. For Austronesian-speaking societies, these larger decreases in complexity may have led to some of the less politically complex Western Malayo-Polynesian societies in island South-East Asia, such as the Iban. Historically, a number of large states and empires such as Srivijaya, Sumatra and Majahapit, Java are known to have waxed and waned over the last 1,500 years, and archaeological evidence also points to the former presence of other large-scale, complex societies before this time[27]. This potential distinction between the patterns of change for increasing and decreasing complexity may be due to it being easier to break up forms of political organization than to develop practices and institutions that enable units to be joined together in a stable manner[25] (see Supplementary Discussion).

The evolution of complex societies since the end of the last ice age has long been a major topic of investigation and debate. Traditional social evolutionary approaches have been criticized for taking an archaic, ladder-like view of evolution[10]. The phylogenetic approach adopted in this study avoids such pitfalls by making explicit the fact that the societies under investigation are contemporaries and enables us to make inferences about the sequence in which the traits of these societies have evolved. The present analyses move us beyond purely verbal arguments, statistically assessing support for different models of political evolution, and finding models with incremental increases in hierarchical organization[1-6,13] to be the best supported by the data. This study highlights the benefits of applying the same kinds of techniques used to study complex systems in nature to investigate long-term human social and cultural evolution[15,19,21,28,29]. Interestingly, these results indicate that political evolution, like biological evolution, tends to proceed through small steps rather than through major jumps in 'design space'[30].

METHODS SUMMARY

Phylogenetic trees. A sample of 1,000 phylogenetic trees was taken from the Bayesian posterior distribution of trees inferred using lexical data from 400 Austronesian languages described in ref. 19. Trees in this sample were pruned to leave 84 languages for which ethnographic data were available. The consensus tree in Fig. 1 provides a visual summary of the 1,000 trees in our sample and posterior probabilities of each node in the consensus tree are shown in Supplementary Fig. 1.

Ethnographic data. Classification of political organization was based on the variable 'jurisdictional hierarchy beyond local community' from Murdock's Ethnographic Atlas. Eight-four Austronesian-speaking societies in the Ethnographic Atlas could be reliably matched to the languages in our sample of phylogenetic trees. The ethonyms, languages and database codes of the societies included in this study are listed in Supplementary Table 1.

Phylogenetic comparative methods. All analyses were conducted using the 'Multistate' Markov chain algorithm implemented in the program BayesTraits (http://www.evolution.reading.ac.uk/SoftwareMain.html). Support for different models of evolution was assessed using 1) Akaike information criterion (AIC) values based on the ML of the data given each model; and 2) the ratio of posterior to prior odds of each model in the RJMCMC analyses (Supplementary Methods and Supplementary Table 2). Confirmatory analyses were conducted to ensure the results were robust (Supplementary Methods, Supplementary Table 3).

Ancestral states. Forms of political organization in ancestral societies were inferred using the best-fitting model of political evolution (unilinear). For each node in the consensus tree pie charts were created showing the proportion of trees in which the node does not appear and the proportional probability of each form of political organization.

1. Diamond, J. *Guns, Germs and Steel* (Vintage, 1997).
2. Johnson, A. W. & Earle, T. *The Evolution of Human Societies* (Stanford Univ. Press, 2000).
3. Service, E. R. *Primitive Social Organization* (Random House, 1962).
4. Flannery, K. V. The cultural evolution of civilizations. *Annu. Rev. Ecol. Evol. Syst.* **3,** 399–426 (1972).
5. Carneiro, R. L. *Evolutionism in Cultural Anthropology* (Westview, 2003).
6. Spencer, C. S. On the tempo and mode of state formation—neoevolutionism reconsidered. *J. Anthropol. Archaeol.* **9,** 1–30 (1990).
7. Keech McIntosh, S. in *Beyond Chiefdoms: Pathways to Complexity in Africa* (ed. Keech McIntosh, S.) 1–30 (Cambridge Univ. Press, 1999).
8. Campbell, R. B. Toward a networks and boundaries approach to early complex polities: the Late Shang case. *Curr. Anthropol.* **50,** 821–848 (2009).
9. Bondarenko, D. M., Grinin, L. E. & Korotayev, A. V. Alternative pathways of social evolution. *Social Evol. Hist.* **1,** 54–79 (2002).
10. Yoffee, N. in *Archaeological Theory: Who Sets the Agenda?* (eds Yoffee, N. & Sherratt, A.) 60–78 (Cambridge Univ. Press, 1993).
11. Yoffee, N. *Myths of the Archaic State: Evolution of the Earliest Cities, States, and Civilizations* (Cambridge Univ. Press, 2005).
12. Tainter, J. A. Archaeology of overshoot and collapse. *Annu. Rev. Anthropol.* **35,** 59–74 (2006).
13. Marcus, J. The archaeological evidence for social evolution. *Annu. Rev. Anthropol.* **37,** 251–266 (2008).
14. Kirch, P. V. & Green, R. C. *Hawaiki, Ancestral Polynesia: An Essay in Historical Anthropology.* (Cambridge Univ. Press, 2001).
15. Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. B* **276,** 1957–1964 (2009).
16. Hage, P. Reconstructing ancestral Oceanic society. *Asian Perspect.* **38,** 200–227 (1999).
17. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401,** 877–884 (1999).
18. Mace, R. & Holden, C. J. A phylogenetic approach to cultural evolution. *Trends Ecol. Evol.* **20,** 116–121 (2005).
19. Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323,** 479–483 (2009).
20. Kirch, P. V. *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact* (Univ. California Press, 2000).
21. Turchin, P. & Gavrilets, S. Evolution of complex hierarchical societies. *Social Evol. Hist.* **8,** 167–198 (2009).
22. Currie, T. E. & Mace, R. Political complexity predicts the spread of ethnolinguistic groups. *Proc. Natl Acad. Sci. USA* **106,** 7339–7344 (2009).
23. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167,** 808–825 (2006).
24. Bellwood, P. in *Origins, Ancestry and Alliance* (eds Fox, J. J. & Sather, C.) 19–41 (Australian National Univ. Press, 1996).
25. Richerson, P. J. & Boyd, R. in *The Origin of Human Social Institutions* (ed. Runciman, W. G.) 197–234 (Oxford Univ. Press, 2001).
26. Wright, H. T. Early state dynamics as political experiment. *J. Anthropol. Res.* **62,** 305–319 (2006).
27. Glover, I. & Bellwood, P. *Southeast Asia: From Prehistory to History* (eds Glover, I. & Bellwood, P.) (2004).
28. Richerson, P. J., Boyd, R. & Bettinger, R. L. Cultural innovations and demographic change. *Hum. Biol.* **81,** 211–235 (2009).
29. Hazen, R. M. & Eldredge, N. Themes and variations in complex systems. *Elements* **6,** 43–46 (2010).
30. Dennett, D. C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (Simon & Schuster, 1996).

Author Contributions T.E.C. conceived and designed the study in conjunction with R.M. S.J.G. and R.D.G. collected the linguistic data and built the phylogenetic trees. T.E.C. collated the ethnographic data and conducted the phylogenetic comparative analyses. T.E.C., S.J.G., R.D.G., T.H. and R.M. wrote the paper and discussed the results and implications and commented on the manuscript at all stages.

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 this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.E.C. (t.currie@ucl.ac.uk).

ARTICLE

The Ndc80 kinetochore complex forms oligomeric arrays along microtubules

Gregory M. Alushin[1], Vincent H. Ramey[1], Sebastiano Pasqualato[2], David A. Ball[3], Nikolaus Grigorieff[4], Andrea Musacchio[2,5] & Eva Nogales[3,6]

The Ndc80 complex is a key site of regulated kinetochore–microtubule attachment (a process required for cell division), but the molecular mechanism underlying its function remains unknown. Here we present a subnanometre-resolution cryo-electron microscopy reconstruction of the human Ndc80 complex bound to microtubules, sufficient for precise docking of crystal structures of the component proteins. We find that the Ndc80 complex binds the microtubule with a tubulin monomer repeat, recognizing α- and β-tubulin at both intra- and inter-tubulin dimer interfaces in a manner that is sensitive to tubulin conformation. Furthermore, Ndc80 complexes self-associate along protofilaments through interactions mediated by the amino-terminal tail of the NDC80 protein, which is the site of phospho-regulation by Aurora B kinase. The complex's mode of interaction with the microtubule and its oligomerization suggest a mechanism by which Aurora B could regulate the stability of load-bearing kinetochore–microtubule attachments.

The Ndc80 complex is a member of the conserved KMN kinetochore network, which also includes the Knl-1 and Mis12 complexes[1]. The Ndc80 complex is the key site for kinetochore–microtubule attachment[1–3] and a landing pad for the spindle assembly checkpoint[4–6]. Although extensively characterized genetically[7,8] and biochemically[1–3,9,10], the mechanisms by which the Ndc80 complex effects and coordinates these activities remain elusive.

The complex is an elongated, 57-nm heterotetramer composed of NDC80 (also known as HEC1), NUF2, SPC24 and SPC25, each having a globular domain connected to a coiled-coil that mediates dimerization: SPC24 with SPC25, and NDC80 with NUF2 (refs 9–12). Tetramerization via the dimerized coiled-coils[9,10] results in a dumbbell architecture, with the SPC24–SPC25 globular head at one end mediating kinetochore association[1,3], and the NDC80–NUF2 head at the other mediating microtubule binding[1–3,12]. The NDC80–NUF2 coiled-coil contains a break, which makes the rod-like complex highly bendable[13].

Crystallographic structures of both globular head domains have been obtained[2,11], as well as that of a chimaeric version of the human complex containing a minimal coiled-coil, where NDC80 was fused to SPC25 and NUF2 to SPC24 (ref. 12). This 17-nm 'bonsai' complex, which we refer to as Ndc80(bonsai), retained microtubule binding and kinetochore localization. Both NDC80 and NUF2 contain a calponin homology domain (CHD), which is also present in other microtubule binding proteins[14,15]. The unstructured, positively charged 80-amino-acid N-terminal tail of the NDC80 protein is required for high-affinity microtubule binding[2,3,12,16,17], probably by interaction with the acidic carboxy-terminal tails of tubulin (also known as 'E-hooks'). This region of the NDC80 protein, the site of phospho-regulation by the Aurora B kinase[3,12,18], is absent from all crystal structures. The key questions of how the Ndc80 complex binds and remains attached to microtubules during microtubule depolymerization and how this attachment is regulated during mitosis remain unanswered.

Ndc80 binds tubulin with a novel monomeric repeat

We used cryo-electron microscopy to obtain a structure of Ndc80(bonsai) (ref. 12), including the N-terminal tail bound to microtubules. We used an implementation[19] of the iterative helical real space reconstruction method[20], which utilizes reference-free classification to sort helical segments based on symmetry and sample quality (Supplementary Figs 1, 2). Class averages show densities protruding from the microtubule (Supplementary Fig. 1b, top right) with the chevron-like orientation also reported for the *Caenorhabditis elegans* complex[1,21]. Power spectra of class averages show layer lines at $1/40\,\text{Å}^{-1}$ (and subsequent orders), corresponding to the spacing of the tubulin monomer (Supplementary Fig. 1b, bottom right), but lack the $1/80\,\text{Å}^{-1}$ layer line typically observed for microtubule-binding proteins which recognize the tubulin heterodimer[22–24]. This result suggested that the Ndc80 complex binds to each tubulin monomer.

A helical reconstruction of the *C. elegans* complex suggested alternating binding of the two CHDs to strong and weak sites present in each tubulin dimer[21]. In order to determine the arrangement of the Ndc80 complex on the microtubule lattice without imposing any symmetry or averaging, we obtained tomographic reconstructions of negatively stained microtubules saturated with Ndc80(bonsai) (Supplementary Fig. 1c). Single volume slices allow us to visualize individual tubulin monomers and bound Ndc80 complexes, which are found with a 40 Å spacing within the thickness of a single protofilament. Thus, the human Ndc80 complex binds both α- and β-tubulin monomers.

Microtubule site recognition of the Ndc80 complex

Using a microtubule as a starting reference (Supplementary Fig. 3), we obtained a reconstruction of the Ndc80 complex-bound microtubule at 8.6 Å resolution (Fourier shell correlation (FSC) 0.143 criterion, Supplementary Fig. 4a), allowing us to visualize secondary structure (Fig. 1a; Supplementary Figs 3, 4). The asymmetric units of reconstructed

[1]Biophysics Graduate Group, University of California, Berkeley, California 94720, USA. [2]Department of Experimental Oncology, European Institute of Oncology, 20139 Milan, Italy. [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Howard Hughes Medical Institute, Rosenstiel Basic Medical Research Center, Brandeis University, Waltham, Massachusetts 02453, USA. [5]Research Unit of the Italian Institute of Technology at the IFOM-IEO Campus, 20139 Milan, Italy. [6]Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.
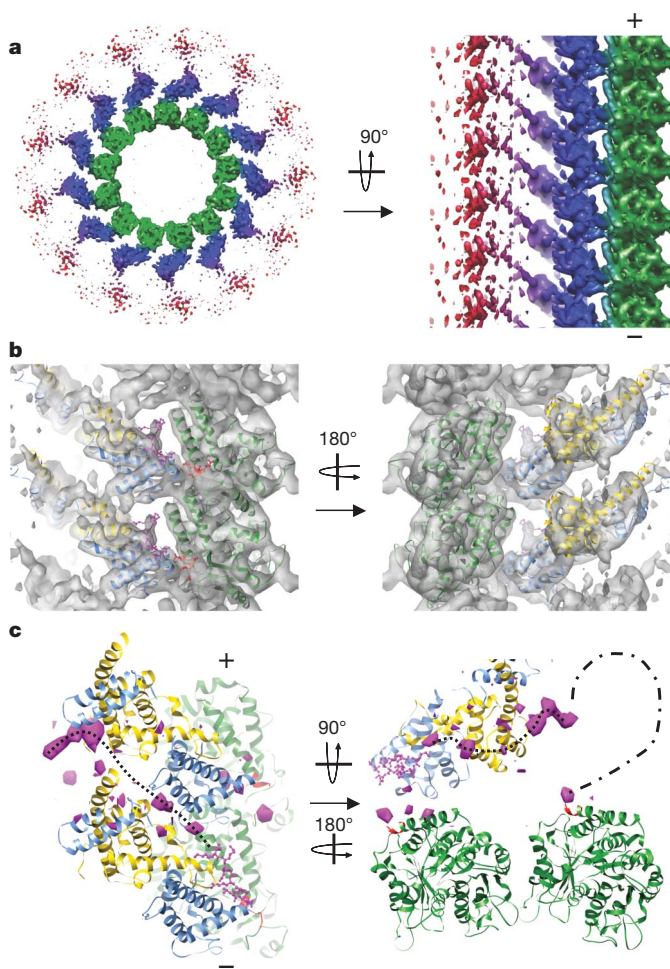
Figure 1 | Structure of the Ndc80 complex–microtubule interface. a, End on (from the plus-end) and side views of the microtubule–Ndc80 complex cryo-electron microscopy reconstruction (tubulin, green; NDC80–NUF2 head, blue; disordered SPC24–SPC25 head, red). **b**, Orthogonal views of docked crystal structures (NDC80, blue; NUF2, gold; tubulin, green). In ball and stick representation are residues adjacent to the absent N terminus of NDC80 in Ndc80(bonsai(ΔN)) (magenta), and ordered residues in tubulin preceding the E-hooks (red). The region of the map occupied by NUF2 (right panel) is further from the Ndc80 complex–microtubule interface and thus is of lower resolution. **c**, Orthogonal views of the positive difference density (magenta) between the cryo-electron microscopy reconstruction and the docked crystal structures, contoured at 2.5σ. Dotted lines show a proposed path for the NDC80 N terminus.

13 and 14 protofilament microtubules were essentially identical (Supplementary Fig. 5). Docking of the crystal structures of Ndc80(bonsai) lacking the N terminus of the NDC80 protein—which we refer to as Ndc80(bonsai(ΔN)); ref. 12—and tubulin[25] allowed us to build a pseudo-atomic model of this interface (Fig. 1b). The excellent correspondence between our map and the crystal structure of Ndc80(bonsai(ΔN)) argues against major rearrangements in the globular domains of the complex upon microtubule binding.

The surface of the Ndc80 complex binding the microtubule lattice is minimal, and includes the C-terminal end of helix G and the short helices B and F in the NDC80 protein. This 'toe' recognizes a site between two tubulin monomers, a 'toe-print' present at both intra- and inter-dimer interfaces. The toe-print is composed of the short helix H11′ in one tubulin subunit and the loop connecting helix H8 with β-strand S7 in the other (Fig. 2a, purple and orange, respectively). Superposition of the crystal structures of α- and β-tubulin and multiple-sequence alignments show the toe-print to be highly conserved between the two tubulin monomers (Fig. 2a, bottom),

consistent with the ability of the Ndc80 complex to recognize them both. The end of the tubulin crystal structure marking the beginning of the disordered E-hook of tubulin (Fig. 2a, red), required for high-affinity Ndc80 complex binding[1,12], is adjacent to the toe-print, from where we propose it extends and acts as a second, distinct binding site.

The interface we observe is largely in agreement with mutagenesis analysis of NDC80 residues important for binding (for example, K123, K166, H176)[12]. Interestingly, the NUF2 CHD is not in contact with the surface of the microtubule, yet mutations in this region also disrupt binding[12]. The positively charged surface of this domain is approximately 15 Å from the E-hook of a laterally adjacent tubulin monomer and thus could be engaged in an electrostatic interaction.

The Ndc80 toe is a tubulin conformation sensor

During microtubule disassembly, protofilaments bend outwards by kinking of tubulin at intra- and inter-dimer interfaces. In spite of its small size, the toe-print spans a proposed hinge point between tubulin monomers[26] and could thus be disrupted by tubulin bending (Fig. 2a). We therefore hypothesized that the NDC80 toe acts as a sensor of the conformational state of tubulin, and that the complex would bind preferentially to straight protofilaments.

Using co-sedimentation assays, we investigated the ability of Ndc80(bonsai) to bind vinblastine-induced tubulin spirals (Supplementary Fig. 6), a polymer analogous to the peels observed at microtubule ends[27]. Because spirals retain the E-hooks, which make a major contribution to binding affinity, we expected the interaction to be reduced, not eliminated. Indeed, we observed a modest but statistically significant ($P < 0.05$, Student's t-test) reduction in affinity for this form of tubulin versus the straight microtubule conformation (Fig. 2b top, Fig. 2c). We next sought to delineate the relative contributions of the toe versus the N terminus of NDC80 to tubulin binding. Ndc80(bonsai(7D)), a phosphomimetic construct of the seven Aurora B phosphorylation sites confirmed *in vitro*[12] (Online Methods, Supplementary Fig. 7), showed significantly reduced affinity towards the straight microtubule conformation and negligible affinity towards the vinblastine-induced, bent conformation; this reduction reflects the increased relative contribution of the toe-print interaction to affinity when the N terminus–E-hook interaction is impaired (Fig. 3b, c). This result is consistent with a bipartite binding mechanism, with the NDC80 N terminus providing affinity without conformation sensitivity, and the toe providing a smaller contribution to affinity that is exquisitely sensitive to tubulin conformation.

We next investigated the effect of the complex's small bias in affinity towards straight tubulin. We polymerized tubulin into dynamic microtubules, then initiated depolymerization by cooling in the presence or absence of Ndc80(bonsai). Cold-stable microtubules and straight tubulin sheets were observed only in the presence of the Ndc80 complex(Fig. 3d, e). Together, these results are consistent with the Ndc80 complex favouring a straight tubulin conformation, and with this specificity being mediated by the toe. Our studies also indicate that the complex has a stabilizing effect on microtubules.

Ndc80 self-assembles on microtubules

In our cryo-electron microscopy reconstruction, connections appear between Ndc80 complexes along protofilaments (Fig. 1a, b), supporting and explaining the observed cooperative binding of the complex to microtubules[12,17]. We collected tomograms of negatively stained microtubules with non-saturating amounts of Ndc80 complex to test for the presence of Ndc80–Ndc80 complex interactions under more physiological conditions. We found that the complex forms clusters along protofilaments that retain tubulin monomer spacing (Fig. 3a, e; Supplementary Fig. 8). Decoration was heterogeneous, with some microtubules approaching saturation and others almost undecorated (Fig. 3a, Supplementary Fig. 9a); this is a direct manifestation of cooperativity. We do not observe ordered self-association of the complex in the absence of microtubules (Supplementary Fig. 10). Analysis
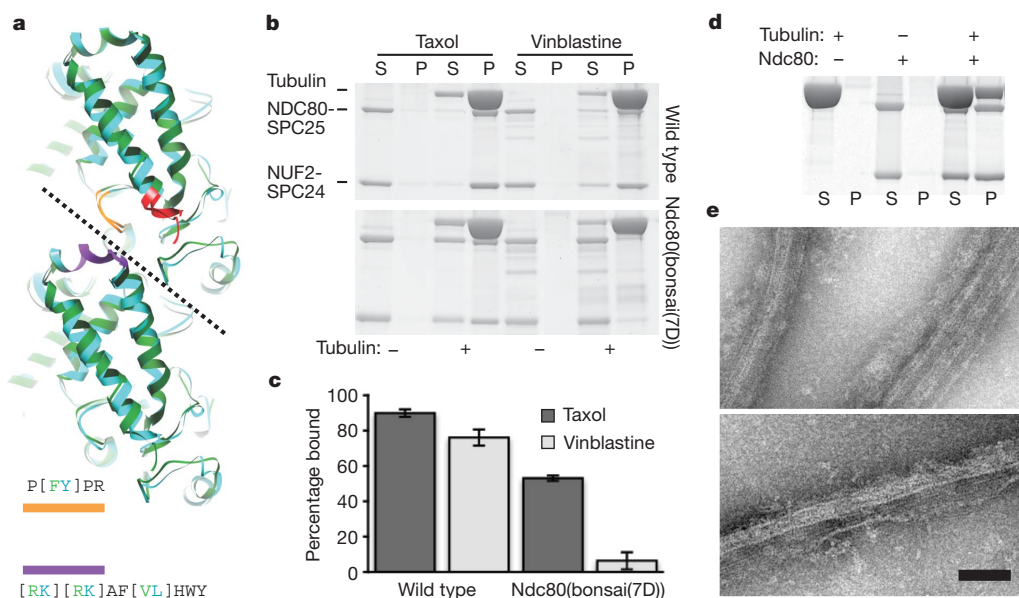
Figure 2 | The NDC80 toe-print is a tubulin conformation sensor.
a, Superposition of the tubulin intra-dimer and inter-dimer interfaces, viewed from the outside of the microtubule (α-tubulin, green; β-tubulin, blue; ordered residues adjacent to the E-hooks, red; conserved toe-print segments, purple and orange). Consensus sequences are indicated, with deviations between the monomers in parenthesis. **b**, SDS–PAGE of co-sedimentation assays with straight (taxol) and curved (vinblastine) tubulin polymers with the indicated Ndc80(bonsai) constructs. S, supernatant; P, pellet. [Tubulin monomer] = 6 μM, [Ndc80(bonsai)] = 0.5 μM. **c**, Quantification of **b**. Error bars ±s.d., $n = 3$. **d**, SDS–PAGE of co-sedimentation assays after cold-induced depolymerization of dynamic microtubules in the presence and absence of Ndc80(bonsai). **e**, Negative-stain electron microscopy of Ndc80 complex-induced cold-stable microtubules and straight tubulin sheets. Scale bar, 50 nm.

of the pooled distributions of cluster number and size indicates that it is most probable for a complex to be found in a cluster of four molecules or in large clusters of more than ten molecules (Fig. 3i, Supplementary Fig. 9b). Given the reported number (6–8) of Ndc80 complexes per microtubule at the kinetochore, the larger clusters are probably not physiologically relevant[28,29]. Our data are therefore consistent with the formation of two or three clusters per kinetochore microtubule.

Analysis of our docking results shows densities not accounted for by the crystal structures. We attempted to visualize these regions by calculating a difference map between the experimental density map and the docked crystal structures (Fig. 1c; Supplementary Fig. 11). The map shows positive density peaks adjacent to the C-terminal end of the tubulin crystal structure, corresponding to the extension of the H12 helices in each monomer. Stabilization of these residues upon Ndc80 complex binding is consistent with this region of tubulin contributing to affinity[1,12]. Small extra density is also observed by the toe-print. The largest difference density is between the globular domain of NUF2 and the αH–αI helical hairpin of NDC80, with a series of small peaks running along the groove between two longitudinally adjacent complexes, connecting this region with the N-terminal end of the Ndc80(bonsai(ΔN)) crystal structure.

The most parsimonious explanation of this path of density is that it corresponds to the N terminus of NDC80. We propose that the N terminus mediates weak contacts between the globular heads of NDC80 and NUF2 before making a strong contact, corresponding to the major density peak. It could then bind the C terminus of tubulin (necessary for high affinity), probably in an adjacent protofilament. The total mass (80 residues) of the N terminus is not accounted for by these densities, which probably correspond to ordered points of contact formed by this mostly unstructured polypeptide.

The N terminus of Ndc80 mediates regulated self-assembly

In agreement with this proposal, we found that Ndc80(bonsai(ΔN)) was deficient in cluster formation (Fig. 3b, f, i). The presence of some clusters of two complexes suggests that the N terminus is the major

but not sole molecular determinant of clustering. We found that Ndc80(bonsai(7D)) showed a clustering phenotype similar to Ndc80(bonsai(ΔN)), with a cluster size of one molecule being the most probable, despite a higher average surface density, and a few clusters of moderate size (Fig. 3c, g, i). We therefore conclude that the N terminus of NDC80 mediates Ndc80–Ndc80 complex interactions, as previously speculated[12], and that phosphorylation at Aurora B sites is capable of modulating this binding.

Next, we dissected the differential contributions of the N terminus to affinity and cooperativity. Subtilisin cleavage of the E-hooks from tubulin significantly reduces the affinity of the Ndc80 complex for microtubules[1,12,17]. As expected, after this treatment we found a significantly lower average surface density of Ndc80 complexes compared to uncleaved microtubules (Fig. 3d, h, i). If the N terminus of NDC80 were only contributing indirectly to cluster formation by increasing microtubule-binding affinity, the cluster size distribution for this condition should be similar to the N terminus mutants. Instead, we observe a most probable cluster size of three molecules, despite an average surface density lower than that of Ndc80(bonsai(7D)) (Fig. 3i). This observation strongly supports our hypothesis that the NDC80 N terminus forms specific intermolecular interactions between complexes on the microtubule surface. It also supports cluster formation occurring after microtubule binding, as the average cluster size on subtilisin-cleaved microtubules is smaller than on uncleaved microtubules.

Role of Ndc80 clusters in mitosis

It was recently demonstrated that Aurora B's activity is governed by spatial localization rather than directly by tension[30], confirming previous proposals[31,32]. On the basis of this finding, we suggest a scheme that starts with phosphorylated Ndc80 complexes in an unattached kinetochore (Fig. 4a)[33]. When the kinetochore encounters a microtubule, individual NDC80-NUF2 heads would initially bind the microtubule through a low-affinity interaction. The bound heads would then escape the Aurora B phosphorylation zone, provided that intra-kinetochore stretching increased as microtubule binding by additional heads allowed the site to come under tension[34]. This could
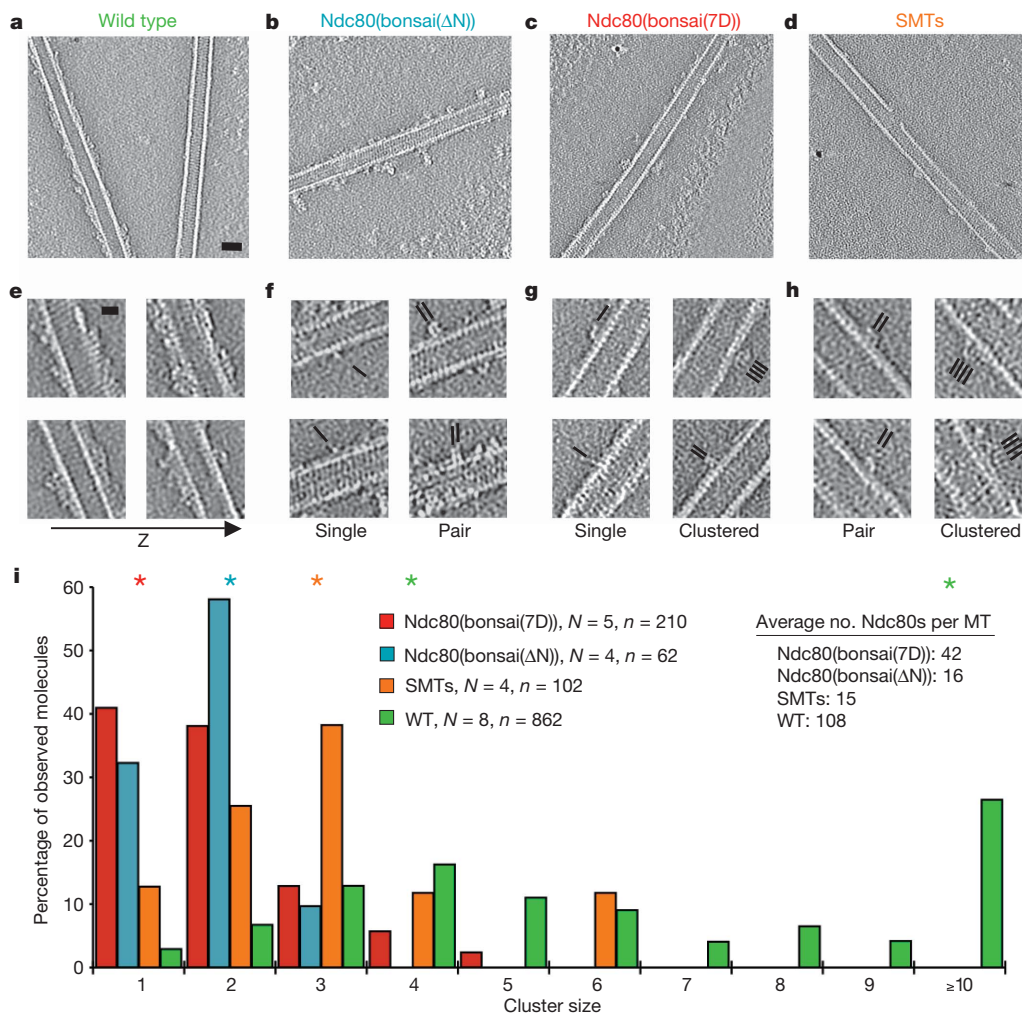
Figure 3 | Cluster formation requires the N terminus of the NDC80 protein. a–d, Central slices of tomograms of microtubule-bound Ndc80(bonsai) constructs under subsaturating conditions. Ndc80(bonsai):tubulin monomer ratio was 1:2 for wild-type and subtilisin-cleaved microtubules (SMTs) and 2:1 for NDC80 N terminus mutants. Scale bar, 25 nm; black dots, gold fiducials. **e**, Serial slices of the wild-type reconstruction 4.5 nm apart show cooperative binding and cluster formation only along (not between) protofilaments. Scale bar, 10 nm. **f–h**, Selected views of clusters in Ndc80(bonsai(ΔN)), Ndc80(bonsai(7D)), and SMT reconstructions, respectively. Black lines indicate position and orientation of Ndc80 complexes. **i**, Quantification of cluster size populations. *N*, number of reconstructed microtubule segments; *n*, total number of Ndc80 complexes observed; MT, microtubule; WT, wild type. Asterisks, the most probable cluster size for each of the populations (wild-type has two peaks). See Supplementary Table 1 for pair-wise statistical comparison of these distributions.

occur via straightening of the flexible Ndc80 complex itself[13] (Fig. 4a) or by the stretching of a compliant link in the inner kinetochore[35]. As the phosphorylation zone is cleared, phosphatase activity (possibly KMN-network-localized Protein phosphatase 1; ref. 36) would dephosphorylate the N terminus of NDC80, resulting in the formation of high-affinity Ndc80 complex clusters.

Once assembled, the cluster arrangement is consistent with the complex maintaining load-bearing attachments via biased diffusion[37] (Fig. 4b), as recently proposed on the basis of functional studies[38]. An alternative model, based on shuffling, is also consistent with our data (Supplementary Discussion, Supplementary Fig. 12). Our results support a biased-diffusion model with significant differences from previous studies that presumed either a continuous sleeve[37] or sites uncoupled from the subunit spacing of the microtubule[38]. We propose that Ndc80 complexes oligomerize into linear arrays, resulting in a set of identical, coupled binding sites along protofilaments that match the longitudinal subunit spacing of the microtubule. This arrangement would arise only after the kinetochore comes under tension. Consistent with this hypothesis, outer kinetochore rearrangement on microtubule interaction has been observed *in vivo*[39].

A cluster could diffuse on the microtubule lattice, but its diffusion would become biased at a microtubule end. Thus, a shrinking microtubule would pull the attached chromatid polewards. Cluster diffusion should be facilitated by the 40 Å rather than 80 Å spacing of the Ndc80 complexes, with a shorter distance to the transition state between binding sites during diffusion. Thus, monomer binding may have evolved to enable a biased-diffusion mechanism with sufficiently rapid kinetics (40 Å is the smallest step a microtubule-binding element with a specific footprint can take longitudinally along the microtubule). Our finding that the NDC80 toe serves as a tubulin conformation sensor suggests that microtubule subunit loss is not required to bias diffusion: the curving of protofilaments at a depolymerizing end would suffice[40]. Depolymerizing ends have been observed at metaphase kinetochores *in vivo*, apparently stabilized by attached filaments of unknown identity (Fig. 4b)[41].

Whereas the KMN network and the Ndc80 complex are conserved from yeast to humans, the mechanism of microtubule attachment may not be. We propose that the fungal Ndc80 complex has diverged to act as a coupler to the Dam1 complex, and our structural results, coupled with conservation analysis, suggest a possible binding site for
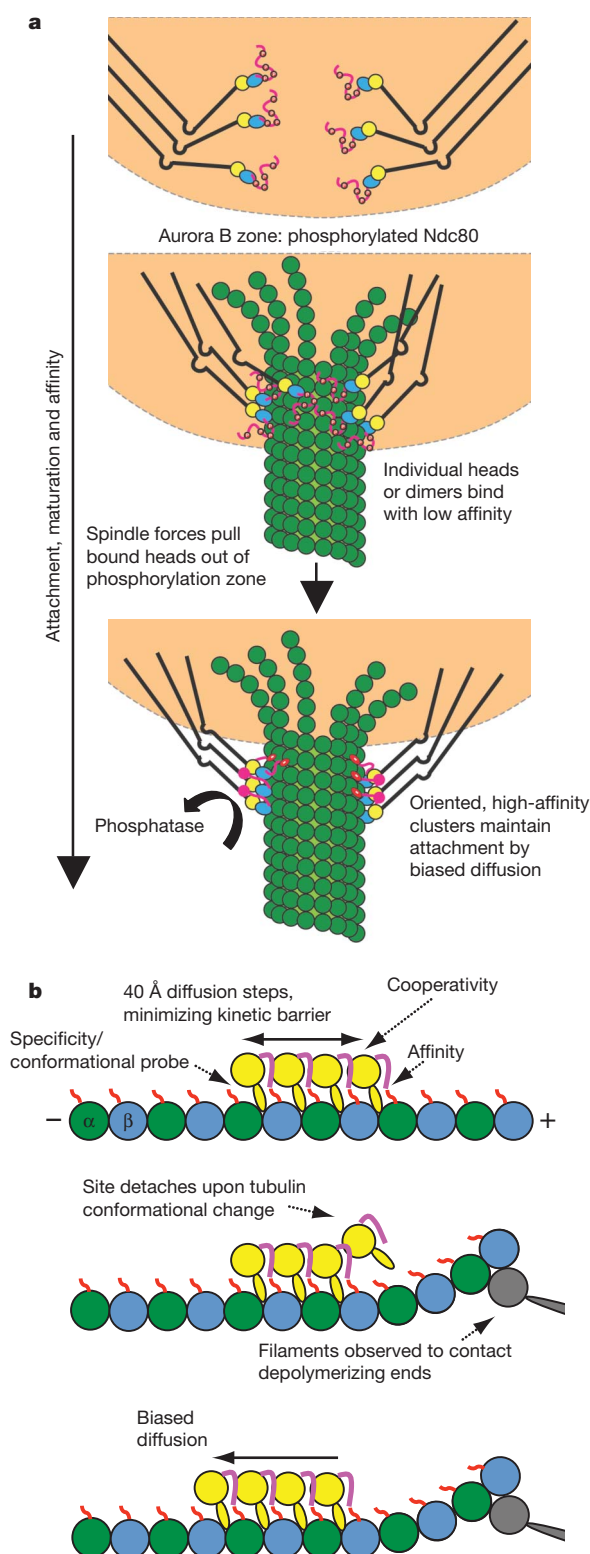
Figure 4 | Proposed models of attachment maturation and biased diffusion. a, Cartoon illustrating the phospho-regulated formation of Ndc80 complex clusters *in vivo* concomitant with stable kinetochore–microtubule attachment (colours as Fig. 1b). In this diagram, the assembly of the cluster biased diffusion machinery is regulated by the spatial localization of Aurora B and a counterbalancing phosphatase rather than directly by tension. **b,** Diagrams of the proposed biased diffusion process for coupling chromosome movement to microtubule depolymerization via the Ndc80 complex. Colours as in Fig. 1b, except that the NDC80–NUF2 head is shown in gold and unidentified filaments are shown in grey.

the Dam1 complex on microtubule-bound Ndc80 (Supplementary Discussion, Supplementary Fig. 13).

We have used cryo-electron microscopy and molecular docking to define the interface at the heart of metazoan kinetochore–microtubule attachment. Our studies also demonstrate the microtubule-mediated oligomerization of the Ndc80 complex, which directly involves the Aurora B-regulated N terminus of the NDC80 protein.

METHODS SUMMARY

Cryo-electron microscopy and helical reconstruction. Taxol-stabilized microtubules were decorated with Ndc80(bonsai) after application to glow-discharged C-flat grids (Protochips), then plunge-frozen in ethane slush. Images were collected on Kodak SO-163 film with a Tecnai F20 electron microscope operating at 200 kV at a nominal magnification of 50,000× with a dose of 15 electrons per $Å^2$. Micrographs were digitized with a Nikon Super CoolScan 8000 scanner with a step size of 6.35 μm. Image processing and projection-matching alignment were carried out using programs from the EMAN, IMAGIC and SPIDER packages, and final refinement and CTF correction was performed with a version of FREALIGN adapted for this work to implement helical symmetry. Visualization and molecular docking was performed with UCSF Chimera. Amplitude-weighted difference maps were calculated using the program DIFFMAP (http://emlab.rose2.brandeis.edu/software). References for all image analysis and visualization software can be found in the Online Methods section.

Tubulin co-sedimentation assays. Tubulin polymerized with or without conformation-stabilizing drugs was mixed with Ndc80(bonsai), then layered on to a 50% glycerol cushion supplemented with additives and pelleted by ultracentrifugation in a Beckman TLA100 rotor at 312,530g for 10 min (ref. 12). Supernatant and pellet fractions were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE).

Electron tomography. Samples were prepared on C-flat grids augmented with a continuous carbon layer, and stained with uranyl formate. Tilt series were collected on 2k × 2k CCD cameras using a JEOL 3100 microscope operating at 300 kV with the SerialEM package, or a Phillips CM200 microscope operating at 200 kV with Digital Micrograph (Gatan). Processing was carried out with programs from the EMAN, SPIDER and IMOD software packages. References for all image analysis software can be found in the Online Methods section.

1. Cheeseman, I. M., Chappie, J. S., Wilson-Kubalek, E. M. & Desai, A. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell* **127,** 983–997 (2006).
2. Wei, R. R., Al-Bassam, J. & Harrison, S. C. The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nature Struct. Mol. Biol.* **14,** 54–59 (2007).
3. DeLuca, J. G. *et al.* Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. *Cell* **127,** 969–982 (2006).
4. Martin-Lluesma, S., Stucke, V. M. & Nigg, E. A. Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. *Science* **297,** 2267–2270 (2002).
5. DeLuca, J. G. *et al.* Nuf2 and Hec1 are required for retention of the checkpoint proteins Mad1 and Mad2 to kinetochores. *Curr. Biol.* **13,** 2103–2109 (2003).
6. Kemmler, S. *et al.* Mimicking Ndc80 phosphorylation triggers spindle assembly checkpoint signalling. *EMBO J.* **28,** 1099–1110 (2009).
7. Chen, Y., Riley, D. J., Chen, P. L. & Lee, W. H. HEC, a novel nuclear protein rich in leucine heptad repeats specifically involved in mitosis. *Mol. Cell. Biol.* **17,** 6049–6056 (1997).
8. Wigge, P. A. *et al.* Analysis of the Saccharomyces spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141,** 967–977 (1998).
9. Wei, R. R., Sorger, P. K. & Harrison, S. C. Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proc. Natl Acad. Sci. USA* **102,** 5363–5367 (2005).
10. Ciferri, C. *et al.* Architecture of the human ndc80-hec1 complex, a critical constituent of the outer kinetochore. *J. Biol. Chem.* **280,** 29088–29095 (2005).
11. Wei, R. R. *et al.* Structure of a central component of the yeast kinetochore: the Spc24p/Spc25p globular domain. *Structure* **14,** 1003–1009 (2006).
12. Ciferri, C. *et al.* Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell* **133,** 427–439 (2008).
13. Wang, H. W. *et al.* Architecture and flexibility of the yeast Ndc80 kinetochore complex. *J. Mol. Biol.* **383,** 894–903 (2008).
14. Hayashi, I., Wilde, A., Mal, T. K. & Ikura, M. Structural basis for the activation of microtubule assembly by the EB1 and p150Glued complex. *Mol. Cell* **19,** 449–460 (2005).
15. Slep, K. C. & Vale, R. D. Structural basis of microtubule plus end tracking by XMAP215, CLIP-170, and EB1. *Mol. Cell* **27,** 976–991 (2007).

16. Guimaraes, G. J., Dong, Y., McEwen, B. F. & Deluca, J. G. Kinetochore-microtubule attachment relies on the disordered N-terminal tail domain of Hec1. *Curr. Biol.* **18,** 1778–1784 (2008).

17. Miller, S. A., Johnson, M. L. & Stukenberg, P. T. Kinetochore attachments require an interaction between unstructured tails on microtubules and Ndc80(Hec1). *Curr. Biol.* **18,** 1785–1791 (2008).

18. Cheeseman, I. M. *et al.* Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell* **111,** 163–172 (2002).

19. Ramey, V. H., Wang, H. W. & Nogales, E. Ab initio reconstruction of helical samples with heterogeneity, disorder and coexisting symmetries. *J. Struct. Biol.* **167,** 97–105 (2009).

20. Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157,** 83–94 (2007).

21. Wilson-Kubalek, E. M., Cheeseman, I. M., Yoshioka, C., Desai, A. & Milligan, R. A. Orientation and structure of the Ndc80 complex on the microtubule lattice. *J. Cell Biol.* **182,** 1055–1061 (2008).

22. Mizuno, N., Narita, A., Kon, T., Sutoh, K. & Kikkawa, M. Three-dimensional structure of cytoplasmic dynein bound to microtubules. *Proc. Natl Acad. Sci. USA* **104,** 20832–20837 (2007).

23. Hoenger, A. & Gross, H. Structural investigations into microtubule-MAP complexes. *Methods Cell Biol.* **84,** 425–444 (2008).

24. des Georges, A. *et al.* Mal3, the *Schizosaccharomyces pombe* homolog of EB1, changes the microtubule lattice. *Nature Struct. Mol. Biol.* **15,** 1102–1108 (2008).

25. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of alpha beta-tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313,** 1045–1057 (2001).

26. Wang, H.-W. & Nogales, E. Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature* **435,** 911–915 (2005).

27. Wilson, L., Jordan, M. A., Morse, A. & Margolis, R. L. Interaction of vinblastine with steady-state microtubules *in vitro. J. Mol. Biol.* **159,** 125–149 (1982).

28. Joglekar, A. P., Bouck, D. C., Molk, J. N., Bloom, K. S. & Salmon, E. D. Molecular architecture of a kinetochore-microtubule attachment site. *Nature Cell Biol.* **8,** 581–585 (2006).

29. Joglekar, A. P. *et al.* Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. *J. Cell Biol.* **181,** 587–594 (2008).

30. Liu, D., Vader, G., Vromans, M. J., Lampson, M. A. & Lens, S. M. Sensing chromosome bi-orientation by spatial separation of aurora B kinase from kinetochore substrates. *Science* **323,** 1350–1353 (2009).

31. Tanaka, T. U. *et al.* Evidence that the Ipl1-Sli15 (Aurora kinase-INCENP) complex promotes chromosome bi-orientation by altering kinetochore-spindle pole connections. *Cell* **108,** 317–329 (2002).

32. Andrews, P. D. *et al.* Aurora B regulates MCAK at the mitotic centromere. *Dev. Cell* **6,** 253–268 (2004).

33. Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *EMBO J.* **28,** 2511–2531 (2009).

34. Maresca, T. J. & Salmon, E. D. Intrakinetochore stretch is associated with changes in kinetochore phosphorylation and spindle assembly checkpoint activity. *J. Cell Biol.* **184,** 373–381 (2009).

35. Wan, X. *et al.* Protein architecture of the human kinetochore microtubule attachment site. *Cell* **137,** 672–684 (2009).

36. Liu, D. *et al.* Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. *J. Cell Biol.* **188,** 809–820 (2010).

37. Hill, T. L. Theoretical problems related to the attachment of microtubules to kinetochores. *Proc. Natl Acad. Sci. USA* **82,** 4404–4408 (1985).

38. Powers, A. F. *et al.* The Ndc80 kinetochore complex forms load-bearing attachments to dynamic microtubule tips via biased diffusion. *Cell* **136,** 865–875 (2009).

39. Dong, Y., Vanden Beldt, K. J., Meng, X., Khodjakov, A. & McEwen, B. F. The outer plate in vertebrate kinetochores is a flexible network with multiple microtubule interactions. *Nature Cell Biol.* **9,** 516–522 (2007).

40. Lombillo, V. A., Stewart, R. J. & McIntosh, J. R. Minus-end-directed motion of kinesin-coated microspheres driven by microtubule depolymerization. *Nature* **373,** 161–164 (1995).

41. McIntosh, J. R. *et al.* Fibrils connect microtubule tips with kinetochores: a mechanism to couple tubulin dynamics to chromosome motion. *Cell* **135,** 322–333 (2008).

METHODS

Mutagenesis. Ndc80(bonsai) point mutations were generated using standard procedures for site-directed mutagenesis. The Ndc80(bonsai(7D)) mutant was generated as follows: Synonymous mutations were performed that destroyed the StuI restriction site at base pair (bp) 81–86 of the NDC80–SPC25 protein coding sequence and subsequently reintroduced it at bp 250–255. An NdeI restriction site was also created 5′ to the beginning of the coding sequence. This construct thus features a swappable NDC80 N terminus cassette. A sequence coding for the seven *in vitro* verified[12] Aurora B phosphorylation sites mutated to aspartic acid (S4D, S8D, S15D, S44D, S55D, S62D, S69D) was synthesized (GENEART) and cloned into the swappable N terminus construct using standard protocols.

Biochemical sample preparation. Ndc80(bonsai) complex was purified as described[12]. 10 mg ml^{-1} bovine brain tubulin (Cytoskeleton) was polymerized in CB1 buffer (80 mM PIPES pH 6.8, 1 mM EGTA, 1 mM MgCl$_2$, 1 mM GTP, 10% glycerol) for 15 min at 37 °C before addition of 160 μM taxol, followed by 30–60 min of further incubation. We have found this procedure primarily produces 13 and 14 protofilament microtubules. Microtubules were pelleted at 17,000*g* in a tabletop microcentrifuge for 20 min, then resuspended in room temperature EM buffer (80 mM PIPES pH 6.8, 1 mM EGTA, 1 mM MgCl$_2$, 1 mM DTT, 0.05% Nonidet P-40) supplemented with 160 μM taxol. Tubulin concentration was assayed by A_{280} after depolymerization on ice with 50 mM CaCl$_2$.

Ndc80(bonsai) at 2 mg ml^{-1} was rapidly thawed before storage on ice, diluted 1:1 with EM buffer, then desalted into EM buffer using a Zeba spin desalting column (Pierce). The removal of salt resulted in a rapid precipitation of the sample, which was clarified by ultracentrifugation in a Beckman TLA100 rotor for 20 min at 61,734*g* at 4 °C. The sample was then warmed to 25 °C for approximately 15 min before electron microscopy grid preparation, which resulted in a second round of precipitation/self-association. Analysis of the sample at this stage by negative-stain electron microscopy did not reveal any ordered assemblies (Supplementary Fig. 10). The sample was once again clarified for 3 min at 17,000*g* in a tabletop microcentrifuge. Protein concentration was estimated with the Coomassie Plus protein assay reagent (Pierce) using bovine serum albumin (BSA) as a reference. It is notable that the amount of precipitation was dependent on the state of the NDC80 N terminus: the Ndc80(bonsai(ΔN)) and Ndc80(bonsai(7D)) constructs were significantly more stable under low-salt conditions.

Cryo-sample preparation. C-flat grids (Protochips) were glow-discharged using an Edwards Carbon Evaporator. Taxol-stabilized microtubules were diluted to 0.25 mg ml^{-1} in EM buffer supplemented with 20 μM taxol, and 2 μl was applied to the grid in the humidity chamber of a Vitrobot (Maastricht Instruments). After 1 min, 4 μl of ~0.7 mg ml^{-1} Ndc80(bonsai) was added, corresponding to a molar ratio of ~2:1 Ndc80 complex:tubulin monomer, and incubated for 1 min. The grid was then briefly blotted before a second 4 μl addition of Ndc80 complex and 1 min incubation. 2 μl was then removed from the grid, which was then blotted for 2 s and plunged into ethane slush. This protocol is essentially similar to that used for visualization of the *C. elegans* complex[21], with minor modification.

Cryo-electron microscopy. 100 micrographs were collected on Kodak SO-163 film with a Tecnai F20 electron microscope operating at 200 kV at a nominal magnification of 50,000× with a dose of 15 electrons per Å^2, between 1.2 μm and 2.2 μm underfocus. Micrographs were digitized with a Nikon Super CoolScan 8000 scanner with a step size of 6.35 μm. After digitization, the power spectra of carbon present in each image was examined, and images lacking Thon rings to at least 8 Å were excluded.

Image processing and IHRSR. We carried out helical processing essentially as described[19], with some modifications. Unless otherwise indicated, all processing steps were implemented with the SPIDER package[42]. The pixel size of the digitized micrographs was calibrated to be 1.24 Å by examination of the power spectrum of samples of tobacco mosaic virus (TMV) imaged under identical conditions. CTF parameters were estimated for each image using the program CTFFIND3[43]. Images were divided into three approximate defocus groups (1.2 μm, 1.7 μm, 2.2 μm). Segments 768 pixels long were extracted from the micrographs with the helix option in BOXER[44], using 90% overlap[20]. Pixel intensities were normalized, and large-scale gradients in intensity due to variations in ice thickness were subtracted. A two-dimensional Wiener filter was then applied to each segment, with the intention of aiding the detection of highly ordered segments, which should diffract to high resolution. Subsequent experience has demonstrated that this step is not necessary for robust classification based on helical quality, and may in fact be detrimental in cases with a low signal-to-noise ratio in the low spatial frequency regime, as amplification of high-resolution spatial frequency components in raw images drives alignment based on noise rather than signal. The segments were then masked such that each image contained approximately 2 turns of the 1-start helix, corresponding to 78 asymmetric units, and decimated threefold to a pixel size of 3.72 Å.

We then subjected the data to reference-free two-dimensional classification as described[19]. This method successfully sorts segments on the basis of protofilament number and helical quality, as is apparent in Supplementary Fig. 2. As microtubules can incorporate varying numbers of protofilaments *in vitro*[45], and segments vary in their quality, this method is used as an alternative to assessing each individual microtubule by manual inspection of Moiré patterns and power spectra. Out of an initial set of 10,253 segments, only 1,475 were members of classes that corresponded to well-ordered 13 protofilament microtubules and were selected for further processing.

Segments from defocus groups were combined before IHRSR with SPIDER. Using a naked microtubule as a reference, the reconstruction converged after 10 rounds of refinement. Particles with cross-correlation scores less than one standard deviation below the mean were excluded, corresponding to approximately 30% of the data. The final refined helical parameters correspond to a rise of 42.7 Å per tubulin monomer. This is slightly greater than the reported value of 41.7 Å for taxol-stabilized microtubules[46], but could result from an error in our pixel size estimation, as TMV was not observed *in situ*. The full three-start symmetry of the microtubule was applied after refinement was complete by real-space averaging. The resolution of the reconstruction at this stage was limited to approximately 20 Å ('bronze' reconstruction, Supplementary Fig. 3), as filtering the volume at higher resolution did not reveal any additional features.

This limit in resolution was found to be at least partially due to limited pixel sampling. With no further alignment of the particles, simply reducing the decimation factor of the original CTF-corrected data from threefold to twofold (2.48 Å per pixel) resulted in a discrete jump in resolution, to approximately 12–15 Å resolution ('silver' reconstruction, Supplementary Fig. 3). The mask on each particle was removed at this stage, increasing the number of asymmetric units incorporated into the reconstruction, as not all segments incorporated into the reconstruction were from adjacent positions in microtubule filaments. Approximately 50,000 asymmetric units were incorporated at this stage, and secondary structure elements began to be visualized when a B-factor of −450 Å^2 was applied with the program bfactor (http://emlab.rose2.brandeis.edu/software) using the cosine edge mask option with a radius of 8.5 Å. This model was used as the input for final refinement in Frealign.

Final refinement with Frealign. We used the computer program Frealign[47] to refine our reconstruction. Frealign was originally designed to work with single particle images. To work with helical particles, we implemented a helical symmetry operator (symmetry operator H in Frealign) and followed a reconstruction algorithm described previously[48]. Masking of the helical sections is done using a rectangular mask that is aligned with the helical axis of the particle in each section. Using the helical symmetry operator, a set of Euler angles and shifts producing equivalent views for each section can be calculated. Using this set of alignment parameters, each section is inserted multiple times into the reconstruction. The number of insertions is matched to the degree of overlap between helical sections such that it corresponds to the number of symmetry-related subunits contained in the non-overlapping part of each section. In the case of the Ndc80 complex-decorated microtubules, the unique, non-overlapping area of the filament was chosen to contain approximately one helical turn of 13 subunits. In addition to this, the 3-start helical symmetry of the 13 protofilament microtubule increased the number of unique subunits per segment threefold, to 39. Therefore, in the present case, each helical segment is inserted into the three-dimensional reconstruction 39 times using Euler angles and shifts that represent equivalent views of the particle. The final reconstruction is masked using a cylindrical mask with user-defined radius, and the helical symmetry is imposed in real space (option BEAUTIFY in Frealign) to remove the small density gradient along the helical axis due to the reconstruction algorithm[48].

For the refinement of Euler angles and shifts for each helical section, the reference projections from the helical reference are masked with a soft-edged mask to reset the image density to background level in a 10% margin near the edges of the image. This is necessary to avoid truncation of reference projections at the edges when applying the small shifts typically observed during alignment. Furthermore, the refinement is carried out using the weighted correlation coefficient[49], but without the use of the absolute value. For helical structures, the low-resolution correlation terms are essentially insensitive to shifts along the helical axis and, therefore, if the absolute value is used, the high-resolution layer lines may be rendered out of phase during alignment. Finally, the Euler angles describing in-plane and out-of-plane alignments of each section are restrained as previously described[50], but using separate statistics for each segment. No restraints were used for the shifts. The final reconstruction was scaled to the same amplitude profile as the input reconstruction, and is shown in gold in Supplementary Fig. 3.

Difference map calculation. The experimental map was segmented, and an appropriate number of Ndc80(bonsai(ΔN)) and tubulin crystal structures were docked into the map to approximately account for the total mass with UCSF

Chimera[51]. These crystal structures were converted into SPIDER volume format using the CP FROM PDB command, then filtered to 8 Å resolution. Although this procedure retained the orientation of the crystal structures relative to each other, this volume had to be re-docked into the experimental map, which was accomplished using a cross-correlation search with SPIDER. The amplitude-weighted difference map was calculated using the program DIFFMAP (http://emlab.rose2.brandeis.edu/diffmap). Amplitude weighting was critical for a meaningful comparison, as the amplitudes of the volume derived from the crystal structures were down-weighted several hundred fold in all resolution shells.

Tubulin co-sedimentation assays. For biochemical assays Ndc80(bonsai) proteins were desalted into binding buffer (80 mM PIPES pH 6.8, 1 mM EGTA, 1 mM MgCl$_2$, 1 mM DTT, 5% sucrose). We found that this buffer increased the solubility of the protein in the absence of salt while allowing for accurate determination of protein concentration by A_{280}. The phenyl group of Nonidet P-40 results in significant absorbance of EM buffer at 280 nm, which is refractory to reproducible protein concentration measurement by this method. The presence of sucrose significantly reduces contrast in the electron microscope, and thus binding buffer was not used for any of the imaging experiments shown. In control imaging experiments, we have found no differences in Ndc80 complex–microtubule interactions or tubulin polymer behaviour between the two buffers (data not shown).

To generate vinblastine spirals, we diluted tubulin stored in CB1 buffer to 3 mg ml^{-1} in binding buffer supplemented with 1 mM vinblastine sulphate (Sigma). After two hours incubation at 25 °C, robust formation of spiral aggregates was observed (Supplementary Fig. 6).

Taxol microtubules or vinblastine spirals (6 μM tubulin monomer) and Ndc80(bonsai) proteins (0.5 μM) were mixed in a 50 μl reaction volume and incubated at 25 °C for 15 min in the presence of 20 μM taxol or 1 mM vinblastine. The binding reactions were layered on to a 100 μl 50% glycerol cushion containing buffer components and the appropriate drug and polymers were pelleted by ultracentrifugation at 312,530g. in a Beckman TLA 100 rotor for 10 min at 25 °C, essentially as described[12]. Supernatant and pellet fractions were collected and precipitated in 90% ethanol at −20 °C for 16 h before analysis by SDS–PAGE. Gels were stained with Flamingo fluorescent gel stain (Bio-Rad), and imaged with a Typhoon Trio (GE Healthcare). Apparent slight degradation of the Ndc80 complex was observed in the presence of 1 mM vinblastine sulphate (Fig. 2b), but this did not cause spurious pelleting of the complex in the absence of tubulin. Quantification was performed with ImageJ[52].

To test for cold stabilization, 10 mg ml^{-1} tubulin in CB1 buffer was polymerized for 45 min at 37 °C. A binding reaction containing Ndc80(bonsai) in binding buffer was set up on ice, and then heated to 37 °C for 1 min. Dynamic microtubules were then added to this reaction for a final concentration of 20 μM tubulin and 3.3 μM NDC80 and incubated at 37 °C for 10 min. The reaction was then shifted to ice for 30 min, and subsequently analysed by pelleting assay as above, except ultracentrifugation was performed at 4 °C.

To analyse the outcomes of the described experiments by negative stain electron microscopy, we repeated them substituting EM buffer for binding buffer. Samples were prepared on continuous carbon grids, stained with 2% uranyl acetate, and imaged on a 1k × 1k CCD camera with a Tecnai T12 electron microscope operating at 120 kV between approximately 1 μm and 2 μm underfocus.

Negative stain sample preparation for tomography. Samples were prepared as described for cryo-electron microscopy, except that the grids were augmented with a layer of thin carbon and treated with gold fiducial markers (British BioCell International) before sample application. To achieve sub-saturating binding, wild-type Ndc80(bonsai) was diluted to 0.15 mg ml^{-1} and only a single application was performed, corresponding to a 1:2 Ndc80 complex:tubulin monomer ratio. To generate SMTs, taxol-stabilized microtubules were digested with 1.5 μM subtilisin A (Calbiochem) in EM buffer for 30 min at 37 °C. The reaction was stopped with 2 mM PMSF, and the SMTs were pelleted and resuspended in EM buffer before sample preparation.

To achieve saturation in the case of wild-type and to observe binding in the case of Ndc80(bonsai(ΔN)) and Ndc80(bonsai(7D)), double-application of 0.6–0.7 mg ml^{-1} Ndc80 complex was performed, corresponding to a 2:1 Ndc80 complex:tubulin monomer ratio. Samples were stained with 2% uranyl formate.

Electron tomography and cluster quantification. The tomograms displayed in Supplementary Fig. 1 were derived from tilt series collected on a 2k × 2k CCD camera from −60° to 60° using a JEOL 3100 microscope operating at 300 kV at approximately 2 μm underfocus with a nominal magnification of 40,000×. The acquisition was performed semi-automatically using a version of SerialEM (http://bio3d.colorado.edu) adapted for operating JEOL microscopes. Raw tomographic image stacks were aligned either manually with the eTomo suite of programs, or automatically using the software RAPTOR[53]. Tomographic reconstructions were constructed using the eTomo suite of programs and visualized using the IMOD software package[54].

The tomograms displayed in Fig. 3 were derived from tilt series collected from −65° to 65° on a 2k × 2k CCD camera using a Phillips CM200 microscope operating at 200 kV at approximately 2.5 μm underfocus at a nominal magnification of 39,000×. Images were filtered to 25–30 Å, before the first phase-inversion of the contrast-transfer function, and decimated twofold before reconstruction as described above.

Cluster quantification was performed by manual inspection of the reconstructions. The raw data are shown in Supplementary Fig. 9b, and consist of number of counts versus cluster size. Clusters ≥10 molecules in size were ignored in the case of the wild type, as we believe they are not physiologically relevant and would bias comparisons between the distributions towards spurious difference. Pair-wise comparisons were performed using Welch's t-test[55], which is appropriate for samples featuring both different numbers of observations and possibly unequal variances (which the unbiased estimator of the variance of the distributions suggested in this case), and P-values are shown in Supplementary Table 1. In all but one of the cases $P < 0.0015$, suggesting that the data derived from each of the conditions shown in Fig. 3 and Supplementary Fig. 9 do indeed sample different distributions. The exception was Ndc80(bonsai(7D)) versus Ndc80(bonsai(ΔN)), where we find a probability of 0.61 of sampling the same distribution. This supports our assertion that these two mutants phenocopy each other.

Conservation analysis and structure alignments. We performed multiple-sequence alignments using CLUSTALW2[56], and mapped this analysis on to the Ndc80(bonsai(ΔN)) crystal structure using the CONSURF server[57]. The alignments included 48 fungi sequences and 39 metazoan sequences in the case of NDC80 and 52 fungi sequences and 22 metazoan sequences in the case of NUF2, and are available on request. Sequences of α- and β-tubulin from six representative organisms (*Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Xenopus laevis*, *Bos taurus* and *Homo sapiens*) were compared. Since the Ndc80(bonsai(ΔN)) crystal structure is of the human complex, fungal sequences had to be threaded on to the structure by alignment with the human sequence. As we observe a similar pattern of conservation at the NDC80 toe, we believe this procedure was successful. All superpositions of crystal structures were performed with the MatchMaker function in UCSF Chimera[51].

42. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
43. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
44. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
45. Wade, R. H., Chretien, D. & Job, D. Characterization of microtubule protofilament numbers. How does the surface lattice accommodate? *J. Mol. Biol.* **212,** 775–786 (1990).
46. Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5,** 900–908 (1995).
47. Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157,** 117–125 (2007).
48. Sachse, C. *et al.* High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J. Mol. Biol.* **371,** 812–835 (2007).
49. Stewart, A. & Grigorieff, N. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* **102,** 67–84 (2004).
50. Chen, J. Z. *et al.* Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl Acad. Sci. USA* **106,** 10644–10648 (2009).
51. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure* **13,** 473–482 (2005).
52. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics International* **11,** 36–42 (2004).
53. Amat, F. *et al.* Markov random field based automatic image alignment for electron tomography. *J. Struct. Biol.* **161,** 260–275 (2008).
54. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116,** 71–76 (1996).
55. Welch, B. L. The generalisation of student's problems when several different population variances are involved. *Biometrika* **34,** 28–35 (1947).
56. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
57. Landau, M. *et al.* Consurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).

LETTER

Gas accretion as the origin of chemical abundance gradients in distant galaxies

G. Cresci[1,2], F. Mannucci[1], R. Maiolino[3], A. Marconi[4], A. Gnerucci[4] & L. Magrini[1]

It has recently been suggested[1,2] that galaxies in the early Universe could have grown through the accretion of cold gas, and that this may have been the main driver of star formation and stellar mass growth[3–5]. Because the cold gas is essentially primordial, it has a very low abundance of elements heavier than helium (referred to as metallicity). If funnelled to the centre of a galaxy, it will result in the central gas having an overall lower metallicity than gas further from the centre, because the gas further out has been enriched by supernovae and stellar winds, and not diluted by the primordial gas. Here we report chemical abundances across three rotationally supported star-forming galaxies at redshift $z \approx 3$, only 2 Gyr after the Big Bang. We find 'inverse' gradients, with the central, star-forming regions having lower metallicities than less active ones, which is opposite to what is seen in local galaxies[6,7]. We conclude that the central gas has been diluted by the accretion of primordial gas, as predicted by 'cold flow' models.

We selected three Lyman-break galaxies[8] with $z > 3$ from the AMAZE[9] and LSD[10] samples—SSA22a-C16 ($z = 3.065$), CDFa-C9 ($z = 3.219$) and SSA22a-M38 ($z = 3.288$)—in order to test for the presence of metallicity gradients beyond $z = 3$, when the Universe was only 14–16% of its present age and only a small fraction (~5–10%; ref. 11) of the total stellar mass in the Universe had been assembled. This epoch ($z > 3$) is particularly interesting, as it precedes the peak of the cosmic star formation density[12]. We selected these three galaxies because they show a remarkably symmetric velocity field in the [O III] emission line, which traces the ionized gas kinematics (Fig. 1). Such kinematics indicates that these are rotationally supported disks (A.G. *et al.*, manuscript in preparation), with no evidence for more complex merger-induced dynamics. Assuming an exponential disk mass model to reproduce the dynamical properties of the three sources, we derive dynamical masses in the range ~$(1–20) \times 10^{10} M_\odot$ (where $M_\odot$ indicates the solar mass). By using the Hβ line flux[13], we infer star formation rates as high as ~$120 M_\odot \, \mathrm{yr}^{-1}$ after correcting for dust extinction[9]. We carefully checked for the presence of active galactic nuclei in our galaxies, as they would invalidate the calibrations derived for star-forming galaxies because of the different ionizing continua involved (Supplementary Information).

Near-infrared spectroscopic observations of the galaxies were obtained with the integral field spectrometer SINFONI[14] on the Very Large Telescope of the European Southern Observatory. The spectra cover the 1.45–2.41 μm wavelength range; strong rest frame optical emission lines of the observed galaxies, such as [O III] 5,007 Å, [O III] 4,959 Å, Hβ, [Ne III] 3,870 Å and [O II] 3,727 Å, are redshifted to this range. The flux ratio between the (rest-frame) optical nebular lines observed in SINFONI spectra depend on gas metallicity, and have been calibrated for local low- and high-metallicity galaxies[15]. In particular, we use a combination of three independent metallicity diagnostics[9], [O III] 5,007 Å/Hβ, [O III] 5,007 Å/[O II] 3,727 Å and, when available, [Ne III] 3,870 Å/[O II] 3,727 Å, to derive the gas phase metallicity in our spectra. Even if a single line ratio can constrain the metallicity in a broad interval of ~0.4 dex or more, the use of three independent

diagnostics helps us in removing the degeneracies and reducing the uncertainties; we can determine for each spatial pixel the combination of metallicity and extinction that best fit the observed line ratios with our calibrations (see Supplementary Information for details).

The derived metallicity maps are shown in Fig. 1. An unresolved region with lower metallicity is evident in each map, surrounded by a more uniform disk of higher metal content. In CDFa-C9, the lower metallicity region is coincident with the galaxy centre, as traced by the continuum peak, whereas it is offset by ~0.60″ (4.6 kpc) in SS22a-C16 and ~0.45″ (3.4 kpc) in SS22a-M38. On the other hand, in all the galaxies the area of lower metallicity coincides with or is closer than 0.25″ (1.9 kpc, which is half the point spread function full-width at half-maximum) to the regions of enhanced line emission, which trace the more active star forming regions. The average difference between high and low metallicity regions in the three galaxies is 0.55 in units of 12+log(O/H), larger than the ~0.2–0.4 dex gradients measured in the Milky Way and other local spirals[16] on the same spatial scales. The measured gas phase abundance variations have a significance between 98% and 99.8% (Table 1).

The observed line ratios depend on metallicity both directly and indirectly through the ionization parameter U (the ratio between the flux of ionizing photons and the gas density), with lower metallicity star forming regions generally having higher U (ref. 15). In principle, the observed variation of line ratios across galaxies could still be due to variations of U unrelated to the metallicity distribution. As local galaxies show a broad range of ionization parameter, this dependence on U is responsible for the spread included in our calibrations. In order to exclude the possibility that the line ratios in our galaxies are biased owing to significantly different values of ionization parameter at high redshift, we have examined their dependence on U based on the latest photo-ionization models available[17,18,19]. These models are not able to simultaneously reproduce the different observed line ratios with a variation of the ionization parameter alone: at constant metallicity, the observed line ratios differ from models predictions by ~0.2 dex. This demonstrates that different metallicities are required to reproduce the observations, and even if other effects are present, they are not dominant.

Observed metallicity gradients in local disk galaxies show metallicities decreasing outward from the galactic centres. In contrast, our high

Table 1 | Properties of the sample galaxies

Name	z	$\log(M_*/M_\odot)$	Metallicity, whole galaxy (12+log(O/H))	Metallicity, low-Z region (12+log(O/H))	Metallicity, high-Z region (12+log(O/H))
SSA22a-C16	3.065	$10.68^{+0.16}_{-0.54}$	$8.36^{+0.06}_{-0.06}$	$8.18^{+0.13}_{-0.14}$	$8.52^{+0.14}_{-0.07}$
CDFa-C9	3.219	$10.03^{+0.40}_{-0.08}$	$8.36^{+0.06}_{-0.05}$	$7.98^{+0.14}_{-0.16}$	$8.56^{+0.12}_{-0.12}$
SSA22a-M38	3.288	$10.86^{+0.18}_{-0.41}$	$8.26^{+0.09}_{-0.11}$	$7.84^{+0.22}_{-0.23}$	$8.59^{+0.05}_{-0.07}$

Here z is the source redshift, M_* the stellar mass derived from multi band photometry and spectral energy distribution fitting assuming a Chabrier[30] initial mass function[9], while the last three columns show the metallicity integrated for the whole galaxy, for the low metallicity regions and for the high metallicity regions respectively, in units of 12+log(O/H). The amplitude and significance of the detected gradients were evaluated by extracting one-dimensional spectra integrated over apertures comparable to the point spread function (FWHM = 0.5″) in both the lowest and highest metallicity region, and using them to measure the line ratios (Supplementary information). The uncertainties are the 1σ confidence intervals.

[1]INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy. [2]Max-Planck-Institut für extraterrestrische Physik (MPE), Giessenbachstrasse 1, 85748 Garching, Germany. [3]INAF – Osservatorio Astronomico di Roma, via di Frascati 33, 00040 Monte Porzio Catone, Italy. [4]Dipartimento di Fisica e Astronomia, Università di Firenze, Largo E. Fermi 2, 50125 Firenze, Italy.
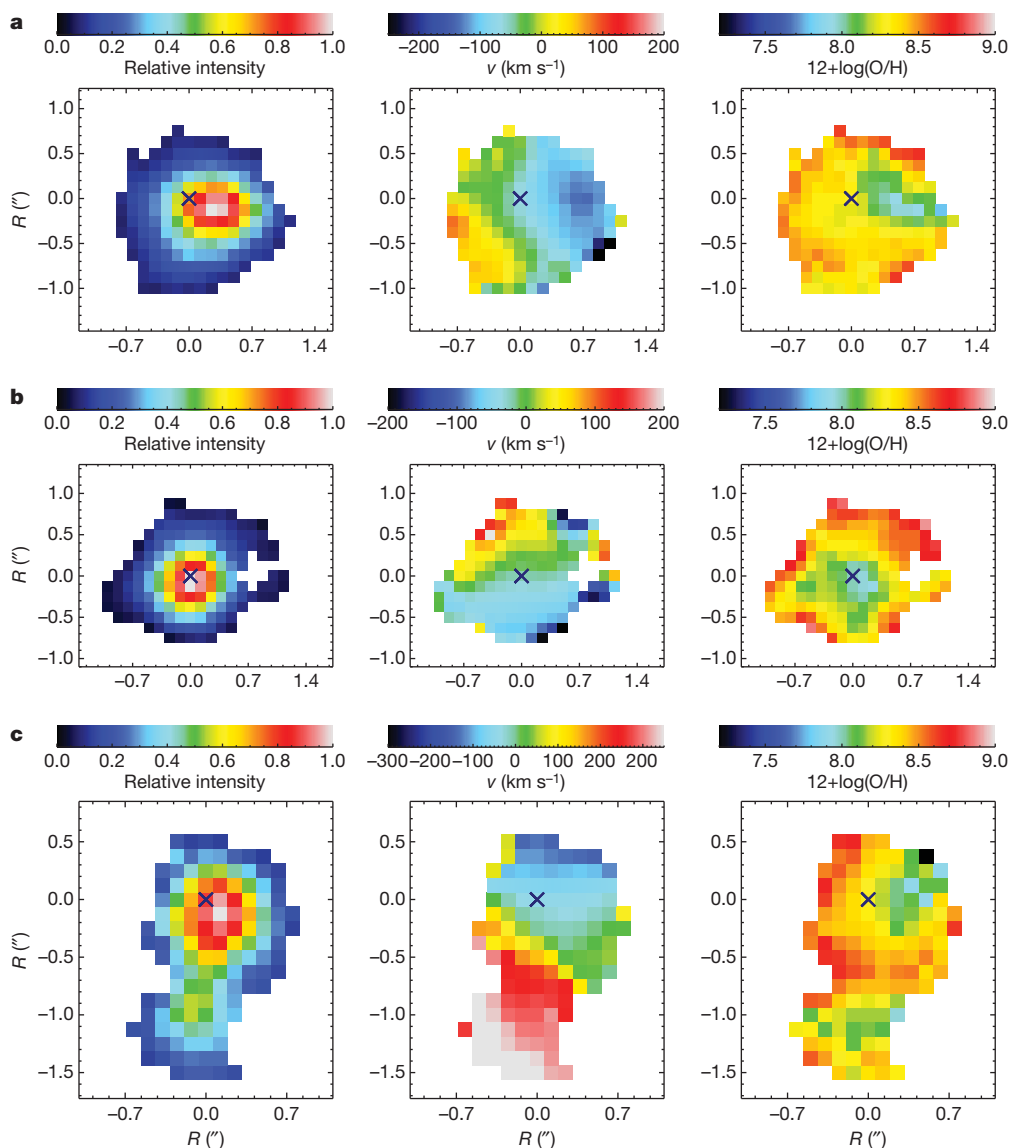
Figure 1 | Surface brightness and velocity of the [O III] 5,007 Å line, and metallicity maps. The data for SSA22a-C16 ($z = 3.065$; **a**), CDFa-C9 ($z = 3.219$; **b**) and SSA22a-M38 ($z = 3.288$; **c**) were obtained with the SINFONI spectrograph using the $0.125'' \times 0.250''$ pixel scale in seeing-limited conditions, resulting in a spatial resolution of $\sim 0.5''$ (full-width at half-maximum (FWHM) of the point spread function). The maps were extracted from the SINFONI datacube after a Gaussian smoothing with FWHM = 3 pixels ($0.375''$). The left panels in each row show normalized surface brightness in the [O III] 5,007 Å emission line. The same line has been used to derive the velocity maps shown in the middle panels; the observed gas kinematics is compatible with a rotating disk, with no evidence of merger-induced complex dynamics. The right panels show maps of gas phase metallicity, as relative abundances of oxygen and hydrogen parameterized in units of $12 + \log(\mathrm{O/H})$. Lower metallicity regions, corresponding to a higher [O III]/[O II] ratio, are surrounded by a more enriched disk. The crosses in each panel mark the position of the continuum peak.

redshift sources show 'inverted' gradients, with lower metallicity in a well defined active region closer to the galaxy centre. Current models of chemical enrichment in galaxies[20,21] cannot reproduce our observations, as they assume radially isotropic gas accretion onto the disk and the instantaneous recycling approximation[22]. Nevertheless, the detected gradients can be explained in the framework of the cold gas accretion scenario[23] recently proposed to explain the properties of gas rich, rotationally supported galaxies observed at high redshift[3,24]. In this scenario, the observed low metallicity regions are created by the local accretion of metal-poor gas in clumpy streams[1]; this gas penetrates deep into the galaxy following the potential well, and sustains the observed high star formation rate in the pre-enriched disk. Stream-driven turbulence is then responsible for the fragmentation of the disks into giant clumps, as observed at $z \gtrsim 2$ (refs 10 and 25), that are the sites of efficient star formation and possibly the progenitors of the central spheroid. This scenario is also in agreement with the dynamical properties of our galaxies, which appear to be dominated by gas rotation in a disk with no evidence of the kinematical asymmetries typically induced by mergers.

If infall and outflows are not present, metallicity is directly related to the fraction of gas present in the galaxy. For this reason, additional hints about the presence of infall and outflows can be obtained considering these two quantities. We can derive an estimate of gas fractions in each single region of our galaxies assuming that the relation between star formation rate and gas density observed in the local

Universe (the Schmidt–Kennicutt relation[26]) also applies to high redshift galaxies[27]. The gas fractions obtained for the two regions in each galaxy are shown in Fig. 2, plotted against metallicities Z. It is evident that our results do not support the closed box model (dashed black line), in which galaxies have no contribution from infalls and outflows. Our data can be reproduced only by including either outflows of very enriched material, such as the ejecta from supernovae, which extract metals and decrease Z, or infalls of pristine gas which increase the gas fraction and decrease Z. In particular, if a strong infall of pristine gas is present, we would expect higher gas fractions in the low metallicity regions, as observed. Using simple models[10,28], we estimate the amount of infall and outflow required to reproduce the observations. We take into account two opposing scenarios. If inflows are dominant, we require ~ 3.5 times more pristine gas infall in the low metallicity regions than in the high metallicity part. On the other hand, if outflows are dominant, high metallicity regions can be explained with an outflow ~ 2 times the star formation rate, whereas unrealistic outflows ~ 12 times the star formation rate would be needed to explain the properties of low metallicity regions. We therefore conclude that the differences in gas fractions as derived by our estimates are not only reasonable given the measured metallicity differences, but are also consistent with expectations from massive infalls of low metallicity gas in the central regions of the galaxies. They are unlikely to be explained using outflows only.
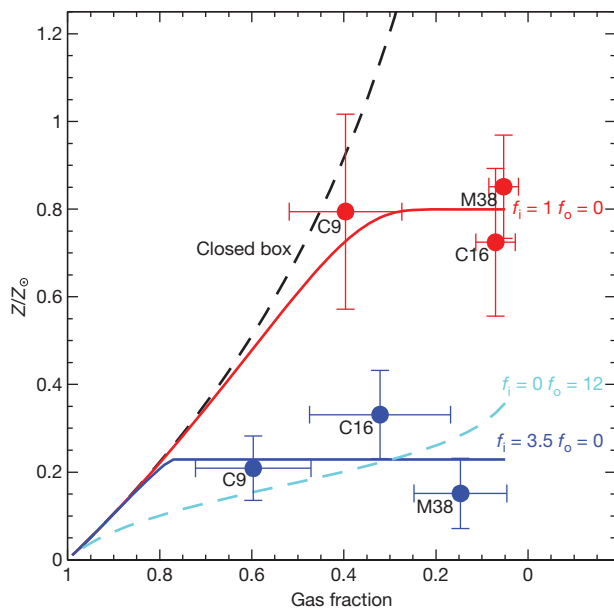
Figure 2 | Gas fractions for the different regions in the observed galaxies.
To estimate gas masses in the low and high metallicity regions of the galaxies, we used the Hβ flux corrected for extinction. We converted measured star formation rate densities from Hβ to a gas mass by inverting the Kennicutt-Schmidt law[26]. To estimate stellar masses, we have to rely on the continuum flux from our SINFONI data, as we are still missing deep and high resolution continuum images of the sources. The continuum is detected in the datacubes of our galaxies, although at a lower signal-to-noise ratio than the emission lines, making this estimate more uncertain. We have to further assume that the stellar mass scales as the K-band continuum light (V-band rest frame). For each galaxy, we thus measure the total K-band continuum flux for the whole galaxy and the flux in the lower metallicity regions, using these relative flux measurements to rescale the total stellar mass from SED (spectral energy distribution) fitting[9], and estimate the stellar mass in the metal poor and metal rich regions. We plot the derived gas fraction against the metallicity in solar units, $Z/Z_\odot$. Filled blue and red circles show respectively the low and high metallicity regions of each galaxy; black dashed line, expectations from a closed-box model. The difference between the regions can be explained by a simple model including only infall of gas f_i, which is, respectively, 3.5 (blue solid line) and 1 times (red solid line) the star formation rate for the regions of lower and higher metallicity. Dashed cyan line, a model of pure outflow f_o, 12 times the star formation rate. The observed difference in gas fraction is consistent with the expectation from a massive infall of low metallicity gas in the central regions. For brevity, C9, C16 and M38 are used to represent respectively CDFa-C9, SSA22a-C16 and SSA22a-M38. The error bars are 1σ formal uncertainties on the measurements, given the assumptions discussed.

Our observations of low metallicity regions in three galaxies at $z \approx 3$ therefore provide evidence for the actual presence of accretion of metal-poor gas in massive high-z galaxies, capable of sustaining high star formation rates without frequent mergers of already evolved and enriched sub-units. This picture was already indirectly suggested by recent observational studies of gas rich disks at $z \approx 1$–2 (refs 3, 5), and by the study of the variation of metallicity as a function of gas fraction[10] and star formation rate[29]. The data presented here directly reveal the metal-poor accreted gas responsible for fuelling the star formation rate in massive galaxies in the early Universe.

Although the three galaxies discussed here are the only $z \approx 3$ sources for which a metallicity gradient has been measured, the detection of lower metallicity clouds in all three galaxies suggests that metal poor gas accretion is probably a common feature in actively rotationally supported star forming galaxies at this early cosmic epoch. However, it is difficult to speculate about how dominant this accretion mechanism is in the global star forming population, as the relative number of rotationally supported and more dynamically complex galaxies is still unknown, and similar observations are still missing for different classes of galaxies.

1. Dekel, A. *et al.* Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).
2. Bournaud, F. & Elmegreen, B. G. Unstable disks at high redshift: evidence for smooth accretion in galaxy formation. *Astrophys. J.* **694**, L158–L161 (2009).
3. Förster Schreiber, N. M. *et al.* The SINS survey: SINFONI integral field spectroscopy of $z \sim 2$ star-forming galaxies. *Astrophys. J.* **706**, 1364–1428 (2009).
4. Daddi, E. *et al.* Multiwavelength study of massive galaxies at $z \sim 2$. I. Star formation and galaxy growth. *Astrophys. J.* **670**, 156–172 (2007).
5. Tacconi, L. J. *et al.* High molecular gas fractions in normal massive star forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).
6. Garnett, D. R., Shields, G. A., Skillman, E. D., Sagan, S. P. & Dufour, R. J. Interstellar abundance gradients in NGC 2403: comparison to M33. *Astrophys. J.* **489**, 63–86 (1997).
7. Magrini, L., Sestito, P., Randich, S. & Galli, D. The evolution of the Galactic metallicity gradient from high-resolution spectroscopy of open clusters. *Astron. Astrophys.* **494**, 95–108 (2009).
8. Steidel, C. C. *et al.* Lyman break Galaxies at redshift $z \sim 3$: survey description and full data set. *Astrophys. J.* **592**, 728–754 (2003).
9. Maiolino, R. *et al.* AMAZE. I. The evolution of the mass-metallicity relation at $z > 3$. *Astron. Astrophys.* **488**, 463–479 (2008).
10. Mannucci, F. *et al.* LSD: Lyman-break galaxies stellar populations and dynamics – I. Mass, metallicity and gas at $z \sim 3.1$. *Mon. Not. R. Astron. Soc.* **398**, 1915–1931 (2009).
11. Pozzetti, L. *et al.* The VIMOS VLT deep survey. The assembly history of the stellar mass in galaxies: from the young to the old universe. *Astron. Astrophys.* **474**, 443–459 (2007).
12. Hopkins, A. M. & Beacom, J. F. On the normalization of the cosmic star formation history. *Astrophys. J.* **651**, 142–154 (2006).
13. Kennicutt, R. C. Jr. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–231 (1998).
14. Eisenhauer, F. *et al.* SINFONI – Integral field spectroscopy at 50 milli-arcsecond resolution with the ESO VLT. *Proc. SPIE* **4841**, 1548–1561 (2003).
15. Nagao, T., Maiolino, R. & Marconi, A. Gas metallicity diagnostics in star-forming galaxies. *Astron. Astrophys.* **459**, 85–101 (2006).
16. van Zee, L., Salzer, J. J., Haynes, M. P., O'Donoghue, A. A. & Balonek, T. J. Spectroscopy of outlying H II regions in spiral galaxies: abundances and radial gradients. *Astron. J.* **116**, 2805–2833 (1998).
17. Dopita, M. A. *et al.* Modeling the pan-spectral energy distribution of starburst galaxies. III. Emission line diagnostics of ensembles of evolving H II regions. *Astrophys. J.* **167** (Suppl.), 177–200 (2006).
18. Levesque, E. M., Kewley, L. J. & Larson, K. L. Theoretical modeling of star-forming galaxies. I. Emission-line diagnostic grids for local and low-metallicity galaxies. *Astron. J.* **139**, 712–727 (2010).
19. Martín-Manjón, M. L., García-Vargas, M. L., Mollá, M. & Díaz, A. I. POPSTAR evolutionary synthesis models. II: optical emission-line spectra from giant HII regions. *Mon. Not. R. Astron. Soc.* **403**, 2012–2032 (2010).
20. Molla, M., Ferrini, F. & Diaz, A. I. Evolution of spiral galaxies. VII. Time evolution of the radial distributions of abundances. *Astrophys. J.* **475**, 519–533 (1997).
21. Hou, J. L., Prantzos, N. & Boissier, S. Abundance gradients and their evolution in the Milky Way disk. *Astron. Astrophys.* **362**, 921–936 (2000).
22. Tinsley, B. M. Stellar lifetimes and abundance ratios in chemical evolution. *Astrophys. J.* **229**, 1046–1056 (1979).
23. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
24. Cresci, G. *et al.* The SINS survey: modeling the dynamics of $z \sim 2$ galaxies and the high-z Tully-Fisher relation. *Astrophys. J.* **697**, 115–132 (2009).
25. Genzel, R. *et al.* From rings to bulges: evidence for rapid secular galaxy evolution at $z \sim 2$ from integral field spectroscopy in the SINS survey. *Astrophys. J.* **687**, 59–77 (2008).
26. Kennicutt, R. C. Jr. The global Schmidt law in star-forming galaxies *Astrophys. J.* **498**, 541–552 (1998).
27. Bouché, N. *et al.* Dynamical properties of $z \sim 2$ star-forming galaxies and a universal star formation relation. *Astrophys. J.* **671**, 303–309 (2007).
28. Erb, D. K. A model for star formation, gas flows, and chemical evolution in galaxies at high redshifts. *Astrophys. J.* **674**, 151–156 (2008).
29. Mannucci, F., Cresci, G., Maiolino, R., Marconi, A. & Gnerucci, A. A fundamental relation between mass, SFR and metallicity in local and high redshift galaxies. *Mon. Not. R. Astron. Soc.* (in the press); preprint at ⟨http://arXiv.org/abs/1005.0006⟩ (2010).
30. Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pacif.* **115**, 763–795 (2003).

LETTER

A collision in 2009 as the origin of the debris trail of asteroid P/2010 A2

Colin Snodgrass[1,2], Cecilia Tubiana[1], Jean–Baptiste Vincent[1], Holger Sierks[1], Stubbe Hviid[1], Richard Moissl[1], Hermann Boehnhardt[1], Cesare Barbieri[3], Detlef Koschny[4], Philippe Lamy[5], Hans Rickman[6,7], Rafael Rodrigo[8], Benoît Carry[9], Stephen C. Lowry[10], Ryan J. M. Laird[10], Paul R. Weissman[11], Alan Fitzsimmons[12], Simone Marchi[3] & the OSIRIS team*

The peculiar object P/2010 A2 was discovered[1] in January 2010 and given a cometary designation because of the presence of a trail of material, although there was no central condensation or coma. The appearance of this object, in an asteroidal orbit (small eccentricity and inclination) in the inner main asteroid belt attracted attention as a potential new member of the recently recognized[2] class of main-belt comets. If confirmed, this new object would expand the range in heliocentric distance over which main-belt comets are found. Here we report observations of P/2010 A2 by the Rosetta spacecraft. We conclude that the trail arose from a single event, rather than a period of cometary activity, in agreement with independent results[3]. The trail is made up of relatively large particles of millimetre to centimetre size that remain close to the parent asteroid. The shape of the trail can be explained by an initial impact ejecting large clumps of debris that disintegrated and dispersed almost immediately. We determine that this was an asteroid collision that occurred around 10 February 2009.

P/2010 A2 orbits much closer to the Sun (its semi-major axis is 2.29 astronomical units, AU) than the previously discovered main-belt comets, the activity of which seems to be driven by episodic ice sublimation[2]. The discovery of a parent body a few arcseconds (~1,500 km) away from the trail[4,5] implied that it was debris from a recent collision rather than the tail of a comet, although Earth-based observations alone are consistent with a comet model[6]. It was suggested that the trail formed between January and August 2009, and was comprised of relatively large (diameter >1 mm) grains[7]. Here we use the term 'trail' to describe a tail made up of large particles, rather than dust from a currently active comet. Hubble Space Telescope observations refine the diameter of the parent body to 120 m and the date to February/March 2009 (ref. 3).

We obtained an improved three-dimensional description of the trail geometry by observing it with the OSIRIS Narrow Angle Camera[8] on board the European Space Agency's Rosetta spacecraft on 16 March 2010. Rosetta was approaching the asteroid belt for its July 2010 fly-by of asteroid 21 Lutetia, and at the time of observation was 1.8 AU from the Sun and 10° out of P/2010 A2's orbital plane. From this vantage point the separation between the anti-velocity (orbit) angle and the anti-Sun (comet tail) direction was much larger than was possible to observe from Earth. We also obtained reference images of P/2010 A2 from Earth using the 3.6 m New Technology Telescope (NTT) at the European Southern Observatory's La Silla observatory and the 200″ Hale telescope at Palomar Mountain. Figure 1 displays images of P/2010 A2 at four epochs, from the Earth and from Rosetta. We measured the position angle of the trail and extracted the flux profile along the trail axis at each epoch (Fig. 2).

We simulate the shape of the observed trail at each epoch by modelling the trajectories of dust grains, as is commonly done for comet tails[9,10]. The motion depends on the grains' initial velocity and the ratio β between solar radiation pressure and solar gravity, which is related to the size of the grains[11]. Owing to the small phase angle as viewed from Earth it is not possible to find a unique solution for the dust ejection epochs from the ground-based observations alone: the best estimate indicates that particles must have been emitted before August 2009, and should be of at least millimetre size to account for the low dispersion and their apparent position close to the projected anti-velocity vector. The higher phase angle of the OSIRIS observations allows a more precise simulation of the trail, and consequently we obtained a very narrow time frame for the emission of the dust. The grains must have been released around 10 February 2009, plus or minus 5 days, with the uncertainty being due to the measurement of the position angle of the faint trail in the OSIRIS images. To account for the position angle and the length of the trail, we must consider grains ranging from millimetre to centimetre size and larger. The particle sizes from this model, together with the brightness profile shown in Fig. 2, allow us to measure the size distribution of grains, and from this derive a total mass of the ejecta of 3.7×10^8 kg, or approximately 16% of a parent body of diameter 120 m, assuming a density of 2,500 kg m^{-3} and an albedo of 15% for both the asteroid and the grains.

The shape of the trail cannot be reproduced with a traditional comet-tail model, even when considering a longer timescale for the event. Cometary models all produce tail geometries in the OSIRIS image with a fan that reaches a point at the nucleus and becomes wider farther from it (see Supplementary Information for examples). All images of P/2010 A2 show a distinctive broad edge at the 'nucleus' end and then a trail with parallel edges. From the Rosetta observing geometry this edge is even broader than it is from Earth. This shape can be reproduced by a number of parallel synchrones: contours in the model that show the location of dust produced at the same time. In this model, an initial dust cloud is formed (presumably by a collision) in February 2009, which initially does not spread much (less than 1,000 km) but over a year solar gravity and radiation pressure expand this small trail to its observed width and length, respectively. Higher-resolution images from the Hubble Space Telescope[3] show the presence of parallel striae in the trail, very well aligned with the synchrone representing the original event as estimated from our simulations. These striae indicate that some areas of higher densities existed in the original cloud: larger clumps of material that fragmented and dispersed as they were ejected. The width of the broad front end of the trail from these different geometries can be used to constrain the speed of particles in the original ejecta cloud to less than 1 m s^{-1}. Impact

[1]Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany. [2]European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Santiago 19, Chile. [3]University of Padova, Department of Astronomy, Vicolo dell'Osservatorio 3, 35122 Padova, Italy. [4]Research and Scientific Support Department, European Space Agency, Keplerlaan 1, Postbus 229, 2201 AZ Noordwijk ZH, The Netherlands. [5]Laboratoire d'Astrophysique de Marseille, UMR6110 CNRS/Université Aix-Marseille, 38 rue Frédéric Joliot-Curie, 13388 Marseille cedex 13, France. [6]Department of Astronomy and Space Physics, Uppsala University, Box 516, 75120 Uppsala, Sweden. [7]PAS Space Research Center, Bartycka 18A, 00-716 Warszawa, Poland. [8]Instituto de Astrofísica de Andalucía, CSIC, Box 3004, 18080 Granada, Spain. [9]LESIA, Observatoire de Paris—Meudon, 5 place Jules Janssen, 92195 Meudon cedex, France. [10]Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK. [11]Jet Propulsion Laboratory, 4800 Oak Grove Drive, MS 183-301, Pasadena, California 91101, USA. [12]Astrophysics Research Centre, Queen's University Belfast, BT7 1NN, UK.
*Lists of participants and affiliations appear at the end of the paper.

Image

16 Feb 2010 UT NTT | a

16 Mar 2010 UT Rosetta | b

4 April 2010 UT Palomar | c

6 April 2010 UT NTT | d

Image + model

e

2mm

1cm 5mm

f

1 mm

5 mm

1cm

g

1.5mm

1cm 5mm

h

1.5mm

5mm

1cm

0   50   100   150   200
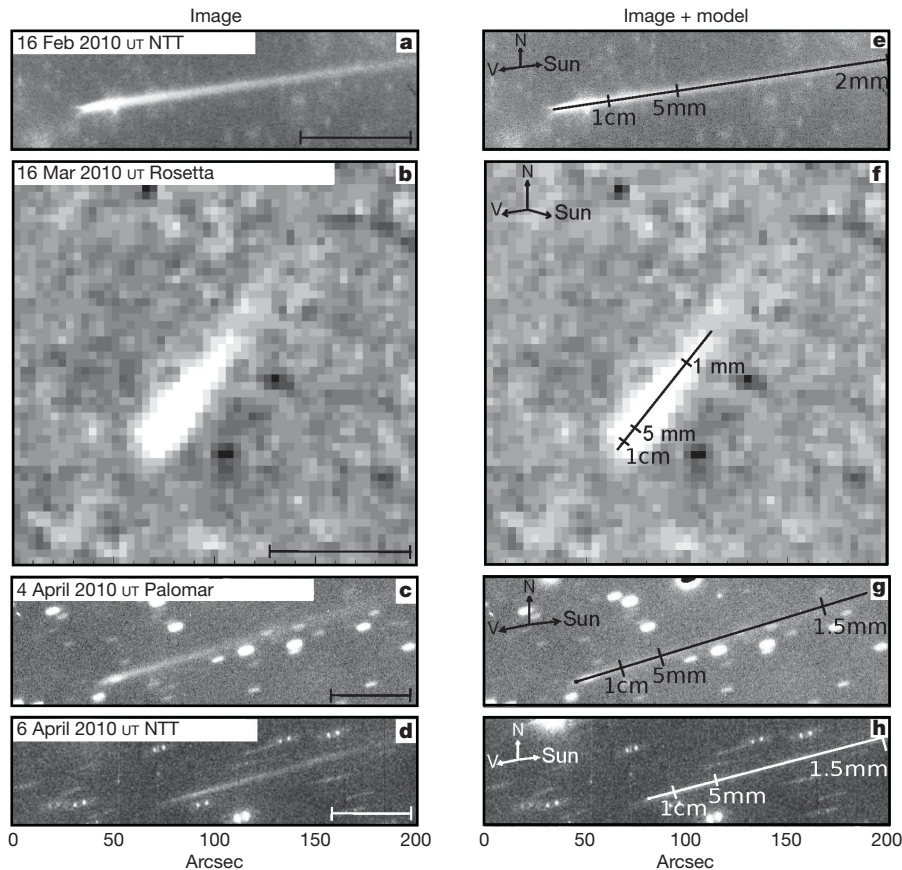Arcsec

0   50   100   150   200
Arcsec

Figure 1 | Images of P/2010 A2 at four epochs.
These are, from top to bottom, from the NTT
(February), Rosetta (March), Palomar (April) and
the NTT (April), respectively. The scale bars in the
lower right of **a–d** show a projected distance of
5×10^4 km. When possible, we median-combined
images centred on the object to increase the signal-
to-noise ratio (relative to a single exposure) of the
trail and remove background stars. To isolate the
faint dust trail in the OSIRIS data we first
subtracted an image of the background star field
from each frame before shifting the frame on the
basis of the motion of the object and then median-
combining. On the right we show the same images
overlaid with synchrones generated from the
Finson–Probstein model. Numbers indicate
estimates of the particle size distribution along the
synchrones, derived from the model. The
orientation of the images is North up, East left. The
compass in the top left of panels **e** to **h** shows the
direction of the heliocentric velocity vector (orbit)
V and the direction to the Sun. The advantage of
the Rosetta observing geometry is clear, with the
broad head of the trail and obvious difference
between the observed position angle and the anti-
velocity vector apparent in the OSIRIS image.
Models based on a period of cometary activity
(rather than a single event) or smaller particle sizes
produce a significantly different pattern of
synchrones in **f** (see Supplementary Figs 2–4) that
does not fit the observations. The same models all
produce similar synchrones to those in the impact
model for **e**, **g** and **h**, and therefore cannot be ruled
out on the basis of Earth-based data alone.
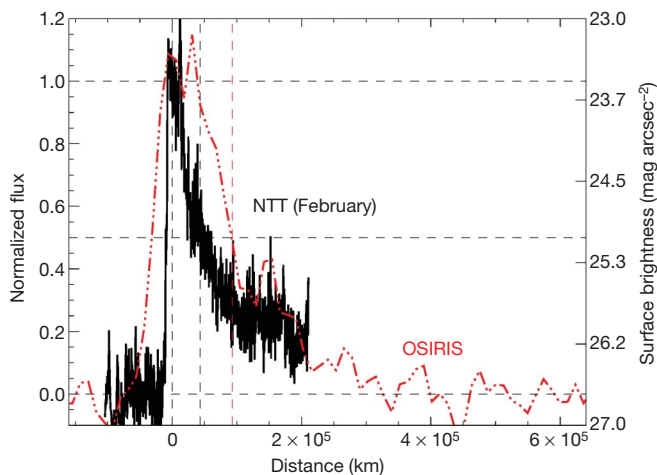
Figure 2 | Flux profiles along the trail. The normalized profiles for the
February NTT (solid black line) and the OSIRIS data sets (dot-dashed red line)
are shown. The x axis is in kilometres along the trail, with the conversion from
the projected scale in arcseconds on sky based on the geometry derived from
our model. The vertical dashed lines indicate the half-maximum of the profiles,
used to measure the scale length of the trails in these images with different
sensitivities. The two profiles have scale lengths of 4.3×10^4 and 9.3×10^4 km
along the trail. The right y axis shows the calibrated surface brightness of the
NTT profile in R-band magnitudes per square arcsecond. The flux profiles from
the other Earth-based observations match the NTT one, but are omitted for
clarity because they have higher noise owing to the shorter integration times.
We derive a size distribution using the NTT flux profile and the size of particles
as a function of distance along the trail from the Finson–Probstein model. This
is done by converting the total flux across the trail at each distance to a reflecting
area (assuming an albedo of 15%), and finding the corresponding number of
particles of the appropriate size. The resulting cumulative size distribution is
shown in Supplementary Fig. 6, and has a slope that matches the prediction for
a population of collisional remnants[24].

experiments[12] find that such a low velocity implies a parent body of low
strength and high porosity, although recent computer simulations
suggest that impacts on such a small asteroid will lead to low-velocity
ejecta independently of porosity[13].

Previously, asteroid collision models have been used to explain the
dust trails associated with main-belt comets[14], but the longer-lasting
dust production and repeated activity of comet Elst–Pizarro at each
perihelion[15,16] rule out recent (within the past few years) collisions.
Collisions inferred from asteroid families[17] or large-scale denser
regions in the zodiacal dust cloud[18] have ages of 10^4 to 10^9 years.
Our observations show direct evidence for a collision that is recent
in observational terms, with a debris trail that is still evolving. From
estimates of the population of the main asteroid belt[19,20] and an esti-
mated impactor diameter of 6–9 m (ref. 21), we expect roughly one
impact of this size every 1.1 billion years for a parent body of diameter
120 m, or approximately one every 12 years somewhere in the asteroid
belt. This is in agreement with a single detection by the LINEAR
survey; we expect that more small collisions will be detected by next-
generation surveys. Collisions of this size therefore contribute around
3×10^7 kg yr^{-1} of dust to the zodiacal cloud, which is negligible com-
pared with comets and the total required to maintain a steady state[22], in
agreement with recent models[23].

1. Birtwhistle, P., Ryan, W. H., Sato, H., Beshore, E. C. & Kadota, K. Comet P/2010 A2
   (LINEAR). *IAU Circ.* **9105** (2010).
2. Hsieh, H. H. & Jewitt, D. A population of comets in the main asteroid belt. *Science*
   312, 561–563 (2006).
3. Jewitt, D., Weaver, H., Agarwal, J., Mutchler, M. & Drahus, M. A recent disruption of
   the main-belt asteroid P/2010 A2. *Nature* doi: 10.1038/nature09456 (this issue).
4. Licandro, J., Tozzi, G. P., Liimets, T., Haver, R. & Buzzi, L. Comet P/2010 A2
   (LINEAR). *IAU Circ.* **9109** (2010).
5. Jewitt, D., Annis, J. & Soares-Santos, M. Comet P/2010 A2 (LINEAR). *IAU Circ.* **9109**
   (2010).
6. Moreno, F. *et al.* Water-ice driven activity on main-belt comet P/2010 A2 (LINEAR)?
   Astrophys. J. **718**, L132–L136 (2010).

7. Sekanina, Z. Comet P/2010 A2 (LINEAR). *IAU Circ.* **9110** (2010).
8. Keller, H. U. *et al.* OSIRIS, the scientific camera system onboard Rosetta. *Space Sci. Rev.* **128**, 433–506 (2007).
9. Finson, M. & Probstein, R. A theory of dust comets. 1. Model and equations. *Astrophys. J.* **154**, 327–380 (1968).
10. Beisser, K. & Boehnhardt, H. Evidence for the nucleus rotation in streamer patterns of comet Halley's dust tail. *Astrophys. Space Sci.* **139**, 5–12 (1987).
11. Burns, J. A., Lamy, P. L. & Soter, S. Radiation forces on small particles in the solar system. *Icarus* **40**, 1–48 (1979).
12. Michikami, T., Moriguchi, K., Hasegawa, S. & Fujiwara, A. Ejecta velocity distribution for impact cratering experiments on porous and low strength targets. *Planet. Space Sci.* **55**, 70–88 (2007).
13. Jutzi, M., Michel, P., Benz, W. & Richardson, D. C. Fragment properties at the catastrophic disruption threshold: the effect of the parent body's internal structure. *Icarus* **207**, 54–65 (2010).
14. Lien, D. J. Asteroid debris trails: evidence for recent collisions in the asteroid belt. *Bull. Am. Astron. Soc.* **30**, 1035 (1998).
15. Hsieh, H. H., Jewitt, D., Lacerda, P., Lowry, S. C. & Snodgrass, C. The return of activity in main-belt comet 133P/Elst-Pizarro. *Mon. Not. R. Astron. Soc.* **403**, 363–377 (2010).
16. Bagnulo, S., Tozzi, G. P., Boehnhardt, H., Vincent, J.-B. & Muinonen, K. Polarimetry and photometry of the peculiar main-belt object 7968 = 133P/Elst-Pizarro. *Astron. Astrophys.* **514**, A99 (2010).
17. Nesvorný, D., Bottke, W. F., Dones, L. & Levison, H. F. The recent breakup of an asteroid in the main-belt region. *Nature* **417**, 720–771 (2002).
18. Nesvorný, D. *et al.* Candidates for asteroid dust trails. *Astron. J.* **132**, 582–595 (2006).
19. Bottke, W. F. *et al.* Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* **179**, 63–94 (2005).
20. Marchi, S. *et al.* The cratering history of asteroid (2867) Steins. *Planet. Space Sci.* **58**, 1116–1123 (2010).
21. Holsapple, K. A. & Housen, K. R. A crater and its ejecta: an interpretation of deep impact. *Icarus* **187**, 345–356 (2007).
22. Sykes, M. V., Grün, E., Reach, W. T. & Jenniskens, P. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 677–693 (Univ. Arizona Press, 2004).
23. Nesvorný, D. *et al.* Cometary origin of the zodiacal cloud and carbonaceous micrometeorites. implications for hot debris disks. *Astrophys. J.* **713**, 816–836 (2010).
24. Dohnanyi, J. W. Collisional model of asteroids and their debris. *J. Geophys. Res.* **74**, 2531–2554 (1969).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The OSIRIS team

M. A'Hearn[13], F. Angrilli[14], A. Barucci[9], J.-L. Bertaux[15], G. Cremonese[16], V. Da Deppo[17], B. Davidsson[6], S. Debei[14], M. De Cecco[18], S. Fornasier[9], P. Gutiérrez[8], W.-H. Ip[19], H. U. Keller[20], J. Knollenberg[21], J. R. Kramm[1], E. Kuehrt[21], M. Kueppers[22], L. M. Lara[8], M. Lazzarin[3], J. J. López-Moreno[8], F. Marzari[23], H. Michalik[20], G. Naletto[24], L. Sabau[25], N. Thomas[26] & K.-P. Wenzel[4]

[13]University of Maryland, Department of Astronomy, College Park, Maryland 20742-2421, USA. [14]Department of Mechanical Engineering—University of Padova, Via Venezia 1, 35131 Padova, Italy. [15]LATMOS, CNRS/UVSQ/IPSL, 11 Boulevard d'Alembert, 78280 Guyancourt, France. [16]INAF—Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy. [17]CNR-IFN UOS Padova LUXOR, Via Trasea 7, 35131 Padova, Italy. [18]UNITN, Università di Trento, Via Mesiano 77, 38100 Trento, Italy. [19]National Central University, Institute of Astronomy, 32054 Chung-Li, Taiwan. [20]Institut für Datentechnik und Kommunikationsnetze der TU Braunschweig, Hans-Sommer-Strasse 66, 38106 Braunschweig, Germany. [21]DLR Institute for Planetary Research, Rutherfordstrasse 2, 12489 Berlin, Germany. [22]ESA-ESAC, Camino bajo del Castillo S/N, 28691 Villanueva de la Cañada, Madrid, Spain. [23]Department of Physics—University of Padova, Via Marzolo 8, 35131 Padova, Italy. [24]Department of Information Engineering—University of Padova, Via Gradenigo, 6/B I, 35131 Padova, Italy. [25]Instituto Nacional de Tecnica Aeroespacial, Carretera de Ajalvir, p.k. 4, 28850 Torrejon de Ardoz, Madrid, Spain. [26]Physikalisches Institut, Abteilung Weltraumforschung und Planetologie, Universität Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.

A recent disruption of the main-belt asteroid P/2010 A2

David Jewitt[1,2,3], Harold Weaver[4], Jessica Agarwal[5], Max Mutchler[6] & Michal Drahus[1]

Most inner main-belt asteroids are primitive rock and metal bodies in orbit about the Sun between Mars and Jupiter. Disruption, through high-velocity collisions or rotational spin-up, is believed to be the primary mechanism for the production and destruction of small asteroids[1,2] and a contributor to dust in the Sun's zodiacal cloud[3], while analogous collisions around other stars feed dust to their debris disks[4]. Unfortunately, direct evidence about the mechanism or rate of disruption is lacking, owing to the rarity of the events. Here we report observations of P/2010 A2, a previously unknown inner-belt asteroid with a peculiar, comet-like morphology. The data reveal a nucleus of diameter approximately 120 metres with an associated tail of millimetre-sized dust particles. We conclude that it is most probably the remnant of a recent asteroidal disruption in February/March 2009, evolving slowly under the action of solar radiation pressure, in agreement with independent work[5].

Object P/2010 A2 was first detected on 6 January 2010 in data from the US Airforce Lincoln Near Earth Asteroid Research (LINEAR) survey telescope[6] and was immediately classified as a short-period comet, based on the orbit and a diffuse appearance presumably caused by ejected dust. The orbital elements, however, are those of a main-belt asteroid (semimajor axis $a = 2.290$ astronomical units (AU), eccentricity $e = 0.1244$, inclination $i = 5.25°$), placing P/2010 A2 in the class of rare objects known as main-belt comets[7].

Ground-based observations taken in early January[8,9] revealed a peculiar morphology that was unlike any previously observed comet, suggesting an origin other than by the normal cometary process of water-ice sublimation. The dust appeared in a parallel-sided tail (sometimes called a 'trail') detached from the nucleus, whereas typical cometary tails have their origin in a dust coma surrounding the nucleus and are fan-shaped. Hubble Space Telescope images taken at higher angular resolution (Supplementary Table 1, Fig. 1) show a point-like nucleus (N) at the leading edge of a diffuse tail in which are embedded crossed filamentary

structures (AA and BB). The filaments are the source of particles for the tail, including several dust streaks barely resolved even at Hubble Space Telescope resolution (F). Several persistent but faint and diffuse subnuclei (C) appear along the filaments. The filament morphology did not change as the Earth crossed the orbital plane of P/2010 A2 on 9 February 2010 (for example, compare images from 29 January and 22 February in Fig. 2, at plane angles −0.9 and +0.9, respectively), showing that the filaments and sub-nuclei are not confined to the plane.

Unlike other main-belt comets[7], P/2010 A2 orbits in the inner regions of the belt where S-type asteroids are most abundant[10]. The S-type asteroids are refractory rocks, dominated by materials formed at high temperatures, not by ice. Indeed, ice is thermodynamically unstable at the 184 K temperature expected of an isothermal blackbody at 2.29 AU. The primary nucleus N is unresolved and fades in accordance with the inverse square law (see Supplementary Information). These properties together show that the nucleus is inert, with an estimated radius of about 60 m (Supplementary Information).

P/2010 A2 shows modest morphological evolution on timescales of months (Fig. 2), owing to changes in the distance (and resolution), the observing perspective and intrinsic changes in the object. We note a steady change in the position angle and a narrowing of the tail with time (Supplementary Table 3). From a model of dust motions including solar gravity and radiation pressure[11], we calculate the expected tail position angle as a function of the date of emission of the dust particles from the nucleus. Particles emitted at a given time with negligible relative speed lie on straight lines emerging from the nucleus (synchrones), with larger particles being closer to the nucleus. The position angles of these synchrones are plotted in Fig. 3, from which we infer dates of ejection in February/March 2009. The dust dynamical model is a function of β, the ratio of the radiation pressure force on a particle to the solar gravitational attraction. We find that particles in the field of view of the Hubble Space Telescope observations have $\beta < 2 \times 10^{-4}$,
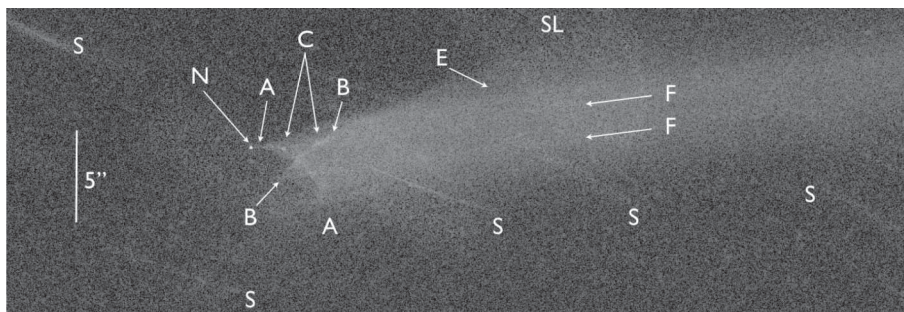
Figure 1 | Key to the major features in P/2010 A2 on UT 2010 Jan 25. The principal nucleus N leads an arcuate dust feature, AA. A second arcuate feature, BB, crosses AA at a large angle. Objects at C are distinct but diffuse features detected at more than one epoch. Particles emitted along AA and BB define the width of the main dust tail. A separate and very diffuse dust structure, E, extends beyond the boundaries of the tail. Linear dust streaks (striae) are visible embedded within the tail at F. Their narrowness shows that they emanate from discrete sources within the AA and BB arcs with negligible initial velocity. Interfering field stars are marked S, and SL is a band of internally scattered instrumental light which could not be removed by image processing.

[1]Department of Earth and Space Sciences, 3713 Geology Building, UCLA, 595 Charles Young Drive East, California 90095, USA. [2]Institute for Geophysics and Planetary Physics, UCLA, 595 Charles Young Drive East, California 90095, USA. [3]Department of Physics and Astronomy, UCLA, 430 Portola Plaza, Box 951547, Los Angeles, California 90095-1547, USA. [4]The Johns Hopkins University Applied Physics Laboratory, Johns Hopkins University, Johns Hopkins Road, Laurel, Maryland 20723, USA. [5]ESA-ESTEC, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands. [6]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.
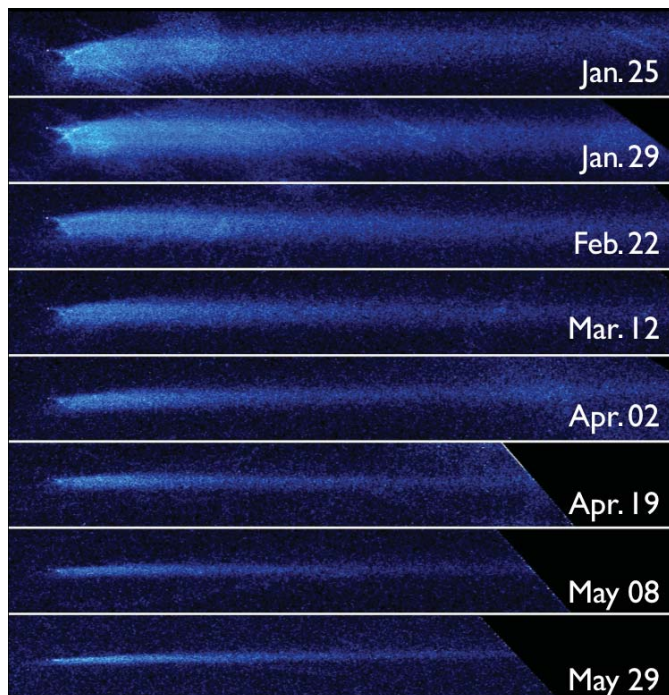
Figure 2 | Hubble Space Telescope images of P/2010 A2 at the eight indicated epochs. Images in each panel have been rotated so that the tail lies approximately horizontally. The images, from Wide Field Camera 3 on the Hubble Space Telescope, have 0.04-arcsecond pixels and are combinations of images with total integration times of about 2,600 s through the F606W filter. Each panel subtends 10 arcseconds in height. Numerous cosmic rays and trailed background objects have been removed from the data. Residual streaks in some panels (such as the diagonal streaks on 25 and 29 January 2010) are due to the incomplete removal of trailed background stars and galaxies.

corresponding to particle sizes larger than 1 mm rising to ~1 cm near the crossed filaments. The narrowing of the tail (Fig. 2) occurs because particles launched perpendicularly to the orbit reach maximum height above the orbit plane one quarter-orbit (10 months) after ejection. The width of the dust tail implies out-of-plane dust velocities $\delta v \approx 0.2 \, \text{m s}^{-1}$. Relative velocities measured between the nucleus N and sub-nuclei in the filaments (for example, between N and C in Fig. 1) are $\delta v < 0.2 \, \text{m s}^{-1}$.

The effective scattering cross-section of the dust tail is comparable to the area of a circle of radius $r_e = 2,100$ m. If contributed by particles in the millimetre to centimetre size range, this cross-section corresponds to a dust mass $M = (6-60) \times 10^7$ kg, equivalent to a sphere of the same density and having a radius $r = 17-36$ m (see Supplementary Information).

One possibility is that P/2010 A2 was disrupted by rotational bursting, perhaps caused by spin-up under the action of radiation torques (the timescale for spin-up is very uncertain but it can be less than a hundred thousand years for a sub-kilometre body[12,13]). If the dust following P/2010 A2 was produced by an impact, r gives an upper limit to the radius of the projectile, r_p, because, in a hypervelocity impact, orders-of-magnitude more mass is ejected from the target than is delivered by the projectile. We infer that the projectile was of the order of a few metres in radius, tiny compared to the primary nucleus. The velocity dispersion among asteroids in the main belt is $\Delta V \approx 5 \, \text{km s}^{-1}$ (ref. 1). From these parameters we infer that the energy per unit target mass in the impact responsible was $E/M = \frac{1}{2}(r_p/r_n)^3 \Delta V^2 \approx (10^3-10^4) \, \text{J kg}^{-1}$, where r_n is the radius of the nucleus. This range encompasses the E/M needed for catastrophic fragmentation in a direct impact[14]. Hypervelocity impact experiments[15] and calculations[16] show that most mass is displaced at low velocities, consistent with the speeds measured.

The expected interval between collisional disruptions of 0.1-km-diameter asteroids in the main belt is about one year[17], whereas damaging but non-disruptive impacts should be more frequent. Because the duration of visibility of the P/2010 A2 debris cloud exceeds
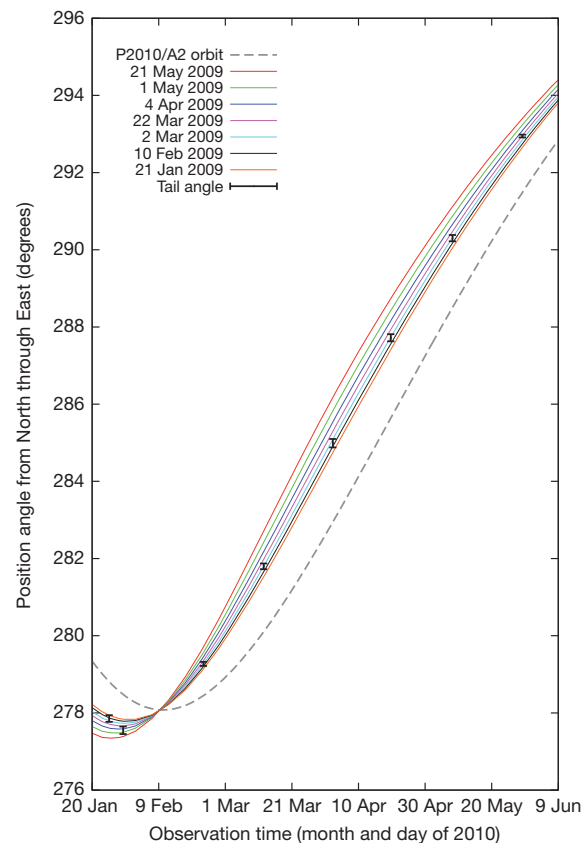
Figure 3 | Position angle of the tail as a function of time showing changes caused by the viewing geometry. Measured position angles of the tail (black symbols) are shown with error bars denoting one standard deviation. Calculated position angles of different synchrones (colour-coded curves) are shown as functions of the epoch of observation. The position angle of the projected orbit is shown as a dashed grey line. To measure the difference between the position angles of the tail and of the projected orbit, we rotated the images so as to align the x axis with the projected orbit. At constant intervals, we obtained profiles perpendicular to the orbit by averaging over 200 pixels parallel and 10 pixels perpendicular to the orbit. To each profile we fitted a Gaussian function. We then fitted a linear function to the peak of the Gaussian versus the distance from the nucleus. The slope and root-mean-square of the slope give us the position angle of the tail and the corresponding error bars. The coloured curves indicate the position angles of specific synchrones, that is, dust emitted at a specific date in 2009 (see synchrone labels) with zero relative velocity. Simulations demonstrate that dust emitted at a given time with zero speed is seen in projection along a straight line starting from the nucleus and with the distance to the nucleus proportional to the radiation pressure coefficient β with larger particles (with smaller β) closer to the nucleus for a given release time. For a given observation date, the position angle of the synchrones is a unique function of the time of emission. The coloured lines show the change of the synchrone position angles with time, primarily owing to the changing viewing geometry. In particular, all synchrones were projected to the south of the orbit before the Earth crossed the orbital plane of the comet on 9 February 2010, and to the north afterwards. The measured position angles of the tail are best matched by the 2 March 2009 synchrone and are inconsistent with synchrones more than a few weeks before or after that date.

one year, we should expect to find one or more similar objects at any time, in any all-sky survey with sensitivity equal to that of LINEAR or greater. Comparable disruption events occurring annually will release about 2 to 20 kg s^{-1} of dust into the zodiacal cloud, on average. This is only 0.1% to 1% of the 600–1,000 kg s^{-1} mass injection rate needed to keep the zodiacal cloud in steady state[18], suggesting that most of the mass comes from comets[19] or another source.

1. Bottke, W. F., Nolan, M. C., Greenberg, R. & Kolvoord, R. A. Velocity distributions among colliding asteroids. *Icarus* **107**, 255–268 (1994).

2. Holsapple, K. A. Spin limits of solar system bodies: from the small fast-rotators to 2003 EL61. *Icarus* **187,** 500–509 (2007).
3. Nesvorný, D. *et al.* Origin of the near-ecliptic circumsolar dust band. *Astrophys. J.* **679,** L143–L146 (2008).
4. Wyatt, M. C. Evolution of debris disks. *Annu. Rev. Astron. Astrophys.* **46,** 339–383 (2008).
5. Snodgrass, C. *et al.* A collision in 2009 as the origin of the debris trail of asteroid P/2010 A2. *Nature* doi: 10.1038/nature09453 (this issue).
6. Birtwhistle, P., Ryan, W. H., Sato, H., Beshore, E. C. & Kadota, K. Comet P/2010 A2 (LINEAR). *Cent. Bur. Electron. Telegr.* **2114** (2010).
7. Hsieh, H. H. & Jewitt, D. A population of comets in the main asteroid belt. *Science* **312,** 561–563 (2006).
8. Jewitt, D., Annis, J. & Soares-Santos, M. Comet P/2010 A2 (LINEAR). *IAU Circ.* **9109** (2010).
9. Licandro, J., Tozzi, G. P., Liimets, T., Cabrera-Lavers, A. & Gomez, G. Comet P/2010 A2 (LINEAR). *Cent. Bur. Electron. Telegr.* **2134** (2010).
10. Gradie, J. & Tedesco, E. Compositional structure of the asteroid belt. *Science* **216,** 1405–1407 (1982).
11. Agarwal, J. *et al.* The dust trail of Comet 67P/Churyumov-Gerasimenko between 2004 and 2006. *Icarus* **207,** 992–1012 (2010).
12. Rubincam, D. P. Radiative spin-up and spin-down of small asteroids. *Icarus* **148,** 2–11 (2000).
13. Taylor, P. A. *et al.* Spin rate of asteroid (54509) 2000 PH5 increasing due to the YORP effect. *Science* **316,** 274–277 (2007).
14. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142,** 5–20 (1999).
15. Michikami, T., Moriguchi, K., Hasegawa, S. & Fujiwara, A. Ejecta velocity distribution for impact cratering experiments on porous and low strength targets. *Planet. Space Sci.* **55,** 70–88 (2007).
16. Jutzi, M., Michel, P., Benz, W. & Richardson, D. C. Fragment properties at the catastrophic disruption threshold: the effect of the parent body's internal structure. *Icarus* **207,** 54–65 (2010).
17. Bottke, W. F. *et al.* Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* **179,** 63–94 (2005).
18. Leinert, C., Roser, S. & Buitrago, J. How to maintain the spatial distribution of interplanetary dust. *Astron. Astrophys.* **118,** 345–357 (1983).
19. Nesvorný, D. *et al.* Cometary origin of the zodiacal cloud and carbonaceous micrometeorites: implications for hot debris disks. *Astrophys. J.* **713,** 816–836 (2010).

LETTER

Laser cooling of a diatomic molecule

E. S. Shuman[1], J. F. Barry[1] & D. DeMille[1]

It has been roughly three decades since laser cooling techniques produced ultracold atoms[1–3], leading to rapid advances in a wide array of fields. Laser cooling has not yet been extended to molecules because of their complex internal structure. However, this complexity makes molecules potentially useful for a wide range of applications[4]. For example, heteronuclear molecules possess permanent electric dipole moments that lead to long-range, tunable, anisotropic dipole–dipole interactions. The combination of the dipole–dipole interaction and the precise control over molecular degrees of freedom possible at ultracold temperatures makes ultracold molecules attractive candidates for use in quantum simulations of condensed-matter systems[5] and in quantum computation[6]. Also, ultracold molecules could provide unique opportunities for studying chemical dynamics[7,8] and for tests of fundamental symmetries[9–11]. Here we experimentally demonstrate laser cooling of the polar molecule strontium monofluoride (SrF). Using an optical cycling scheme requiring only three lasers[12], we have observed both Sisyphus and Doppler cooling forces that reduce the transverse temperature of a SrF molecular beam substantially, to a few millikelvin or less. At present, the only technique for producing ultracold molecules is to bind together ultracold alkali atoms through Feshbach resonance[13] or photoassociation[14]. However, proposed applications for ultracold molecules require a variety of molecular energy-level structures (for example unpaired electronic spin[5,9,11,15], Omega doublets[16] and so on). Our method provides an alternative route to ultracold molecules. In particular, it bridges the gap between ultracold (submillikelvin) temperatures and the ~1-K temperatures attainable with directly cooled molecules (for example with cryogenic buffer-gas cooling[17] or decelerated supersonic beams[18]). Ultimately, our technique should allow the production of large samples of molecules at ultracold temperatures for species that are chemically distinct from bialkalis.

The laser cooling of atoms has allowed unprecedented access to ultracold temperatures. The power of laser cooling generally derives from the ability of certain atoms to continuously scatter photons from a laser. For instance, Doppler laser cooling relies on small but repetitive momentum kicks resulting from the absorption of red-detuned photons propagating counter to the motion of an atom. Doppler cooling of an atom of mass $m > 20$ AMU from room temperature to ultracold temperatures using visible light requires $>10^4$ photon scatters. To scatter this many photons, an atom must have a closed cycling transition in which each photon absorption is always followed by spontaneous decay back to the initial state.

However, there are no completely closed transitions in any real physical systems, and inevitably spontaneous decays to other states occur, usually before ultracold temperatures can be reached. Each additional populated level requires a 'repump' laser to return population to the main cycle, so that photon scattering can continue. Cycling transitions requiring one or two repump lasers are common in atomic systems, but they are quite difficult to find in molecules because of their vibrational and rotational degrees of freedom. Control over vibrational states is particularly problematic because there is no strict selection rule governing the branching ratios for the decay of an excited electronic state into different vibrational levels.

Instead these branching ratios are governed by the molecule's Franck–Condon factors (FCFs), which describe the overlap of the vibrational wavefunctions for different electronic states. For a typical molecule, the probability of returning to the original vibrational level after 10^4 photon scatters is extremely small. Furthermore, decay from a single excited state can populate up to three rotational levels per vibrational level, because rotational selection rules generally only require that $\Delta N = 0, \pm 1$. (We use N, v, F and M as the rotational, vibrational, total angular momentum and Zeeman quantum numbers respectively.) Because each substantially populated level requires an individually tunable repump laser, laser cooling of a molecule can easily require so many lasers as to be experimentally challenging.

Here we experimentally demonstrate direct laser cooling of a diatomic molecule by reducing the transverse velocity spread of a cryogenic beam of SrF in one dimension. Cryogenic buffer-gas beam sources produce highly directional beams with large fluxes for a variety of molecules, as described elsewhere[19,20]. Our scheme for eliminating the rotational and vibrational branching, necessary to ensure optical cycling, has been described previously[12]. We briefly recount the main points in the context of the experiment reported here. We use the $X^2\Sigma^+ \to A^2\Pi_{1/2}$ electronic transition of SrF for cycling. Use of the first excited state, $A^2\Pi_{1/2}$, ensures that no other electronic states can be populated by spontaneous decay[21]. The $A^2\Pi_{1/2}$ state has a large spontaneous decay rate, $\Gamma = 2\pi \times 7$ MHz (ref. 22), which allows the application of strong optical forces. We have chosen SrF primarily because its favourable FCFs dictate that only three vibrational levels will be significantly populated after 10^5 photon scatters[23], which in principle is more than sufficient to stop molecules in our cryogenic beam. Rotational branching is eliminated by driving an $N = 1 \to N' = 0$-type transition[24], where the prime indicates the excited state. Transitions of this type lead to optical pumping into dark ground-state Zeeman sublevels not excited by the laser[25]. For example, in this system the $F = 2$, $|M| = 2$ sublevels are dark when driven by linear laser polarization. We eliminate these dark states by applying a magnetic field oriented at an angle θ_B with respect to the fixed linear laser polarization, forcing the dark states to Larmor precess into bright states. Finally, radio-frequency sidebands on the lasers address all ground-state hyperfine and spin–rotation substructure. With this scheme, only three cooling lasers (one main pump laser, with $\lambda_{00} = 663.3$ nm, and two vibrational repump lasers, with $\lambda_{10} = 686.0$ nm and $\lambda_{21} = 685.4$ nm) are required, as shown in Fig. 1. (Here λ_{ij} is the wavelength of the $X(v = i) \to A(v' = j)$ transition.)

Use of the $N = 1 \to N' = 0$-type transition leads to a lower photon scattering rate than a traditional two-level atomic cycling transition[12,24]. To overcome this problem, we use an elongated transverse cooling region (Supplementary Information, section 1, and Supplementary Fig. 1). The SrF beam is intersected by the three cooling laser beams at nearly 90°. The laser beams are reflected back and forth at a slight angle such that they intersect the SrF beam ~75 times in the 15-cm-long cooling region. At the end of the cooling region, the beams are nearly retroreflected, resulting in the formation of standing waves. The combination of magnetic field remixing of Zeeman sublevels and standing waves can lead to Sisyphus forces in addition to Doppler forces, as has been observed in atomic systems[26–28]. Ten centimetres

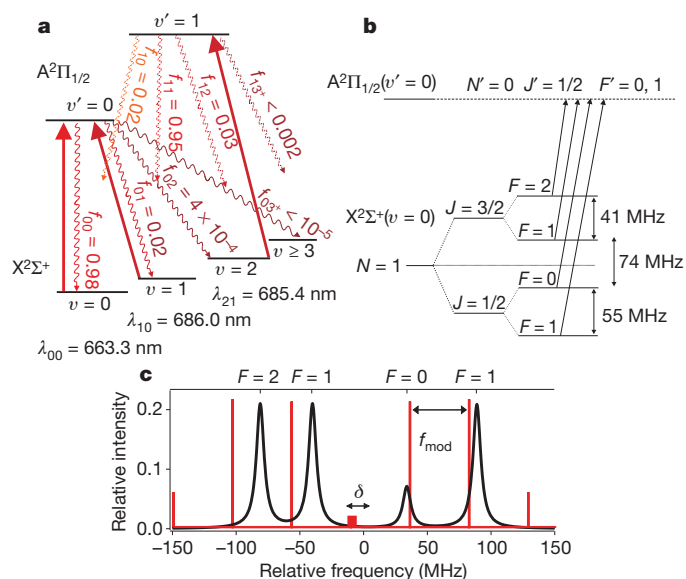[1]Department of Physics, Yale University, PO Box 208120, New Haven, Connecticut 06520, USA.

Figure 1 | Energy-level structure in SrF. a, Relevant electronic and vibrational structure in SrF. Solid upward lines indicate laser-driven transitions in the experiment at wavelengths $\lambda_{v'v}$. Wavy lines indicate spontaneous decays from the A($v = 0$) state (solid) and the A($v = 1$) state (dashed) with calculated FCFs $f_{v'v}$. **b**, Relevant rotational energy levels, splittings[29] and transitions (vertical arrows) in the SrF cycling scheme. SrF has an unpaired electron spin $S = 1/2$ that splits the X($N = 1$) level into $J = N \pm S$ levels through spin–rotation interaction, and ^{19}F has nuclear spin $I = 1/2$ that splits J into $F = J \pm 1/2$ levels through ground-state hyperfine (HFS) interactions. HFS splitting is not resolved in the A state[30]. **c**, Laser (vertical lines) and molecular (curved) spectra for addressing all HFS ground-state sublevels. Molecular lines are shown with a power-broadened linewidth of 1.5Γ and experimentally determined line strength. Electro-optic modulation of the laser frequency with index $m = 2.6$ and frequency $f_{\mathrm{mod}}^{\mathrm{p}} \approx 43$ MHz produces four sidebands of nearly equal strengths that almost match the HFS splitting. The laser spectrum shown has $f_{\mathrm{mod}}^{\mathrm{p}} = 46.4$ MHz and $\delta_{\mathrm{p}} = -1.5\Gamma$, which are experimentally determined to be optimal λ_{00} laser parameters for Doppler cooling. The laser carrier frequency corresponding to $\delta_{\mathrm{p}} = 0$ is defined experimentally by the frequency (thick central line in laser spectrum) that produces maximal LIF, and δ_{p} corresponds to the frequency difference of the carrier laser frequency from this value.

downstream from the end of the cooling region, laser-induced fluorescence (LIF) is imaged to obtain the spatial distribution of the molecular beam. This spatial distribution maps onto the velocity distribution of the molecules with a resolution of ~1 m s^{-1}, so from such images we can extract information about the velocity-dependent forces applied to the molecules, as well as the beam's transverse temperature, T.

We find that there are two cooling regimes with qualitatively different features that depend critically on the magnitude of the applied magnetic field, B. In Fig. 2a and Fig. 2b, we show data for $B = 5$ and 0.6 G, respectively, that are characteristic of these regimes. In both regimes, we find that θ_B is unimportant if $\theta_B \neq 0°$, 90°. In all cases, we observe that the total integrated LIF signal is constant to within the experimental reproducibility (~5%); therefore, changes in the molecular spatial distribution accurately reflect changes in the velocity distribution of the molecular beam. The top and bottom plots in Fig. 2a, b show representative molecular beam data for two different main pump laser carrier frequencies, $\delta_{\mathrm{p}} = \pm 1.5\Gamma$. Here δ_{p} is not a detuning in the usual sense; in this system, there is not a single, well-defined value of the detuning for the pump and repump lasers because each laser contains several sideband frequencies that each interact with multiple transitions between the ground and excited states. Instead, we define $\delta_{\mathrm{p}} = 0$ and $\delta_{\mathrm{r}} = 0$ for the pump and repump lasers experimentally by determining the pump and repump laser carrier frequencies that produce maximal LIF. For $\delta_{\mathrm{p}} = \pm 1.5\Gamma$, the detuning of the nearest sideband from each transition ranges from zero to around $\pm 1.5\Gamma$,
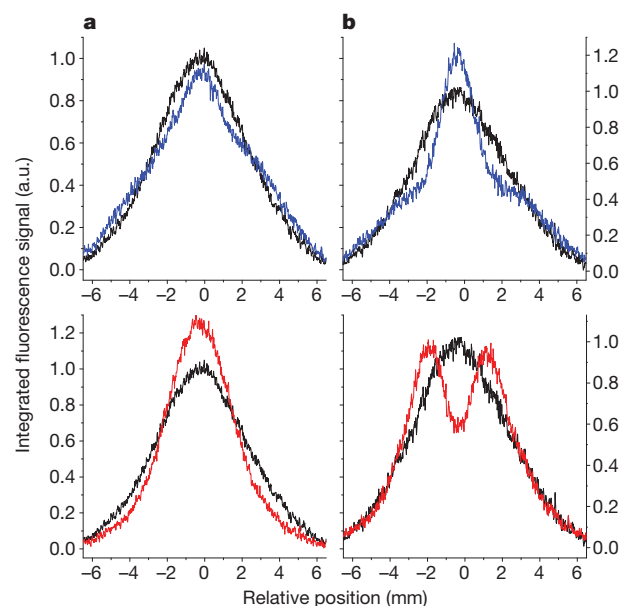
Figure 2 | Laser cooling of SrF. a, b, LIF in the probe region without cooling lasers in the interaction region (black), with cooling lasers and main pump laser red-detuned by $\delta_{\mathrm{p}} = -1.5\Gamma$ (red), and with cooling lasers and main pump laser blue-detuned by $\delta_{\mathrm{p}} = +1.5\Gamma$ (blue), for $f_{\mathrm{mod}}^{\mathrm{p}} = 46.4$ MHz, $B = 5$ G and $\theta_B = 60°$ (**a**) and for $f_{\mathrm{mod}}^{\mathrm{p}} = 43.2$ MHz, $B = 0.6$ G and $\theta_B = 30°$ (**b**). In **a**, cooling or heating of the beam is observed for red or, respectively, blue detuning, both in accordance with expectations for Doppler forces. In **b**, cooling or heating is observed for blue or, respectively, red detuning, as is expected for Sisyphus forces. In all cases, the total integrated signal is the same to within 5%. The blue-detuned curves systematically have the lowest total integrated signal, probably because some molecules have been pushed outside the imaging region by the Doppler heating force. We observe qualitatively similar behaviour in both cases for pump laser modulation frequencies 42 MHz $< f_{\mathrm{mod}}^{\mathrm{p}} <$ 47 MHz. However, the $f_{\mathrm{mod}}^{\mathrm{p}}$ values shown here were chosen because they produced the clearest Doppler and Sisyphus effects. This plot shows raw data, with no rescaling applied. a.u., arbitrary units.

indicating that $\delta_{\mathrm{p}} < 0$ corresponds to net average red detuning and $\delta_{\mathrm{p}} > 0$ corresponds to net average blue detuning. (Here and throughout the rest of the paper, both vibrational repump lasers have modulation frequencies $f_{\mathrm{mod}}^{\mathrm{r}} = 43$ MHz and $\delta_{\mathrm{r}} = 0$. The main pump laser has the modulation frequencies, $f_{\mathrm{mod}}^{\mathrm{p}}$, indicated in the figure legends.)

For a red-detuned main pump laser, $\delta_{\mathrm{p}} = -1.5\Gamma$, and $B = 5$ G, we observe significant narrowing of the molecular beam and enhancement of molecules with low transverse velocity, v_{t}, as shown in the bottom plot in Fig. 2a. This corresponds to a reduction in the spread of v_{t}, and is a signature of Doppler cooling. Also, it is evident that the entire molecular beam experiences cooling forces, indicating that the cooling force is significant for all velocities v_{t} in the molecular beam. The molecular beam is constrained by collimating apertures to have $v_t < v_{\mathrm{Max}}^{\mathrm{Beam}} \approx 4$ m s^{-1}. For a blue-detuned main pump laser, $\delta_{\mathrm{p}} = +1.5\Gamma$, and $B = 5$ G, we observe depletion of low-v_{t} molecules and broadening of the molecular beam as shown in the top plot of Fig. 2a, as expected for Doppler heating. Under these conditions, there is also a small but noticeable sharp feature in the centre of the molecular beam, indicating some residual cooling of the remaining molecules with low v_{t}.

In Fig. 2b, we show data characteristic of a small applied field, $B = 0.6$ G, which are strikingly different from the data in Fig. 2a. The most significant difference between Fig. 2a and Fig. 2b is that cooling occurs for detunings of opposite sign. For a red-detuned main pump laser, $\delta_{\mathrm{p}} = -1.5\Gamma$, we observe (Fig. 2b, bottom plot) two relatively sharp peaks, neither of which is centred around $v_{\mathrm{t}} = 0$. This indicates the heating of molecules with low $|v_{\mathrm{t}}|$ values and accumulation around two stable velocity points with $v_{\mathrm{t}} \neq 0$. For a blue-detuned

main pump laser, $\delta_p > +1.5\Gamma$, we observe (Fig. 2b, bottom plot) a sharp central spike and a large enhancement of low-v_t molecules. This feature results from the strong cooling of molecules over a small range of v_t around zero. Meanwhile, molecules with larger $|v_t|$ values experience small heating forces, resulting in a very slight enhancement at large $|v_t|$.

A complete characterization of the detailed cooling forces responsible for these observations would require the solution of the optical Bloch equations for this system. All relevant quantities (detunings, Larmor frequencies, Rabi frequencies and Γ) are the same within factors of order one, so the full 44-level system, driven by twelve laser frequencies, each interacting with multiple levels, must be solved. Such a calculation is beyond the scope of this work; however, the main features we have observed are common to any system with magnetically remixed dark sublevels driven by a standing wave[26–28]. The simplest such system is an $F = 1 \rightarrow 0$ transition driven by a single laser frequency, and it provides substantial insight into our observations. A linearly polarized, blue- or red-detuned laser of wavelength λ produces an alternating-current Stark shift that respectively attracts or repels the $F' = 0$, $M = 0$ and $F' = 1$, $M' = 0$ levels while leaving the $F = 1$, $M = \pm 1$ levels unperturbed to first order, as shown schematically in Supplementary Fig. 2. In a standing wave, the $F = 1$, $M = 0$ level undergoes a spatially periodic alternating-current Stark shift with period $\lambda/2$. For blue or red detuning, molecules move up a potential hill in this level, respectively losing or gaining kinetic energy, before they are pumped at a rate $\gamma_p < \Gamma$ into the $F = 1$, $M = \pm 1$ sublevels that are dark. If a magnetic field is applied at $\theta_B \neq 0°, 90°$, then the molecules precess from the dark $M = \pm 1$ sublevels back into the $M = 0$ levels at the nodes of the standing wave at a rate ω_B. The Sisyphus force is maximized for $v_t = v_{\mathrm{Max}}^{\mathrm{Sis}} \approx (\lambda/4)\gamma_p$, and $\omega_B^{\mathrm{Sis}} \approx \gamma_p$ (Supplementary Information, section 2). The effective velocity range of the Sisyphus force is defined by $v_{\mathrm{Max}}^{\mathrm{Sis}}$. Larger values of ω_B produce a much smaller Sisyphus force because the molecules are pumped back and forth between bright and dark states at random points in the standing wave. Because the molecules spend more time in the bright states, the photon scattering rate is higher and the Doppler force is larger. Simple arguments suggest that $\omega_B^{\mathrm{Dop}} \approx \Gamma$ for maximum Doppler cooling forces, and that Doppler forces occur over a larger range of velocities, $0 < v_t < v_{\mathrm{Max}}^{\mathrm{Dop}} = \lambda\Gamma/2\pi = 4\ \mathrm{m\,s}^{-1}$ (Supplementary Information, section 2).

This qualitative discussion provides substantial insight into our observations. At low magnetic fields, we expect to observe Sisyphus forces. These forces are characterized by cooling for blue detuning and heating for red detuning, for molecules with low v_t values, as observed in Fig. 2b. At higher magnetic fields, we expect to observe Doppler forces. These forces are characterized by cooling for red detuning and heating for blue detuning. Because $v_{\mathrm{Max}}^{\mathrm{Beam}} \approx v_{\mathrm{Max}}^{\mathrm{Dop}}$, Doppler cooling forces should affect the entire molecular beam as observed in the bottom plot of Fig. 2a. In Supplementary Information, section 2, we provide the argument for the estimate $\gamma_p \approx \Gamma/42$, which yields $v_{\mathrm{Max}}^{\mathrm{Sis}} \approx 0.2\ \mathrm{m\,s}^{-1}$. This is much smaller than $v_{\mathrm{Max}}^{\mathrm{Beam}} \approx v_{\mathrm{Max}}^{\mathrm{Dop}} \approx 4\ \mathrm{m\,s}^{-1}$, and is consistent with our observations.

Of course, neither regime can be characterized purely by either Sisyphus or Doppler forces. In the moderate-B regime, residual Sisyphus forces lead to a slight additional broadening or narrowing in the low-velocity part of the distribution for red or, respectively, blue detuning. This residual narrowing is clear in the top plot of Fig. 2a, but the broadening is too small to observe in the bottom plot. In the low-B regime, residual Doppler forces result in small heating or cooling over a large range of velocities for blue or, respectively, red detuning. This gives rise to non-zero unstable or stable velocities, $v_t = \pm v_0$, where the net force is zero and population is depleted or, respectively, accumulates. These stable points, with $v_t = \pm v_0$, are clear in the bottom plot of Fig. 2b. In the top plot of Fig. 2b, the unstable velocities are depleted and there is some residual Doppler broadening in the wings of the molecular beam. Although the arguments presented here derive from

an $F = 1 \rightarrow F' = 0$ example system, they are common to any $F \rightarrow F' \leq F$ system. Furthermore, the effects described here have been observed in several such atomic systems[26–28].

To clearly illustrate the magnetic field dependence of the cooling force, we show the magnetic field dependence of the cooling forces for a red-detuned pump laser, $\delta_p = -1.5\Gamma$, in Fig. 3a. As shown for very small magnetic fields, the width of the molecular beam increases owing to Sisyphus heating effects. This increase is followed by a sharp decrease in the molecular beam width that corresponds to Doppler cooling. The width of the molecular beam is quite insensitive to the magnetic field over a range of intermediate field amplitudes between 2 and 6 G. Finally, at magnetic fields higher than 6 G the Doppler forces are reduced because the magnetic field artificially broadens the transitions, resulting in lower scattering rates and lower Doppler forces. We estimate values for ω_B in Supplementary Information, section 2, that yield $B_{\mathrm{Dop}} \approx 5\ \mathrm{G}$ and $B_{\mathrm{Sis}} \approx 0.2\ \mathrm{G}$ for the maximum Sisyphus and Doppler forces, in reasonable agreement with our observations. Once again, detailed comparison would require the full solution of the optical Bloch equations for our system.

As an example of the complex features present in this system, we show the frequency dependence of the width of the SrF beam under Doppler-force-dominated conditions of $B = 5\ \mathrm{G}$, for various pump laser detunings, δ_p, in Fig. 3b. The frequency dependence of the width is substantially more complicated than that of a typical two-level system. As shown, the force oscillates many times between heating and cooling for $-250\ \mathrm{MHz} < \delta_p < 250\ \mathrm{MHz}$. However, this complicated structure is amenable to a simple interpretation. As the pump laser frequency is varied, the laser frequency nearest to each molecular transition oscillates between red detuned (cooling) and blue detuned (heating) with a variation that is nearly periodic in the sideband frequency. The frequency dependence in the Sisyphus regime (Fig. 3c) is slightly more complicated, but for small detunings, $-50\ \mathrm{MHz} < \delta_p < 50\ \mathrm{MHz}$, the Sisyphus force has the opposite sign to the Doppler forces as expected. In Supplementary Information, section 4, and Supplementary Fig. 4, we
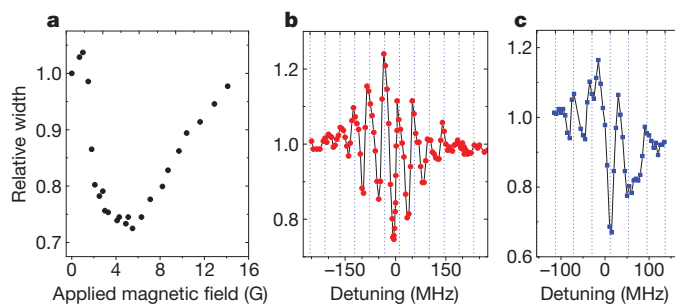

Figure 3 | Magnetic field and frequency dependence of the cooling forces.
a, Magnetic field dependence of the cooling force. The width of the molecular beam is plotted for $f_{\mathrm{mod}}^P = 46.4\ \mathrm{MHz}$, $\delta_p = -1.5\Gamma$ and various magnetic fields at $\theta_B = 30°$. For $B < 2\ \mathrm{G}$, the beam is broader owing to Sisyphus forces. At $B \approx 2\ \mathrm{G}$, we see a sharp transition between Sisyphus-dominated heating forces and Doppler-dominated cooling forces. For $B > 6\ \mathrm{G}$, the width increases as described in the main text. **b, c**, Frequency dependence of the cooling forces under Doppler-dominated ($B = 5\ \mathrm{G}$; **b**) and Sisyphus-dominated ($B = 0.6\ \mathrm{G}$; **c**) conditions. The solid lines between data points are drawn as guides to the eye. The vertical dashed lines are spaced by the sideband frequencies of $46.4\ \mathrm{MHz}$ (**b**) and $43.2\ \mathrm{MHz}$ (**c**), and illustrate the dependence of the oscillations on sideband frequency. Describing the molecular beam simply by a width is potentially problematic in the Sisyphus-dominated regime where more than one velocity class is present. Nonetheless, the total widths of the beams for the red-detuned and blue-detuned pump lasers are larger and, respectively, smaller than the unperturbed beam width, and the width thus provides a reasonable description of the molecular beam. The asymmetry around $\delta_p = 0$ in **b** and **c** underscores the complex nature of this system and is probably representative of the fact that every sideband does not have the same detuning from the nearest transition. Therefore, the detuning of each sideband from the nearest transition is not zero for $\delta_p = 0$. Furthermore, the absolute values of the individual sideband detunings are not the same for $+\delta_p$ and $-\delta_p$.

also show and discuss the power dependences of the Sisyphus and Doppler cooling forces, which are both in reasonable agreement with expectations.

Finally, we discuss our determination of the temperature, T, of the molecules after they are cooled. The unperturbed beam is constrained by collimating apertures to have temperature $T_0 = 50$ mK. The long interaction region prevents a precise determination of T for the cooled beam, because the beam continues to expand as it experiences an imperfectly known distribution of cooling forces throughout the interaction region. To estimate T, we calculate the molecular beam profile using a Monte Carlo simulation of classical particles subjected to the qualitatively expected force–velocity profile (Supplementary Information, section 3, and Supplementary Fig. 3). Using these simulations, we find that $T = T_{Sis} \approx 300\ \mu K$ for the Sisyphus regime in the top plot of Fig. 2b. We also estimate a conservative upper limit on the temperature of $T < T_{Sis}^{Max} = 5$ mK.

For the Doppler regime, we find that $T \approx T_{Dop} \approx 5$ mK and that $T < T_{Dop}^{Max} = 15$ mK. These values of T for the Doppler regime are consistent with the final temperature expected if the molecules are subjected to $N_{sc} \approx 500$–$1,000$ photon scatters. This value of N_{sc} agrees with expectations based on the previously observed scattering rate for this system and the roughly known interaction time. Importantly, the total integrated signal of the Doppler-cooled beam and the unperturbed beam are the same to within the level of experimental reproducibility ($\sim 5\%$), indicating that our cycling scheme is highly closed.

Our results have immediate implications for a number of future experiments. For example, the one-dimensional cooling and optical cycling demonstrated here could drastically improve the statistical sensitivity of searches for electron electric dipole and nuclear anapole moments[9,11], by providing more-collimated molecular beams and enhanced detection efficiency in these experiments. In addition, the combination of one-dimensional cooling and a highly closed cycling transition is a step towards the laser cooling of molecules in three dimensions. Given the calculated FCFs, a large fraction of molecules should scatter the $\sim 40,000$ photons necessary to bring a beam of SrF to a stop and subsequently load the molecules into a trap. Furthermore, the experimentally determined loss rate of molecules in this system is $<5\%$ for $N_{sc} \approx 1,000$, implying that $>10\%$ of the molecules can be brought to rest, given sufficient interaction time. The laser cooling techniques presented here are limited, from a practical standpoint, to those molecules that have closed electronic transitions with diagonal FCFs and therefore require relatively few lasers. For this reason, these techniques are applicable only to a small fraction of diatomic molecules. However, because the set of diatomic molecules is very large, this subset contains a significant number of molecules[23]. We are aware of ~ 10 diatomic molecules with a wide range of internal structures that seem amenable to laser cooling with similar methods. Laser cooling such molecules to ultracold temperatures would open the door to the study of a wide variety of new physical phenomena.

1. Chu, S. The manipulation of neutral particles. *Rev. Mod. Phys.* **70,** 685–706 (1998).
2. Cohen-Tannoudji, C. N. Manipulating atoms with photons. *Rev. Mod. Phys.* **70,** 707–719 (1998).
3. Phillips, W. D. Laser cooling and trapping of neutral atoms. *Rev. Mod. Phys.* **70,** 721–741 (1998).
4. Carr, L., DeMille, D., Krems, R. & Ye, J. Cold and ultracold molecules: science, technology and applications. *N. J. Phys.* **11,** 055049 (2009).
5. Pupillo, G. Micheli, A., Büchler, H.-P. & Zoller, P. in *Cold Molecules: Theory, Experiment, Applications* (eds Krems, R., Friedrich, B. & Stwalley, W. C.) Ch. 12 (CRC Press, 2009).
6. DeMille, D. Quantum computation with trapped polar molecules. *Phys. Rev. Lett.* **88,** 067901 (2002).
7. Balakrishnan, N. & Dalgarno, A. Chemistry at ultracold temperatures. *Chem. Phys. Lett.* **341,** 652–656 (2001).
8. Krems, R. V. Cold controlled chemistry. *Phys. Chem. Chem. Phys.* **10,** 4079–4092 (2008).
9. Tarbutt, M., Hudson, J., Sauer, B. & Hinds, E. Prospects for measuring the electric dipole moment of the electron using electrically trapped polar molecules. *Faraday Discuss.* **142,** 37–56 (2009).
10. Flambaum, V. V. & Kozlov, M. G. Enhanced sensitivity to the time variation of the fine-structure constant and m_p/m_e in diatomic molecules. *Phys. Rev. Lett.* **99,** 150801 (2007).
11. DeMille, D., Cahn, S. B., Murphree, D., Rahmlow, D. A. & Kozlov, M. G. Using molecules to measure nuclear spin-dependent parity violation. *Phys. Rev. Lett.* **100,** 023003 (2008).
12. Shuman, E. S., Barry, J. F., Glenn, D. R. & DeMille, D. Radiative force from optical cycling on a diatomic molecule. *Phys. Rev. Lett.* **103,** 223001 (2009).
13. Ni, K.-K. *et al.* A high phase-space-density gas of polar molecules. *Science* **322,** 231–235 (2008).
14. Sage, J. M., Sainis, S., Bergeman, T. & DeMille, D. Optical production of ultracold polar molecules. *Phys. Rev. Lett.* **94,** 203001 (2005).
15. André, A. *et al.* A coherent all-electrical interface between polar molecules and mesoscopic superconducting resonators. *Nature Phys.* **2,** 636–642 (2006).
16. Vutha, A. C. *et al.* Search for the electric dipole moment of the electron with thorium monoxide. *J. Phys. At. Mol. Opt. Phys.* **43,** 074007 (2010).
17. Weinstein, J. D., deCarvalho, R., Guillet, T., Friedrich, B. & Doyle, J. M. Magnetic trapping of calcium monohydride molecules at millikelvin temperatures. *Nature* **395,** 148–150 (1998).
18. Bethlem, H. L. *et al.* Electrostatic trapping of ammonia molecules. *Nature* **406,** 491–494 (2000).
19. Maxwell, S. E. *et al.* High-flux beam source for cold, slow atoms or molecules. *Phys. Rev. Lett.* **95,** 173201 (2005).
20. Patterson, D. & Doyle, J. M. Bright, guided molecular beam with hydrodynamic enhancement. *J. Chem. Phys.* **126,** 154307 (2007).
21. Allouche, A. R., Wannous, G. & Aubért-Frecon, M. A ligand-field approach for the low-lying states of Ca, Sr and Ba monohalides. *Chem. Phys.* **170,** 11–22 (1993).
22. Dagdigian, P. J., Cruse, H. W. & Zare, R. N. Radiative lifetimes of the alkaline earth monohalides. *J. Chem. Phys.* **60,** 2330–2339 (1974).
23. Di Rosa, M. D. Laser-cooling molecules. *Eur. Phys. J. D* **31,** 395–402 (2004).
24. Stuhl, B. K., Sawyer, B. C., Wang, D. & Ye, J. A magneto-optical trap for polar molecules. *Phys. Rev. Lett.* **101,** 243002 (2008).
25. Berkeland, D. J. & Boshier, M. G. Destabilization of dark states and optical spectroscopy in Zeeman-degenerate atomic systems. *Phys. Rev. A* **65,** 033413 (2002).
26. Emile, O. *et al.* Magnetically assisted Sisyphus effect. *J. Phys. II Fr.* **3,** 1709–1733 (1993).
27. Gupta, R., Padua, S., Xie, C., Batelaan, H. & Metcalf, H. Simplest atomic system for sub-Doppler laser cooling. *J. Opt. Soc. Am. B* **11,** 537–541 (1994).
28. Plimmer, M. D. *et al.* 2D laser collimation of a cold Cs beam induced by a transverse B field. *JETP Lett.* **82,** 17–21 (2005).
29. Childs, W. J., Goodman, L. S. & Renhorn, I. Radio-frequency optical double-resonance spectrum of SrF: the $X^2\Sigma^+$ state. *J. Mol. Spectrosc.* **87,** 522–533 (1981).
30. Kändler, J., Martell, T. & Ernst, W. Electric dipole moments and hyperfine structure of SrF $A^2\Pi$ and $B^2\Sigma^+$. *Chem. Phys. Lett.* **155,** 470–474 (1989).

An amorphous solid state of biogenic secondary organic aerosol particles

Annele Virtanen[1], Jorma Joutsensaari[2], Thomas Koop[3], Jonna Kannosto[1], Pasi Yli-Pirilä[4], Jani Leskinen[4], Jyrki M. Mäkelä[1], Jarmo K. Holopainen[4], Ulrich Pöschl[5], Markku Kulmala[6,7], Douglas R. Worsnop[2,6,8,9] & Ari Laaksonen[2,9]

Secondary organic aerosol (SOA) particles are formed in the atmosphere from condensable oxidation products of anthropogenic and biogenic volatile organic compounds (VOCs)[1–7]. On a global scale, biogenic VOCs account for about 90% of VOC emissions[1,8] and of SOA formation (90 billion kilograms of carbon per year)[1–4]. SOA particles can scatter radiation and act as cloud condensation or ice nuclei[1,2,5,9,10], and thereby influence the Earth's radiation balance and climate[1,2,5,9,10]. They consist of a myriad of different compounds with varying physicochemical properties, and little information is available on the phase state of SOA particles. Gas–particle partitioning models usually assume that SOA particles are liquid[1,5,11], but here we present experimental evidence that they can be solid under ambient conditions. We investigated biogenic SOA particles formed from oxidation products of VOCs in plant chamber experiments and in boreal forests within a few hours after atmospheric nucleation events. On the basis of observed particle bouncing in an aerosol impactor and of electron microscopy we conclude that biogenic SOA particles can adopt an amorphous solid—most probably glassy—state. This amorphous solid state should provoke a rethinking of SOA processes because it may influence the partitioning of semi-volatile compounds, reduce the rate of heterogeneous chemical reactions, affect the particles' ability to accommodate water and act as cloud condensation or ice nuclei, and change the atmospheric lifetime of the particles[12–15]. Thus, the results of this study challenge traditional views of the kinetics and thermodynamics of SOA formation and transformation in the atmosphere and their implications for air quality and climate.

The plant chamber experiments were carried out using living Scots pine seedlings as natural sources of biogenic VOCs, because terpenoids emitted from living plants are more representative of natural atmospheric conditions than artificial VOCs (for example, α-pinene) or the mixtures of VOCs commonly used in smog chamber experiments[16,17]. In this study we focused on the physical and morphological properties of aerosol particles formed from both O3-initiated and OH-initiated oxidation of pine VOC emissions. The evolution of the aerosol particle size distributions were measured by two scanning mobility particle sizers (SMPSs) and one electrical low-pressure impactor (ELPI); see Methods. The twelve-stage cascade impactor of the ELPI is constructed such that larger particles deposit on the upper impactor plates, and smaller particles deposit on the subsequent lower stages. The particles are charged, so the current induced in each stage by collected particles can be measured.

One general problem with impactor measurements, however, is the efficiency with which particles of different size, shape, morphology and physical state are collected during the collision process[18,19]. For example, liquid particles or other easily deformed materials usually tend to adhere to the impaction plates upon collision (Fig. 1c), whereas solid particles are captured less efficiently owing to bouncing[18]

(Fig. 1b). In a cascade impactor, the aerosol particle bounce perturbs the measurement because larger particles are transferred to lower stages when they bounce, thus tilting the inferred size distribution to smaller particles (see Supplementary Information). This characteristic feature of an impactor is usually considered an undesired artefact, but here we use it as a tool for studying the physical state of the particles.

Particle bounce was observed during particle growth in the plant chamber, because a large excess current of particles was measured in the back-up filter and the lowest impactor stages with cut-off diameters smaller than 30 nm. This was the case even when the maximum of the size distribution was as high as 100 nm and no small particles were present, as confirmed by simultaneous SMPS data. To ensure that the measured excess current was indeed due to particle bounce, the smooth collection plates were replaced by more effective porous plates (moistened with oil to increase adhesion), resulting in a nearly complete removal of bounce (see Supplementary Figs 1b and 3). To
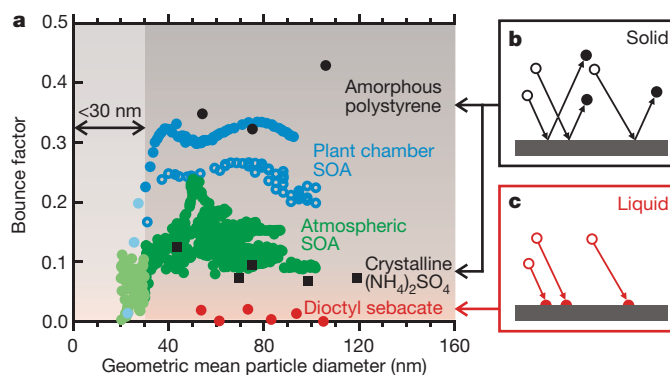
Figure 1 | Observations and interpretation of particle bounce of SOA and reference particles. a, The bounce factor of aerosol particles is shown as a function of mode geometric mean diameter for pine-derived particles nucleated and grown in the chamber (blue circles). The particle bounce was detected for both O3-initiated (solid blue circles) and OH-initiated oxidation (open blue circles). In the latter case SO2 was also added to the chamber and sulphate ions were detected in the particles by the aerosol mass spectrometer. Relative humidity was $30 \pm 5\%$ during the chamber experiments. The particle bounce factor for recently formed atmospheric aerosol particles measured in Hyytiälä (sampling relative humidity: 10–20%) during 15–18 May 2009 is also shown (green solid circles). For comparison, the bounce factors measured for materials known to be liquid (red circles), crystalline (black squares), and amorphous solid (black circles) are shown. Bounce factor data for geometric mean diameter less than 30 nm cannot be determined reliably by our method; see Supplementary Information. Therefore, newly nucleated sub-30-nm particles might well be in a solid state despite the apparently small bounce factor values (lightly shaded data in **a**). **b, c,** Schematics indicating the impact process of solid (**b**) and liquid (**c**) particles onto a smooth surface are also shown. Solid particles are more likely to bounce, while liquid particles tend to adhere to the surface.

[1]Department of Physics, Tampere University of Technology, P O Box 692, 33101 Tampere, Finland. [2]Department of Physics and Mathematics, University of Eastern Finland, Kuopio Campus, P O Box 1627, 70211 Kuopio, Finland. [3]Department of Chemistry, Bielefeld University, 33615 Bielefeld, Germany. [4]Department of Environmental Science, University of Eastern Finland, Kuopio Campus, P O Box 1627, 70211 Kuopio, Finland. [5]Biogeochemistry Department, Max Planck Institute for Chemistry, 55128 Mainz, Germany. [6]Department of Physics, University of Helsinki, P O Box 64, 00014 Helsinki, Finland. [7]Department of Applied Environmental Science, Stockholm University, 10691 Stockholm, Sweden. [8]Aerodyne Research, 45 Manning Road, Billerica, Massachusetts 08121-3976, USA. [9]Finnish Meteorological Institute, PO Box 503, 00101 Helsinki, Finland.

quantify the observed bounce, we define the fraction of excess current (resulting from the charges carried by the bounced particles) in the lowest impactor stages (those with cut-off diameters smaller than 30 nm and the back-up filter) relative to the idealized signal inferred from SMPS data as the 'bounce factor' (see Supplementary Information for derivations).

In Fig. 1a, the bounce factor is shown as a function of the particles' geometric mean diameter for pine-derived particles nucleated and grown in the chamber (blue circles). A clear bounce was observed in all cases, and for particles larger than 30 nm the bounce factor seems to be almost unaffected by particle size or particle age. This is consistent with chemical information obtained simultaneously by a quadrupole aerosol mass spectrometer, which showed that the composition of the particles remains essentially constant during the growth process. According to the aerosol mass spectrometer spectra the plant chamber particles represent less oxidized organic aerosol (see Supplementary Information). The observed bounce strongly suggests that the SOA particles are in a solid rather than liquid state. To verify this conclusion we have performed experiments with reference aerosol particles of known physical state. As expected, bounce is negligible for liquid droplets of moderate viscosity (dioctyl sebacate; red circles in Fig. 1a). In contrast, crystalline $(NH_4)_2SO_4$ particles (at a relative humidity of 12%) and amorphous solid polystyrene particles do show significant bounce (black squares and circles, respectively, in Fig.1a). We verified that the 50-nm polystyrene particles were in an amorphous solid state by determining their glass transition temperature at about 97 °C (see Supplementary Information). The bounce factor of SOA particles measured in the plant chamber closely matches that of polystyrene particles, suggesting that the SOA particles are also in a glassy (amorphous solid) state at the moment of impact.

To confirm these observations by an independent method, SOA particles were collected on electron microscopy grids. Figure 2a shows a scanning electron microscope (SEM) image of approximately 100-nm-sized particles formed from the OH-initiated oxidation products of pine emissions in the presence of SO_2, that is, of those particles that showed clear bounce behaviour (Fig. 1). The particles reveal an almost spherical shape, and some particles are agglomerates formed from two or more primary particles. Although the agglomeration could have taken place in either the aerosol phase or on the collection grid, it does confirm that the particles were solid at the moment of collection (which took place at room temperature and pressure), thus excluding the possibility of a glassy state induced by evaporation of water or other semi-volatile components when the particles are exposed to reduced pressures in the impactor or in the microscope.

SEM images of newly formed particles (Fig. 2b; SMPS peak diameter 28 nm) are very similar. The particle size distributions derived from the SEM images are very close to those from SMPS data obtained during collection of the particles (Fig. 2e and f). This indicates that the particles were neither significantly flattened on the collection grid nor did considerable evaporation take place during the analysis. Moreover, the absence of electron diffraction patterns in transmission electron microscopy of the particles indicates that they were in a non-crystalline, amorphous state (see Supplementary Information). While such a non-crystalline state of multi-component organic particles is to be expected from thermodynamic considerations[20], all the above experiments suggest that the SOA particles in the plant chamber were in an amorphous solid state rather than in a liquid state. The amorphous solid state of SOA particles is in line with the observed presence of oligomers or other organic compounds with high molar mass in SOA particles[21], because those are known to be more prone to glass formation than smaller molecules[13]. Hydrogen bonding between oxygenated compounds and functional groups may also favour the formation of solid phases[15], and the potential influence of oxidation on the phase state of organic aerosol particles should be explored in future studies.

To corroborate these results for naturally occurring SOA particles in the atmosphere, we performed aerosol measurements in a boreal forest environment during spring 2009 (Hyytiälä Forestry Field Station, SMEAR II, Finland[22]) by following the growth of freshly nucleated SOA particles with identical instruments (see Supplementary Information). Although the bounce factor of these newly formed atmospheric SOA particles (green circles in Fig. 1a) is lower than that of plant chamber SOA particles, it is significantly larger than that of the liquid dioctyl sebacate particles, and even slightly larger than that of crystalline $(NH_4)_2SO_4$ particles. In addition, SEM samples collected 4–7 h after the start of the nucleation bursts (see Supplementary Fig. 5) again reveal nearly spherical particles (see Fig. 2c and d) and size distributions similar to the SMPS data (Fig. 2g and h). The observed bounce factor, as well as the SEM image analysis, confirms that also in the atmosphere SOA particles adopt a solid physical state within at least a few hours after nucleation begins.
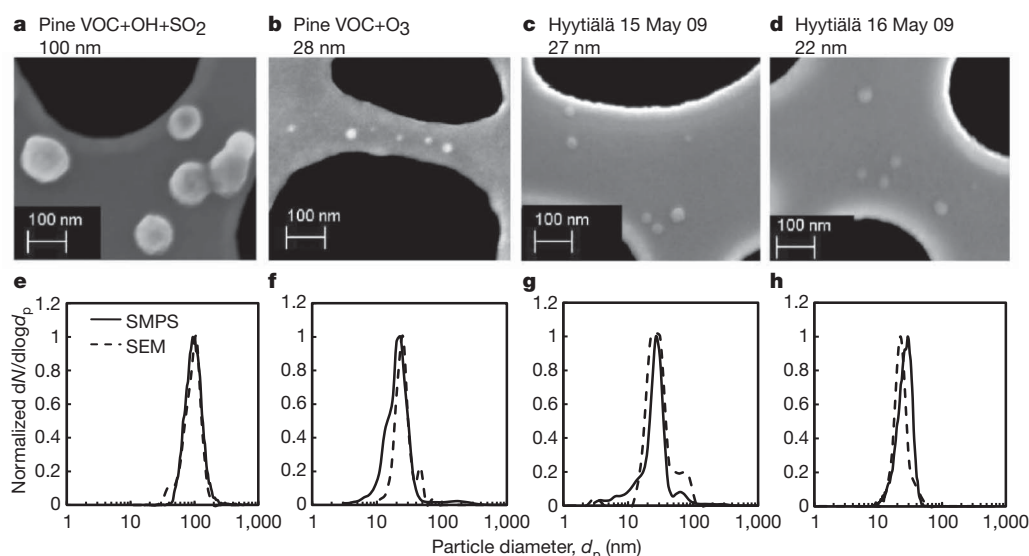
Figure 2 | SEM images of plant chamber SOA particles and atmospheric SOA particles. **a**, 100-nm-sized particles formed from the OH-initiated oxidation products of pine emissions in the presence of SO_2. **b**, The 28-nm-sized particles formed from the O_3-initiated oxidation of pine emissions (in the absence of SO_2). **c, d**, Images of atmospheric SOA particles collected 4–7 h after the start of the nucleation. **e–h**, Each size distribution is positioned below its corresponding SEM image. The SMPS size distributions measured during the SEM sample collection (solid line) are compared to those derived from SEM image analysis (dashed line). For comparison, images of liquid dioctyl sebacate particles are shown in the Supplementary Information. N is number concentration.

Although recent studies[13–15] have suggested, on the basis of laboratory experiments, that organic-rich atmospheric aerosols might be in an amorphous solid state, the results presented here are the first to show (to our knowledge) that this is indeed the case for both laboratory-grown as well as natural SOA particles in a boreal forest environment. The solid amorphous state of SOA particles has important implications for a number of atmospheric processes. First of all, it might reshape our general understanding of atmospheric SOA formation. The entire gas-to-particle conversion process of natural VOC oxidation products may need to be re-evaluated, because our results suggest that it is more similar to adsorptive re-sublimation than to condensation into a liquid. There are models of gas–particle partitioning during SOA formation that describe the partitioning as an adsorptive process[23], but it is more common to apply models that assume absorption of gases into a well-mixed organic liquid particle state[11]. An amorphous solid state of SOA particles, however, suggests strong deviations from such thermodynamic equilibrium partitioning because mixing within the particle bulk will suffer from kinetic limitations, given that diffusion is strongly inhibited in glassy or highly viscous particles.

Moreover, the water uptake of solid SOA particles might also be diminished or even fully inhibited at low to medium relative humidity[13–15], with implications for the particles' size, scattering coefficients, and direct radiative effect on climate. Nevertheless, significant water uptake may occur at high relative humidity owing to a moisture-induced glass transition resulting in a liquid state of the particles[13,15]. In the relative humidity range 71% to 90%, the hygroscopic growth factor (defined as wet diameter divided by dry diameter) of 20–50-nm particles formed in nucleation events is usually in the range of ~1.1 to 1.3 (refs 24 and 25), indicating that the particles are deliquesced droplets. On the other hand, our ELPI and SEM measurements show that that the particles were solid at 30% relative humidity. Therefore, the solid SOA particles appear to become deliquescent in the relative humidity range between 30% and 70% (ref. 15), implying that the gas-to-particle partitioning process responsible for particle growth may change from adsorptive to absorptive, very probably allowing for condensation of more volatile species once the transition has taken place. We note that in other experiments and environments SOA particles may be liquid depending on precursor compounds, chamber humidity and VOC gas phase concentrations[26]. For example, high humidity and high VOC gas phase concentrations may enhance the partitioning of low-molar-mass species into the condensed phase, thereby suppressing solidification (plasticizer effect)[15].

Finally, chemical reactions are impeded in viscous aerosol particles[12] and might even be inhibited completely in glassy aerosol particles, because mass transport (diffusion) of reactants within the aerosol particle bulk becomes the rate-limiting step. For the same reasons, chemical reactions in highly viscous and solid particles are typically surface-limited, that is, heterogeneous reactions occur only on the particle surfaces[27]. We suggest that glassy/ultraviscous particle cores or coatings in SOA might be the reason that aged atmospheric aerosols usually contain species that—according to laboratory studies—should have very short chemical lifetimes in liquid phase particles. For example, the diffusion coefficient of water in several aqueous carbohydrates is about 10^{-10} cm^2 s^{-1} in the proximity of the glass transition at room temperature[28]. Model calculations using the same value for ozone showed that the atmospheric chemical lifetime of oleic acid can increase from a few minutes in liquid particles to several hours in ultraviscous particles[29]. Because diffusion is further diminished at lower temperatures, we expect oxidation reactions to be slowed even more in the free and upper troposphere.

All the effects discussed above, both individually and in combination, may substantially increase the atmospheric lifetime of organic chemicals and of SOA particles. Because of these potentially important atmospheric implications, we strongly encourage more detailed studies in the laboratory and in the atmosphere, the better to characterize, comprehend and model the occurrence and effects of glassy atmospheric aerosol particles.

METHODS SUMMARY

The plant chamber experiments are described in detail in the Supplementary Information. Briefly, the system consists of a plant enclosure, a reaction chamber (volume 6 m^3), and a gas and particle measurement system. Plant-emitted VOCs were conducted by purified and dried air into the reaction chamber, and O$_3$ was separately added to the reaction chamber. The size distributions of the newly formed particles were measured by two SMPSs with measurement ranges from 3–60 nm and 10–700 nm. In the atmospheric measurements in Hyytiälä the particle size distribution (3–1,000 nm) was measured by a differential mobility particle sizer. The ELPI used in this study had a measurement range of 7–6,000 nm. In the chamber experiments the ELPI was equipped with an additional impactor stage with a cut-off size of 17 nm (placed between the back-up filter and the first impactor stage with a cut-off size of 32 nm). In the ambient measurements the normal ELPI set-up was used. When the measured ELPI currents were compared to ideal currents not affected by bounce, the fraction of excess current in the lowest impactor stages can be estimated, and we define this fraction as the 'bounce factor' (see Supplementary Information for exact derivations).

The electron microscopy samples were collected onto copper grids covered by holey carbon film. The collection took place at a sample flow rate of 0.2 litres per minute at room temperature and pressure. The samples were analysed using a field emission gun scanning electron microscope (FEG-SEM, Zeiss ULTRAplus). Before analysis, the samples were carbon-coated with a JEOL JEE-4X vacuum evaporator (Tokyo) using spectrographically pure carbon rods. The size distributions of particles were determined from the SEM images by image analysis (see Supplementary Information). Furthermore, the crystallinity of the particles was studied by analysing the electron diffraction pattern with transmission electron microscopy (Jeol JEM 2010).

1. Hallquist, M. et al. The formation, properties and impact of secondary organic aerosol: current and emerging issues. Atmos. Chem. Phys. **9**, 5155–5236 (2009).
2. Kanakidou, M. et al. Organic aerosol and global climate modelling: a review. Atmos. Chem. Phys. **5**, 1053–1123 (2005).
3. Kavouras, I. G., Mihalopoulos, N. & Stephanou, E. G. Formation of atmospheric particles from organic acids produced by forest. Nature **395**, 683–686 (1998).
4. Claeys, M. et al. Formation of secondary organic aerosols through photooxidation of isoprene. Science **303**, 1173–1176 (2004).
5. Jimenez, J. L. et al. Evolution of organic aerosols in the atmosphere. Science **326**, 1525–1529 (2009).
6. Laaksonen, A. et al. The role of VOC oxidation products in continental new particle formation. Atmos. Chem. Phys. **8**, 2657–2665 (2008).
7. Tunved, P. et al. High natural aerosol loading over boreal forests. Science **312**, 261–263 (2006).
8. Guenther, A. et al. A global model of natural volatile organic compound emissions. J. Geophys. Res. **100** (D5), 8873–8892 (1995).
9. Fuzzi, S. et al. Critical assessment of the current state of scientific knowledge, terminology, and research needs concerning the role of organic aerosols in the atmosphere, climate, and global change. Atmos. Chem. Phys. **6**, 2017–2038 (2006).
10. Intergovernmental Panel on Climate Change (IPCC). Climate Change 2007: The Physical Science Basis Ch. 2 161–177, Ch. 7 556–564 (Cambridge University Press, 2007).
11. Pankow, J. F. An absorption model of the gas/aerosol partitioning involved in the formation of secondary organic aerosol. Atmos. Environ. **28**, 189–193 (1994).
12. Zahardis, J. & Petrucci, G. A. The oleic acid-ozone heterogeneous reaction system: products, kinetics, secondary chemistry, and atmospheric implications of a model system—a review. Atmos. Chem. Phys. **7**, 1237–1274 (2007).
13. Zobrist, B., Marcolli, C., Pedernera, D. A. & Koop, T. Do atmospheric aerosols form glasses? Atmos. Chem. Phys. **8**, 5221–5244 (2008).
14. Murray, B. J. Inhibition of ice crystallisation in highly viscous aqueous organic acid droplets. Atmos. Chem. Phys. **8**, 5423–5433 (2008).
15. Mikhailov, E., Vlasenko, S., Martin, S. T., Koop, T. & Pöschl, U. Amorphous and crystalline aerosol particles interacting with water vapor: conceptual framework and experimental evidence for restructuring, phase transitions and kinetic limitations. Atmos. Chem. Phys. **9**, 9491–9522 (2009).
16. Joutsensaari, J. et al. Nanoparticle formation by ozonolysis of inducible plant volatiles. Atmos. Chem. Phys. **5**, 1489–1495 (2005).
17. Mentel, Th. F. et al. Photochemical production of aerosols from real plant emissions. Atmos. Chem. Phys. **9**, 4387–4406 (2009).
18. Dahneke, B. Capture of aerosol particles by surfaces. J. Colloid Interf. Sci. **37**, 342–347 (1971).
19. Stein, S. W., Turpin, J. B., Cai, X., Huang, P. F. & McMurry, P. H. Measurements of relative humidity dependent bounce and density for atmospheric particles using DMA-impactor technique. Atmos. Environ. **28**, 1739–1746 (1994).
20. Marcolli, C., Luo, B. P. & Peter, T. Mixing of the organic aerosol fractions: liquids as the thermodynamically stable phases. J. Phys. Chem. A **108**, 2216–2224 (2004).
21. Kalberer, M. et al. Identification of polymers as major components of atmospheric organic aerosols. Science **303**, 1659–1662 (2004).
22. Hari, P. & Kulmala, M. Station for measuring ecosystem-atmosphere relations (SMEAR II). Boreal Environ. Res. **10**, 315–322 (2005).

23. Pankow, J. F. Review and comparative analysis of the theories of partitioning between the gas and aerosol particulate phases in the atmosphere. *Atmos. Environ.* **21,** 2275–2283 (1987).

24. Ehn, M. *et al.* Hygroscopic properties of ultrafine particles in the boreal forest: diurnal variation, solubility and the influence of sulfuric acid. *Atmos. Chem. Phys.* **7,** 211–222 (2007).

25. Hämeri, K. & Väkevä, M. Ultrafine aerosol particle hygroscopicity and volatility in boreal forest. *Rep. Ser. Aerosol Sci.* **47,** 47–59 (2000).

26. Bahreini, R. *et al.* Measurements of secondary organic aerosol from oxidation of cycloalkanes, terpenes, and m-xylene using an aerodyne aerosol mass spectrometer. *Environ. Sci. Technol.* **39,** 5674–5688 (2005).

27. Maria, S. F., Russell, L. M., Gilles, M. K. & Myneni, S. C. B. Organic aerosol growth mechanisms and their climate-forcing implications. *Science* **306,** 1921–1924 (2004).

28. Parker, R. & Ring, S. G. Diffusion in maltose-water mixtures at temperatures close to the glass transition. *Carbohydr. Res.* **273,** 147–155 (1995).

29. Shiraiwa, M., Pfrang, C. & Pöschl, U. Kinetic multi-layer model of aerosol surface and bulk chemistry (KM-SUB): the influence of interfacial transport and bulk diffusion on the oxidation of oleic acid by ozone. *Atmos. Chem. Phys.* **10,** 3673–3691 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

LETTER

Seismic evidence of negligible water carried below 400–km depth in subducting lithosphere

Harry W. Green II[1], Wang-Ping Chen[2] & Michael R. Brudzinski[3]

Strong evidence exists that water is carried from the surface into the upper mantle by hydrous minerals in the uppermost 10–12 km of subducting lithosphere, and more water may be added as the lithosphere bends and goes downwards. Significant amounts of that water are released as the lithosphere heats up, triggering earthquakes and fluxing arc volcanism[1]. In addition, there is experimental evidence for high solubility of water in olivine, the most abundant mineral in the upper mantle, for even higher solubility in olivine's high-pressure polymorphs, wadsleyite and ringwoodite[2], and for the existence of dense hydrous magnesium silicates that potentially could carry water well into the lower mantle[3] (deeper than 1,000 km). Here we compare experimental and seismic evidence to test whether patterns of seismicity and the stabilities of these potentially relevant hydrous phases are consistent with a wet lithosphere. We show that there is nearly a one-to-one correlation between dehydration of minerals and seismicity at depths less than about 250 km, and conclude that the dehydration of minerals is the trigger of instability that leads to seismicity. At greater depths, however, we find no correlation between occurrences of earthquakes and depths where breakdown of hydrous phases is expected. Lastly, we note that there is compelling evidence for the existence of metastable olivine (which, if present, can explain the distribution of deep-focus earthquakes[4–7]) west of and within the subducting Tonga slab[8] and also in three other subduction zones, despite metastable olivine being incompatible with even extremely small amounts of water (of the order of 100 p.p.m. by weight[7]). We conclude that subducting slabs are essentially dry at depths below 400 km and thus do not provide a pathway for significant amounts of water to enter the mantle transition zone or the lower mantle.

Lithosphere is formed at mid-ocean ridges where magma upwells and freezes to form new oceanic crust. Interaction between cold water of the deep ocean ($\sim$0 °C) and the extreme heat of magma ($>$1,100 °C) results in widespread cracking of rocks and pervasive hydrothermal circulation that drives sea water many kilometres below the surface, producing extensive hydrous alteration of the oceanic crust and the uppermost few kilometres of the mantle. Away from the mid-ocean ridges, little more happens to average lithosphere as it moves along under the ocean until reaching the oceanic trench, bending sharply and descending back into the mantle (Fig. 1). Near the trench, numerous normal faults are created in association with bending of the lithosphere, whose thickness is typically on the order of 100 km by then. The vast majority of such faults dip towards the trench, many of them reaching depths of 20 km or more[9]. These faults could potentially provide a pathway for additional water to enter the down-going lithosphere.

At the trench, large amounts of water are expelled back into the ocean from the down-going sediments[10]. Meanwhile, numerous earthquakes associated with thrust faulting—some of which, such as the great Sumatran earthquake of 2004, are extremely large in magnitude—occur along the interface between the two converging plates to depths of about 25–35 km. Below this depth, the interface ceases to

generate earthquakes; instead, seismicity occurs within the subducting lithosphere, first in the hydrated oceanic crust and subsequently in the mantle portion of the slab[11]. As shown in Fig. 1, along many subduction zones there is also a lower seismic zone deep within the slab[12]. With increasing depth, the two zones progressively approach each other and eventually merge at depths of about 200–250 km.

Within experimental error, these double seismic zones in subducting slabs seem to follow the dehydration boundary of antigorite serpentine[12,13] (the chemical composition of serpentine is essentially hydrated olivine, which constitutes at least 60% of the upper mantle). Recent high-resolution studies beneath Japan have shown that double seismic zones coincide with regions of marked changes in the ratio between speeds of P and S waves within the lithosphere (V_P/V_S), probably owing to active dehydration of hydrous phases[14]. Thus, dehydration and seismicity in subducting lithosphere apparently go hand-in-hand to a depth of around 300 km.

Geochemical tracers and compositions of volcanic arc magmas demonstrate that after dehydration large amounts of water leave the slab, migrate upwards into the overlying mantle wedge, where temperatures are higher, and provide the flux for arc volcanism. Fluids released from the subducting lithosphere also may hydrate the cooler portions of the mantle wedge above the descending slab, resulting in serpentinization near the continental Moho[15]; whether hydration occurs immediately above subducting slabs at greater depths is unknown.

Figure 1c shows that the overall number of earthquakes falls exponentially from depths of around 50 km to a minimum at $\sim$300 km. Concentrations of rare hydrous phases in subducting lithosphere that are stable to greater pressures than antigorite probably dehydrate and could account for the very few earthquakes at depths of 250–350 km. The lull in seismicity persists until about 450 km, where the seismicity rebounds towards a secondary maximum near $\sim$600 km and then drops precipitously to zero just before the boundary between the mantle transition zone (MTZ, between depths of about 410 and 660 km) and the lower mantle. We note that there are no gaps in focal depths below 410 km (ref. 16): all subducting slabs that produce earthquakes below this depth show a continuous distribution in focal depths.

Experiments show that at high pressures, a critical prerequisite to developing a shearing instability that could lead to an earthquake is to generate a small amount of a low-viscosity phase. Such a phase could be a true fluid (for example water, as discussed above[17–19]) or an extremely fine-grained, 'superplastic' solid that can flow readily by grain-boundary sliding at seismic strain rates[4–6,20]. Thus, a continuous distribution of earthquakes in the MTZ requires a locus of mineral reactions that continuously produces such a 'fluid' phase down to $\sim$660 km and abruptly ceases there, despite the interpretation of seismic tomography which suggests that some subducting slabs (for example the southern Marianas) penetrate directly into the lower mantle.

Such a correlation between dehydration of down-going slabs and deep earthquakes does not persist below 300 km. The principal potential carriers for water at these depths and below would be nominally anhydrous phases, principally olivine ($\sim$1 wt% maximum[2]), which is

[1]Institute of Geophysics and Planetary Physics and Department of Earth Sciences, University of California, Riverside, California 92521, USA. [2]Department of Geology, University of Illinois, Urbana-Champaign, Illinois 61801, USA. [3]Department of Geology, Miami University, Oxford, Ohio 45056, USA.
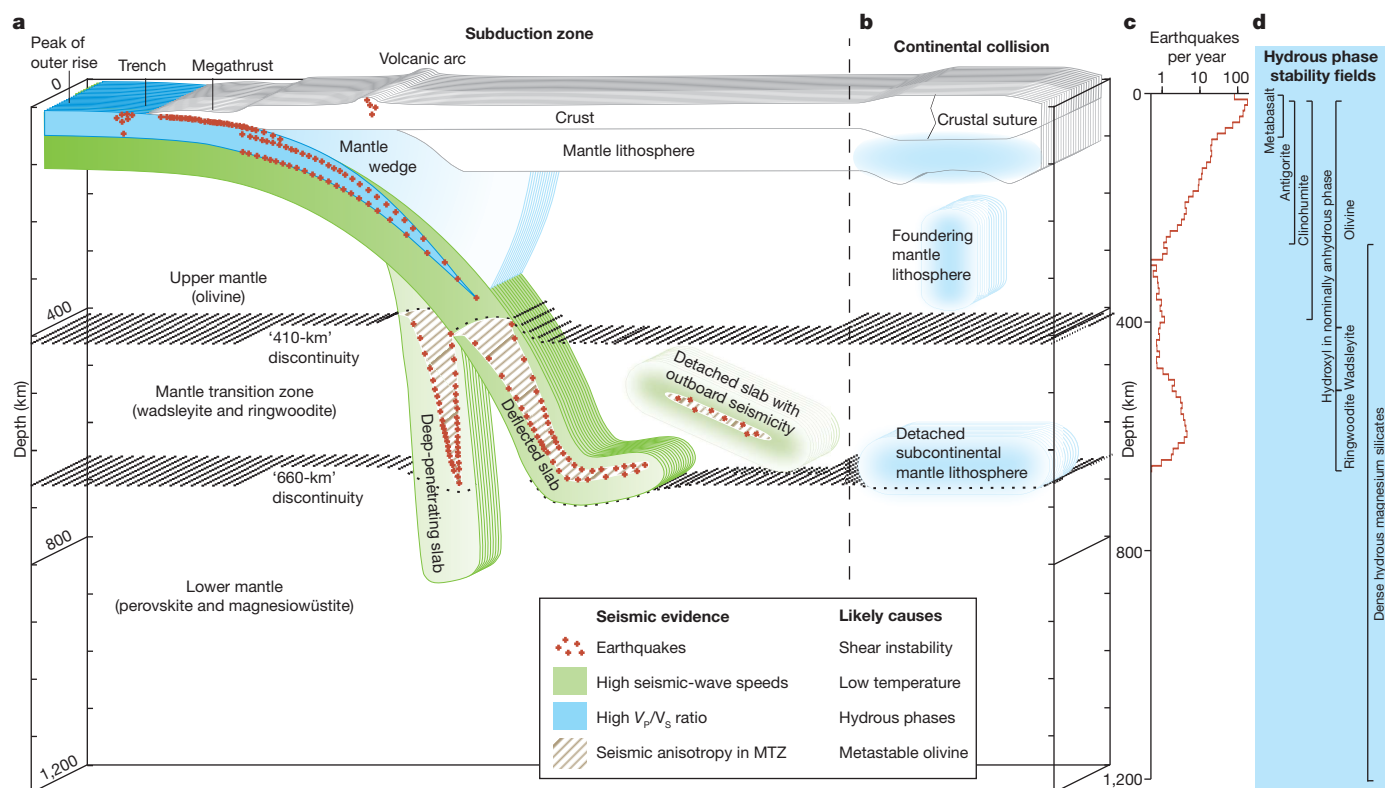
Figure 1 | Schematic summarizing key points presented in this study.
a, Salient features of subduction zones and overall structures of the upper mantle and the mantle transition zone, with key seismic observations and their likely causes. The most prominent example of deflected and detached slabs is under the back-arc region of the Tonga–Kermadec subduction zone[8], whereas the slab along the steeply dipping Mariana subduction zone seems to penetrate into the lower mantle[27]. Globally, not a single earthquake has been recorded below a depth of about 680 km. **b**, Salient features of a continental collision zone, such as the Himalaya–Tibet collision zone, where Rayleigh–Taylor instability causes thickened subcontinental mantle lithosphere to founder. Before collision, the subcontinental mantle lithosphere was hydrated as part of

the mantle wedge along an Andean-type continental margin[33]. **c**, Globally averaged number of earthquakes (body-wave magnitude, ≥ 5) per year as a function of depth. Notice that the horizontal scale is logarithmic. **d**, Summary of stability of various hydrous phases, emphasizing the effect of depth (or, equivalently, pressure). There is a sequence of dehydration reactions that can account for the concentration of seismicity above depths of about 350 km. In contrast, there is no corresponding dehydration for the concentration of seismicity at greater depths. Moreover, there is no seismicity associated with expected dehydration of nominally anhydrous olivine polymorphs and dense hydrous magnesium silicates at greater depths.

stable to ~400 km, wadsleyite (~2–3 wt% maximum[21,22]), which is stable to ~550 km, and ringwoodite (~1–3 wt% maximum[21,22]), which is stable to ~700 km in subduction zones (Fig. 1d). There is also a series of dense hydrous magnesium silicate (DHMS) phases, which, in principle, could be stable at depths from ~250 km to deep in the lower mantle[3]. However, to stabilize these 'alphabet' phases to any significant degree, the coexisting phase with olivine stoichiometry would have to be saturated with water (see Supplementary Information for additional details). This is an extremely unlikely possibility given the significant solubility of water in olivine and its polymorphs, their great abundance ($\geq 60\%$ of the rock), the pervasive prior dehydration in the upper mantle and the lack of evidence for high V_P/V_S ratios around active subduction zones within the MTZ[8] (Fig. 1a). Thus, the maximum possible amount of water entering the MTZ within the subducting lithosphere could only trivially exceed that carried by saturated olivine, yielding greatly undersaturated wadsleyite and, subsequently, undersaturated ringwoodite.

At present, major subduction zones have convergence rates of ~100–200 mm yr^{-1}, so it would take an element of oceanic lithosphere at the surface only 3.5–7 million years to reach 700 km, where the pressure is ~25 GPa (250,000 times the atmospheric pressure at sea level). Rocks are very poor thermal conductors; hence, within a subducting slab pressure climbs much more rapidly than temperature. This is especially true for fast subducting slabs, which are coldest and therefore have the greatest chance of retaining serpentine, which can carry water deep enough for olivine or phase A possibly to take over as

the principal carrier. Thus, as a subducting slab approaches and passes through the MTZ, relevant dehydration reactions are those that occur with increasing pressure. Yet at such depths in cold slabs, there are no known dehydration reactions that could take place, considering all likely compositions of the mantle lithosphere and ranges of temperatures in subducting slabs (Fig. 1d and Supplementary Figure 1).

Indeed, the solubility of water in olivine, wadsleyite and ringwoodite is known to increase with pressure, and fragmentary data indicate that it decreases only moderately with increasing temperature[22]. Similarly, the sequence of DHMS phases, which potentially could be present at progressively greater depths if sufficient water were to exist[3], would not release fluid either, because each successively higher-pressure phase contains even greater amounts of water in its crystal structure along a cold slab geotherm (Supplementary Fig. 1). Thus, no dehydration reactions that could produce earthquakes are expected for slabs in the MTZ. One possible exception is when wadsleyite transforms to ringwoodite, but only if wadsleyite already contains more water than can dissolve in ringwoodite, a possibility that is rendered unlikely by the expected undersaturation of olivine with water, described above. Moreover, were such a seismogenic reaction to occur, a marked spike in seismicity at depths around 550 km would be expected; but such a pattern is not observed (Fig. 1c), further ruling out this possibility. The lack of spikes in the distribution of earthquakes deeper than 400 km further argues against transient release of water from DHMS as a seismogenic mechanism in the MTZ, because such dehydration reactions would be tied to particular pressures.

Lastly, ringwoodite has a much greater water solubility than phases it gives rise to in the lower mantle, namely magnesium silicate perovskite and magnesiowüstite[22]. Therefore, if wet ringwoodite were present in a down-going slab that passes into the lower mantle (for example along the southern end of the Marianas slab), dehydration should result in a flurry of earthquakes in the topmost lower mantle; however, such earthquakes are not observed in any subduction zone. Alternatively, it is conceivable that phase D, the stable DHMS phase at the highest pressures, could form along with perovskite and magnesiowüstite from wet ringwoodite as it enters the lower mantle, and prevent any free fluid from forming (hence no earthquakes). However, should this happen, then when phase D breaks down at greater depth in the lower mantle[3], a flurry of earthquakes should occur at that depth—a prediction that is also contrary to observation.

At such depths, it is possible that stresses inside subducted slabs become too low to cause earthquakes, but this is an extremely unlikely situation for several reasons. Foremost, it is well established that the lower mantle is significantly more viscous than the upper mantle[23]. Therefore, one would expect greater resistance to subduction and, hence, larger tectonic stress with increasing depth. Moreover, when regional tectonic stresses are absent in either intermediate (100–300-km) or greater (400–650-km) depths, large earthquakes persist[16]; apparently the juxtaposition of material that can trigger earthquakes and material from which the unstable component recently has been removed by reaction (for example the boundary between serpentine-bearing and serpentine-lacking peridotite) generates sufficient local stresses to maintain the production of earthquakes. So if a subducted slab were to carry phase D to great depths, dehydration of phase D would be expected to produce earthquakes in the lower mantle.

If dehydration cannot explain earthquakes in the MTZ, there is a known alternative. More than 30 years ago, it was suggested that if the interior of subduction zones were sufficiently cold olivine would be kinetically inhibited from transforming to wadsleyite or ringwoodite in the MTZ[6,23], implying the presence of a wedge of metastable olivine within cold slabs that could, in principle, extend to the bottom of the MTZ. It was subsequently shown by experiments[4,5] that, under stress, a previously unknown shearing instability develops at the threshold of kinetic enablement of this reaction[24]. For this mechanism to be activated, metastable olivine must be present in the core of subduction zones. Shortly after this instability was discovered, seismic evidence supporting it was reported from the Izu-Bonin[23] slab (see also ref. 6), and not long after that a deep double seismic zone as predicted by this mechanism was observed in the same place[20] and also in Tonga[25].

Since then, experimental work has placed additional constraints on the conditions under which subduction zones could influence the olivine–wadsleyite reaction. In particular, given estimated temperature distributions within subducting slabs, even very small amounts of water dissolved in olivine or its polymorphs ($\sim$100 p.p.m. by weight) would flux the polymorphic reactions and thereby prohibit metastable olivine in the MTZ[7,26]. Furthermore, there is now abundant *in situ* evidence, based on a variety of mutually complementary observations, including strong seismic anisotropy (Fig. 1a), for metastable olivine in the Tonga subduction zone[8], the Marianas[27] and beneath Japan[28]. Together, the experimental and seismic results strongly suggest that subduction zones are extremely dry below $\sim$400 km.

The preceding discussion shows that in the earthquake-generating portion of cold slabs below 400 km, the water content is drastically reduced to no more than 100 p.p.m. on the basis of evidence for metastable olivine. A possibility not yet addressed is whether the uppermost part of the slab that dehydrated at shallow depths could be rehydrated at greater depths by water rising from dehydration occurring deeper within the slab. One possibility that could result in rehydration, suggested by laboratory experiments, is stabilization of phase A and clinohumite[29,30]. However, this mechanism is unlikely to trap significant quantities of water. First, phase A is precluded because it would have to occur in regions where seismic evidence shows metastable olivine, whose presence is incompatible with hydrous phases. Second, formation of clinohumite in the presence of olivine at less than 8 GPa (for subsequent subduction to greater depth), requires titanium or fluorine, neither of which is abundant enough in peridotite to stabilize significant amounts of clinohumite (see Supplementary Information for additional details).

There may be other potential pathways for recycling water into the MTZ or deeper, where constraints from deep earthquakes do not apply. As discussed in the previous paragraph, some of the water expelled from slabs could hydrate cooler regions of the mantle wedge above the slab; and such hydrous phases (for example clinohumite) could be entrained by subduction and carried to greater depths, escaping detection by not triggering shear instability as they progressively dehydrate, if the local stresses outside the slab are sufficiently low (see ref. 19 for a demonstration of the competition between faulting and flow during dehydration depending on the level of stress). Such entrainment could be an alternative interpretation for geophysical observations that suggest the presence of melt/fluid 'ponding' on top of the 410-km discontinuity[31]. Other hypothetical possibilities would be addition of water to the MTZ by sinking of negatively buoyant hydrous magmas from the upper mantle or rising of buoyant hydrous magmas from the lower mantle[32].

One practical mechanism that does seem capable of transferring water from near the surface to the MTZ is convective removal or delamination of hydrous mantle lithosphere during continental collision[33] (Fig. 1b). Whether repeated cycles of continental collision, the main process for assembling supercontinents, could be sufficient to build up a wet MTZ over geological time is not yet known. This mechanism and other potential pathways hold the key to fully deciphering the Earth's deep water cycle because subduction of oceanic lithosphere is unlikely to be an effective pathway for recycling water into the deep mantle. (See additional discussion in Supplementary Information.)

1. Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction related magmatism. *Science* **268**, 858–861 (1995).
2. Smyth, J. R., Frost, D. J., Nestola, F., Holl, C. M. & Bromiley, G. Olivine hydration in the deep upper mantle: effects of temperature and silica activity. *Geophys. Res. Lett.* **33**, 1530 L15301 (2006).
3. Angel, R. J. *et al.* Stabilities and equations of state of dense hydrous magnesium silicates. *Phys. Earth Planet. Inter.* **127**, 181–196 (2001).
4. Green, H. W. II & Burnley, P. C. A new self-organizing mechanism for deep-focus earthquakes. *Nature* **341**, 733–737 (1989).
5. Green, H. W. II, Young, T. E., Walker, D. & Scholz, C. H. Anticrack-associated faulting at very high pressure in natural olivine. *Nature* **348**, 720–722 (1990).
6. Green, H. W. II & Houston, H. The mechanics of deep earthquakes. *Annu. Rev. Earth Planet. Sci.* **23**, 169–213 (1995).
7. Diedrich, T., Sharp, T. G., Leinenweber, K. & Holloway, J. R. The effects of small amounts of H$_2$O on olivine to ringwoodite transformation growth rates and implications for subduction of metastable olivine at 18 GPa. *Chem. Geol.* **262**, 87–99 (2009).
8. Chen, W.-P. & Brudzinski, M. Evidence for large-scale remnant of subducted lithosphere beneath Fiji. *Science* **292**, 2475–2478 (2001).
9. Ranero, C. R., Phipps Morgan, J., McIntosh, K. & Reichert, C. Bending-related faulting and mantle serpentinization at the Middle America trench. *Nature* **425**, 367–373 (2003).
10. Dixon, T. H. & Moore, J. C. (eds) *The Seismogenic Zone of Subduction Thrust Faults* 147–315 (Columbia Univ. Press, 2007).
11. Hacker, G. R., Peacock, S. M., Abers, G. A. & Holloway, S. D. Subduction factory: 2. Are intermediate depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108**, 2030 (2003).
12. Brudzinski, M. R., Thurber, C. H., Hacker, B. R. & Engdahl, R. Global prevalence of double Benioff zones. *Science* **316**, 1472–1474 (2007).
13. Peacock, S. M. Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* **29**, 299–302 (2001).
14. Zhang, H. *et al.* High-resolution subducting-slab structure beneath northern Honshu, Japan, revealed by double-difference tomography. *Geology* **32**, 361–364 (2004).
15. Bostock, M. G., Hyndman, R. D., Rondenay, S. & Peacock, S. M. An inverted continental Moho and serpentinization of the forearc mantle. *Nature* **417**, 536–538 (2002).
16. Brudzinski, M. R. & Chen, W.-P. Earthquakes and strain in subhorizontal slabs. *J. Geophys. Res.* **110**, B08303 (2005).

17. Raleigh, C. B. & Paterson, M. S. Experimental deformation of serpentinite and its tectonic implications. *J. Geophys. Res.* **70,** 3965–3985 (1965).
18. Jung, H., Green, H. W. II & Dobrzhinetskaya, L. F. Intermediate-depth earthquake faulting by dehydration embrittlement with negative volume change. *Nature* **428,** 545–549 (2004).
19. Zhang, J., Green, H. W. II, Bozhilov, K. & Jin, Z.-M. Faulting induced by precipitation of water at grain boundaries in hot subducting oceanic crust. *Nature* **428,** 633–636 (2004).
20. Iidaka, T. & Furukawa, Y. Double seismic zone for deep earthquakes in the Izu-Bonin subduction zone. *Science* **263,** 1116–1118 (1994).
21. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. Solubility of water in the α, β and γ phases of (Mg,Fe)2SiO4. *Contrib. Mineral. Petrol.* **123,** 345–357 (1996).
22. Bolfan-Casanova, N., McCammon, C. A. & Mackwell, S. J. *Water in Transition Zone and Lower Mantle Minerals* 57–68 (Geophys. Monogr. 168, American Geophysical Union, 2006).
23. Iidaka, T. & Suetsugu, D. Seismological evidence for metastable olivine inside a subducting slab. *Nature* **356,** 593–595 (1992).
24. Green, H. W. & Marone, C. J. in *Plastic Deformation of Minerals and Rocks* (eds Wenk, H.-R. & Karato, S.) 181–199 (Mineralogical Society of America, 2002).
25. Wiens, D. A., McGuire, J. J. & Shore, P. J. Evidence for transformational faulting from a deep double seismic zone in Tonga. *Nature* **364,** 790–793 (1993).
26. Mosenfelder, J. L. *et al.* Experimental constraints on the depth of olivine metastability in subducting lithosphere. *Phys. Earth Planet. Inter.* **127,** 165–180 (2001).
27. Kaneshima, S., Okamoto, T. & Takenaka, H. Evidence for a metastable olivine wedge inside the subducted Mariana slab. *Earth Planet. Sci. Lett.* **258,** 219–227 (2007).
28. Jiang, G., Zhao, D. & Zhang, G. Seismic evidence for a metastable olivine wedge in the subducting Pacific slab under Japan Sea. *Earth Planet. Sci. Lett.* **270,** 300–307 (2008).
29. Stalder, R. & Ulmer, P. Phase relations of a serpentine composition between 5 and 14 GPa: significance of clinohumite and phase E as water carriers into the transition zone. *Contrib. Mineral. Petrol.* **140,** 670–679 (2001).
30. Iizuka, Y. & Nakamura, E. Experimental study of the slab-mantle interaction and implications for the formation of titanoclinohumite at deep subduction zone. *Proc. Jpn. Acad. B* **71,** 159–164 (1996).
31. Toffelmier, D. A. & Tyburczy, J. A. Electromagnetic detection of a 410-km-deep melt layer in the southwestern United States. *Nature* **447,** 991–994 (2007).
32. Hirschmann, M. M. Water, melting and the deep Earth water cycle. *Annu. Rev. Earth Planet. Sci.* **34,** 629–653 (2006).
33. Tseng, T.-L. & Chen, W.-P. Discordant contrasts of *P*- and *S*-wave speeds across the 660-km discontinuity beneath Tibet: a case for hydrous remnant of sub-continental lithosphere. *Earth Planet. Sci. Lett.* **268,** 450–462 (2008).

Author Contributions The concepts underlying this paper have developed through discussions among the authors over the last few years. Interpretation of experimental results is primarily due to H.W.G.; global seismic results and artwork are primarily due to W.-P.C. and M.R.B. H.W.G. wrote the first draft of the manuscript; all authors participated in manuscript revision.

LETTER

xnd-1 regulates the global recombination landscape in *Caenorhabditis elegans*

Cynthia R. Wagner[1]†, Lynnette Kuervers[2], David L. Baillie[2] & Judith L. Yanowitz[1,3]

Meiotic crossover (CO) recombination establishes physical linkages between homologous chromosomes that are required for their proper segregation into developing gametes, and promotes genetic diversity by shuffling genetic material between parental chromosomes. COs require the formation of double strand breaks (DSBs) to create the substrate for strand exchange. DSBs occur in small intervals called hotspots[1–3] and significant variation in hotspot usage exists between and among individuals[4]. This variation is thought to reflect differences in sequence identity and chromatin structure, DNA topology and/or chromosome domain organization[1,5–9]. Chromosomes show different frequencies of nondisjunction (NDJ)[10], reflecting inherent differences in meiotic crossover control, yet the underlying basis of these differences remains elusive. Here we show that a novel chromatin factor, *X non-disjunction factor 1* (*xnd-1*), is responsible for the global distribution of COs in *C. elegans*. *xnd-1* is also required for formation of double-strand breaks (DSBs) on the X, but surprisingly XND-1 protein is autosomally enriched. We show that *xnd-1* functions independently of genes required for X chromosome-specific gene silencing, revealing a novel pathway that distinguishes the X from autosomes in the germ line, and further show that *xnd-1* exerts its effects on COs, at least in part, by modulating levels of H2A lysine 5 acetylation.

The *C. elegans* genome has an unusual organization with regard to recombination and gene organization. Recombination preferentially occurs towards the chromosome ends, which contain fewer and less highly expressed genes than the central gene cluster[11] (D.L.B., results not shown). The X chromosome shows a similar distribution of crossovers[12], but genes are more evenly distributed along its length. The X is also transcriptionally silent in the germline except for a brief period during late oogenesis[13]. Thus, during the critical period of CO formation, the X chromosome is replete with histone post-translational modifications (HPTMs) indicative of heterochromatin. Although recombination can occur in heterochromatin, closed chromatin generally presents an unfavourable environment for DSB formation. This unique configuration of an entire chromosome provides an advantage of the nematode system for understanding the relationship between higher order DNA structure and CO regulation[9].

To identify meiotic recombination regulatory proteins, we screened over 350 known and putative chromatin binding proteins by RNA interference (RNAi) for segregation of triply marked chromosomes (see Methods). RNAi against C05D2.5 resulted in a marked increase in recombinant Dpy progeny (Fig. 1a, left panel). A second triply marked chromosome confirmed these results (Fig. 1a, right panel), indicating that C05D2.5 affected recombination rather than expression of the genetic markers. C05D2.5 was renamed *xnd-1* (*X chromosome nondisjunction-1*) on the basis of its primary phenotypes. Further analysis was performed on two deletion mutations, *ok708* and *ok709* (Supplementary Fig. 1).

Single nucleotide polymorphism (SNP) analysis was used to confirm a role for *xnd-1* in CO regulation. In wild type, COs occurred in more gene-poor regions towards the autosome ends[11] (Fig. 1b and Supplementary Tables 1 and 2). In *xnd-1* mutants, the global crossover landscape on autosomes was inverted; most COs occurred abnormally in the gene-rich region of chromosome I (Fig. 1b and Supplementary Tables 1 and 2). This dramatic shift in CO placement occurred without a change in the overall map size (that is, total number of crossovers per chromosome) (Fig. 1b and Supplementary Tables 1 and 2). These data therefore indicate that *xnd-1* is required for the normal distribution of meiotic crossovers on autosomes in *C. elegans*.

Because the X has a different chromatin state than autosomes, we also mapped COs on X. Like on autosomes, CO distribution is altered and occurs preferentially in the central region in *xnd-1* mutants (Fig. 1c). In contrast to autosomes, loss of *xnd-1* reduced the genetic map of the X to half that observed in wild type (24.4 map units (m.u.) versus 51.1 m.u. respectively; Fig. 1c and Supplementary Table 3), indicating a twofold decrease in recombination frequency. Thus, *xnd-1* is required for CO distribution genome-wide, but is selectively required for the normal CO frequency on the X.

The increase in non-recombinant X's was accompanied by an increase in the percentage of male progeny (XO genotype) arising from nondisjunction in XX hermaphrodites (~20% males versus <0.2% in wild type; Supplementary Table 4), consistent with defects in meiotic CO formation[14]. Autosomal aneuploidy manifests as embryonic lethality, and on average, 35% of *xnd-1* embryos do not hatch (Supplementary Table 4). To determine if both the X and autosomes were defective in CO formation, we analysed diakinesis oocytes when homologues are held together at the CO and revealed as six DAPI (4′,6-diamidino-2-phenylindole) staining bivalents in wild type (Fig. 2a). In *xnd-1*, seven DAPI-staining bodies were evident in ~ 50% of oocytes: five bivalents and two smaller univalents are observed (Fig. 2b), revealing that a single chromosome pair had frequently not received a CO. Fluorescence *in situ* hybridization (FISH) confirmed that the achiasmate chromosome was the X in all instances (Fig. 2a–c), demonstrating that *xnd-1* is required for wild-type levels of X chromosome COs. Because no achiasmate autosomes were observed, the embryonic lethality associated with *xnd-1* seems to be independent of its role in CO formation (see Methods).

The ability of *xnd-1* to affect CO formation and distribution raised the possibility that *xnd-1* might be involved in regulating DSBs. Meiotic DSBs are catalysed by SPO-11 (ref. 15) and the failure to form bivalents in *spo-11* mutants can be rescued by exogenous breaks induced by ionizing radiation[16] (Fig. 2d). Ionizing radiation significantly reduces the number of nuclei with seven DAPI staining bodies at diakinesis in *xnd-1* (Fig. 2d), indicating that the deficit in X COs is either the result of a failure to make DSBs or a failure to make CO-competent DSBs.

To determine whether *xnd-1* is required for DSB formation, we assayed recruitment of the DSB repair protein RAD-51 to the X chromosome[9]. By co-labelling for the autosomal-associated HTZ-1 protein[17] (Supplementary Fig. 2), we assessed the frequency that RAD-51 associated with the (unlabelled) X. RAD-51 foci were detected on X in ~95%

[1]Carnegie Institution of Washington, Department of Embryology, Baltimore, Maryland 21218, USA. [2]Simon Fraser University, Dept. of Molecular Biology and Biochemistry, Burnaby, British Columbia V5A 1S6, Canada. [3]Magee-Womens Research Institute and University of Pittsburgh School of Medicine, Dept. of Obstetrics, Gynecology and Reproductive Sciences, Pittsburgh, Pennsylvania 15213, USA.
†Current address: University of Maryland, Baltimore County, Dept. of Biological Sciences, Baltimore, Maryland 21250, USA.
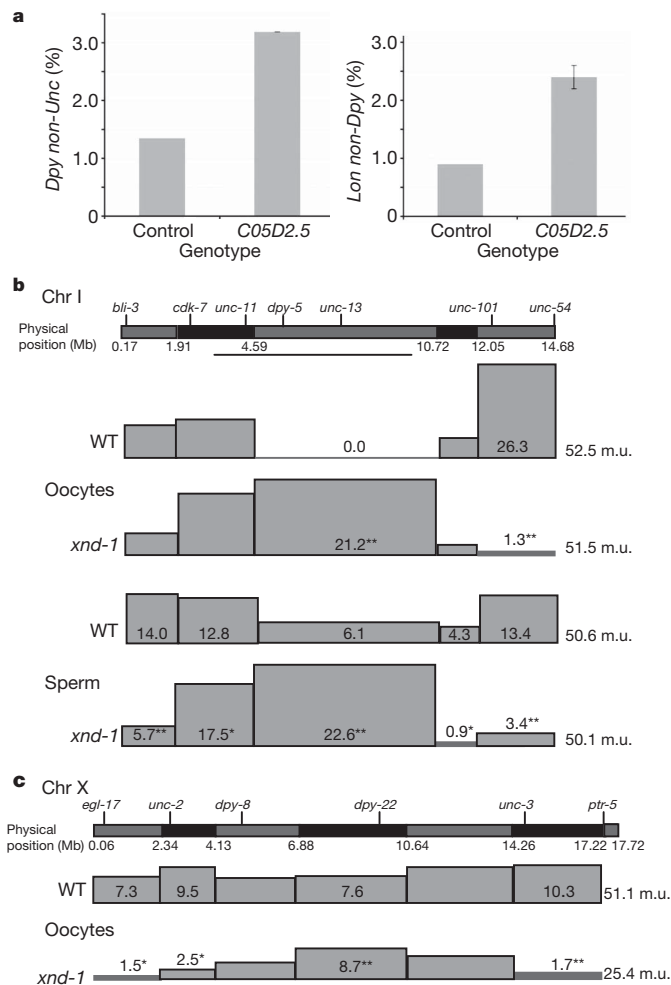
Figure 1 | xnd-1 is needed for the normal recombination landscape in C. elegans. a, C05D2.5 was identified by its increase in the number of recombinant progeny with the phenotype of the middle of three genetic markers, either *unc-45 dpy-18 unc-64* (wild type, n = 8750, s.d. = 0.06; *xnd-1*, n = 3980, s.d. = 0.15) (left panel) or *dpy-1 lon-1 dpy-18* (wild type, n = 2229, s.d. = 0; *xnd-1*, n = 1863, s.d. = 0.21) (right panel). Error bars represent the standard deviation from three or two independent experiments, respectively. **b**, Chromosome I (Chr I) recombination maps for wild-type (WT) and *xnd-1* oocytes (top) and sperm (bottom). **c**, X chromosome recombination maps for wild-type and *xnd-1* oocytes (data for **b**, **c** are found in Supplementary Tables 1–3). Genetic and physical markers are shown above and below the graphic representation of each chromosome, respectively. Boxes represent the relative map size for each interval as determined by SNP analysis (see Methods). Significant differences between wild type and *xnd-1* are marked (*$P < 0.05$; **$P < 0.001$).

of wild-type nuclei (Fig. 2e and Supplementary Movie 1). In contrast, only ~50% of *xnd-1* mutant nuclei had detectable RAD-51 foci on X (Fig. 2e, and Supplementary Movie 2), indicating that *xnd-1* has a crucial role in recruiting DSBs to the X.

In *C. elegans*, pairing and synapsis are independent of DSB formation[16]. Therefore we reasoned these processes should be unaffected by *xnd-1* loss. FISH and immunofluorescence revealed that the kinetics and extent of pairing and synapsis did not differ between *xnd-1* and wild type (Supplementary Fig. 3a, b). Thus, the synaptonemal complex (SC) was fully polymerized between all homologues upon entry into pachytene (Supplementary Fig. 3b and data not shown). Nevertheless, by late pachytene, SYP-1 and SYP-2 were no longer visible on the X in half of the nuclei (Supplementary Fig. 3b, c) and a subset of X's prematurely separated (Supplementary Fig. 3a). Irradiation prevented desynapsis and separation (Supplementary Fig. 3d, e), indicating that these phenotypes are a consequence of insufficient DSB formation. These observations suggest that the formation of DSBs is communicated to the SC, allowing
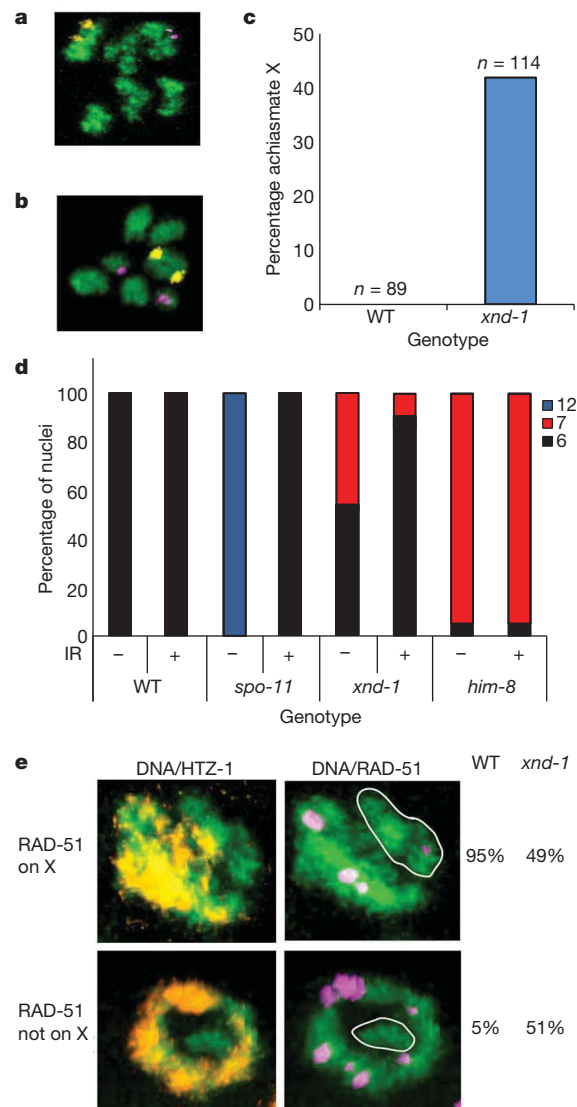
Figure 2 | xnd-1 is required for efficient DSB formation on the X chromosome. a–c, Achiasmate chromosomes are observed at diakinesis in *xnd-1*. FISH probes mark chromosomes V (yellow) and X (magenta); DNA is stained with DAPI (green). **a**, Six DAPI-staining bodies indicate that all chromosomes have recombined, **b**, seven reveal the achiasmate X's in *xnd-1* oocytes. **c**, Quantification of achiasmate X frequency. Chromosome V was never achiasmate in wild type or *xnd-1*. **d**, Ionizing radiation (IR) rescues the CO defects of *xnd-1*. Quantification of ionizing radiation rescue as assessed by the number of DAPI staining bodies at diakinesis 24 h post-irradiation. (Wild type ± ionizing radiation, n = 91, 82; *spo-11* ± ionizing radiation, n = 73, 100; *xnd-1* ± ionizing radiation, n = 190, 157; *him-8* ± ionizing radiation, n = 83, 65). **e**, Fewer RAD-51 foci are observed on the X chromosome (white circles) in *xnd-1*. *rad-54(RNAi)*-treated animals[9] were dissected and germlines co-stained for DNA (green), HTZ-1 (which marks autosomes[17], yellow), and the DNA repair protein, RAD-51 (magenta), which acts as a marker of DSBs[9]. The percentage of nuclei in which RAD-51 foci were observed on the X was quantified (wild type, n = 369; *xnd-1*, n = 391).

for its stabilization. These results complement our understanding that DSBs influence SC polymerization in *C. elegans*[18], revealing an additional role for DSBs in SC maintenance.

To understand *xnd-1* function further, we investigated its subcellular localization. Labelling with anti-XND-1 antibodies revealed nuclear staining from the most distal region of the gonad through mid- to late pachytene (Fig. 3a). XND-1 was not detected in somatic tissues or in *xnd-1* mutant gonads (Supplementary Fig. 4). Closer examination of germline nuclei showed that a single chromosome pair lacked XND-1
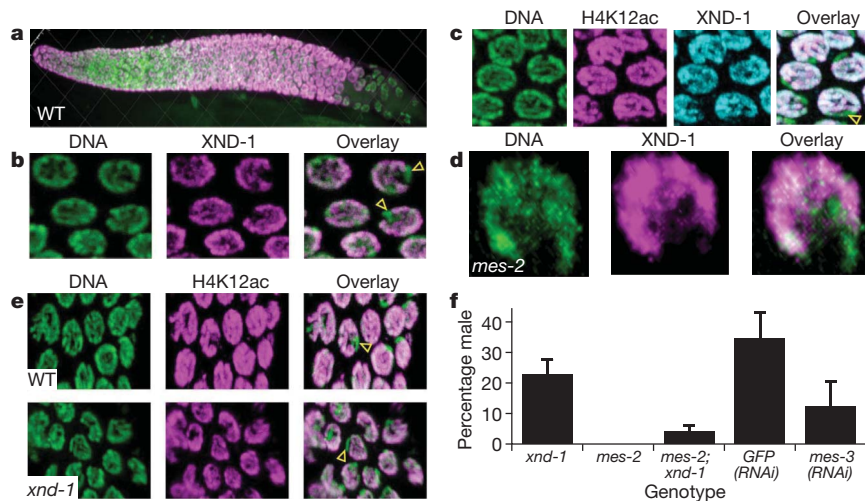
Figure 3 | XND-1 is an autosomal protein that regulates X chromosome crossing over. a, Anti-XND-1 antibody staining of a wild-type hermaphrodite germline. **b**, Close-up of wild-type nuclei reveals the absence of staining on one chromosome (yellow arrowheads). **c**, Co-staining of wild-type pachytene nuclei with anti-XND-1 and anti-H4K12Ac reveals that these proteins are coincident, indicating that XND-1 is enriched on autosomes. A yellow arrowhead indicates the unstained X. **d**, Localization of XND-1 is independent of the X chromosome silencing gene *mes-2*. XND-1 antibody staining in *mes-2 (M⁻Z⁻)* mutants with rare pachytene nuclei reveals normal XND-1 localization. **e**, Activating HPTMs remain excluded from the X in *xnd-1* mutants (yellow arrowheads). Histone H4K12ac (magenta) is enriched on autosomes in wild type (top) and *xnd-1* (bottom). **f**, Suppression of *xnd-1* HIM phenotype by *mes-2(bn11)* and *mes-3(RNAi)* (*xnd-1*, n = >8,000; *mes-2* = >2,000; *mes-2; xnd-1*, n = 733; *GFP(RNAi)*, n = 192; *mes-3(RNAi)*, n = 283). Error bars represent standard deviation from at least three experiments.

labelling (Fig. 3b and Supplementary Fig. 5). Remarkably, co-staining germline nuclei with autosome-specific HPTMs indicated that the unstained chromosome was the X (Fig. 3c, Supplementary Movie 3). Thus, XND-1 is an autosomally enriched protein required for efficient DSB formation on the X.

The localization of XND-1 is strikingly similar to MES-4 and MRG-1, which together with the X-enriched proteins, MES-2, -3 and -6, regulate X chromosome gene silencing[19]. These factors could function to exclude XND-1 from the X. However, in *mes-2 (M⁻Z⁻)* germlines, XND-1 protein did not relocalize to the X (Fig. 3d and Methods). XND-1 localization was also not altered in other mutants affecting gene silencing, meiotic DNA structure or X-specific behaviours, nor in mutants that affect pairing, axis and SC formation, DSB formation and repair (Supplementary Table 5). Thus, XND-1 localization seems to be independent of factors required for X chromosome silencing and vice versa: HPTMs associated with transcriptional repression remain on the X and those associated with activation are not recruited to the X (Fig. 3e and data not shown). Furthermore, *xnd-1* is not required for the similar process of transgene silencing[20] (Supplementary Table 6). Thus, while XND-1 enrichment on autosomes is similar to MES-4 and MRG-1, it confers a novel and independent level of sex-chromosome-specific behaviour.

One possibility for the deficit in COs on the X is the establishment of a more closed, inaccessible chromatin environment. If this were the case, mutations that lead to more open chromatin on the X, such as *mes-2, -3, -4* and *-6*, might counteract the effect of *xnd-1* mutation. We observed that *mes-2* or *mes-3* partially suppressed the high incidence of males (HIM) phenotype of *xnd-1* (Fig. 3f), indicating that *xnd-1* and the *mes* genes have opposing effects on X chromosome COs: *xnd-1* ensuring that it receives COs and the *mes* genes inhibiting CO formation through the imposition of heterochromatic gene silencing. These results suggest that *xnd-1* acts at the level of chromatin structure to control recombination.

To determine whether *xnd-1* has a direct effect on germline chromatin, we examined HPTMs in mutant gonads. In *xnd-1* mutants, a striking increase in H2A lysine 5 acetylation (H2AK5ac) was observed (Fig. 4a, b). In wild type, H2AK5ac labelling appeared faint and punctate whereas in *xnd-1*, H2AK5ac labelling was more intense and uniform (Fig. 4b). No differences were observed for related HPTMs, including H4K8ac, H4K12ac, H2BK5ac and H3K4me2 or me3. Our data suggest that the H2AK5ac mark may be critical for the regulation of meiotic crossovers in *C. elegans*.

In vitro, the human Tip60 protein (also known as KAT5) can acetylate H2AK5 (ref. 21). We reasoned that if H2AK5ac is responsible for the CO defects in *xnd-1*, then mutating the *C. elegans* homologue of Tip60, *mys-1* (ref. 22), might suppress the mutant. No meiotic defects were observed in *mys-1(RNAi)*-treated or wild-type controls as assessed
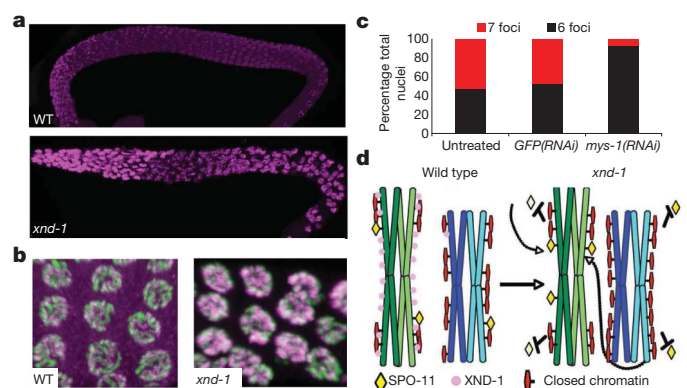
Figure 4 | Germline chromatin is altered in *xnd-1* mutants. a, Histone H2A lysine 5 acetylation is increased in the *xnd-1* mutant. Wild-type (upper) and *xnd-1* (lower) mutant gonads were stained on one slide and imaged identically, revealing the more intense anti-histone H2AK5ac staining in the *xnd-1* gonad. **b**, Mid-pachytene nuclei from wild-type and *xnd-1* gonads (DNA, green; anti-H2AK5ac, magenta) showing the more intense and uniformly distributed staining in the mutant. The gain in the wild-type image has been increased to reveal the modification in these nuclei. **c**, *mys-1(RNAi)* suppresses the X chromosome CO defect of *xnd-1* mutants. Quantification of DAPI staining bodies at diakinesis: 6 foci, grey; 7 foci, red (untreated, n = 94 nuclei; *GFP(RNAi)*, n = 124; *mys-1(RNAi)*, n = 288). **d**, Proposed model for XND-1 function. The presence of XND-1 on autosomes allows SPO-11 to gain access to the transcriptionally silent, heterochromatin-like X chromosomes. In the absence of XND-1, SPO-11 targets the gene-rich clusters (more transcriptionally active regions) on autosomes and titrates DSBs away from the silent X chromosomes and the less transcriptionally active autosomal arms.

by analysing diakinesis figures. As compared to *xnd-1; GFP(RNAi)* or untreated *xnd-1* animals, which each reveals ~50% of nuclei with seven foci, the *xnd-1; mys-1(RNAi)*-treated worms showed only ~12% of oocytes in this class (Fig. 4c). *mys-1(RNAi)* also decreased the amount of the H2AK5ac mark observed in these animals (Supplementary Fig. 6). These data strongly support the conclusion that *xnd-1* is influencing meiotic crossover formation in *C. elegans* through modulation of germline chromatin.

Characterization of *xnd-1* suggests that chromatin structure is a major determinant of chromosome-specific meiotic behaviours and may explain why chromosomes have different propensities to missegregate[10]. We identified XND-1 as an autosome-associated factor required for regulation of X chromosome recombination in *C. elegans*. Because *xnd-1* and *mes* seem to act antagonistically with regard to X COs, we suggest *xnd-1* may have evolved to ensure the proper disjunction of the nearly heterochromatic X, and in so doing, altered the distribution of COs genome-wide. *xnd-1* may limit association (or action) of the DSB machinery with the most permissive sites in the genome, thereby increasing the pool of DSB machinery available to interact with the X chromosome and the gene-sparse regions on autosome arms (Fig. 4i).

xnd-1 expands our understanding of genes involved in CO control. In *C. elegans*, the (as yet uncloned) *rec-1* mutant shows a similar redistribution of CO into the gene cluster[23], but does not affect the frequency of COs on the X. Thus, redistribution of COs to the gene clusters, *per se*, does not cause a deficit in COs on the X. Double-mutant analysis with *xnd-1* will likely inform how these gene products collaborate to regulate global chromatin organization and CO distribution. In humans, a significant fraction of hotspot usage can be explained by variants of PRDM9, a zinc-finger containing histone methyltransferase[24–26], and divergent 13mer motifs to which it binds[27]. Although PRDM9 promotes H3K4 trimethylation[28], it alone cannot explain hotspot choice because this modification localizes to a majority of transcriptionally active loci. Our work implicates another HPTM, H2AK5ac, as a critical determinant of DSB distribution in *C. elegans*. Little is known about the *in vivo* role of H2AK5ac; our study provides the first functional evidence that H2AK5ac has a deterministic role in the meiotic DSB landscape. Our data suggests that Tip60/MYS-1 modulates H2AK5ac levels and thereby functions in CO control. In mammals, Tip60 is recruited to sites of DSB repair by phosphorylated H2AX and leads to acetylation of histone H4, a hallmark of recombinational repair[7]. In *C. elegans*, however, H2AX is not conserved. This raises the possibility that DSB initiation and repair may be coupled by XND-1 through the recruitment or regulation of MYS-1 and/or H2AK5 acetylation.

METHODS SUMMARY

Worm strains and growth conditions. All strains were grown and maintained under standard conditions at 20 °C (ref. 29). *xnd-1(ok708)III* and *xnd-1(ok709)III* were obtained from the *C. elegans* Gene Knockout Consortium, outcrossed at least six times, and maintained balanced over hT2::GFP. The mutations display maternal rescue (10% HIM in M^+Z^- animals (*M*, maternal load; *Z*, zygotic synthesis) compared to ~20% in M^-Z^-) and maternally deposited XND-1 protein can be seen in M^+Z^- germ cells. Therefore, our analyses of *xnd-1* were restricted to M^-Z^- progeny.

xnd-1(ok708) and *(ok709)* gave equivalent percentages of males compared to co-suppression lines, suggesting that they are loss of function alleles (Supplementary Table 1). We present analysis of *ok709*, but similar results have been obtained with *ok708*. Approximately 40% of the *xnd-1* population is 'clear' steriles, which are easily distinguished from fertile animals on adult day 1. These animals were analysed separately for meiotic defects and show no differences in pairing, SC formation or DSB formation compared to wild type. However, these animals have a cellularization defect in which multiple pachytene nuclei are aberrantly packaged into single oocytes creating a disorganized diakinesis region (for example, see Supplementary Fig. 7). These defects result in oocytes and eggs with an increased mass of DAPI staining material and massive aneuploidy and likely account for observed frequency of embryonic lethality.

Immunostaining. Fixation and treatment of gonads for XND-1N staining was performed according to established procedures[30] using 1% paraformaldehyde fix for 5 min before freeze-cracking and ethanol treatment. For XND-1C, gonads were dissected in PBS, immediately freeze-cracked, and fixed in methanol (20 min) and acetone (10 min). Antibody concentrations are provided in Methods.

Irradiation. L4 larvae aged for 20 h and exposed to 2 krad of ionizing radiation using a ^{137}Cs source (MARK I irradiator, J.L. Shepard and Associates). Animals were further aged for 8–24 h before fixation and staining. To determine if ionizing radiation suppressed desynapsis, samples were analysed 8–16 h post-ionizing radiation; for analysis of DAPI staining bodies, 24 h post-ionizing radiation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Gerton, J. L. *et al.* Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 11383–11390 (2000).
2. Buhler, C., Borde, V. & Lichten, M. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in *Saccharomyces cerevisiae*. *PLoS Biol.* **5**, e324 (2007).
3. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485 (2008).
4. Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398 (2008).
5. Robine, N. *et al.* Genome-wide redistribution of meiotic double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol* **27**, 1868–1880 (2007).
6. Mieczkowski, P. A., Dominska, M., Buck, M. J., Lieb, J. D. & Petes, T. D. Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae*. *DNA Repair (Amst.)* **104**, 3955–3960 (2007).
7. Buard, J., Barthes, P., Grey, C. & de Massy, B. Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J.* **28**, 2616–2624 (2009).
8. Borde, V. *et al.* Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* **28**, 99–111 (2008).
9. Mets, D. G. & Meyer, B. J. Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. *Cell* **139**, 73–86 (2009).
10. Hassold, T., Hall, H. & Hunt, P. The origin of human aneuploidy: where we have been, where we are going. *Hum. Mol. Genet.* **16**, R203–R208 (2007).
11. Barnes, T. M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**, 159–179 (1995).
12. Rockman, M. V. & Kruglyak, L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000419 (2009).
13. Kelly, W. G. *et al.* X-chromosome silencing in the germline of *C. elegans*. *Development* **129**, 479–492 (2002).
14. Hodgkin, J., Horvitz, H. R. & Brenner, S. Nondisjunction mutants of the nematode *Caenorhabditis elegans*. *Genetics* **91**, 67–94 (1979).
15. Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).
16. Dernburg, A. F. *et al.* Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* **94**, 387–398 (1998).
17. Petty, E. L., Collette, K. S., Cohen, A. J., Snyder, M. J. & Csankovszki, G. Restricting dosage compensation complex binding to the X chromosomes by H2A.Z/HTZ-1. *PLoS Genet.* **5**, e1000699 (2009).
18. Smolikov, S., Schild-Prufert, K. & Colaiacovo, M. P. *CRA-1* uncovers a double-strand break-dependent pathway promoting the assembly of central region proteins on chromosome axes during *C. elegans* meiosis. *PLoS Genet.* **4**, e1000088 (2008).
19. Fong, Y., Bender, L., Wang, W. & Strome, S. Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. *Science* **296**, 2235–2238 (2002).
20. Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct requirements for somatic and germline expression of a generally expressed *Caenorhabditis elegans* gene. *Genetics* **146**, 227–238 (1997).
21. Ikura, T. *et al.* Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463–473 (2000).
22. Ceol, C. J. & Horvitz, H. R. A new class of *C. elegans* synMuv genes implicates a Tip60/NuA4-like HAT complex as a negative regulator of Ras signaling. *Dev. Cell* **6**, 563–576 (2004).
23. Zetka, M. C. & Rose, A. M. Mutant *rec-1* eliminates the meiotic pattern of crossing over in *Caenorhabditis elegans*. *Genetics* **141**, 1339–1349 (1995).
24. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
25. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).
26. Parvanov, E. D., Petkov, P. M. & Paigen, K. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
27. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).

28. Hayashi, K., Yoshida, K. & Matsui, Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**, 374–378 (2005).
29. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
30. Chan, A. & Meyer, B. J. Protocol 21: antibody staining of *C. elegans* gonads in *Wormbook* 〈http://www.wormbook.org/chapters/www_intromethodscellbiology/intromethodscellbiology.html#d0e2311〉 (2006).

METHODS

Worm strains and growth conditions. All strains were grown and maintained under standard conditions at 20 °C[29]. xnd-1(ok708)III and xnd-1(ok709)III were obtained from the *C. elegans* Gene Knockout Consortium, outcrossed at least six times, and maintained balanced over hT2::GFP. The mutations display maternal rescue (10% HIM in M^+Z^- animals (M, maternal load; Z, zygotic synthesis) compared to ~20% in M^-Z^-) and maternally deposited XND-1 protein can be seen in M^+Z^- germ cells. Therefore, our analyses of xnd-1 were restricted to M^-Z^- progeny.

xnd-1(ok708) and (ok709) gave equivalent percentages of males compared to co-suppression lines, suggesting that they are loss of function alleles (Supplementary Table 1). We present analysis of ok709, but similar results have been obtained with ok708. Approximately 40% of the xnd-1 population is 'clear' steriles, which are easily distinguished from fertile animals on adult day 1. These animals were separately analysed for meiotic defects and show no differences in pairing, SC formation or DSB formation compared to wild type. However, these animals have a cellularization defect in which multiple pachytene nuclei are aberrantly packaged into a single oocytes creating a disorganized diakinesis region (for example, see Supplementary Fig. 7). These defects result in oocytes and eggs with an increased mass of DAPI staining material and massive aneuploidy and therefore likely explains the increase in embryonic lethality.

Strains used for these analyses are listed in Supplementary Table 6 and below:

LGII: mes-2(bn11) unc-4(e120)/mnC1[dpy-10(e128)unc-52(e444)].

LGIII CB185: lon-1(e185). CB364 dpy-18(e364). QP224 unc-45(e286), dpy-18(e364), unc-64(e246). QP231 dpy-1(e1), lon-1(e185), dpy-18(e364) III. RB868 xnd-1(ok708). RB869 xnd-1(ok709).

LGIV: AV106 spo-11(ok79) IV/nT1[unc-?(n754) let-?](IV;V). AV157 spo-11(me44)/nT1[unc-?(n754) let-? qIs50](IV;V). CA257 him-8(tm611)

LG V: JK2663 dpy-11(e224) mes-4(bn67) V/nT1[unc-?(n754) let-? qIs50](IV;V); zuIs178 [his-72(1kb)::HIS-72::GFP]; unc-119(ed3)]

LG X: CB678 lon-2(e678).

Extrachromosomal array: let-868::GFP[20].

Recombination screen. Triply marked mutant hermaphrodites, either unc-45, dpy-18, unc-64III or dpy-1, lon-1, dpy-18 III, were crossed to N2 males to produce F1 progeny heterozygous for each of the three markers. Approximately five L4 hermaphrodite F1 progeny per plate were fed dsRNA for gfp (control) or chromatin-associated genes according to established protocols[31]. Plates were scored 3–5 days later for F2 progeny that were either Dpy non-Unc or Lon non-Dpy, respectively, indicating they were homozygous for the middle marker and heterozygous or wild type for flanking markers. Because distinguishing the flanking Unc or Dpy mutant phenotypes from each other was not possible, the assessment of recombination rates was expressed as a fraction of the total progeny. Three to ten plates per genotype were scored, with a minimum of two trials per genotype.

Characterization of xnd-1 gene and transcripts. To confirm the endpoints of the deletion alleles ok708 and ok709, primers flanking the putative breakpoints were used to amplify the genomic region from homozygous animals. PCR products were sequenced using primers spanning the region. ok708 is a simple deletion and ok709 a more complex deletion/inversion. Breakpoints of both deletions are shown in Supplementary Fig. 1. Additional sequence information is available from WormBase.

Seven cDNA clones were obtained from Y. Kohara, and were sequenced using standard procedures. Five of the seven cDNAs matched the reported gene structure, according to WormBase. Two cDNAs revealed the presence of a novel splice variant (Supplementary Fig. 1).

XND-1 antibody production. A GST fusion protein to the XND-1 amino terminus (amino acids 1–329) was expressed from pGex-6P-1 (GE Healthcare) in BL21 DE3 cells. Protein was induced with 100 μM isopropylthiogalactoside (IPTG) for 2 h at room temp before cell lysis and protein purification. The protein was run over a glutathione-agarose column and cleaved with PreScission protease (GE Healthcare) to isolate untagged protein for inoculation. A GST-tagged construct of the XND-1 carboxy terminus, amino acids 530–702, was also made and protein purification performed and both sets of proteins were inoculated into animals for antibody production by the Proteintech Group. The XND-1N protein was inoculated into guinea pigs, XND-1C protein into rabbits.

Immunostaining. Fixation and treatment of gonads for XND-1N staining was performed according to established procedures[30] using 1% paraformaldehyde fix for 5 min before freeze cracking and ethanol treatment. For XND-1C, gonads were dissected in PBS, immediately freeze-cracked, and fixed in methanol (20 min) and acetone (10 min).

XND-1 antibodies were pre-absorbed against xnd-1(ok708) animals[32]. XND-1N was used at a final concentration of 1:600; XND-1C at 1:200. Antibodies were generously provided to us by numerous labs and were used at the following concentrations: guinea pig anti-HTP-3(1:500) and guinea pig anti-HIM-8

(1:500) from A. Dernburg; rabbit anti-HIM-3 (1:200) from A. Villeneuve; and guinea pig anti-SYP-1N (1:200), rabbit anti-SYP-1C (1:200), and rabbit anti-SYP2 (1:200) antibodies from M. Colaiacovo; rabbit anti-RAD-51 (1:250 to 1:500) from A. la Volpe; rat anti-HTZ-1 (1:2,000) antibody from G. Csankovszki. Rabbit anti-RAD-51 (1:10,000) and anti-REC-8 (1:10,000) antibodies were purchased from SDI; rabbit anti-histone H4 (ac12) (1:1,000) from AbD Serotec; rabbit anti-dimethyl histone H3 (Lys 4) (1:1,000) from Upstate; rabbit anti-acetyl H2A (Lys 5) (1:500; 1:2,500) from Cell Signaling/Millipore and Serotec. Secondary antibodies were Alexa- conjugated guinea pig 488 or 561; rabbit 561; or rat 488 (Invitrogen Molecular Probes).

For staining mes M^-Z^- germlines, the first F1 daughters of heterozygous mes mothers were individually plated and their first daughters were analysed. By this method, we selected for granddaughters with the greatest deposition of (grand)maternal product which enhanced our ability to isolate F2 progeny with germ cells. We excluded from our analysis animals in which germ cell morphology became indistinct.

FISH analysis. FISH XR, XL and 5S probes were synthesized and labelled as described previously[33]. FISH gonad preparation, fixation and hybridization were performed according to published protocols[16]. FISH in combination with antibody staining (SYP1C, 1:200; XND-1, 1:600) was performed as described[34].

Imaging. Except where noted, all studies were analysed by confocal microscopy on either a Leica TCS SP2 or SP5 microscope followed by three-dimensional analysis using Imaris software (Bitplane).

Irradiation. L4 larvae aged for 20 h and were exposed to 2 krad of ionizing radiation using a ^{137}Cs source (MARK I irradiator, J.L. Shepard and Associates). Animals were further aged for 8–24 h before fixation and staining. To determine if ionizing radiation suppressed desynapsis, samples were analysed 8–16 h post-ionizing radiation; for analysis of DAPI staining bodies, 24 h post-ionizing radiation.

Analysis of break formation on the X chromosome. rad-54 or gfp RNAi were performed as described[9]. L4 larval progeny were dissected, fixed and stained with DAPI, rat anti-HTZ-1 and rabbit anti-RAD-51 antibodies. Quantification of X chromosomal RAD-51 foci was performed using Imaris 3-D software and only mid- to late pachytene nuclei with eight or more foci were evaluated.

Analysis of pairing and synapsis. xnd-1 and wild-type germlines were dissected and co-labelled with 5S and XL FISH probes (described above) followed by labelling with anti-SYP-1 antibody and DAPI. Confocal images of full germlines were analysed by dividing the transition zone to end of pachytene region into four approximately equal parts and assessing the presence or absence of apposed FISH probes (indicative of pairing chromosomes) and the staining of SYP-1 along the length of a DNA fragment adjacent to the marked chromosome.

SNP mapping. xnd-1(ok708) was introgressed six times into the CB4658 Hawaiian background and the presence of Hawaiian-specific polymorphisms was confirmed by genotyping. Spontaneous males from this stock were crossed to lon-2(e678); xnd-1(ok709) hermaphrodites and non-Lon, xnd-1 trans-heterozygous hermaphrodites were subsequently crossed to GFP+ males. GFP+ L4 hermaphrodites were individually plated and genotyped as described for chromosome X[35] and chromosome I[36]. Chromosome I primers correspond to location 1, −19, −12, −6, −1, 5, 13, 26 (ref. 36). To map sperm recombination, the Lon, trans-heterozygous males from the above cross were mated to dpy-18 hermaphrodites and non-Dpy cross progeny were individually plated and their progeny were genotyped as described above for chromosome I markers. Wild type controls for both sperm and egg were prepared simultaneously and X chromosome mapping was described previously[35]. Chromosome I mapping was performed as described above.

Hatching rates. Individual L4 larvae were plated onto to freshly seeded NGM plates and transferred every 12 h until the cessation of egg-laying. Eggs were counted immediately after transferring and adult animals were counted by picking worms off the plates 60–96 h later. The number of wild-type hermaphrodite (XX), male (XO), and DPY hermaphrodites (XXX) progeny were counted.

RNA interference. RNAi was performed according to established protocols[9]. For mys-1(RNAi), xnd-1 or wild type L4 larvae were placed on bacteria expressing the dsRNA for 48 h and allowed to recover for 48 h before fixation and DAPI staining. Diakinesis stage oocytes were assayed as described above for the number and appearance of the DAPI-staining bodies.

31. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395,** 854 (1998).
32. Colaiácovo, M. P. *et al.* Synaptonemal complex assembly in *C. elegans* is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Dev. Cell* **5,** 463–474 (2003).
33. Phillips, C. M. *et al.* HIM-8 binds to the X chromosome pairing center and mediates chromosome-specific meiotic synapsis. *Cell* **123,** 1051–1063 (2005).
34. Tsai, C. J. *et al.* Meiotic crossover number and distribution are regulated by a dosage compensation protein that resembles a condensin subunit. *Genes Dev.* **22,** 194–211 (2008).

35. Lim, J. G., Stine, R. R. & Yanowitz, J. L. Domain-specific regulation of recombination in *Caenorhabditis elegans* in response to temperature, age and sex. *Genetics* **180,** 715–726 (2008).

36. Davis, M. W., Hammarlund, M., Harrach, T., Hullett, S. O. & Jorgensen, E. M. Rapid single nucleotide polymorphism mapping in *C. elegans. BMC Genomics* **6,** 118 (2005).

LETTER

The structural basis for autonomous dimerization of the pre–T-cell antigen receptor

Siew Siew Pang[1], Richard Berry[1], Zhenjun Chen[2], Lars Kjer-Nielsen[2], Matthew A. Perugini[3], Glenn F. King[4], Christina Wang[1], Sock Hui Chew[1], Nicole L. La Gruta[2], Neal K. Williams[1], Travis Beddoe[1], Tony Tiganis[1], Nathan P. Cowieson[1]†, Dale I. Godfrey[2], Anthony W. Purcell[3], Matthew C. J. Wilce[1], James McCluskey[2] & Jamie Rossjohn[1]

The pre-T-cell antigen receptor (pre-TCR), expressed by immature thymocytes, has a pivotal role in early T-cell development, including TCR β-selection, survival and proliferation of CD4⁻CD8⁻ double-negative thymocytes, and subsequent αβ T-cell lineage differentiation[1–3]. Whereas αβTCR ligation by the peptide-loaded major histocompatibility complex initiates T-cell signalling[4], pre-TCR-induced signalling occurs by means of a ligand-independent dimerization event[5]. The pre-TCR comprises an invariant α-chain (pre-Tα) that pairs with any TCR β-chain (TCRβ) following successful TCR β-gene rearrangement[6]. Here we provide the basis of pre-Tα–TCRβ assembly and pre-TCR dimerization. The pre-Tα chain comprised a single immunoglobulin-like domain that is structurally distinct from the constant (C) domain of the TCR α-chain[7]; nevertheless, the mode of association between pre-Tα and TCRβ mirrored that mediated by the Cα–Cβ domains of the αβTCR. The pre-TCR had a propensity to dimerize in solution, and the molecular envelope of the pre-TCR dimer correlated well with the observed head-to-tail pre-TCR dimer. This mode of pre-TCR dimerization enabled the pre-Tα domain to interact with the variable (V) β domain through residues that are highly conserved across the Vβ and joining (J) β gene families, thus mimicking the interactions at the core of the αβTCR's Vα–Vβ interface. Disruption of this pre-Tα–Vβ dimer interface abrogated pre-TCR dimerization in solution and impaired pre-TCR expression on the cell surface. Accordingly, we provide a mechanism of pre-TCR self-association that allows the pre-Tα chain to simultaneously 'sample' the correct folding of both the V and C domains of any TCR β-chain, regardless of its ultimate specificity, which represents a critical checkpoint in T-cell development. This unusual dual-chaperone-like sensing function of pre-Tα represents a unique mechanism in nature whereby developmental quality control regulates the expression and signalling of an integral membrane receptor complex.

How the extracellular domain of the pre-Tα chain is able to 'report' productive TCR β-gene rearrangement and proper β-chain folding through pairing with any successfully rearranged TCR β-chain, comprising variable V, D and J segments, was unclear. Therefore, we expressed and purified the extracellular domains of pre-Tα–TCRβ and determined its crystal structure (Supplementary Table 1). The pre-Tα domain adopts an immunoglobulin fold (Fig. 1a), best described as a C-type topology, and comprises a compact β-sandwich of two antiparallel β-sheets tethered by the canonical intramolecular disulphide bond, with the amino and carboxy termini at opposite ends of the molecule (Fig. 1a). The pre-Tα domain was more structurally similar to the C domain of an antibody light chain (IgλC; 34% sequence similarity; root mean squared deviation of 5.1 Å for 73 Cα atoms) in comparison with the Cα domain (26% sequence similarity; root mean squared deviation

12.6 Å for 57 Cα atoms) (Supplementary Fig. 1a). Indeed, the pre-Tα fold contrasted with that of the atypical immunoglobulin fold of the αβTCR's Cα domain[7] (Fig. 1b), which is characterized by one β-sheet that abuts a number of loosely packed random coils. Moreover, the pre-Tα domain was notably elongated (42 Å) in comparison with its Cα counterpart (33 Å), an elongation that is attributable to longer A, B and E β-strands, and the E–F and A–B loops (Fig. 1). Although the 'top' of the pre-Tα domain is unusually hydrophobic, flat and featureless (Supplementary Fig. 1b, c), a large cluster of hydrophobic residues is present on one face of the β-sheet (Supplementary Fig. 1c) and a Trp residue is prominently exposed on the other face (Supplementary Fig. 1b,c). These are features that typically mediate protein–protein interactions, and indeed they were observed to have a role in mediating pre-TCR assembly (see below).

Within the pre-Tα–TCRβ complex, the conformation of the TCR β-chain closely matched that observed in the parental LC13 αβTCR[8], revealing that the lack of a Vα domain in the pre-TCR does not result in the crumpling of the Vβ domain (Fig. 2a, b). Notably, this asymmetrical shape leads to a large, surface-exposed hydrophobic patch on the unpaired Vβ domain within the pre-Tα–TCRβ complex, leaving Tyr 35 and Phe 108 solvent exposed (Fig. 2a). Within the pre-TCR, the pre-Tα domain interacts exclusively with the Cβ domain in a clasp-like interaction (Fig. 2a). Collectively, this forms a continuous and extensive interface with a buried surface area (BSA) of ~1,770 Å², albeit with poor
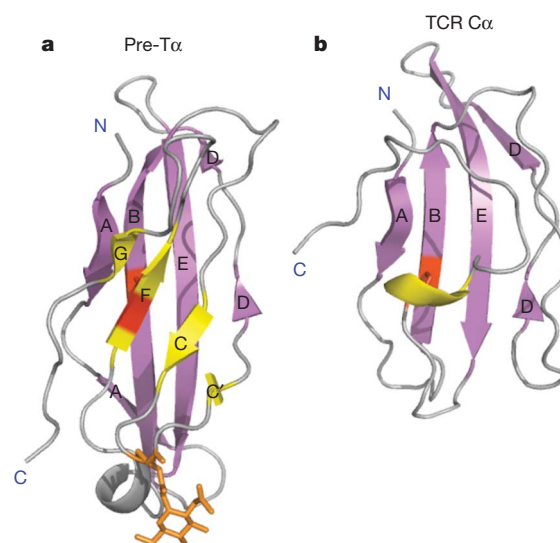
Figure 1 | The pre-Tα structure. Immunoglobulin folds of pre-Tα (**a**) and the Cα domain (**b**), with disulphide bond in red.

[1]The Protein Crystallography Unit, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia. [2]Department of Microbiology & Immunology, University of Melbourne, Parkville, Victoria 3010, Australia. [3]Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. [4]Division of Chemistry & Structural Biology, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia. †Present address: The Australian Synchrotron, Blackburn Road, Monash, Clayton, Victoria 3168, Australia.
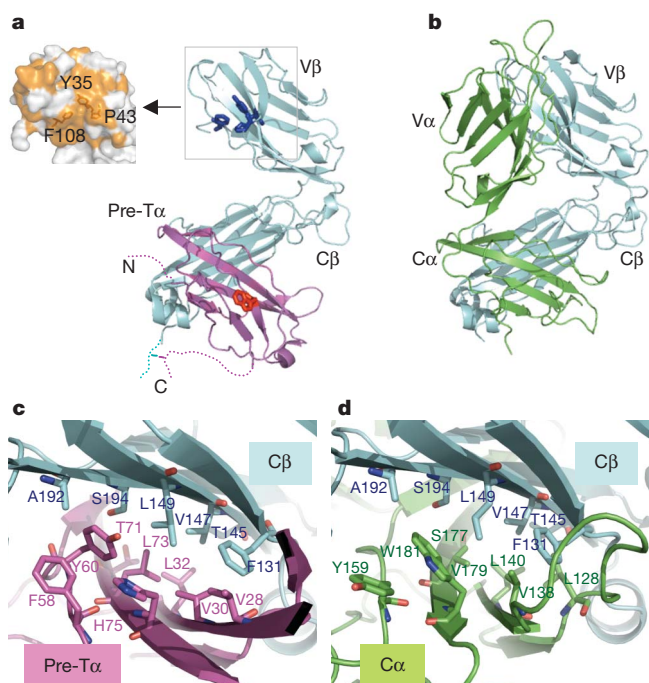
Figure 2 | Comparison between the pre-TCR and αβTCR structures. a, Pre-Tα (magenta) interacts with the Cβ domain. The hydrophobic patch is shown as a blue stick model (also see inset); Trp 46, red. Dotted lines represent the N and C termini, which are not visible in the crystal structure. **b,** αβTCR (Protein Data Bank ID, 1KGC[9]) contains two chains (α, green; β, cyan), with Vα interacting with Vβ and Cα interacting with Cβ. **c, d,** The contact interfaces between pre-Tα (magenta) and Cβ (cyan) of pre-TCR (**c**) and between Cα (green) and Cβ (cyan) of αβTCR (**d**). The contact residues are shown as stick models.

shape complementarity, of 0.57 (Fig. 2c and Supplementary Table 2). The core of the pre-Tα–Cβ interface is dominated by van der Waals contacts, including interactions mediated by a number of small polar or aliphatic residues along the surface of the β-sheets and a cluster of aromatic residues (Fig. 2c and Supplementary Table 2). The location of the pre-Tα–Cβ interface coincides with that of the Cα–Cβ interface, although the latter interface is much more extensive[9] (BSA ≈ 2,500 Å^2; Fig. 2d). The larger Cα–Cβ interface arises mainly from the longer D–E loop of the Cα domain. Nevertheless, core interactions at the Cα–Cβ interface are mimicked at the pre-Tα–Cβ interface, involving 17 common Cβ contact residues that include a series of interlocking small hydrophobic residues between the β-sheets and a small aromatic cluster (Fig. 2d and Supplementary Table 2), which suggests that hydrophobic forces drive pre-Tα–TCRβ assembly[9]. The larger BSA and greater shape complementarity (0.67) at the Cα–Cβ interface is consistent with the ability of the TCR α-chain to compete off the pre-Tα domain effectively during αβTCR development.

The pre-TCR is known to self-associate on the cell surface of CD4$^-$CD8$^-$ double-negative thymocytes, and spontaneously dimerizes *in vitro* and *in vivo*, with the dimerization being mediated by residues within the pre-Tα domain[5,10]. In the absence of any structural data, this has led to speculative models of the pre-TCR dimeric assembly. Within the crystal lattice, two potential pre-TCR dimers were observed. The first pre-TCR dimer involved a side-by-side arrangement in which the total BSA at the interface was small (BSA ≈ 1,380 Å^2), and involved contacts that were mediated mainly through polar residues (Supplementary Fig. 2). This side-by-side pre-TCR dimer would sterically block any interactions the Cβ domain is reported to make with the CD3 complex[11]. Furthermore, the shape of this side-by-side pre-TCR dimer did not match what was observed in solution (see below). Accordingly, the side-by-side pre-TCR dimer was considered to be attributable solely to the crystal contacts and was discounted from further analyses. The
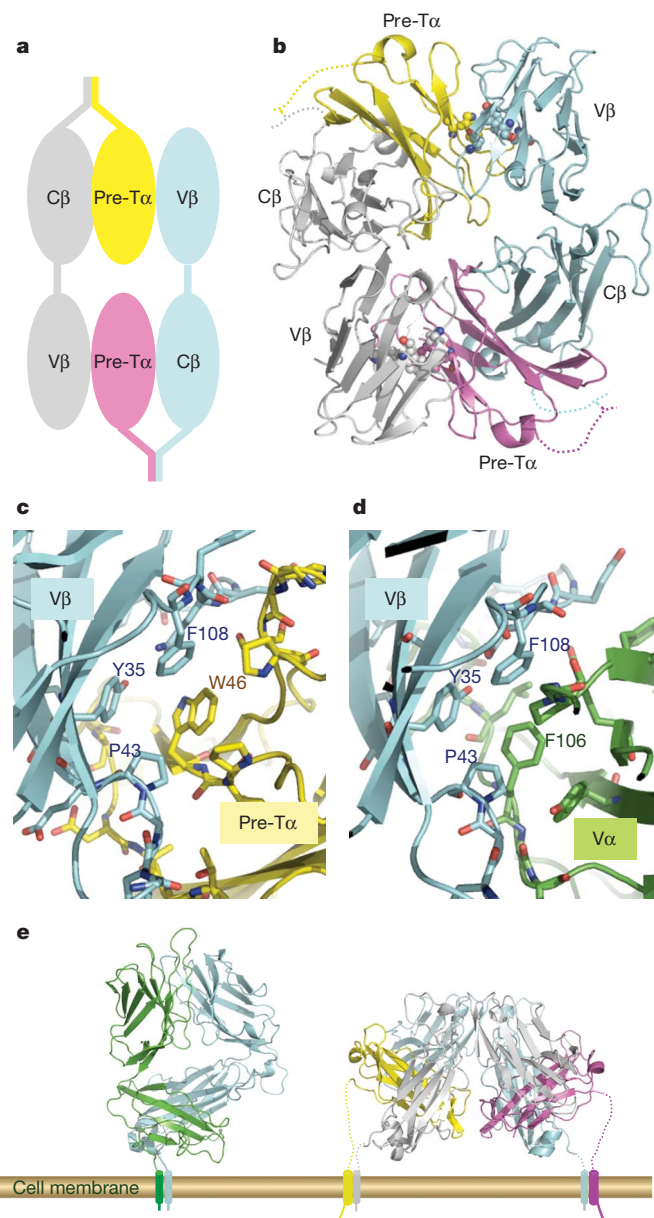
Figure 3 | The pre-TCR dimer. a, Pre-TCR (pre-Tα, magenta; β-chain, cyan) interacts with a second receptor (pre-Tα, yellow; β-chain, grey) in a head-to-tail arrangement. **b,** Dimerization is mediated mainly between the pre-Tα and the unpaired Vβ domains. Some of the residues involved at these interfaces are shown as space-filling models. **c, d,** Contact interface between pre-Tα (yellow) and Vβ (cyan) of pre-TCR (**c**) and between Vα (green) and Vβ (cyan) of αβTCR (**d**). The highly conserved interacting residues are labelled and shown as stick models. **e,** Cartoon showing orientation of the TCRs on the cell surface: left, mature αβTCR; right, pre-TCR.

second pre-TCR dimer (Fig. 3) observed in the crystal lattice was markedly more biophysically and biologically convincing, and involved a head-to-tail dimeric arrangement of the pre-TCR heterodimers. Here the pre-Tα domain of one protomer was sandwiched between the Cβ and Vβ domains (Fig. 3a, b). The hydrophobic and featureless 'top' of the pre-Tα domain was critical for this head-to-tail dimeric arrangement, as large, bulky and charged residues would have prevented this mode of dimerization. The pre-TCR head-to-tail dimerization creates two new interfaces, each involving the C'CGF β-sheet of pre-Tα from one receptor interacting with the β-sheet of the unpaired Vβ domain from the other (Fig. 3b and Supplementary Table 3). This arrangement occludes the otherwise surface-exposed hydrophobic patch on the Vβ domain (Fig. 2a) and provides immediate insight into how and why the

pre-TCR autonomously dimerizes in a *cis*-configuration on the cell surface (Fig. 3e). By pairing with the Vβ domain, the location of the pre-Tα domain mimics the positioning of the Vα domain at the Vα-Vβ interface, although different faces of the domains are used (Fig. 2b). The total BSA at the head-to-tail pre-TCR dimer interface was large (BSA ≈ 3,000 Å^2) and comprised a large number of van der Waals interactions (Supplementary Table 3), as is typically observed within oligomeric protein–protein complexes. Unlike in previous theories of pre-TCR dimerization[5], charged residues within the pre-Tα domain did not have a role in mediating pre-TCR dimerization (Supplementary Fig. 3), and thus may be involved in other ways, such as in pre-TCR–CD3 assembly. Consistent with this, some of the mutations known to affect pre-TCR function reside within the A–B loop of the pre-Tα domain (Asp 22 and Lys 24), a loop that is implicated in αβTCR–CD3 interactions[12]. Importantly, the pre-Tα–Vβ interface principally involved residues located on the C′CGF-sheet of the pre-Tα domain (Fig. 3c), a region that is structurally dissimilar from the Cα domain, thereby providing an explanation of why the corresponding Cα domain cannot act as a suitable surrogate for the pre-Tα domain in mediating this head-to-tail dimeric assembly and, hence, pre-TCR signalling. Within this interface, which covers ≈1,260 Å^2, the invariant pre-Tα Trp 46 is positioned at the core (Fig. 3c and Supplementary Table 3). This ball-and-socket interaction mirrored what was observed at the LC13 Vα–Vβ interface (BSA ≈ 1,490 Å^2), in which Phe 106 from Vα is the corresponding central residue; furthermore, there were 12 common Vβ contact points between pre-Tα–Vβ and Vα–Vβ (Fig. 3d and Supplementary Table 3). Notably, pre-Tα makes contact with residues encoded by the Vβ domain (Tyr 35 and Pro 43) and the Jβ region (Phe 108), thus providing a mechanism that enables the pre-Tα domain to 'sense' the correct folding of the various components derived from disparate gene segments that create the TCR β-chain. Within the TCR β-locus, there are 42 Vβ gene segments, but of these two aforementioned Vβ contact points, position 43 is invariably occupied by a short hydrophobic residue (Pro, Leu, Ile or Val), and Tyr 35 is invariant. Furthermore, Phe 108 is almost invariant within the 12 Jβ gene segments (Supplementary Fig. 4). This conserved constellation of interacting residues provides a framework for understanding how pre-Tα can indiscriminately pair with any successfully rearranged TCR β-chain.

To further verify that the pre-TCR could dimerize *in vitro*, we investigated the behaviour of the pre-TCR in solution. As judged by sedimentation velocity analytical ultracentrifugation (SV-AUC) and small-angle X-ray scattering (SAXS), the pre-TCR displayed the hallmarks of a weak self-associating system, typical of many immunological receptors (Fig. 4)[13]. Specifically, using SAXS analysis, we observed the pre-TCR to undergo a concentration-dependent increase in molecular mass, radius of gyration (R_g) and maximal particle dimension (Supplementary Table 4). These effects cannot be accounted for by the presence of non-specific protein aggregation, as judged by the linearity of SAXS Guinier plots (Supplementary Fig. 5). Furthermore, using SV-AUC, two distinct species were observed (Fig. 4a–c). On the basis of hydrodynamic calculations, these species correspond to pre-Tα–TCRβ (apparent sedimentation coefficient, s_{app} = 3.1 S; frictional ratio, f/f_0 = 1.44) and the pre-TCR dimer (s_{app} = 5.0 S, f/f_0 = 1.37) (Fig. 4a–c, e and Supplementary Table 5). Consistent with the AUC results, the SAXS measurements show an increase in average particle size with increasing protein concentration, consistent with a monomer–dimer equilibrium (Supplementary Fig. 6a). To gain an insight into the shape of the pre-TCR in solution, low-resolution models were restored from the SAXS scattering curves *ab initio* using the DAMMIF[14] and GASBOR[15] programs (Supplementary Fig. 6a). As the molecular weight determined by SAXS suggests that the pre-TCR is fully dimeric at 15 mg ml^{-1}, we fitted the pre-TCR dimer into this SAXS model. In contrast to the side-by-side pre-TCR dimer (Supplementary Fig. 6b), the head-to-tail pre-TCR dimer convincingly fitted the central globular region of the SAXS model (Fig. 4f and
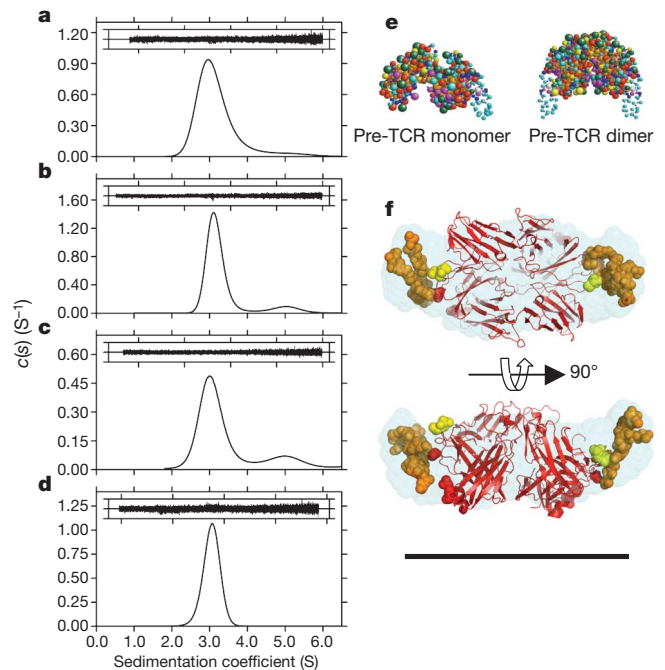
Figure 4 | Pre-TCR dimerization in solution. a–d, Continuous size distributions, determined by SV-AUC, of pre-TCR at 1 mg ml^{-1} (**a**), 3 mg ml^{-1} (**b**) and 8.7 mg ml^{-1} (**c**), and of W46αR pre-TCR mutant at 1.4 mg ml^{-1} (**d**). Insets show the corresponding residuals. **e,** Bead models generated from the pre-TCR structure for the monomer (left) and the head-to-tail dimer (right). **f,** Overlay of the pre-TCR head-to-tail dimer (red ribbon) and the DAMMIF SAXS model (cyan spheres), shown in two orientations. Red dots indicate C-terminal residues of the pre-Tα and LC13 β-chains, orange dots indicate modelled C-terminal tails and yellow dots indicate N-linked glycosylation sites. Scale bar, 100 Å.

Supplementary Fig. 6c). Furthermore, in the head-to-tail pre-TCR dimer, the regions that are missing from the crystal structure (principally the C-terminal stalks of both pre-Tα and the TCR β-chains) are present at opposing poles, where they are well positioned to occupy the areas left vacant by the crystal structure. Thus, the head-to-tail pre-TCR dimer observed in the crystal structure correlated well with the differing solution-based measurements. This *cis*-dimeric arrangement suggests that the pre-TCR sits 'flat', almost parallel to the membrane surface, in contrast to the αβTCR, which is usually depicted in an upright orientation on the membrane (Fig. 3e). Consistent with this, the relative dependency of the various CD3 subunits within the pre-TCR complex may differ from that within the αβTCR–CD3 complex[16].

Next we used three distinct approaches to further validate the pre-TCR dimer observed in the crystal structure and in solution. First, we used crosslinking/mass spectrometry studies of recombinant pre-TCR to directly assess the configuration of the dimer (Supplementary Fig. 7 and Supplementary Table 6; see Methods). Notably, the only cross-linked species between the pre-Tα chain and the TCR β-chain that we observed was between the N-terminal sequence (head) of the Vβ domain and the C-terminal sequence (tail) of the pre-Tα chain (Supplementary Fig. 7 and Supplementary Table 6). This crosslinked species is incompatible with the pre-Tα–TCRβ and the side-by-side pre-TCR dimer, but is fully consistent with a head-to-tail pre-TCR dimer configuration. Second, we generated a W46αR mutant in the pre-Tα chain and assessed its impact on pre-TCR dimerization as judged by SV-AUC. In contrast to the wild-type (WT) pre-TCR, sedimentation velocity analysis of the W46αR mutant shows that it is monomeric in solution (s_{app} = 3.1 S) with no evidence of dimerization (Fig. 4d). Third, to assess the pre-TCR dimer conformation on the cell surface, we transfected WT and mutant pre-TCRs into the TCR αβ-negative,
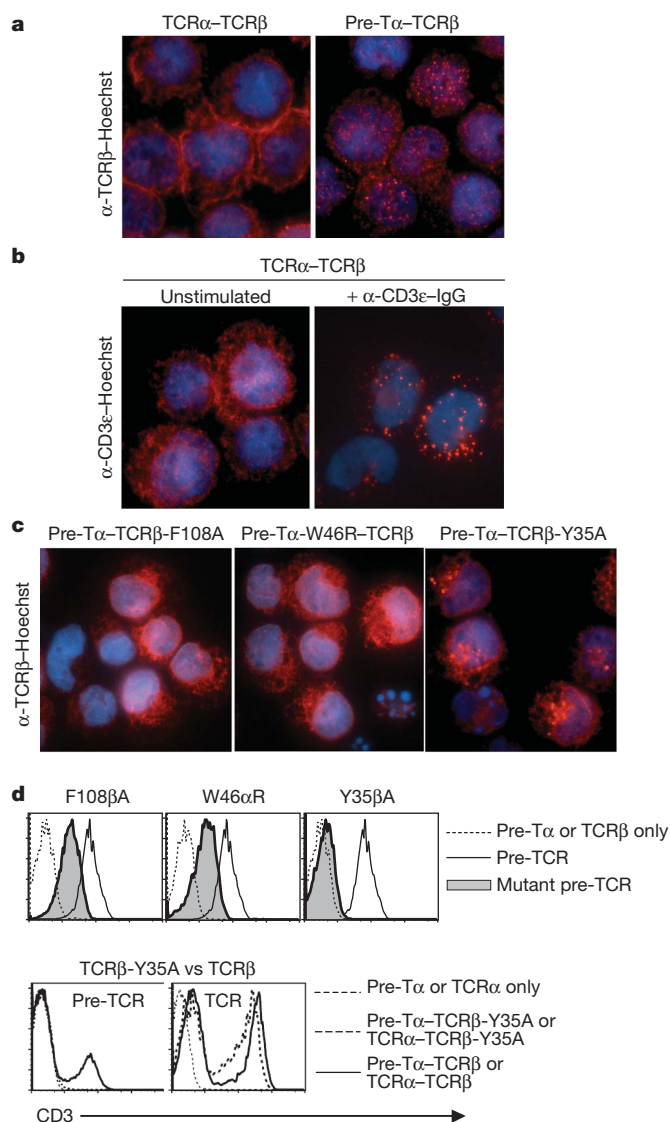
Figure 5 | Expression of pre-TCR mutants. a, TCRβ staining of SKW3 T cells expressing TCRα–TCRβ or pre-Tα–TCRβ. Hoechst, Hoechst stain. **b**, CD3ε staining of unstimulated and activated SKW3 T cells expressing TCRα–TCRβ. IgG, immunoglobulin G. **c**, TCRβ staining of T cells expressing the indicated pre-Tα and TCRβ mutations. **d**, Flow staining (CD3ε) of T cells expressing pre-Tα and TCRVβ mutations. The top row shows negative-control T cells (either pre-Tα or LC13β alone; thin dashed lines), positive-control pre-Tα–LC13β transfectants (solid lines) and SKW3.pre-Tα+LC13β mutants (grey shading). The bottom row shows unsorted SKW3 T cells expressing either pre-Tα–LC13β-Y35A, pre-Tα alone or pre-Tα–LC13β (left), and unsorted SKW3 T cells expressing LC13α–LC13β-Y35A, LC13α–LC13β or TCRα alone (negative control; right). See also Supplementary Fig. 8.

CD3-positive SKW3 T-cell line and visualized cell surface and intracellular pre-TCR localization (Fig. 5). As a control, we also transfected the parental LC13 αβTCR into the SKW3 cell line. We demonstrated, using immunofluorescence, that transfection of the WT pre-TCR resulted in a punctate expression pattern, typical of pre-TCR clustering on the cell surface and consequent endocytosis, which is in direct contrast to the linear cell surface expression pattern of the mature αβTCR in unstimulated cells (Fig. 5a) and the large puncta resulting from crosslinking the αβTCR on the cell surface with a CD3ε-reactive antibody (Fig. 5b). Next we examined the impact of three pre-TCR mutants, Y35βA, F108βA and W46αR, on cell surface expression and intracellular localization using flow cytometric analysis and immunofluorescence microscopy. The F108βA and W46αR pre-TCR mutants were expressed at lower levels on the cell surface and led to a greater retention

of these mutants in an intracellular compartment typical of the endoplasmic reticulum (Fig. 5c, d). Furthermore, the Y35βA TCR–Vβ mutant completely abrogated cell surface expression of the pre-TCR (Fig. 5d) and resulted in accumulation of the expressed pre-TCR Y35βA mutant in perinuclear puncta reminiscent of the Golgi complex and the endoplasmic reticulum, consistent with defective trafficking to the cell surface (Fig. 5c). However, the Y35βA mutant did not affect LC13 αβTCR cell surface expression (Fig. 5d), indicating that this mutant does not affect the fold of the TCR β-chain *per se*. This finding highlights the fact that pre-TCR cell surface expression is finely tuned to pre-Tα–Vβ interactions. Collectively, the crosslinking experiments in solution, coupled with the effect of the mutants on pre-TCR dimerization in solution and on the cell surface, are fully consistent with the head-to-tail pre-TCR dimer configuration.

In vertebrates there are three classes of TCR, namely αβTCR, γδTCR and the pre-TCR. Although structural data exists on the architecture of αβTCR[7] and γδTCRs[17], information is lacking on the pre-TCR. Furthermore, whereas αβTCR and γδTCR interact with defined ligands, the pre-TCR is unique in that a ligand-independent mechanism underpins its role in T-cell development. Our findings formally demonstrate that the pre-Tα chain functions much more than a simple surrogate TCR α-chain, whereby the pre-Tα domain simultaneously binds to the Cβ and Vβ domains within the head-to-tail pre-TCR dimer. The role of the pre-Tα domain is to act as a checkpoint that assesses correct rearrangement and expression of the TCR β-chain on the cell surface. The observed mode of dimerization can only occur when both the variable and constant domains of the TCR β-chain are correctly folded; this is a previously unrecognized determinant for β-selection. Thus, the pre-Tα has a dual chaperone-sensing role that is synchronously linked to dimerization and pre-TCR signal transduction. Many proteins require chaperones for correct folding, and subsequent function, with intricate cellular machines typically being required to aid in this process. The pre-B-cell receptor[18] has a number of components that allow productive expression of the B-cell receptor in a manner that is quite distinct from the individual pre-TCR. The evolutionarily conserved pre-Tα has achieved its role by adopting dual functionality and, as such, represents a previously undescribed mechanism in nature that ensures the integrity of protein assembly and function.

METHODS SUMMARY

1. von Boehmer, H. Unique features of the pre-T-cell receptor α-chain: not just a surrogate. *Nature Rev. Immunol.* **5,** 571–577 (2005).
2. Saint-Ruf, C. *et al.* Different initiation of pre-TCR and γδTCR signalling. *Nature* **406,** 524–527 (2000).

3. Borowski, C., Li, X., Aifantis, I., Gounari, F. & von Boehmer, H. Pre-TCRα and TCRα are not interchangeable partners of TCRβ during T lymphocyte development. *J. Exp. Med.* **199,** 607–615 (2004).
4. Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248,** 701–702 (1974).
5. Yamasaki, S. *et al.* Mechanistic basis of pre-T cell receptor-mediated autonomous signaling critical for thymocyte development. *Nature Immunol.* **7,** 67–75 (2005).
6. Saint-Ruf, C. Analysis and expression of a cloned pre-T cell receptor gene. *Science* **266,** 1208–1212 (1994).
7. Garcia, K. C. *et al.* An αβ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274,** 209–219 (1996).
8. Kjer-Nielsen, L. *et al.* A structural basis for the selection of dominant αβ T cell receptors in antiviral immunity. *Immunity* **18,** 53–64 (2003).
9. Kjer-Nielsen, L. *et al.* The 1.5 Å crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. *Structure* **10,** 1521–1532 (2002).
10. Yamasaki, S. & Saito, T. Molecular basis for pre-TCR-mediated autonomous signaling. *Trends Immunol.* **28,** 39–43 (2007).
11. Kuhns, M. S. & Davis, M. M. Disruption of extracellular interactions impairs T cell receptor-CD3 complex stability and signaling. *Immunity* **26,** 357–369 (2007).
12. Beddoe, T. *et al.* Antigen ligation triggers a conformational change within the constant domain of the αβ T cell receptor. *Immunity* **30,** 777–788 (2009).
13. van der Merwe, P. A. & Davis, S. J. Molecular interactions mediating T cell antigen recognition. *Annu. Rev. Immunol.* **21,** 659–684 (2003).
14. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42,** 342–346 (2009).
15. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80,** 2946–2953 (2001).
16. Dave, V. P. Hierarchical role of CD3 chains in thymocyte development. *Immunol. Rev.* **232,** 22–33 (2009).
17. Allison, T. J., Winter, C. C., Fournie, J. J., Bonneville, M. & Garboczi, D. N. Structure of a human γδ T-cell antigen receptor. *Nature* **411,** 820–824 (2001).
18. Melchers, F. The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. *Nature Rev. Immunol.* **5,** 578–584 (2005).
19. Semenyuk, A. V. & Svergun, D. I. GNOM - a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24,** 537–540 (1991).
20. Brookes, E., Demeler, B., Rosano, C. & Rocco, M. The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur. Biophys. J.* **39,** 423–435 (2010).

Author Contributions S.S.P. solved the structure, undertook analysis, performed experiments and contributed to manuscript preparation. J.M., D.I.G. and J.R. contributed to the design and interpretation of experiments, project management and the writing of the manuscript. R.B., Z.C., L.K.-N., M.A.P., G.F.K., N.L.L.G., T.T. and A.W.P. performed experiments and contributed to the writing of the manuscript; C.W., S.H.C., N.K.W., T.B., N.P.C. and M.C.J.W. performed experiments. J.M. and J.R. were the joint senior authors—they co-led the investigation, devised the project, analysed the data and wrote the manuscript.

METHODS

Pre-TCR cloning, expression and purification. Genes encoding the ectodomains (including the interchain cysteine) of the human pre-Tα and TCR LC13β were cloned into the pFastBac Dual vector (Invitrogen). The genes were cloned downstream of the baculovirus GP67 secretory signal sequence and were under the control of different promoters for co-expression. The C terminus of pre-Tα was modified to include an enterokinase cleavage site, a Jun zipper, a BirA tag and a hexahistidine (6xHis) tag, and the LC13 β-chain has an enterokinase site and a Fos zipper. The pre-TCRs were expressed and secreted by High Five insect cells (Invitrogen) as a covalently linked dimer. At the end of expression, the insect cell supernatant was collected and diafiltrated with TBS by tangential flow filtration. The hexahistidine-tagged proteins were purified using Ni Sepharose High Performance resin (GE Healthcare) and by gel filtration (HiLoad 16/60 Superdex 200, GE Healthcare). For protein crystallization, the C-terminal tag regions were removed by enterokinase digestion (GenScript). Final purification and buffer exchange into 20 mM Tris-HCl, (pH 8.0) and 0.15 M NaCl was carried out by gel filtration. The pooled protein was concentrated to ∼10 mg ml^{-1}, snapped frozen and stored at −80 °C in small aliquots. The typical protein yield of the cleaved pre-TCR is about 1–3 mg per litre of expression culture.

Crystallization and X-ray data collection. Nanoscale crystallization trials of pre-TCR were carried out using sitting-drop vapour diffusion at room temperature (19 °C) and at 4 °C. A few hits were identified and reproduced manually using hanging-drop vapour diffusion. The optimized crystallization conditions were produced by mixing equal volumes of pre-TCR (6 mg ml^{-1}) with the reservoir buffer (0.1 M sodium cacodylate (pH 7.0), 0.1 M calcium acetate and 13–16% PEG 1500). Rod-shaped pre-TCR crystals appeared quickly after overnight incubation at 4 °C and continued to grow in size for the next few days. For cryoprotection, the rod-shaped crystals were soaked in the reservoir buffer plus 15% glycerol overnight at 4 °C. The soaked crystals were then transferred into reservoir buffer containing 25% glycerol and flash-frozen in liquid nitrogen. X-ray data to 2.8 Å were collected using the Micro Crystallography (MX2) beamline at the Australian Synchrotron. The diffraction data were processed and analysed using MOSFILM 6.2.6 and SCALA (CCP4 suite)[21]. A summary of data collection statistics is provided in Supplementary Table 1.

Structure determination and refinement. Initial analysis of the pre-TCR crystals indicated that the X-ray data could be scaled into any of the primitive hexagonal space groups, suggesting crystal twinning. Further examination using The Merohedral Crystal Twinning Server (http://nihserver.mbi.ucla.edu/Twinning/), CCP4 programs and PHENIX.XTRIAGE[22] confirmed perfect crystal merohedral twinning. The structure was solved by molecular replacement (PHASER from the CCP4 suite) using the β-chain of LC13 TCR (PDB ID, 1KGC) as the search model. Solutions were found in a number of space groups (six molecules for P3$_2$, three molecules for P3$_2$21 or P3$_2$12, and one molecule for P6$_2$22). Each of the solutions was examined and refined using PHENIX[23], taking into consideration the twinning factor and operator (0.516, −k, h + k, l; k, h and l are Miller indices) After initial rounds of refinement, the P3$_2$21 solution gave a reasonable R$_{factor}$ and an unbiased density of the missing pre-Tα chains was observed, suggesting the correct solution. To avoid bias due to the presence of data twinning, the R$_{free}$ set was assigned in PHENIX.REFINE. Further restrained refinements were performed using PHENIX.REFINE after initial rounds of rigid-body refinement and stimulated annealing. Each round of refinement was interspersed with electron density inspection and manual model building with COOT[24]. The progress of refinement was monitored by the R$_{free}$ value and the final refinement statistics for pre-TCR are given in Supplementary Table 1. There were three pre-TCR complexes in the asymmetric unit, all of which were highly similar to each other, and structural analyses were thus confined to one pre-TCR complex. The final model comprises residues 7–110 from the pre-Tα chain and residues 2–245 from the TCR β-chain. Disordered regions include six residues from the N terminus and nine from the C terminus of pre-Tα, and one N- and three C-terminal residues from TCRβ, respectively. The missing regions include the intermolecular disulphide bond at the flexible C-terminal end of the polypeptides.

SAXS. SAXS data were collected at the Australian Synchrotron using a 1M Pilatus detector. Buffers/samples were loaded into 1-mm quartz capillaries and continuously flowed through the beam during data collection. Multiple 1-s exposures were collected and compared to control for radiation damage. Data were collected using a single camera length with a 1.5-m sample-to-detector distance to cover a momentum transfer interval of 0.0056 Å^{-1} < q < 0.3 Å^{-1}. The modulus of the momentum transfer is defined as $q = 4\pi\sin(\theta/\lambda)$, where 2θ is the scattering angle and λ is the wavelength. Scattering images were radially averaged and blank subtracted using SAXS15ID software (Australian Synchrotron). Additional SAXS data were collected at the Advanced Light Source (ALS), Berkeley, using a MAR165CCD detector that was 1.5 m from the sample, resulting in a q range of 0.01–0.3 Å^{-1}. Circular integration and normalization was performed using in-house software. Molecular mass

estimates were obtained by normalizing scattering to BSA (ALS) or lysozyme (Australian Synchrotron). Data quality was assessed on the basis of the linearity of Guinier plots, and R$_g$ and the pairwise intraparticle distance distribution function were determined using GNOM[19]. Ab initio models were generated using DAMMIF[14] and GASBOR[15] using P1 symmetry in all cases except for the 15 mg ml^{-1} concentration that corresponds to pre-TCR dimer (P2 symmetry). Ten independent DAMMIF/GASBOR runs were aligned, combined and filtered to generate a final model that retained the most consistent features using the DAMMAVER package. The normalized spatial discrepancies between individual 15 mg ml^{-1} models were 1.17–1.22 (GASBOR) and 0.61–0.66 (DAMMIF). High-resolution models of pre-TCR were fitted within ab initio models using PyMOL. Bead models (Fig. 4) generated from the pre-TCR structure for the monomer and head-to-tail pre-TCR dimer include C termini that were manually generated in PyMOL.

AUC. AUC experiments were conducted using a Beckman model XL-I analytical ultracentrifuge at a temperature of 20 °C. Samples (380 μl) of wild-type pre-Tα–LC13β solubilized in 20 mM Tris, 100 mM NaCl and 50 mM Ca acetate (pH 7.1), W46R-mutant pre-Tα–LC13β solubilized in 20 mM Tris, 150 mM NaCl and 1 mM EDTA (pH 7.0), and reference solutions (400 μl) were loaded into a conventional double-sector quartz cell and mounted in a Beckman four-hole An-60 Ti rotor. Intensity data were collected in continuous mode at a single wavelength between 293 and 306 nm using a rotor speed of 40,000 r.p.m. and a step size of 0.003 cm without averaging. Frictional ratios (f/f$_0$) were calculated with SEDNTERP[25] using the measured sedimentation coefficient of the given species, the molecular mass calculated from the amino-acid sequence (41,800-Da pre-TCR and 83,600-Da pre-TCR dimer), a solvent density of 1.005 g ml^{-1}, a viscosity of 1.021 cP and a partial specific volume of 0.722 ml g^{-1}. Sedimentation velocity intensity data at multiple time points were converted to absorbance data and then fitted to a continuous size distribution model using the program SEDFIT[26] with a resolution of 200 species ranging from 1.0 S (or 10 kDa) to 7.2 S (or 110 kDa) and a P value of 0.95.

Mass spectrometry and crosslinking studies. A 2 mg ml^{-1} solution of recombinant pre-TCR was buffer exchanged into PBS and treated with bis(sulfosuccinimidyl) suberate (BS3) crosslinker (5 mM final concentration) for 30 min on ice, and the reaction was quenched with the addition of 200 mM Tris (pH 8.0). The crosslinked material was resolved on a one-dimensional SDS–PAGE gel and the band corresponding to the dimer excised and an in-gel tryptic digest performed. Crosslinked peptides were examined by LC-MS/MS using an AB SCIEX Q-STAR ELITE mass spectrometer and X-QUEST software[27].

Cloning and expression of pre-Tα, LC13α and LC13β. pGEM–LC13α and pMIG–LC13β constructs were generated as previously described[12]. Pre-TCR cDNA was reverse-transcribed from isolated MOLT RNA using Superscript II (Invitrogen). PCR products encoding the pre-Tα gene were cloned into the pGEM–T Easy vector (Promega) and verified by sequencing before being transferred into the retroviral expression vector pMIG. Primers used were as follows: pre-Tα-F, 5′-CGGAATTCACGCGTATGGCCGGTACATGGCTGCT-3′; pre-Tα-R, 5′-CGCTCGAGCACAAAGTGTCAGGCAGCTCCAGCCTGCAGAGG-3′. Mutants were created with pGEM–pre-Tα or pGEM–LC13β plasmid DNA as template by site-directed mutagenesis (QuikChange; Stratagene). Primers used are as follows with underlined letters indicating codons mutated: pre-Tα-W46R-F, 5′-ctt gacagccccatcCGGttctcagccggcaat-3′; pre-Tα-W46R-R, 5′-attgccggctgagaaCCGgat ggggctgtcaag-3′; LC13β-Y35A-F, 5′-CATGTATCCCTTTTTTGGGGCCCAACAGG CCCTGGGG-3′; LC13β-Y35A-R, 5′-CCCCAGGGCCTGTTGGGCCCAAAAAA GGGATACATG-3′; LC13β-F108A-F, 5′-GCCTACGAGCAGTACGCCGGGCCG GGCACCAGG-3′; LC13β-F108A-R, 5′-CCTGGTGCCCGGCCCGGCGTACTG CTCGTAGGC-3′.

Functional assays. SKW3 is a human T-cell leukaemia line that is CD4$^+$CD8$^+$ double positive and TCRα$^-$β$^-$, TCRγ$^-$δ$^-$ deficient[28,29]. Single TCR gene-transfectants SKW3.pre-Tα, SKW3.LC13α and SKW3.LC13β were generated by retroviral transduction of the TCR α- and β-chains into SKW3 cells as described previously[30]. Transfected cells (GFP positive) were enriched by cell sorting with a BD FACSAria cell sorter. Double TCR gene-transfectants were then created by a second round of transduction of SKW3.LC13β with individual complementary pMIG constructs. Each transfectant line was cloned or enriched using a BD FACSAria cell sorter with staining of anti-CD3ε. The double TCR gene-transfectants include: SKW3.LC13αβ (LC13), SKW3.pre-Tα–LC13β (WT), SKW3.pre-Tα.W46R–LC13β (W46R), SKW3.pre-Tα–LC13β.Y35A (βY35A), SKW3.pre-Tα–LC13β.F108A (βF108A).

Control: the conventional LC13 TCR is specific for the FLR epitope (FLRGRAYGL) of Epstein–Barr virus in the context of HLA-B8[8]. These experiments were performed several times with similar results.

Immunofluorescence. Human SKW3 T cells stably expressing WT TCR (TCRα–TCRβ), pre-TCR (pre-Tα–TCRβ) or mutant pre-TCR (pre-Tα–TCRβ-Y35A, pre-Tα-W46R–TCRβ, pre-Tα–TCRβ-F108A) were incubated in RPMI medium

containing 1% (v/v) heat-inactivated FBS and left unstimulated or incubated with mouse anti-human CD3ε (5 μg ml^{-1} OKT3; eBioscience) for 15 min on ice, washed and stimulated by crosslinking with goat anti-mouse immunoglobulin G (20 μg ml^{-1}) at 37 °C for 30 min. Cells were fixed with 0.5% (w/v) paraformaldehyde for 30 min, washed with PBS containing 3% (v/v) FBS and resuspended in PBS containing 5% (v/v) FBS plus 0.1% (w/v) saponin plus primary antibody (OKT3 or anti-TCRβ), and incubated at room temperature for 1 h. Cells were then washed and incubated with the same buffer plus secondary antibody (Alexa Fluor 568 goat anti-mouse; Invitrogen-Molecular Probes) for 30 min. Washed cells were then stained for DNA with Hoechst 33258 (Invitrogen-Molecular Probes) for 10 min, collected onto glass slides by centrifugation (500 g, 5 min) and processed for microscopic analysis using a Zeiss Axioskop 2 MOT Plus epifluorescence microscope (Zeiss). These experiments were performed twice with similar results.

21. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
22. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
23. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426,** 419–435 (2008).
24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
25. Cole, J. L. *et al.* Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol.* **84,** 143–179 (2008).
26. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
27. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nature Methods* **5,** 315–318 (2008).
28. Burger, R., Hansen-Hagge, T. E., Drexler, H. D. & Gramatzki, M. Heterogeneity of T-acute lymphoblastic leukemia (T-ALL) cell lines: suggestion for classification by immunophenotype and T-cell receptor studies. *Leuk. Res.* **23,** 19–27 (1999).
29. Hirano, T. *et al.* In vitro immune response of human peripheral lymphocytes: IV. Specific induction of human suppressor T cells by an antiserum to the T leukemia cell line HSB. *J. Immunol.* **123,** 1133–1140 (1979).
30. Holst, J., Vignali, K. M., Burton, A. R. & Vignali, D. A. A. Rapid analysis of T-cell selection in vivo using T cell-receptor retrogenic mice. *Nature Methods* **3,** 191–197 (2006).

LETTER

ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours

Ping Chi[1,2]*, Yu Chen[1,3]*, Lei Zhang[4], Xingyi Guo[5], John Wongvipat[3], Tambudzai Shamu[3], Jonathan A. Fletcher[6], Scott Dewell[7], Robert G. Maki[1], Deyou Zheng[5,8], Cristina R. Antonescu[4], C. David Allis[2]* & Charles L. Sawyers[3,9]*

Gastrointestinal stromal tumour (GIST) is the most common human sarcoma and is primarily defined by activating mutations in the *KIT* or *PDGFRA* receptor tyrosine kinases[1,2]. *KIT* is highly expressed in interstitial cells of Cajal (ICCs)—the presumed cell of origin for GIST—as well as in haematopoietic stem cells, melanocytes, mast cells and germ cells[2,3]. Yet, families harbouring germline activating *KIT* mutations and mice with knock-in *Kit* mutations almost exclusively develop ICC hyperplasia and GIST[4–7], suggesting that the cellular context is important for *KIT* to mediate oncogenesis. Here we show that the *ETS* family member *ETV1* is highly expressed in the subtypes of ICCs sensitive to oncogenic *KIT* mediated transformation[8], and is required for their development. In addition, *ETV1* is universally highly expressed in GISTs and is required for growth of imatinib-sensitive and resistant GIST cell lines. Transcriptome profiling and global analyses of ETV1-binding sites suggest that ETV1 is a master regulator of an ICC-GIST-specific transcription network mainly through enhancer binding. The ETV1 transcriptional program is further regulated by activated KIT, which prolongs ETV1 protein stability and cooperates with ETV1 to promote tumorigenesis. We propose that GIST arises from ICCs with high levels of endogenous ETV1 expression that, when coupled with an activating KIT mutation, drives an oncogenic ETS transcriptional program. This differs from other *ETS*-dependent tumours such as prostate cancer, melanoma and Ewing sarcoma where genomic translocation or amplification drives aberrant *ETS* expression[9–11]. It also represents a novel mechanism of oncogenic transcription factor activation.

Reasoning that transcription factors are likely to play critical roles in defining the cellular context, we used three expression data sets[12,13] to search for GIST-specific genes that might provide new molecular insights. We identified an 11-gene signature exclusively associated with GIST in all three data sets that included the *ETS* family transcription factor *ETV1* (Fig. 1a and Supplementary Table 1). Examination of individual tumour samples revealed that *ETV1* is highly expressed in all GISTs and at significantly higher levels than any other tumour type (Fig. 1b and Supplementary Fig. 1). *ETV1* was of immediate interest because *ETS* family transcription factors are well-established oncogenes in Ewing sarcoma, melanoma and prostate cancer[9–11].

Next, we assessed messenger RNA (mRNA) and protein levels of ETV1 in GIST and other sarcomas in clinical samples, GIST cell lines (imatinib-resistant GIST48 and imatinib-sensitive GIST882), the U2OS osteosarcoma cell line and the LNCaP prostate cancer cell line known to overexpress ETV1 owing to translocation[14] (Fig. 1c, d). *ETV1* mRNA and protein were highly and exclusively expressed in all GIST tumour samples and GIST cell lines, and in LNCaP cells. As expected, *KIT* mRNA and protein were highly expressed in all GIST tumours and GIST cell lines, but not in other sarcomas or non-GIST

cell lines, and phospho-KIT was only seen in GIST samples with activating *KIT* mutations. Four additional GIST samples amenable to immunohistochemical analysis all showed strong nuclear ETV1 staining whereas a leiomyosarcoma control sample did not (Supplementary Fig. 2). These data show that *ETV1* is universally highly expressed in all GISTs both at transcript and protein levels.

To explore the requirement of ETV1 in GIST pathogenesis, we performed RNAi experiments using two *ETV1*-specific hairpins validated for both protein and mRNA suppression (Supplementary Fig. 3a). Infection with either hairpin resulted in growth inhibition of both GIST cell lines, but did not affect the growth of U2OS cells. Consistent with the level of *ETV1* knockdown, ETV1sh2 was more growth suppressive than ETV1sh1 in both GIST cell lines (Fig. 1e). Cell cycle analysis showed that *ETV1* knockdown resulted in both decreased cell cycle progression and increased apoptosis (Supplementary Fig. 3b). *ETV1* knockdown also impaired the tumorigenicity of GIST882 cells in SCID mouse xenografts, and those tumours that did grow had escaped *ETV1* suppression (Fig. 1f). Collectively, these observations indicate that ETV1 is required for GIST growth and survival.

Next, we addressed the mode of high *ETV1* expression in GIST. Fluorescence *in situ* hybridization on four GIST samples and two GIST cell lines showed no evidence of amplification or 'breakaway' between the 3′ and 5′ ends of *ETV1* locus. Quantitative reverse transcription PCR showed no evidence of differential exon expression, which is expected with *ETV1* translocation (Supplementary Fig. 4). Furthermore, no focal *ETV1* amplification was found in 40 GIST tumours and six GIST cell lines in a recent high-density single nucleotide polymorphism array study[15]. The fact that high levels of *ETV1* expression are consistently observed in the absence of obvious genomic alterations raises the possibility that the ICCs that give rise to GIST may endogenously express *ETV1*.

The musculature of the gastrointestinal tract is organized into two principal layers—the inner circular muscle layer beneath the mucosa and the outer longitudinal muscle layer[16]. In the large intestine, myenteric ICCs (ICC-MY) form a network between the circular muscle and longitudinal muscle layers surrounding the neuronal myenteric plexus, intramuscular ICCs (ICC-IM) are singly dispersed in the circular muscle, and submucosal ICCs (ICC-SMP) form network surrounding the submucosal plexus (Fig. 2a). In the small intestine, ICC-IMs and ICC-SMPs are absent and ICC-DMPs form a network around the deep muscular plexus in the circular muscle close to the mucosa (Supplementary Fig. 5a). All four ICC subtypes are identified by positive membrane expression of Kit[16] (Fig. 2b and Supplementary Fig. 5b). In the large intestine, ICC-MYs and ICC-IMs but not ICC-SMPs stain with nuclear Etv1 (Fig. 2b). In the small intestine, ICC-MYs but not ICC-DMPs stain with nuclear Etv1 (Supplementary Fig. 5b). This finding is further supported by our analysis of a published ICC expression data set from

[1]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, 10065 USA. [2]Laboratory of Chromatin Biology & Epigenetics, The Rockefeller University, New York, New York, 10065 USA. [3]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York, 10065 USA. [4]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York, 10065 USA. [5]Department of Neurology, Albert Einstein College of Medicine, Bronx, New York, 10461 USA. [6]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, 02115 USA. [7]Genomics Resource Center, The Rockefeller University, New York, New York, 10065 USA. [8]Departments of Genetics and Neuroscience, Albert Einstein College of Medicine, Bronx, New York, 10461 USA. [9]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.
*These authors contributed equally to this work.
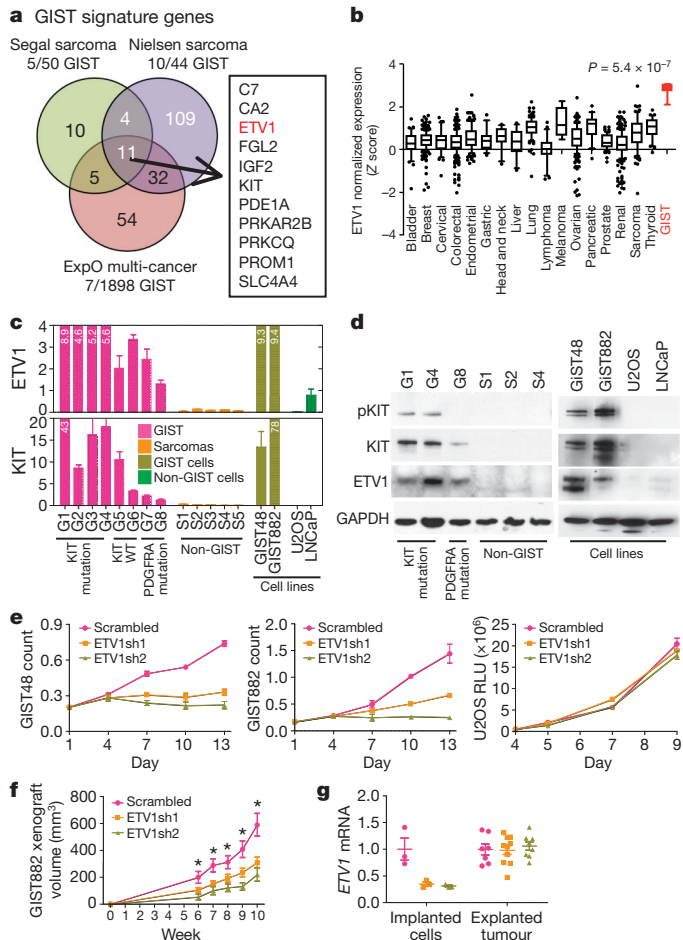
a GIST signature genes

Figure 1 | ETV1 is universally highly expressed and required for tumour growth and survival in GIST. **a**, Venn diagram of GIST-signature genes from three data sets. **b**, Expression of *ETV1* in multiple tumour types from the ExpO data set. Box, 25th–75th percentile; error bar, 10th–90th percentile; dots, outliers. **c**, *ETV1* and *KIT* mRNA levels by qRT–PCR of GIST and non-GIST samples, whose details are described in Methods. Mean ± SD, *n* = 3. **d**, Immunoblotting of selected tumour tissues and cell lines from **c**. **e**, Growth curves of GIST and U2OS cells after shRNA-mediated *ETV1* suppression compared with control. Mean ± s.e.m., *n* = 3. **f**, Tumour volume over time in SCID mice implanted with GIST882 cells after shRNA-mediated *ETV1* suppression compared with scrambled shRNA controls. Mean ± s.e.m., *P* < 0.05; *n* = 7, 10, 8 for scrambled, ETV1sh1 and ETV1sh2, respectively. **g**, *ETV1* mRNA levels of preimplanted GIST882 cells and explanted xenografts at week 10. Mean ± s.d.

mouse small intestine[17] showing that *Etv1* is only highly expressed in ICC-MYs (Supplementary Fig. 5c). Notably, in the *Kit*[A558] mutant mice only ICC-MY and ICC-IM develop hyperplasia whereas ICC-SMP and ICC-DMP do not[8]. These data suggest that ETV1 is a lineage-specific transcription factor for the ICCs that give rise to GIST.

We therefore asked if Etv1 is required for the normal development of ICCs by examining the gastrointestinal tracts of *Etv1*[−/−] mice[18]. Cross-section and reconstructed whole-mount images from *Etv1*[−/−] mice showed significant loss of Kit-positive ICC-IMs and ICC-MYs in the large intestine (Fig. 2c, d, Supplementary Fig. 9 and Supplementary Movies 1 and 2), small intestine, stomach and caecum (Supplementary Figs 6–9 and Supplementary Movies 3–8). In contrast, ICC-DMPs and ICC-SMPs in the small and large intestine, respectively, were preserved, consistent with absent Etv1 expression in these ICC subtypes. These results were confirmed with a second ICC marker Ano1 (ref. 19) (Supplementary Fig. 10). Immunostaining with the neuronal marker Pgp9.5 confirmed the integrity of the myenteric plexus in *Etv1*[−/−] mice (Fig. 2c and Supplementary Figs 6–8 and 11). Collectively, these
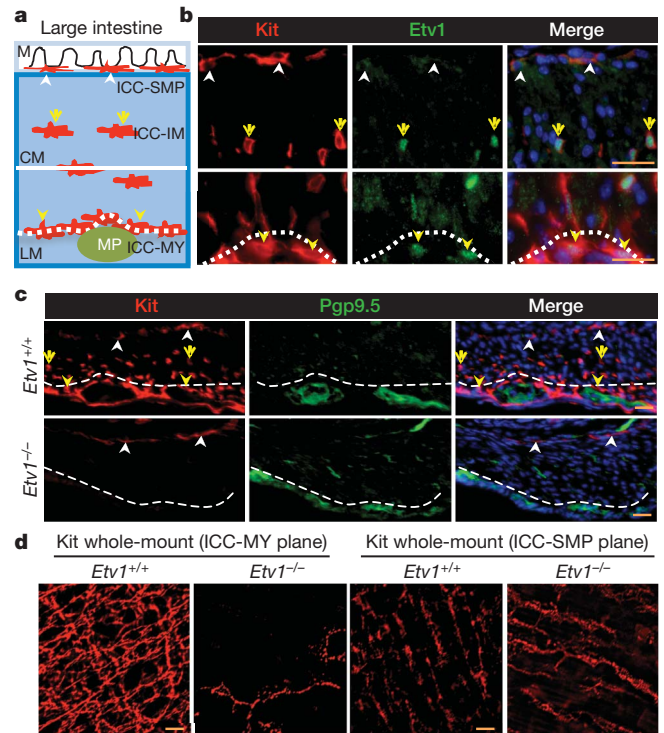
Figure 2 | Etv1 is expressed in the subtypes of ICCs susceptible to oncogenesis and is required for their development. a, Schematic showing localization of ICC-MY (yellow arrowheads), ICC-IM (yellow arrows) and ICC-SMP (white arrowheads) in the large intestine. M, mucosa; CM, circular muscle; LM, longitudinal muscle; MP, myenteric plexus. All three ICC subtypes express Kit (red). **b**, Co-immunofluorescence (divided into two microscopy fields) of Kit (red), Etv1 (green) and 4′,6-diamidino-2-phenylindole (DAPI) (blue) of the large intestine of wild-type mice. **c**, Co-immunofluorescence of Kit (red), Pgp9.5 (green) and DAPI (blue) of the large intestine of *Etv1*[+/+] and *Etv1*[−/−] mice. **d**, Representative deconvoluted whole-mount Kit-immunofluorescence images of the large intestine of *Etv1*[+/+] and *Etv1*[−/−] mice. A single microscopy field focused to the ICC-MY and ICC-SMP planes is shown. The entire Z-stacks are shown in Supplementary Movies 1 and 2. Scale bar, 20 μm.

data indicate that Etv1 is selectively required for development of ICC-MY and ICC-IM and, by implication, a lineage-specific survival factor for the ICC-GIST lineage.

To identify ETV1 target genes in GIST, we analysed the effect of short hairpin RNA (shRNA)-mediated *ETV1* suppression on the transcriptomes of GIST48 and GIST882 cells. The overlap of genes perturbed by both *ETV1*-specific hairpins and across both cell lines was highly statistically significant, suggesting that ETV1 regulates a core set of genes in GIST (Supplementary Fig. 12). To minimize cell-line-specific and off-target effects, we generated a ranked gene list based on the average change in gene expression induced by the two *ETV1*-specific hairpins in both GIST cell lines (Fig. 3a, b). We independently confirmed downregulation of five of these genes using real-time PCR (Supplementary Fig. 13). Among the 48 genes suppressed more than 1.7-fold by *ETV1* knockdown, 32 were expressed at higher levels in human GIST samples relative to other tumour types in the ExpO expression data set (Fig. 3b). We performed gene set enrichment analysis[20] of the 'shETV1' ranked list using more than 3,000 gene sets in the Molecular Signature Database and five custom gene sets defined by GIST-signature genes from the Segal, Nielsen and ExpO data sets and by ICC-MY- and ICC-DMP-signature genes (Supplementary Table 1). All three GIST sets and the ICC-MY set were among the most negatively enriched gene sets whereas the ICC-DMP set was not (Fig. 3c, Supplementary Fig. 14 and Supplementary Table 2). These data suggest that *ETV1* is a master regulator of a transcriptional program conserved in ICC-IM/MYs and GISTs.
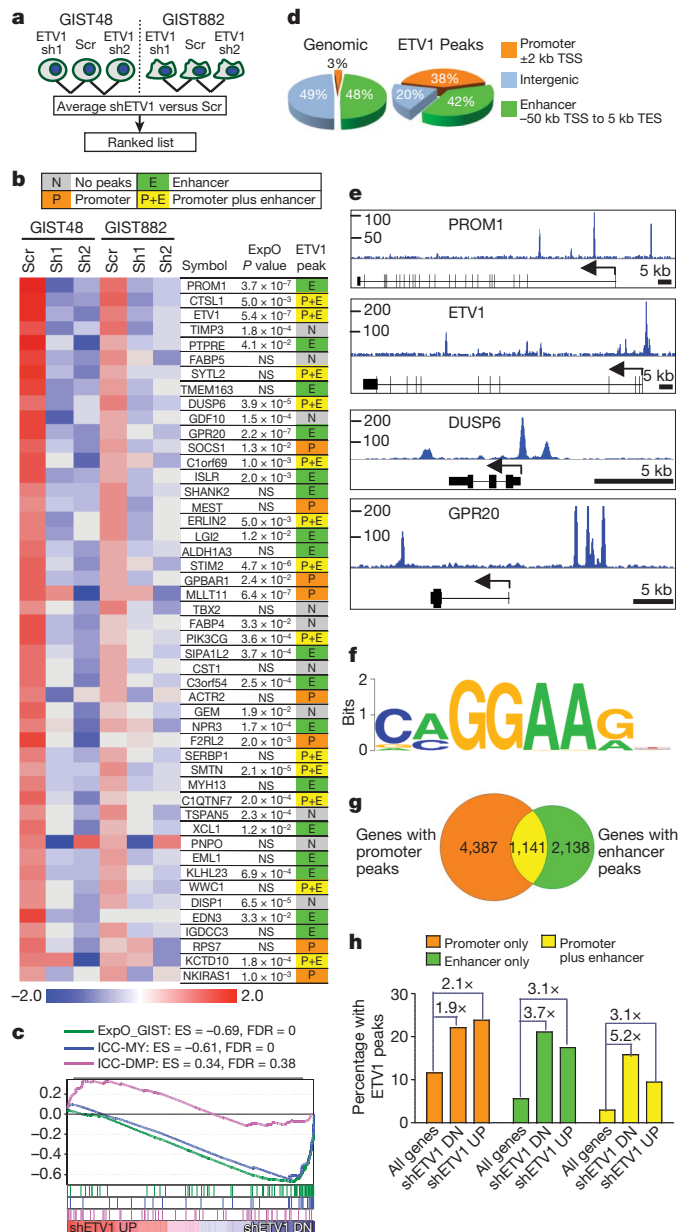
Figure 3 | ETV1 regulates GIST-signature genes predominantly through enhancer binding. a, Ranked list of ETV1 regulated genes was generated based on the average fold-change by the two *ETV1* hairpins in two cell lines. **b**, Heatmap of expression of the 48 genes with average downregulation greater than 1.7-fold. For each gene, table shows *P* value of GIST versus other tumour types from the ExpO data set, calculated by Oncomine (NS: *P* > 0.05); and the presence of ETV1-binding sites from ChIP-Seq analysis. **c**, GSEA plots of the shETV1 ranked list using three gene sets: GIST signature genes from ExpO data set, ICC-MY and ICC-DMP signature genes in mouse small intestine. ES, enrichment score; FDR, false discovery rate. **d**, Pie charts of genomic structure and distribution of ETV1 ChIP-Seq peaks. TSS, transcription start site; TES, transcription end site. **e**, Representative ChIP-Seq reads in top ETV1 target genes. **f**, The consensus sequence motif identified in the ETV1-binding sites by the MEME program. **g**, Pie chart of genes with ETV1-binding sites divided into promoter only, enhancer only and both. **h**, Plot of percentage of all genes, genes averagely downregulated 1.4-fold by shETV1 ($n = 410$) and genes averagely upregulated 1.4-fold by shETV1 ($n = 380$) with promoter only, enhancer only and both promoter and enhancer ETV1 binding. Fold enrichment over all genes is shown above the plots.

To define the direct transcriptional targets of ETV1 in GIST, we performed genome-wide analyses of ETV1-binding sites using chromatin immunoprecipitation and deep sequencing (ChIP-Seq) in GIST48 cells. We identified 14,741 ETV1-binding sites (ETV1 peaks), which are enriched in promoter regions (Fig. 3d). Motif analysis of the peaks identified the ETS core consensus motif, GGAA, in approximately 90% of peaks (Fig. 3f). Integrative analyses of the ETV1 ChIP-Seq data with the transcriptomes from shRNA-mediated *ETV1* suppression in GIST cells showed that 38 of the top 48 shETV1 downregulated genes contain ETV1 peaks (Fig. 3b, e and Supplementary Fig. 15). Analysis of genes with 1.4-fold change by shETV1 knockdown revealed that both shETV1 upregulated and shETV1 downregulated genes are enriched for ETV1 peaks. Furthermore, enhancer binding, in particular enhancer and promoter binding, is highly predicative of transcriptional activation by ETV1 (Fig. 3h). Because enhancers are in general cell-lineage specific[21,22], our data suggest that these ICC–GIST-lineage-specific genes are likely directly regulated by ETV1 binding to their enhancer regulatory elements.

The dual requirement of *KIT* and *ETV1* in normal ICC development and GIST survival raises the possibility that KIT and ETV1 cooperate in GIST oncogenesis. Inhibition of KIT signalling by imatinib in imatinib-sensitive GIST882 cells resulted in rapid loss of ETV1 protein, without significant effect on *ETV1* mRNA levels (Fig. 4a, b and Supplementary Fig. 16). Similar results were observed with the MEK inhibitor PD325901. This loss of ETV1 protein was faster than the natural degradation rate, as revealed by cyclohexamide experiments to inhibit protein synthesis, and was rescued from proteosomal degradation by MG132 (Fig. 4b). Therefore, KIT–MEK signalling stabilizes ETV1 protein. Consistent with this KIT–MEK–ETV1 signalling pathway model, the overlap between genes transcriptionally altered by imatinib treatment (KIT-regulated) and by ETV1 knockdown in GIST882 cells is highly significant (Fig. 4c). Furthermore, these ETV1 transcriptional targets preferentially contain ETV1 enhancer peaks (Fig. 4d), which indicates that KIT signalling influences the ETV1 transcriptional output of the tissue and lineage-specific genes in GIST.

Having established a signalling pathway from KIT to ETV1, we explored their potential cooperativity in tumorigenesis by expressing *ETV1*, wild-type *KIT*, *KIT* harbouring a common GIST mutation (*KITΔ560*) and control vectors in combination in NIH3T3 cells. KIT-dependent stabilization of ETV1 protein was recapitulated in this system (Fig. 4e). In anchorage-independent colony formation assays, ETV1 significantly increased the number and size of colonies in KITΔ560 expressing cells but was insufficient to confer anchorage-independent growth on its own (Supplementary Fig. 17). Furthermore, KITΔ560 and ETV1 strongly cooperated in conferring tumorigenic growth in SCID mice (Fig. 4f, g).

Taken together, these findings establish an oncogenic role for ETV1 in GIST that differs from classical models of *ETS*-driven malignancies where structural alterations (for example, *TMPRSS2-ETV1* translocation in prostate cancer, *ETV1* amplification in melanoma) lead to aberrant expression and promote tumorigenesis[9,11]. Rather, ETV1 expression in GIST is inherited from ICC-MY/IM cells, where ETV1 is also a survival factor. We further established that KIT activity, through MEK, stabilizes ETV1, providing a mechanism for KIT–ETV1 cooperativity (Fig. 4h). These observations provide an explanation for why patients and mice with germline-activating *KIT* mutations develop neoplasia in only the ICC-MY/IM lineage. Although the mechanism of *ETV1*-mediated oncogenesis in GIST differs from other *ETS*-driven cancers, we anticipate that the *ETV1*-dependent transcriptional program defined here may serve as a valuable resource for further understanding of other *ETV1*- and other *ETS*-driven transcriptional programs in various cellular contexts such as prostate cancer.

The fact that ETV1 is universally highly expressed in all GISTs makes it immediately useful as a candidate diagnostic biomarker, because the current standard of KIT immunoreactivity is negative in about 5% of all GISTs[23]. Although transcription factors have classically been considered 'undruggable', reports of successful inhibition of the NOTCH transcription factor complex and androgen receptor activity by blocking coactivator binding have challenged this model[24,25]. Owing to established requirements of ETV1 in subsets of prostate cancer and
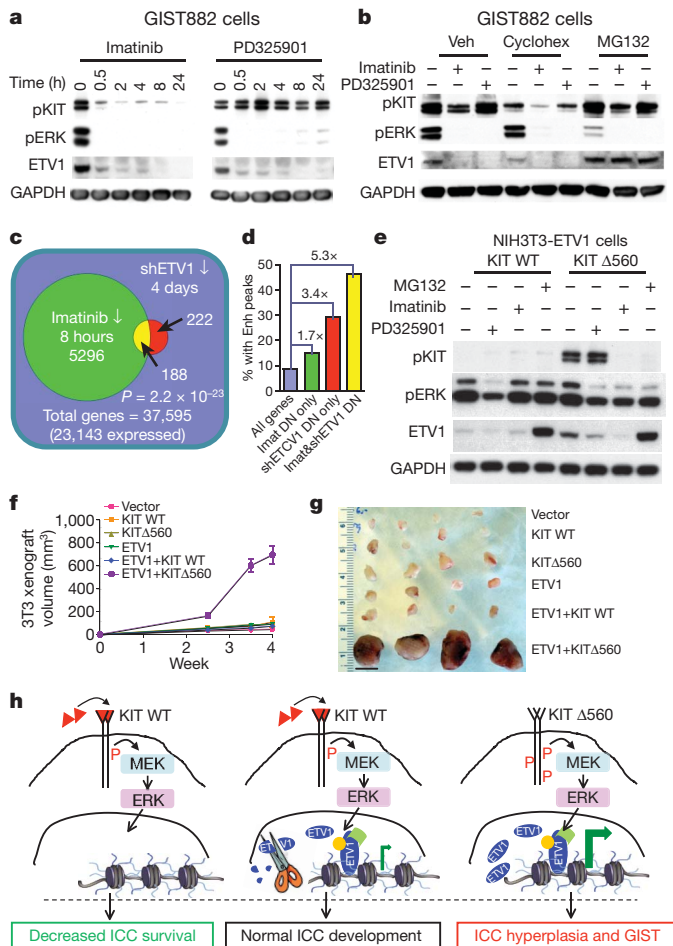
Figure 4 | KIT signalling synergizes with ETV1 in GIST tumorigenesis by stabilization of ETV1 protein. a, Immunoblots of GIST882 cells treated with the imatinib (1 μM) and PD325901 (100 nM) for the indicated time points. **b,** Immunoblots of GIST882 cells treated for 2 h with imatinib or PD325901 in combination with cyclohexamide (10 μg ml^{-1}) or MG132 (10 μM). **c,** Venn diagram of genes downregulated by 1.4-fold by shETV1 and by imatinib in GIST882 cells. P value: Fisher's exact test based on number of expressed genes. **d,** Percentage of all genes, imatinib-downregulated genes, shETV1-downregulated genes and overlapping genes with ETV1 enhancer peaks. **e,** Immunoblot of NIH3T3 cells expressing *ETV1* and either *KIT* WT or *KITΔ560* 2 h after treatment with PD325901, imatinib or MG132. **f,** Growth of xenografts of engineered NIH3T3 cells stably expressing the indicated genes (*n* = 12, mean ± s.e.m.). **g,** Photograph of four representative explanted xenografts at 4 weeks after implanting. Scale bar, 1 cm. **h,** Model of the role of ETV1 in ICC maintenance and GIST oncogenesis. Normal level of KIT activation by KIT ligand (red triangle) stabilizes ETV1 transcription factor through the MAPK pathway, and results in physiological ETV1 transcriptional output critical for ICC development (middle). In the absence of ETV1, there is decreased ICC development, which phenocopies genetic loss of KIT signalling (left). Activating mutation of *KIT* (for example, *KITΔ560*) leads to constitutive activation of the KIT-MAPK signalling pathway, increased ETV1 stability and augmented ETV1 transcriptional output that promotes tumorigenesis (right).

melanoma, efforts to find ETV1 inhibitors are underway and may yield novel therapeutic agents for imatinib-resistant GIST.

METHODS SUMMARY

All mined data sets were downloaded from Gene Expression Omnibus (GSE2109, GSE7809, GSE2719, GSE3443, GSE8167, GSE17743) and were analysed by Oncomine or by using Genespring 10. GIST-signature genes from three data sets containing both GIST and non-GIST malignancies met the following two criteria: (1) $q < 0.05$ and (2) a Z-score expression difference more than 1.5 between GIST and non-GIST tumours. Expression profiling of GIST cell lines with different shRNA conditions was performed in duplicate on Illumina HumanHT-12 array.

Gene set enrichment analysis was performed using MSigDB C2, MSigDB C4 and the GIST and ICC signature gene sets. For ChIP-Seq, sheared chromatin enriched by ETV1 IP was sequenced on Solexa Genome Analyser, aligned using ELAND software. Peaks were identified by Model-based Analysis of ChIP-Seq (MACS) with input DNA as control using a false discovery rate of less than 1%.

Materials. GIST48 and GIST882 cells were established in the Fletcher laboratory[26]. All other cells were obtained from the American Type Culture Collection. *Etv1*$^{-/-}$ mice, with targeted deletion of the ETS domain, were obtained from the Jessell laboratory (Columbia) and CB17-SCID mice were from Taconic. Antibody sources were ETV1, ANO1, PGP9.5 (Abcam), KIT for western blotting, P-Tyr703-KIT (Cell Signaling), P-Tyr204-ERK, GAPDH (Santa Cruz) and anti-mouse Kit for immunofluorescence (clone ACK2, E-Biosciences).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Heinrich, M. C. *et al.* PDGFRA activating mutations in gastrointestinal stromal tumors. *Science* **299**, 708–710 (2003).
2. Hirota, S. *et al.* Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* **279**, 577–580 (1998).
3. Kindblom, L. G., Remotti, H. E., Aldenborg, F. & Meis-Kindblom, J. M. Gastrointestinal pacemaker cell tumor (GIPACT): gastrointestinal stromal tumors show phenotypic characteristics of the interstitial cells of Cajal. *Am. J. Pathol.* **152**, 1259–1269 (1998).
4. Antonescu, C. R. Gastrointestinal stromal tumor (GIST) pathogenesis, familial GIST, and animal models. *Semin. Diagn. Pathol.* **23**, 63–69 (2006).
5. Nakai, N. *et al.* A mouse model of a human multiple GIST family with KIT-Asp820Tyr mutation generated by a knock-in strategy. *J. Pathol.* **214**, 302–311 (2008).
6. Rubin, B. P. *et al.* A knock-in mouse model of gastrointestinal stromal tumor harboring kit K641E. *Cancer Res* **65**, 6631–6639 (2005).
7. Sommer, G. *et al.* Gastrointestinal stromal tumors in a mouse model by targeted mutation of the Kit receptor tyrosine kinase. *Proc. Natl Acad. Sci. USA* **100**, 6706–6711 (2003).
8. Kwon, J. G. *et al.* Changes in the structure and function of ICC networks in ICC hyperplasia and gastrointestinal stromal tumors. *Gastroenterology* **136**, 630–639 (2009).
9. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
10. Mertens, F. *et al.* Translocation-related sarcomas. *Semin. Oncol.* **36**, 312–323 (2009).
11. Jane-Valbuena, J. *et al.* An oncogenic role for ETV1 in melanoma. *Cancer Res* **70**, 2075–2084 (2010).
12. Nielsen, T. O. *et al.* Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* **359**, 1301–1307 (2002).
13. Segal, N. H. *et al.* Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *Am. J. Pathol.* **163**, 691–700 (2003).
14. Tomlins, S. A. *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595–599 (2007).
15. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
16. Ward, S. M. & Sanders, K. M. Physiology and pathophysiology of the interstitial cell of Cajal: from bench to bedside. I. Functional development and plasticity of interstitial cells of Cajal networks. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G602–G611 (2001).
17. Chen, H. *et al.* Differential gene expression in functional classes of interstitial cells of Cajal in murine small intestine. *Physiol. Genomics* **31**, 492–509 (2007).
18. Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene Er81 controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).
19. Gomez-Pinilla, P. J. *et al.* Ano1 is a selective marker of interstitial cells of Cajal in the human and mouse gastrointestinal tract. *Am. J. Physiol. Gastrointest. Liver Physiol.* **296**, G1370–G1381 (2009).
20. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
21. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
22. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
23. Miettinen, M. & Lasota, J. Gastrointestinal stromal tumors: review on morphology, molecular pathology, prognosis, and differential diagnosis. *Arch. Pathol. Lab. Med.* **130**, 1466–1478 (2006).
24. Andersen, R. J. *et al.* Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. *Cancer Cell* **17**, 535–546 (2010).
25. Moellering, R. E. *et al.* Direct inhibition of the NOTCH transcription factor complex. *Nature* **462**, 182–188 (2009).
26. Bauer, S., Yu, L. K., Demetri, G. D. & Fletcher, J. A. Heat shock protein 90 inhibition in imatinib-resistant gastrointestinal stromal tumor. *Cancer Res.* **66**, 9153–9161 (2006).

METHODS

Cell lines, antibodies and reagents. The GIST882 cell line, obtained from an imatinib naive patient, harbours a homozygous exon 13 *KIT* mutation (K642E) and was maintained in RPMI supplemented with 15% FBS, 10 mM HEPES pH 7.5. The GIST48 cell line, obtained from an imatinib-resistant patient, harbours a homozygous exon 11 *KIT* mutation (V560D) and a secondary heterozygous Exon 17 *KIT* mutation (D820A) and was maintained in Ham's F10 media supplemented with 15% FBS, 0.5% MITO + Serum Extender (BD Biosciences), and 30 mg l^{-1} bovine pituitary extract (BD Biosciences)[26]. The U2OS osteosarcoma, LNCaP prostate cancer and NIH-3T3 mouse embryonic fibroblast cells were obtained from the American Type Culture Collection and cultured as recommended.

The following antibodies were used: rabbit anti-ETV1 for western blotting, IHC and ChIP (Abcam), mouse anti-KIT for western blotting (Cell Signaling), rabbit anti-Phospho-tyr703 KIT for western blotting (Cell Signalling), mouse anti-phospho-tyr204 ERK for western blotting (Santa Cruz Biotechnologies), mouse anti-GAPDH for western blotting (Santa Cruz Biotechnologies), ACK2 rat anti-mouse Kit for immunofluorescence (E-Biosciences), rabbit anti-PGP9.5 for immunofluorescence (Abcam) and rabbit anti-ANO1 for immunofluorescence (Abcam). Secondary antibodies were Alexa-488 labelled anti-rabbit antibody and Alexa-594 labelled anti-rat antibody (Invitrogen). PD325901 was synthesized in the Memorial Sloan-Kettering Cancer Center Organic Synthesis Core Facility by O. Ouerfelli, and imatinib was a gift from Novartis.

Expression data mining. Five publically deposited databases containing expression of GIST samples are as follows: Expression Project for Oncology (ExpO, GSE2109) containing seven GIST samples among over 2,000 tumours of all types, NIELSEN_SARCOMA (GSE3443 (ref. 12)) containing 10 GIST samples among 44 sarcoma samples, SEGAL_SARCOMA (GSE2719 (ref. 13)) containing five GIST samples among 50 sarcoma samples, YAMAGUCHI_GIST (GSE8167 (ref. 27)) with 32 GIST samples and OSTROWSKI_GIST (GSE17743 (ref. 28)) with 29 GIST samples. Normalized expression of ETV1 of a given sample was calculated as the Z-score (that is, standard deviations from median) of ETV1 probe signal among all probes. GIST-signature gene sets derived from three data sets (ExpO, Nielsen and Segal) containing both GIST and non-GIST malignancies included genes that met the following two criteria: (1) multiple-hypothesis testing significance of $q < 0.05$ and (2) a mean expression that had a Z-score 1.5 times higher in GIST than non-GIST tumours. The three gene sets, EXPO_GIST_SIG, NIELSEN_GIST_SIG and SEGAL_GIST_SIG, contain 102, 156 and 30 genes, respectively (Supplementary Table 1).

For mouse ICC gene expression, we analysed an expression data set of ICC-MY, ICC-DMP and smooth muscle cells isolated from fluorescence-activated cell sorting of the mouse small intestine (GSE7809)[17]. ICC-MY and ICC-DMP signature genes sets were defined as genes expressed higher in ICC-MY and ICC-DMP compared with muscle by 1.5 Z-score with $P < 0.05$ and contain 61 and 81 genes, respectively (Supplementary Table 1).

Tumour samples. Clinical samples from patients with GIST and other sarcomas were obtained according to Memorial Sloan-Kettering Cancer Center Institutional Review Board protocol, and frozen and paraffin embedded tissue samples were banked. All GIST and non-GIST tumours were pathologically reviewed and confirmed by a sarcoma expert (C.R.A.). All GIST tumours used were sequenced for KIT and PDGFA mutations. For tumours in Fig. 1, G1–G8 are GISTs: G1, G2 and G4 (*KIT* exon11 mutation), G3 (*KIT* exon 9 mutation), G5 and G6 (*KIT* and *PDGFRA* wild-type), G7 and G8 (*PDGFRA* D842V mutation); S1–S5 are leiomyosarcoma, gastrointestinal leiomyoma, gastrointestinal leiomyoma, intra-abdominal desmoid tumour, and gastrointestinal schwannoma, respectively.

Fluorescence *in situ* hybridization of tumour samples. Two bacterial artificial chromosomes, one covering *ETV1* gene body and one on the 3′ side (RP11-138H16, RP11-124L22, CHORI) were labelled with Spectrum Green-dUTP (Vysis) and two bacterial artificial chromosomes covering 5′ upstream region of *ETV1* (RP11-703A4, RP11-115D14) were labelled with Spectrum Orange-dUTP (Vysis). Paraffin-embedded slides were de-waxed with xylene (Fisher), and then heated in 10 mM sodium citrate (pH 6) by microwaving for 5–10 min in a tender cooker (Nordic Ware). Slides were co-denatured with the probes for 6 min at 75 °C using HYbrite (Vysis), and kept in a dark, moist chamber at 37 °C for 16 h. Slides were then washed with 50% formamide (Fisher)/2 × SSC at 45 °C according to standard procedure. Nuclei were stained with DAPI, and slides were mounted with Vectashield (Vector). Results were viewed under a Zeiss microscope, and images were captured with an ISIS camera.

RNA isolation and qRT–PCR. For tissue culture cells, RNA was isolated using RNeasy (Qiagen). For frozen samples of clinical specimens and for explanted xenografts, tissue samples were grounded in 500 µl Trizol (Invitrogen) using a PowerGen homogenizer (Fisher Scientific), followed by addition of 100 µl chloroform. The samples were then centrifuged at 10,000*g* for 15 min. The upper phase

was mixed with an equal volume of 70% ethanol, and the RNA was further purified using an RNeasy column.

For qRT–PCR, RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (ABI) and PCR was run using Power SYBR Master Mix (ABI) on a Realplex machine (Eppendorf). Expression was normalized to the ribosomal protein RPL27 (ref. 29). Primers were designed using primer blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast) except when cited and purchased from Operon. The following primers pairs were used:

CTSL1: F: AGGCGCGTGACTGGTTGAGC, R: TGCATCGCCTTCCACTTG GTCC;

DUSP6: F: TGCCGGGCGTTCTACCTGGA, R: GGCGAGCTGCTGCTACAC GA;

ETV1-Exon23: F: AACAGAGATCTGGCTCATGATTCA, R: CTTCTGCAAG CCATGTTTCCTGTA[9];

ETV1-Exon67: F: CTACCCCATGGACCACAGATTT, R: CTTAAAGCCTTG TGGTGGGAAG[9];

KIT: F: GGGATTTTCTCTGCGTTCTG, R: GATGGATGGATGGTGGAG AC;

PLAT: F: CACTGGGCCTGGGCAAACATA, R: CACGTCAGCCTGCGGT TCTTC;

PROM1: F: GCCACCGCTCTAGATACTGC, R: GCTTTTCCTATGCCAAA CCA;

PTPRE: F: TGCACGCGGAGCAGAAGGTG, R: GCCGTGCATGGTCTGC AGGT;

RPL27: F: CATGGGCAAGAAGAAGATCG, R: TCCAAGGGGATATCCA CAGA;

TIMP3: F: CCAGCGCAAGGGGCTGAACT, R: TAGCCGCCCTTCTGCC GGAT.

Lentiviral knockdown. Two pLKO.1 constructs against *ETV1* (ETV1sh1: TRCN0000013923, targeting GTGGGAGTAATCTAAACATTT in 3′ UTR; and ETV1sh2: TRCN0000013925, targeting CGACCCAGTGTATGAACACAA in exon 7) were purchased from Open Biosystems and pLKO.1 shScr (targeting CCTAAGGTTAAGTCGCCCTCG) was purchased from Addgene[30]. Lentiviruses were generated by co-transfecting the shETV1 hairpin constructs with psPax2 and pVSVG (Addgene) into 293FT cells (Invitrogen) using Lipofectamine 2000 (Invitrogen). Viruses were collected at 48 and 72 h after transfection and concentrated 100-fold using Lenti-X concentrator (Clontech) and stored at −80 °C in aliquots. Infection was performed in 12-well dishes in the presence of 7.5 µg ml^{-1} polybrene (Sigma) by a 1-h centrifugation at 500*g*. To determine virus titre in GIST48 and GIST882 cells, two steps were taken. First, the relative virus titre of shScr, ETV1sh1 and ETV1sh2 was obtained by counting puromycin-resistant colonies after infection of U2OS cells, whose growth is insensitive to ETV1 knockdown. Next, the absolute titre of the shScr was calculated by counting puromyicn-resistant colonies after infection into GIST48 and GIST882 cells.

Stable gene expression. *ETV1* cDNA (Open Biosystems) and EGFP were cloned into MSCV-puro. Human wild-type KIT in MSCV-IRES-EGFP vector was obtained from G. Gilliland and generation of Δ560 mutation was performed using QuikChange II XL site-directed mutagenesis kit (Stratagene). Retrovirus was produced in 293FT cells by standard methods using ecotropic packaging vector and infected into NIH-3T3 cells at multiplicity of infection (MOI) = 5. For ETV1 expression, cells were selected with puromycin (1 µg ml^{-1}). For KIT expression, fluorescence-activated cell sorting of green fluorescent protein demonstrated greater than 90% infection; therefore cells were not further enriched.

Cell growth and cell cycle analysis after shRNA-mediated knockdown. For GIST48 and GIST882 cells, cells were plated at 2×10^5 cells per well in a 24-well plate on day 0 and infected with pLKO.1 hairpin viruses (MOI = 5) on day 1. Cells were not further puromycin selected. Triplicate wells were counted every 3 days until day 13. For U2OS cells, cells were plated at 1×10^5 cells per well in 24-well plates on day 0 and infected at MOI = 5 in 24-well plates on day 1. On day 3, they were split into 96-well plates (2,500 cells per well). Viability was assessed using Cell Titer Glo (Promega). For cell cycle analysis, cells were trypsinized on day 4 after infection, fixed in 70% ethanol and incubated with 20 µg ml^{-1} propidium iodide (Sigma) and 0.2 mg ml^{-1} RNase A (Invitrogen) for 2 h before fluorescence-activated cell sorting analysis on a FACSCalibur (BD Biosciences). Fluorescence intensities of G1 and G2 peaks were normalized between samples, and cell cycle analysis was performed using FlowJo software.

Xenograft growth. For GIST882, cells were infected at MOI = 1 on day 0 and briefly selected with 5 µg ml^{-1} puromycin between days 2 and 4. On day 4, 5×10^6 cells resuspended in 100 µl of 1:1 mix of growth media and Matrigel (BD Biosciences) were subcutaneously injected into CB17-SCID mice (Taconic). Tumour sizes were measured weekly starting 6 weeks after xenografting. After 10 weeks, xenografts were explanted and RNA isolated for qRT–PCR analysis. For NIH-3T3 cells stably expressing combinations of vector, ETV1, KIT and

KITΔ560, 1×10^6 cells were implanted as above and tumour sizes were measured starting 2 weeks after xenografting.

Anchorage independent growth. Transduced NIH-3T3 cells were suspended in soft agar with 20% calf-serum at 5,000 cells per well in a six-well plate in triplicate. After 3 weeks, each well was stained with 0.5% crystal violet, imaged and counted using GelCount (Oxford Optronix).

Immunofluorescence. For ETV1 immunofluorescence of paraffin sections of tumour samples, heat-mediated antigen retrieval was performed followed by staining with ETV1 antibody at $2.5\,\mu g\,ml^{-1}$ and Alexa-594 labelled anti-rabbit secondary antibody.

For immunofluorescence of cryostat sections of the mouse gastrointestinal tract, mouse stomach, small intestine, caecum and large intestine were dissected and fixed in 4% paraformaldehyde for 2 h followed by an overnight incubation in 30% sucrose. They were then embedded in OCT, flash frozen and cut into 5-μm sections using a cryostat. $Etv1^{-/-}$ and $Etv1^{+/+}$ littermate controls were embedded onto the same block to ensure identical processing. Tissue sections were blocked overnight using 5% goat serum, incubated with primary antibodies at 4 °C overnight, and secondary antibody for 2 h at room temperature. Slides were mounted using Prolong Gold (Invitrogen) and images were taken on a Nikon Eclipse TE2000-E microscope using a Photometric Coolsnap HQ camera. Images were taken with ×20 (numerical aperture 0.75) or ×60 (numerical aperture 1.4) objectives. Monochrome images taken with DAPI, FITC and Texas Red filter sets were pseudo-coloured blue, green and red, respectively, and merged using ImageJ (http://rsbweb.nih.gov/ij/). The exposure, threshold and maximum were identical between $Etv1^{-/-}$ and $Etv1^{+/+}$ images.

For whole-mount staining, mouse gastrointestinal tissue was fixed in ice-cold acetone for more than 1 h. It was then rehydrated in PBS, stretched and the mucosa of these gastrointestinal tissues was blunt dissected away. Samples were then blocked in 5% goat serum, incubated with ACK2 rat-anti-KIT antibody or PGP9.5 antibody, followed by secondary antibody. Images were taken with ×20 (numerical aperture 0.75) or ×60 (numerical aperture 1.4) objectives with a Z-step of 1 μm through the muscle layer. The Z-stack of images was deconvoluted using Autoquant X2 software (Media Cybernetics). Movies were generated from deconvoled stacks using ImageJ software.

To quantify ICC subsets, immunofluorescence images from at least four fields each from two $Etv1^{+/+}$ and $Etv1^{-/-}$ mice (at least eight fields in total) of the stomach, small intestine, caecum and large intestine were examined. A percentage of ICCs were obtained by dividing the total number of KIT-positive immunostaining cells by the number of DAPI-positive nuclei. The results are shown as the percentage of residual KIT-positive immunostaining in $Etv1^{-/-}$ mice compared with $Etv1^{+/+}$ controls.

Expression profiling. For profiling of *ETV1* knockdown, GIST48 and GIST882 cells plated in 12-well plates were infected in triplicates with shScr, and ETV1sh1 and ETV1sh2 viruses at MOI = 5. Four days after infection, RNAs were isolated from duplicate samples and protein lysates were obtained from the third for confirmatory ETV1 western blot. The duplicate samples were expression profiled using an Illumina HumanHT-12 array. Raw data was imported into Genespring 10, quartile normalized and $\log_2$ transformed. The log transformed expression value of each probe was averaged across duplicates. For each probe, the average change by ETV1 knockdown is calculated as $((\text{ETV1sh1}_{\text{GIST48}} - \text{ShScr}_{\text{GIST48}}) + (\text{ETV1sh2}_{\text{GIST48}} - \text{ShScr}_{\text{GIST48}}) + (\text{ETV1sh1}_{\text{GIST882}} - \text{ShScr}_{\text{GIST882}}) + (\text{ETV1sh2}_{\text{GIST882}} - \text{ShScr}_{\text{GIST882}}))/4$.

We next generated a ranked list of genes based on the averaged $\log_2$ transformed fold change. The top 48 genes showed an average decrease of greater than 1.7-fold with *ETV1* knockdown (Fig. 2). We defined a 'shETV1 DN' gene set as the 410 genes with an average decrease greater than 1.4-fold and a 'shETV1 UP' gene set as the 380 genes with an average increase of greater than 1.4-fold.

For profiling of imatinib treatment, GIST882 cells were treated with vehicle or 1 μM imatinib for 8 h in triplicate and profiled as above.

Gene set enrichment analysis. Given a ranked list of genes, GSEA determines whether predefined sets of genes (for example, genes within a Gene Ontology category) are disproportionally overrepresented in the top or the bottom of the list instead of randomly across the list[20]. To assess the phenotypic association with *ETV1* knockdown, we used the ranked list of genes generated above from the most upregulated by *ETV1* knockdown to most downregulated by *ETV1* knockdown.

For genes with multiple probes by the Illumina HumanHT-12 array, the median signal was used. We used gene sets from the Molecular Signatures Database C2 version 2.5 with 1,892 curated gene sets of chemical and genetic perturbations and canonical pathways and C5 version 2.5 with 1,454 Gene Ontology gene sets. In addition, we added five custom gene sets including the three GIST signature gene sets EXPO_GIST_SIG, NIELSEN_GIST_SIG and SEGAL_GIST_SIG and two ICC signature gene sets ICC_MY_SIG and ICC_DMP_SIG, defined above.

ChIP-Seq. ChIP-Seq was performed as previously described[31]. Briefly, 10^8 GIST48 cells were crosslinked for 15 min in 1% paraformaldehyde, washed and lysed. Chromatin was sheared using Bioruptor (Diagenode) to fragments of approximately 150 base pairs and was incubated with anti-rabbit IgG Dynabeads (Invitrogen) pre-conjugated with anti-ETV1 antibody, washed and eluted. The eluted chromatin was reverse-cross-linked, and DNA was column purified. The purified ChIP DNA was blunt-ended, ligated to Solexa adaptors, amplified with 18 cycles of PCR and sequenced on a Solexa Genome Analyser. We performed independent quantitative PCR validation of separate ChIP samples using the following primers:
DUSP6 promoter: F: GCCCGCTGTTGCAGCTTGTT, R: GCCGGCTGGAAC AGGTTGTG
GPR20 enhancer: F: CCCTCCCAGGCTCTCCCCAC, R: TCCGGGCCTGCTCT CTGTCC
GAPDH promoter: F: TCCCAAAGTCCTCCTGTTTCA, R: CAGCAGGACAC TAGGGAGTCAA.

All reads (approximately 150 base pairs) were aligned to the human genome (hg18) using the ELAND alignment software within the Illumina Analysis Pipeline. Unique reads mapped to a single best-matching location with no more than two mismatches were kept and used to generate genome-wide distribution of ETV1 binding and for peak identification. Redundant reads were considered once. The software MACS[32] was applied to the ChIP-Seq data with sequencing data from input DNA as control for identifying genomic regions enriched with ETV1 ChIP signals (that is, peaks). For each peak, we subsequently calculated the number of reads (extending 200 nucleotides in the 3′-direction) covering each nucleotide position and then defined the maximal count as the peak height. Using an FDR less than 1% (as defined by MACS) and a height of at least 15 reads, we defined 14,741 peaks. These peaks were classified as promoter peaks (within ±2 kilobases (kb) of transcription start sites), enhancer peaks (from −50 kb of transcription start sites to +5 kb of transcription end sites, but excluding promoter regions) or otherwise intergenic peaks based on their genomic distances to known Refseq gene annotation. To associate an ETV1 peak with a specific gene, we used the following protocol. A promoter peak was assigned to all genes if it was in the promoter regions of multiple genes. An enhancer peak within a gene body was assigned to the host gene, but otherwise was assigned to the nearest gene. Nucleotide sequences of all ETV1 peaks were retrieved from the reference human genome and subjected to motif analysis by the program MEME[33].

To correlate gene expression with ETV1-binding sites, we first limited the analysis to the genes with peaks to those that were also represented on the mRNA microarray (Illumina HumanHT-12). Overlapping analysis of genes whose expression was perturbed by ETV1 knockdown and those with ETV1 binding peaks was performed using Microsoft Excel macros, and significance was calculated using Fisher's exact test.

27. Yamaguchi, U. *et al.* Distinct gene expression-defined classes of gastrointestinal stromal tumor. *J. Clin. Oncol.* **26,** 4100–4108 (2008).
28. Ostrowski, J. *et al.* Functional features of gene expression profiles differentiating gastrointestinal stromal tumours according to KIT mutations and expression. *BMC Cancer* **9,** 413 (2009).
29. de Jonge, H. J. *et al.* Evidence based selection of housekeeping genes. *PLoS ONE* **2,** e898 (2007).
30. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307,** 1098–1101 (2005).
31. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140,** 678–691 (2010).
32. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
33. Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34,** W369–W373 (2006).

LETTER

Anthrax toxins cooperatively inhibit endocytic recycling by the Rab11/Sec15 exocyst

Annabel Guichard[1], Shauna M. McGillivray[2,3]*, Beatriz Cruz-Moreno[1]*, Nina M. van Sorge[2]*, Victor Nizet[2,4] & Ethan Bier[1]

Bacillus anthracis is the causative agent of anthrax in humans and other mammals[1,2]. In lethal systemic anthrax, proliferating bacilli secrete large quantities of the toxins lethal factor (LF) and oedema factor (EF), leading to widespread vascular leakage and shock. Whereas host targets of LF (mitogen-activated protein-kinase kinases) and EF (cAMP-dependent processes)[3] have been implicated in the initial phase of anthrax[1,2], less is understood about toxin action during the final stage of infection. Here we use Drosophila melanogaster to identify the Rab11/Sec15 exocyst, which acts at the last step of endocytic recycling, as a novel target of both EF and LF. EF reduces levels of apically localized Rab11 and indirectly blocks vesicle formation by its binding partner and effector Sec15 (Sec15–GFP), whereas LF acts more directly to reduce Sec15–GFP vesicles. Convergent effects of EF and LF on Rab11/Sec15 inhibit expression of and signalling by the Notch ligand Delta and reduce DE-cadherin levels at adherens junctions. In human endothelial cells, the two toxins act in a conserved fashion to block formation of Sec15 vesicles, inhibit Notch signalling, and reduce cadherin expression at adherens junctions. This coordinated disruption of the Rab11/Sec15 exocyst by anthrax toxins may contribute to toxin-dependent barrier disruption and vascular dysfunction during B. anthracis infection.

B. anthracis, the aetiological agent of anthrax, secretes three factors that are required for systemic virulence[1–3]: the toxic enzymatic moieties LF and EF, and protective antigen (PA), which promotes entry of LF and EF into host cells. LF is a zinc metalloprotease that cleaves and inactivates most human mitogen-activated protein-kinase kinases (MAPKKs)[4,5], and EF is a potent calmodulin-dependent adenylate cyclase[6]. It has been speculated that additional host targets may contribute to mediating the lethal effects of anthrax toxins[7] and interactions between the two toxins remain poorly understood.

We chose *D. melanogaster* as a model system to identify new candidate pathways involved in anthrax pathogenesis. LF and EF act on conserved signalling components, MAPKKs and PKA, respectively, when expressed directly within the cells of transgenic flies, bypassing the need for PA-mediated endocytosis[8]. Here we report that strong expression of either LF or EF in the larval wing primordium also produces new, unexpected phenotypes. These phenotypes, including wing notching and thickened veins (Supplementary Fig. 1a–c and Supplementary Fig. 2a), are typical of mutants in the Notch signalling pathway and were strongest for EF when using the same *GAL4* driver (for example, Fig. 1b, c and h, i). Consistent with these adult phenotypes, high-level expression of either toxin greatly reduced expression of the Notch target genes *wg* and *cut* (Supplementary Fig. 1d–i). We also observed potent dose-sensitive genetic interactions between mutations in Notch pathway components and the expression of LF (Supplementary Fig. 2b–f) or EF (Supplementary Fig. 2g–n).

An important unresolved issue is whether LF and EF, which are both required individually for the pathogenicity of *B. anthracis*, also

interact in some concerted fashion[9]. We tested for toxin synergy by co-expressing them with a weak ubiquitous wing-specific *GAL4* (*wkG4*) driver. Expression of LF alone produced no obvious phenotype (Fig. 1b, compare to wild type in a). Similarly, expression of only EF resulted in mild occasional notching of the wing margin (Fig. 1c), although it also caused an unrelated phenotype consisting of small wings with altered vein spacing. When EF and LF were co-expressed with this *GAL4* driver, however, strong and penetrant wing margin notching phenotypes were superimposed on the EF patterning phenotype (Fig. 1d). Correspondingly, expression of the Notch target gene *wg* (Fig. 1a–d, lower panels) and a Notch reporter construct (not shown) were greatly reduced in LF+EF wing discs. Synergy between LF and EF was also observed using other drivers, such as a strong ubiquitous *GAL4* driver (*stgG4*) (Fig. 1h–j) and the *dpp-GAL4* driver (*dppG4*) (data not shown).

These initial phenotypic observations led us to examine the mechanisms underlying the Notch inhibitory effects of the anthrax toxins. Both LF (Fig. 1e, f) and EF (Fig. 1e, g) greatly reduced levels of the Notch ligand Delta (Dl) at the apical surface of wing discs, which we confirmed by selective staining for extracellular Dl expression (Supplementary Fig. 3a–c), and modestly decreased apical levels of Notch (Supplementary Fig. 3g–i). EF also significantly reduced surface expression of a second Notch ligand, Serrate (Supplementary Fig. 3d, e). We conclude that both LF and EF inhibit trafficking of Dl to the apical cell surface, diminishing Notch signalling.

Activation of Dl requires initial cell surface expression followed by endocytosis and recycling (reviewed in ref. 10), targeting it to the adherens junction, where it engages the Notch receptor[11]. Small GTPases from the Rab family mediate specific steps of this process[12]. We expressed dominant-negative forms of each of the *Drosophila* Rab proteins (Rab(DN))[13] and identified a single dominant-negative factor, Rab11(DN), which produced phenotypes virtually identical to those caused by EF (Fig. 1i, k, Supplementary Fig. 4b, c and Supplementary Table 1). Reciprocally, we co-expressed wild-type forms of each Rab with EF to determine whether increasing their dosage could rescue the Notch-like EF phenotype and found that only Rab11 could suppress EF activity (Fig. 1l, m, compare to i; Supplementary Table 1). Consistent with Rab11 mediating the inhibition of Notch signalling by EF, co-expression of Rab11(DN) with EF greatly enhanced its wing phenotypes (Fig. 1n) and Rab11(DN), like EF, could synergize with LF to produce a stronger phenotype (Fig. 1o, similar to j). Rab11(DN) also mimicked the effect of EF in blocking Dl trafficking to the cell surface (Supplementary Fig. 5d–f). A similar role of Rab11 in recycling endocytosed Dl to the apical cell surface has been demonstrated during sensory organ precursor cell development in *Drosophila*[11,14].

The endogenous Rab11 protein is distributed as small grainy particles just below the apical plasma membrane in wild-type wing discs[11] (Fig. 2a). In EF-expressing discs, the level of apical Rab11 expression was greatly diminished (Fig. 2b) and ectopic Rab11-positive vesicles

[1]Section of Cell and Developmental Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0349, USA. [2]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0687, USA. [3]Department of Biology, Texas Christian University, 2800 South University Drive, Fort Worth, Texas 76129, USA. [4]Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0687, USA.
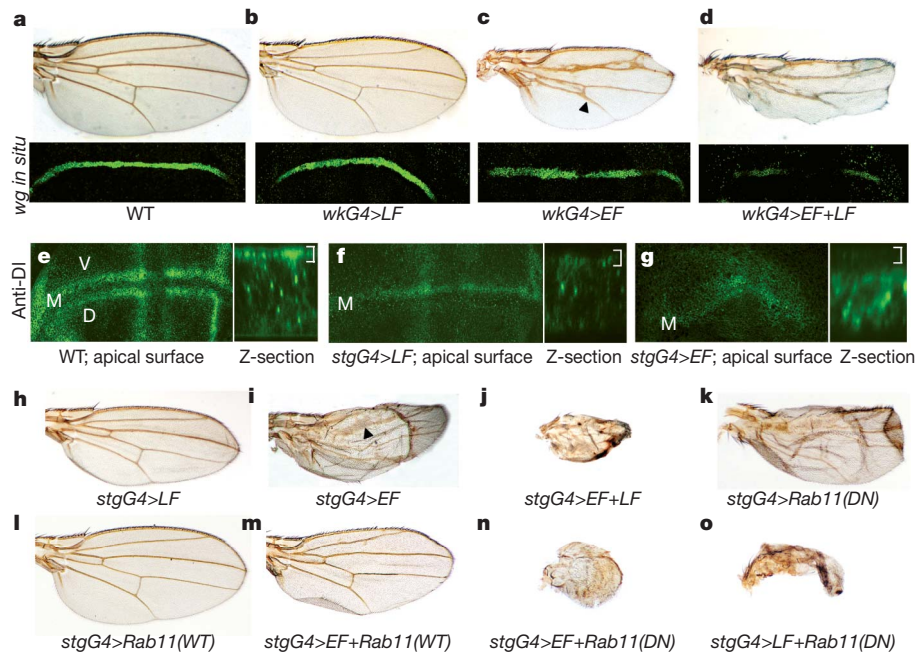*These authors contributed equally to this work.

Figure 1 | LF and EF synergistically inhibit Notch signalling. a–d, Wings (upper panels) and corresponding *wg* expression in wing imaginal discs (bottom panels) of the following genotypes: **a**, wild-type (WT); **b**, *wkG4>LF2X* (*wkG4* refers to the *1348-GAL4* driver); **c**, *wkG4>EF* (wing has anterior/posterior patterning phenotype superimposed on thickened veins (arrowhead) and an occasional small notch at the wing margin); **d**, *wkG4>LF2X+EF* (*LF2X* refers to two copies of the *UAS-LF* transgene). **e–g**, Reticular pattern of Dl staining in wing discs with accompanying Z-sections. The *stgG4* driver (*MS1096GAL4*) is expressed at higher levels on the dorsal (D) surface than on

the ventral (V) surface. **e**, Wild-type Dl expression has both cell surface (bracket) and vesicle-like intracellular components, and is expressed along the future margin (M) in two parallel lines as well as in vein primordia (which intersect the margin in perpendicular stripes). **f**, *stgG4>LF* (here, and in all subsequent panels, *LF* refers to three copies of the *UAS-LF* transgene). **g**, *stgG4>EF*. **h–o**, Wings of the following genotypes: **h**, *stgG4>LF*; **i**, *stgG4>EF* (arrowhead indicates thickened veins); **j**, *stgG4>LF+EF*; **k**, *stgG4>Rab11(DN)*; **l**, *stgG4>Rab11(WT)*; **m**, *stgG4>EF+Rab11(WT)*; **n**, *stgG4>EF+Rab11(DN)*; **o**, *stgG4>LF+Rab11(DN)*.

appeared in basolateral areas of the cytoplasm (quantified in Supplementary Table 2). This finding indicates that EF acts, at least in part, by reducing the amount and/or altering the distribution of the Rab11 GTPase. Because Rab11 has been implicated in targeting other proteins to the adherens junction in addition to Notch signalling components[15], we examined the expression of several adherens junction proteins. In wild-type wing discs, the homophylic adhesion molecule DE-cadherin (DECad) is expressed at points of cell–cell contact (Fig. 2d), where it co-localizes with Dl (Supplementary Fig. 5a–c). In contrast, EF precipitously reduced DECad expression at the adherens junction (Fig. 2e), mimicking the effect of inhibiting Rab11 function (Fig. 2f and Supplementary Fig. 6a, b). This downregulation of DECad by EF could be partially rescued by co-expression with wild-type Rab11 (Supplementary Fig. 5g–i). Similar, albeit less pronounced, reductions were observed for the adherens junction proteins α-Catenin and β-Catenin (Supplementary Fig. 6d, e and g, h, respectively); however, Discs Large (Dlg) expression was unaltered (Supplementary Fig. 6j, k), indicating that EF acts selectively and does not lead to gross disruption of adherens junction integrity per se.

As Rab11 interacts with its effector Sec15 to initiate formation of the exocyst complex, thereby leading to fusion of vesicles from the recycling endosome with the plasma membrane (reviewed in ref. 12), we examined the effect of EF on expression of a *Drosophila* Sec15–GFP fusion protein construct[11]. Sec15–GFP expression has two staining components (Fig. 2g): large round structures near the cell surface and diffuse cytoplasmic staining. Vesicular Sec15–GFP, which corresponds to a late endocytic compartment poised to fuse with the plasma membrane[11,16–18], co-localized with Rab11 (Fig. 2j and Supplementary Fig. 7a), consistent with the known interaction of these two proteins in the exocyst complex[11,15,19]. Expression of EF virtually abolished large Sec15–GFP vesicles (Fig. 2h), and the few that remained were typically smaller than those in wild-type discs and did not co-label as strongly with Rab11 (Supplementary Fig. 7a, b). In contrast, the uniform cytoplasmic component of Sec15–GFP staining was largely unaltered by

EF. EF probably blocks formation of large Sec15–GFP vesicles indirectly, through its effect on Rab11, because inhibition of Rab11 by Rab11(DN) (Fig. 2i) or by RNA-mediated interference (RNAi) (data not shown) had the same effect. Furthermore, co-expression of wild-type Rab11 with EF fully rescued punctate Sec15–GFP expression (Fig. 2k and Supplementary Fig. 8a–c).

Because EF and LF interact synergistically in the wing and both toxins reduce access of Dl to the cell surface, we tested whether LF acted at the same recycling step as EF. Although LF did not appreciably alter Rab11 staining (Fig. 2c) it, like EF, nearly eliminated large Sec15–GFP vesicles (Fig. 2l) and residual small Sec15–GFP vesicles no longer strongly co-labelled with Rab11 (Supplementary Fig. 7a, c). In contrast to EF, however, the loss of Sec15–GFP staining was only weakly rescued by co-expression with wild-type Rab11 (Supplementary Fig. 8e, f, compare to b, c for EF). LF also reduced levels of DECad at the apical cell surface (Supplementary Fig. 6a–c), although not as strongly as EF. Consistent with Sec15 being a mediator of LF Notch inhibitory activity, knockdown of endogenous *sec15* function by RNAi caused Notch-like phenotypes in the wing (Supplementary Fig. 9b, d), although overexpression of wild-type *sec15* had no effect (Supplementary Fig. 9a, c). We conclude that LF and EF converge to inhibit two interacting components of the Rab11/Sec15 exocyst, resulting in reduced Notch signalling and weakened adherens junctions.

We next asked whether EF and LF could disrupt function of the well-conserved Rab11/Sec15 exocyst and its downstream effectors Notch and cadherins in mammalian systems. Established models of endothelial cell function were selected as Notch signalling has a central role in vascular remodelling (reviewed in ref. 20), and cadherins are essential for maintaining vascular integrity[21]. We transfected a rat Sec15–GFP construct into human brain microvascular endothelial cells (hBMECs) to visualize the exocyst and observed large vesicles (Fig. 3a, d) similar to those in *Drosophila* wing discs (for example, Fig. 2g), yeast[17] and various mammalian cell types[18]. As in *Drosophila*, treatment with either EF toxin (EF+PA) (Fig. 3b, h) or LF toxin (LF+PA) (Fig. 3c) greatly
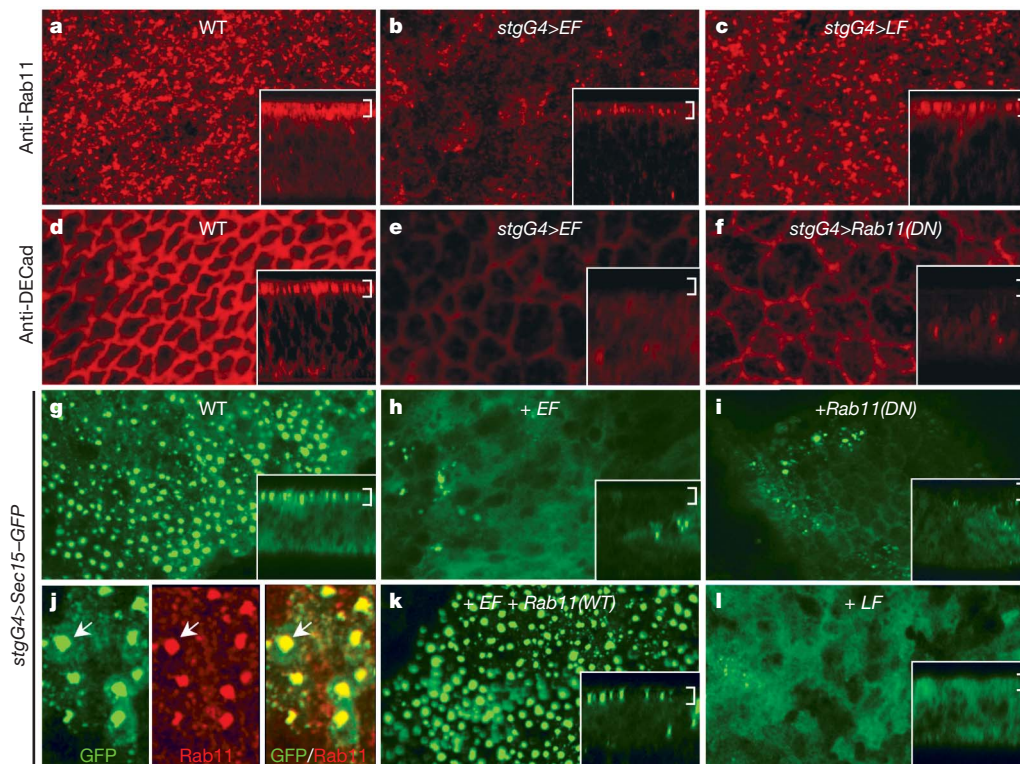
Figure 2 | LF and EF inhibit Rab11/Sec15-dependent recycling.
a–c, Endogenous Rab11 expression in wing imaginal discs detected by immunofluorescence. Insets are Z-sections from regions of the same discs in these and other panels (brackets indicate the cell surface). **a**, Wild type. **b**, *stgG4>EF*. **c**, *stgG4>LF*. **d–f**, DECad expression detected by immunofluorescence. **d**, Wild type. DECad and Dl co-stain in a net-like pattern at points of cell–cell contact (Supplementary Fig. 5a–c). **e**, *stgG4>EF*. **f**, *stgG4>Rab11(DN)*. **g–l**, Expression of a *UAS-Sec15–GFP* construct driven by the *stgG4* driver. **g**, Wild-type wing disc expressing *stgG4>Sec15–GFP* in large

vesicles beneath the cell surface. **h**, *stgG4>EF+Sec15–GFP*. **i**, *stgG4>Rab11(DN)+Sec15–GFP* (vesicular Sec15–GFP expression is similarly dependent on *Rab11 = Sec4* function in yeast[17]). **j**, High-magnification view of Sec15–GFP and Rab11 co-localization in a wild-type wing disc. Left panel, Sec15–GFP (green). Middle panel, Rab11 (red). Right panel, overlap between Sec15–GFP and Rab11 (yellow). Arrows indicate one example of co-localization. **k**, *stgG4>EF+Rab11(WT)+Sec15–GFP*. **l**, *stgG4>LF+Sec15–GFP*. Staining differences are quantified in Supplementary Table 2.

reduced the number and size of the Sec15–GFP vesicles. Moreover, in the case of EF toxin, co-transfection of cells with human RAB11–RFP rescued the formation of large Sec15–GFP vesicles (compare Fig. 3h with i). Mirroring other systems, Rab11 and Sec15 co-localized, both in untreated (Fig. 3e–g) and in EF-treated endothelial cells rescued with RAB11–RFP (Fig. 3i–k). We conclude that the EF and LF toxins function similarly in human and *Drosophila* cells to disrupt formation of the Rab11/Sec15 exocyst.

Looking downstream of the exocyst, we found that treatment with EF toxin disrupted the strong and uninterrupted pan-cadherin (pCAD) expression found at points of cell–cell contact in untreated monolayers of hBMECs (compare Fig. 3b, h with a, d), in primary human dermal microvascular endothelial cells (hDMECs; compare Fig. 3m with l; Supplementary Fig. 10a, b), and in primary human lung microvascular endothelial cells (hMVEC-Ls; Supplementary Fig. 11a, b). LF toxin had no clear effect on pCAD in hBMECs (Fig. 3c), although levels were moderately reduced in hDMECs (Supplementary Fig. 10c), which form more regular borders than hBMECs.

We also examined the effect of anthrax toxins on Notch signalling in mammalian cells. hBMECs were infected with wild-type *B. anthracis* Sterne bacteria, which express both EF and LF, or isogenic mutant bacteria lacking EF (ΔEF), LF (ΔLF), or both toxins (ΔLF+EF or ΔpXO1)[22]. Bacterial anthrax toxin production inhibited hBMEC expression of the Notch target gene *HES1* (Fig. 3n; quantified in Supplementary Fig. 12a) and *RBPJ* (Supplementary Fig. 12b), with EF exerting the dominant effect. Notch-dependent regulation of *HES1* in hBMECs was confirmed using the γ-secretase inhibitor DBZ (Fig. 3n). In addition, EF toxin treatment of hDMECs or hMVEC-Ls triggered formation of large and misshapen intracellular vesicles of the

endothelial-specific Notch ligand Delta-Like 4 (DLL4) (compare Fig. 3m with l and Supplementary Fig. 11a with b) as observed in *Drosophila* (Fig. 1e–g).

Because previous studies have described the effect of anthrax toxins on vascular leakage and pulmonary oedema[9,23,24], we analysed endothelial barrier integrity using *in vitro* and *in vivo* assays during infection. Exposure to wild-type *B. anthracis* increased the permeability of hBMEC transwell monolayers, an effect principally dependent on EF activity (Fig. 3o). Similarly, purified EF toxin induced dose-dependent hBMEC permeability in the same assay (Supplementary Fig. 12c). Next, individual mice were infected subcutaneously in adjacent locations with wild-type and mutant strains of *B. anthracis*[22], followed 6 h later by intravenous injection of Evans blue dye (Miles assay[25,26]). Wild-type *B. anthracis* induced severe vascular effusion at the site of injection (Fig. 3p), and this effect was greatly attenuated in ΔEF mutant bacteria, but only modestly so in ΔLF mutants (Fig. 3p, q). Similarly in the lung, wild-type *B. anthracis* induced pulmonary oedema, indicative of pulmonary endothelial barrier dysfunction, and this effect was also abrogated in ΔEF mutant bacteria (Supplementary Fig. 11c, d).

In summary, LF and EF toxins interact synergistically in *Drosophila* to block Rab11/Sec15-dependent endocytic recycling, resulting in reduced Notch signalling and cadherin-dependent adhesion at the adherens junction, and these toxins produce very similar effects in mammalian cells. Failure to target proteins to the adherens junction may contribute to the loss of endothelial barrier integrity in EF-toxin-treated cells and to the toxin-dependent vascular effusion caused *in vivo* during *B. anthracis* infection (see summary scheme in Fig. 3r). The reduction in Dl/Notch levels in response to anthrax toxin treatment requires further analysis with respect to potential consequences
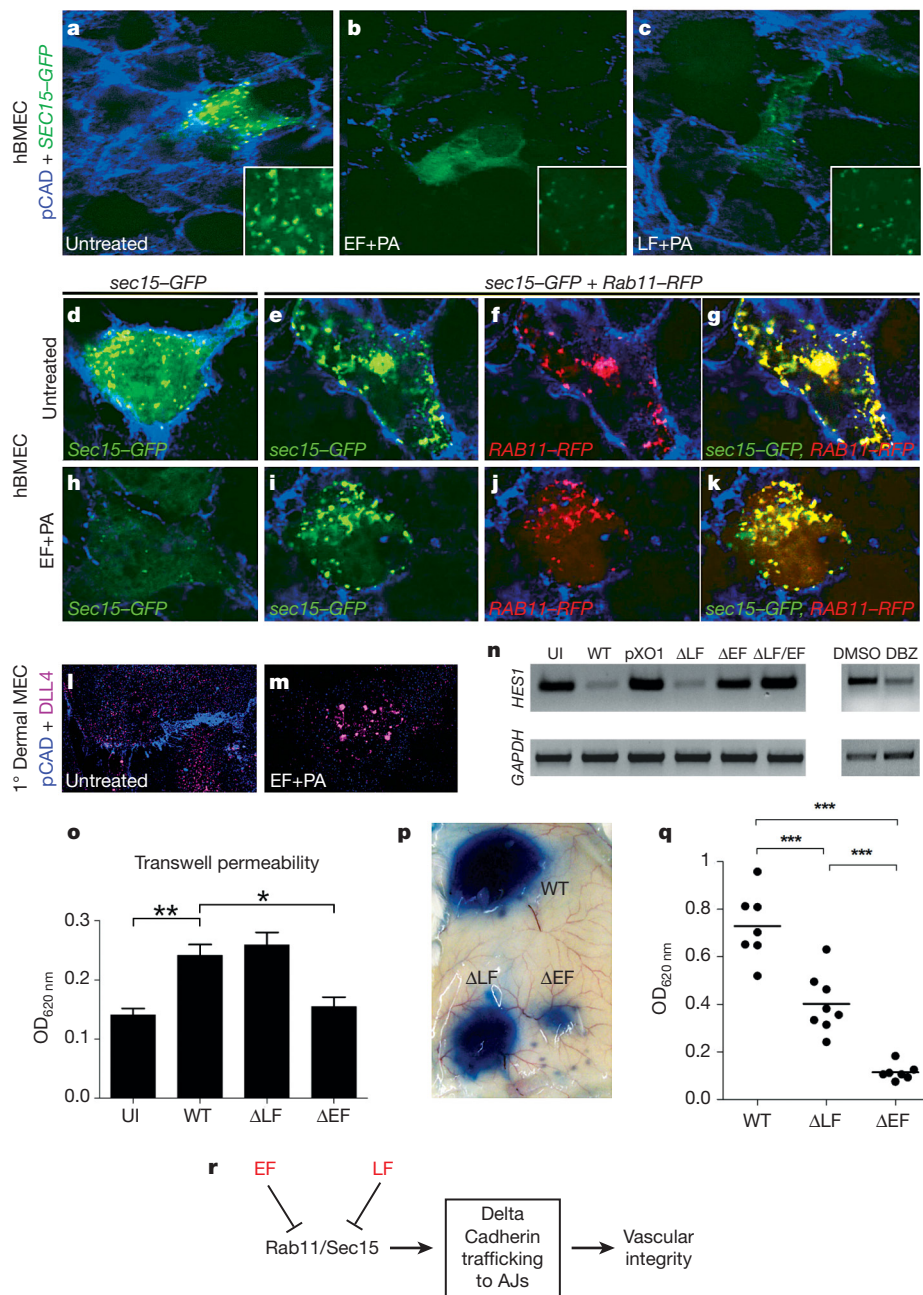
Figure 3 | Conserved activity of anthrax toxins in mammals. a–c, Pan-cadherin (pCAD) staining (blue) in hBMECs transfected with a human *SEC15–GFP* construct (green). **a,** Untreated cells. **b,** Cells treated with EF toxin (3 μg EF and 6 μg PA) for 24 h. **c,** Cells treated with LF toxin (3 μg LF and 6 μg PA) for 24 h. Treatment with doses ranging from 0.3 μg to 3 μg gave similar results. Staining differences are quantified in the legend of Supplementary Table 2. **d–k,** Rescue of Sec15–GFP (green) expression by Rab11–RFP (red) in EF-toxin-treated hBMECs. **d–g,** Untreated cells, and **h–k,** cells treated with 0.3 μg EF and 0.6 μg PA. **d, h,** Cells transfected with *Sec15–GFP* alone. **e–g, i–k,** Cells co-transfected with *sec15–GFP* and wild-type *RAB11–RFP*. Cells in panels **d–k** were also stained for pCAD (blue). **l, m,** Expression of pCAD (blue) and DLL4 (magenta) in untreated hDMECs (**l**), or after treatment with 1 μg EF and 2 μg PA (**m**). **n,** Semi-quantitative analysis of *HES1* and *GAPDH* RNA expression in hBMECs. First set of lanes, hBMECs infected with *B. anthracis* or isogenic mutants. UI, uninfected control; WT, *B. anthracis* Sterne bacteria; ΔpX01, *B. anthracis* lacking the pX01 plasmid; ΔLF, *B. anthracis* with deletion of LF; ΔEF, *B. anthracis* with deletion of EF; ΔLF/ΔEF, *B. anthracis* with deletion of both LF and EF. Second set of lanes, effect of the γ-secretase

inhibitor DBZ (2 μM) or vehicle control (dimethyl sulphoxide (DMSO)) on *HES1* expression in hBMECs. **o,** Transwell permeability assay of hBMECs grown to confluence in a transwell chamber and infected with *B. anthracis* wild-type or isogenic toxin mutants. Leakage across the monolayer was determined 6 h later by measuring Evans blue leakage to the bottom chamber at $OD_{620\,nm}$. Abbreviations for genotypes of bacteria are as in panel **n**. *, $P < 0.05$; **, $P < 0.01$. Mean and standard deviation (represented by error bars) of a representative experiment are shown. **p,** Vascular effusion in response to subcutaneous infection with *B. anthracis* (wild-type *B. anthracis*) or isogenic toxin mutants ΔLF or ΔEF. Effusion was visualized by Evans blue dye leakage. **q,** Quantification of vascular permeability shown in Fig. 3p. ***, $P < 0.001$. **r,** Proposed schematic model for the convergent activity of EF and LF on the exocyst. EF reduces Rab11 levels/activity, which indirectly inhibits formation of Sec15 exocyst complexes, whereas LF acts more directly on Sec15. The combined effect of these two toxins is to reduce cell surface expression of the Notch ligand Dl and cadherins at adherens junctions (AJs), and possibly other adherens junction proteins involved in cell–cell adhesion and barrier maintenance, thereby compromising vascular integrity.

on vascular integrity, which could be direct (for example, mediated by Notch-dependent regulation of factors such as VEGF or by Dl–Notch adhesion) or indirect (for example, mediated by Notch-dependent regulation of cytokine production). The precise mechanisms by which EF and LF cooperatively inhibit Rab11/Sec15 function remain to be elucidated. EF-mediated cAMP production could act on either or both of two known effectors, PKA and Epac, both of which have connections to Rab11 regulation[27,28]. LF may function via cleavage and inactivation of its known MAPKK targets or act on novel targets. It is also unclear why EF has consistently stronger effects than LF in both flies and vertebrates, as they both converge on the exocyst. It may be that Rab11 has additional partners that act in parallel to Sec15. Alternatively, LF may block only a subset of Sec15 functions or may exert competing effects on the exocyst mediated by opposing actions of different MAPKKs or as yet unidentified targets. Future genetic dissection of these impinging pathways will be required to distinguish between these possibilities. Finally, it may be fruitful to examine whether the mammalian exocyst and its downstream effectors could also be targets of other microbial virulence factors known to increase cAMP levels, inhibit MAPKK signalling, or disrupt host barrier integrity.

METHODS SUMMARY

Drosophila genetics. Transgenic lines UAS-LF2X/FM7, UAS-LF3X/FM7 and UAS-EF UAS-Flp/TM3 were described previously[8]. UAS-Sec15–GFP was provided by H. Bellen. UAS-Rab and GAL4 lines were obtained from the Bloomington *Drosophila* stock centre. The UAS-*sec15* RNAi stock was obtained from the Vienna Drosophila RNAi Centre (VDRC).

Immunofluorescence on imaginal discs. Imaginal disc staining involved the following antibodies: anti-Dl (clone C594.9B-c; A. Parks), rat anti-Rab11 (R. Cohen) and rat anti-Serrate (K. Irvine). Other antibodies were obtained from the Developmental Studies Hydridoma Bank (DSHB): anti-Cut (2B10-c), anti-DECad (DCAD2), anti-α-catenin (DCAT-1), anti-β-Catenin (N2 7A1), anti-Discs Large (DLG1) and anti-NotchECD (C458.2H). *In situ* hybridization on wing discs was performed as described[29].

hBMEC, hDMEC and hMVEC-L experiments. hBMECs[30] were infected with *B. anthracis* Sterne (pXO1+, pXO2−) or isogenic mutants[22] and RNA was collected 6 h later[30] for semi-quantitative PCR and quantitative PCR. For immunofluorescence, hBMECs were transfected using 0.5 μg of DNA (rat *Sec15–GFP*[18] plasmid was provided by C. Mitchell and human *RAB11-RFP* plasmid was a gift from M. Colombo) plus 2 μl Fugene (Roche). Purified EF+PA or LF+PA (S. Leppla) were added for 24 h and fixed cells were stained using anti-pan-cadherin (pCAD) (Abcam, ab6528) or anti-Dll4 antibodies (Lifespan). For transwell assays, cells were grown on collagenized transwells (Transwell-COL) for 7 days. hDMECs (Lonza CC-2543) or hMVEC-Ls (Lonza, CC-2527) were treated with purified EF or LF toxin, fixed and stained as described earlier for hBMECs except that hMVEC-Ls were treated with EF toxin for 48 h. Vascular permeability in the skin was assessed using the Miles assay[25,26].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Mourez, M. Anthrax toxins. *Rev. Physiol. Biochem. Pharmacol.* **152**, 135–164 (2004).
2. Tournier, J. N., Quesnel-Hellmann, A., Cleret, A. & Vidal, D. R. Contribution of toxins to the pathogenesis of inhalational anthrax. *Cell. Microbiol.* **9**, 555–565 (2007).
3. Lacy, D. B. & Collier, R. J. Structure and function of anthrax toxin. *Curr. Top. Microbiol. Immunol.* **271**, 61–85 (2002).
4. Duesbery, N. S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734–737 (1998).
5. Vitale, G. *et al.* Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages. *Biochem. Biophys. Res. Commun.* **248**, 706–711 (1998).
6. Leppla, S. H. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proc. Natl Acad. Sci. USA* **79**, 3162–3166 (1982).
7. Moayeri, M. & Leppla, S. H. The roles of anthrax toxin in pathogenesis. *Curr. Opin. Microbiol.* **7**, 19–24 (2004).
8. Guichard, A., Park, J. M., Cruz-Moreno, B., Karin, M. & Bier, E. Anthrax lethal factor and edema factor act on conserved targets in *Drosophila. Proc. Natl Acad. Sci. USA* **103**, 3244–3249 (2006).
9. Pezard, C., Berche, P. & Mock, M. Contribution of individual toxin components to virulence of *Bacillus anthracis. Infect. Immun.* **59**, 3472–3477 (1991).
10. Fortini, M. E. & Bilder, D. Endocytic regulation of Notch signaling. *Curr. Opin. Genet. Dev.* **19**, 323–328 (2009).
11. Jafar-Nejad, H. *et al.* Sec15, a component of the exocyst, promotes notch signaling during the asymmetric division of *Drosophila* sensory organ precursors. *Dev. Cell* **9**, 351–363 (2005).
12. Wu, H., Rossi, G. & Brennwald, P. The ghost in the machine: small GTPases as spatial regulators of exocytosis. *Trends Cell Biol.* **18**, 397–404 (2008).
13. Zhang, J. *et al.* Thirty-one flavors of *Drosophila* rab proteins. *Genetics* **176**, 1307–1322 (2007).
14. Emery, G. *et al.* Asymmetric Rab 11 endosomes regulate delta recycling and specify cell fate in the *Drosophila* nervous system. *Cell* **122**, 763–773 (2005).
15. Langevin, J. *et al. Drosophila* exocyst components Sec5, Sec6, and Sec15 regulate DE-Cadherin trafficking from recycling endosomes to the plasma membrane. *Dev. Cell* **9**, 365–376 (2005).
16. Guo, W., Roth, D., Walch-Solimena, C. & Novick, P. The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis. *EMBO J.* **18**, 1071–1080 (1999).
17. Salminen, A. & Novick, P. J. The Sec15 protein responds to the function of the GTP binding protein, Sec4, to control vesicular traffic in yeast. *J. Cell Biol.* **109**, 1023–1036 (1989).
18. Zhang, X. M., Ellis, S., Sriratana, A., Mitchell, C. A. & Rowe, T. Sec15 is an effector for the Rab11 GTPase in mammalian cells. *J. Biol. Chem.* **279**, 43027–43034 (2004).
19. Wu, S., Mehta, S. Q., Pichaud, F., Bellen, H. J. & Quiocho, F. A. Sec15 interacts with Rab11 via a novel domain and affects Rab11 localization *in vivo. Nature Struct. Mol. Biol.* **12**, 879–885 (2005).
20. Roca, C. & Adams, R. H. Regulation of vascular morphogenesis by Notch signaling. *Genes Dev.* **21**, 2511–2524 (2007).
21. Dejana, E., Tournier-Lasserve, E. & Weinstein, B. M. The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications. *Dev. Cell* **16**, 209–221 (2009).
22. Janes, B. K. & Stibitz, S. Routine markerless gene replacement in *Bacillus anthracis. Infect. Immun.* **74**, 1949–1953 (2006).
23. Firoved, A. M. *et al. Bacillus anthracis* edema toxin causes extensive tissue lesions and rapid lethality in mice. *Am. J. Pathol.* **167**, 1309–1320 (2005).
24. Kuo, S. R. *et al.* Anthrax toxin-induced shock in rats is associated with pulmonary edema and hemorrhage. *Microb. Pathog.* **44**, 467–472 (2007).
25. Gozes, Y., Moayeri, M., Wiggins, J. F. & Leppla, S. H. Anthrax lethal toxin induces ketotifen-sensitive intradermal vascular leakage in certain inbred mice. *Infect. Immun.* **74**, 1266–1272 (2006).
26. Tessier, J. *et al.* Contributions of histamine, prostanoids, and neurokinins to edema elicited by edema toxin from *Bacillus anthracis. Infect. Immun.* **75**, 1895–1903 (2007).
27. Balzac, F. *et al.* E-cadherin endocytosis regulates the activity of Rap1: a traffic light GTPase at the crossroads between cadherin and integrin function. *J. Cell Sci.* **118**, 4765–4783 (2005).
28. Silvis, M. R. *et al.* Rab11b regulates the apical recycling of the cystic fibrosis transmembrane conductance regulator in polarized intestinal epithelial cells. *Mol. Biol. Cell* **20**, 2337–2350 (2009).
29. Kosman, D. *et al.* Multiplex detection of RNA expression in *Drosophila* embryos. *Science* **305**, 846 (2004).
30. van Sorge, N. M. *et al.* Anthrax toxins inhibit neutrophil signaling pathways in brain endothelium and contribute to the pathogenesis of meningitis. *PLoS ONE* **3**, e2964 (2008).

METHODS

***Drosophila* genetics.** Transgenic lines UAS-LF2X/FM7, UAS-LF3X/FM7 and UAS-EF UAS-Flp/TM3 were described previously[8]. UAS-Sec15–GFP was provided by H. Bellen. Wild-type UAS-Rab11 and UAS-Rab11(DN), as well as all UAS-Rab transgenic lines, generated by H. Bellen and M. Scott, were obtained from the Bloomington *Drosophila* stock centre. The UAS-*sec15* RNAi stock is from VDRC (#35161). GAL4 drivers included: *stgG4*, *MS1096-GAL4*; *wkG4*, *1348-GAL4*; *L2G4*, *E-GAL4*; *vgG4*, *dppG4* and *brkG4* have been described previously[31,32] and are available from the Bloomington *Drosophila* stock or can be obtained on request.

Immunofluorescence on imaginal discs. Immunostaining of imaginal discs was performed according to standard protocols, using the following antibodies: anti-Dl (clone C594.9B-c was provided by A. Parks and used at 1:1,000), anti-*Drosophila* Rab11 antibody (a gift from R. Cohen, used at 1:500), anti-Serrate antibodies (provided by K. Irvine, used at 1:1,000). Other antibodies were obtained from DSHB: anti-Cut (clone 2B10-c, 1:100), anti-DECad (DCAD2, 1:500), anti-α-Catenin (D-CAT1, 1:20), anti-β-Catenin (Armadillo, clone N2 7A1, 1:20), anti-Discs Large (DLG1, 1:20) and anti-NotchEC (C458.2H, 1:500). *In situ* hybridization on wing discs was performed using a digoxigenin-labelled *wg* antisense probe as described previously for fluorescent detection[29] and for histochemical staining[33].

hBMEC, hDMEC and hMVEC-L experiments. hBMECs[30] were infected with *B. anthracis* Sterne (pXO1$^+$, pXO2$^-$) or isogenic mutants ΔpXO1, ΔLF, ΔEF, or ΔLF/ΔEF[22] and RNA was collected 6 h later as described previously[30]. Semi-quantitative PCR was performed using 26 cycles and qPCR was performed using iQ SYBR Green supermix (BioRad). For immunofluorescence, hBMECs were treated with indicated amounts of purified EF+PA or LF+PA (provided by S. Leppla) for 24 h and stained using mouse anti-pan-cadherin antibody (pCAD) (Abcam, ab6528, 1:100) or anti-DLL4 antibodies (Lifespan, LS-C19035) and appropriate secondary antibodies. hBMECs were transfected with *Sec15–GFP* plasmid (provided by C. Mitchell) and/or *RAB11–RFP* plasmid (a gift from M. Colombo) using 0.5 μg of DNA and 2 μl Fugene (Roche). Forty hours after transfection, purified EF+PA or LF+PA were added to the wells and cells were fixed and stained 24 h later. For transwell assays, cells were seeded on collagenized transwells (Transwell-COL; Corning-Costar) and grown for 7 days at 37 °C with 5% CO$_2$. Cells were infected for 6 h with bacteria or treated with indicated amounts of EF toxin for 24 h. 0.4% Evans blue was added to the upper chamber and leakage was quantified by measuring the colour change in the bottom chamber containing Hank's buffered salt solution (HBSS) at 620 nm. For DBZ treatment, hBMECs were treated with a final concentration of 2 μM DBZ or vehicle control (DMSO) for 6 h. For transwell assays, cells were seeded on collagenized transwells (Transwell-COL) and grown for 7 days. Cells were infected for 6 h with bacteria or treated with indicated amounts of EF toxin for 24 h. 0.4% Evans blue was added to the upper chamber and leakage was measured at 620 nm. hDMECs (Lonza CC-2543) or hMVEC-Ls (Lonza CC-2527) were treated with purified EF or LF toxin, fixed and stained as described earlier for hBMECs except that hMVEC-Ls were treated with EF toxin for 48 h.

Vascular permeability assay in mice. Vascular permeability in the skin was assessed using the Miles assay[25,26]. Nine-week-old CD-1 female mice were injected with 100 μl PBS containing 1×10^6 colony-forming units (c.f.u.) of *B. anthracis* Sterne, ΔLF and ΔEF bacteria subcutaneously in the hind flank (three spots per mouse). After 6 h, mice were injected intravenously with 0.1 ml of 2% Evans blue in PBS and 30 min later mice were killed, skins were inverted and examined. For leakage quantification, the site of injection was excised and placed in formamide at 65 °C (24 h) for dye extraction (absorbance 620 nm). Statistical significance was assessed using one-way ANOVA.

31. Cook, O., Biehs, B. & Bier, E. *brinker* and *optomotor-blind* act coordinately to initiate development of the L5 wing vein primordium in *Drosophila*. *Development* **131,** 2113–2124 (2004).
32. Lunde, K. *et al.* Activation of the *knirps* locus links patterning to morphogenesis of the second wing vein in *Drosophila*. *Development* **130,** 235–248 (2003).
33. O'Neill, J. W. & Bier, E. Double-label *in situ* hybridization using biotin and digoxigenin-tagged RNA probes. *Biotechniques* **17,** 874–875 (1994).

LETTER

Non–muscle myosin IIA is a functional entry receptor for herpes simplex virus–1

Jun Arii[1,2], Hideo Goto[3], Tadahiro Suenaga[4], Masaaki Oyama[5], Hiroko Kozuka-Hata[5], Takahiko Imai[1], Atsuko Minowa[1], Hiroomi Akashi[2], Hisashi Arase[4,6,7], Yoshihiro Kawaoka[3,8,9,10] & Yasushi Kawaguchi[1]

Herpes simplex virus-1 (HSV-1), the prototype of the α-herpesvirus family, causes life-long infections in humans. Although generally associated with various mucocutaneous diseases, HSV-1 is also involved in lethal encephalitis[1]. HSV-1 entry into host cells requires cellular receptors for both envelope glycoproteins B (gB) and D (gD)[2-4]. However, the gB receptors responsible for its broad host range *in vitro* and infection of critical targets *in vivo*[1] remain unknown. Here we show that non-muscle myosin heavy chain IIA (NMHC-IIA), a subunit of non-muscle myosin IIA (NM-IIA), functions as an HSV-1 entry receptor by interacting with gB. A cell line that is relatively resistant to HSV-1 infection[5] became highly susceptible to infection by this virus when NMHC-IIA was over-expressed. Antibody to NMHC-IIA blocked HSV-1 infection in naturally permissive target cells. Furthermore, knockdown of NMHC-IIA in the permissive cells inhibited HSV-1 infection as well as cell–cell fusion when gB, gD, gH and gL were coexpressed. Cell-surface expression of NMHC-IIA was markedly and rapidly induced during the initiation of HSV-1 entry. A specific inhibitor of myosin light chain kinase, which regulates NM-IIA by phosphorylation[6], reduced the redistribution of NMHC-IIA as well as HSV-1 infection in cell culture and in a murine model for herpes stromal keratitis. NMHC-IIA is ubiquitously expressed in various human tissues and cell types[7] and, therefore, is implicated as a functional gB receptor that mediates broad HSV-1 infectivity both *in vitro* and *in vivo*. The identification of NMHC-IIA as an HSV-1 entry receptor and the involvement of NM-IIA regulation in HSV-1 infection provide an insight into HSV-1 entry and identify new targets for antiviral drug development.

To identify new HSV-1 entry receptor(s), we adopted a tandem affinity-purification approach with a membrane-impermeable cross-linker coupled with mass-spectrometry-based proteomics technology[8] (Supplementary Figs 1 and 2). This approach identified NMHC-IIA as a potential gB receptor (Supplementary Fig. 2). To verify the specific association of HSV-1 gB with NMHC-IIA, co-immunoprecipitation studies were performed. In Vero cells exposed to HSV-1 expressing MEF–gB (a Myc-TEV-Flag tag fused to gB) or wild-type gB, NMHC-IIA co-precipitated with MEF–gB or gB (Fig. 1a and Supplementary Fig. 3a). However, NMHC-IIA did not co-precipitate with gH in Vero cells exposed to HSV-1 expressing MEF–gH (Fig. 1a). Furthermore, the rod domain of NMHC-IIA fused to glutathione S-transferase (GST) pulled down gB from lysates of cells infected with wild-type HSV-1, whereas GST alone did not (Supplementary Fig. 3b). To examine whether gB interacts with NMHC-IIA on the cell surface, cell surface proteins of Vero cells exposed to HSV-1 expressing MEF–gB were bio-tinylated and the immunoprecipitated complex of NMHC-IIA and MEF–gB was then detected with streptavidin (Fig. 1b). NMHC-IIA that

co-precipitated with MEF–gB was biotinylated (Fig. 1b), whereas no biotinylated protein with a molecular mass similar to NMHC-IIA was detectable in cells exposed to HSV-1 expressing MEF–gH (Fig. 1b). To demonstrate further the gB–NMHC-IIA interaction on the cell surface, 293T cells transfected with a plasmid expressing HSV-1 gB or gD were analysed by flow cytometry with the rod domain of NMHC-IIA fused to the immunoglobulin G (IgG) Fc fragment (NMHC-IIA–Fc). Cells
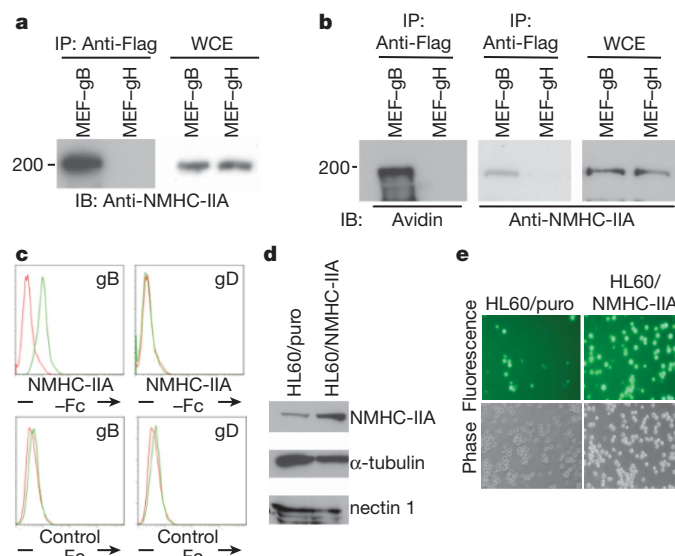

Figure 1 | NMHC-IIA associates with HSV-1 gB and mediates HSV-1 infection. **a**, Vero cells exposed to HSV-1 expressing MEF–gB (MEF–gB) or MEF–gH (MEF–gH) at a multiplicity of infection (MOI) of 50 at 4 °C for 2 h to allow for viral adsorption, followed by a temperature shift to 37 °C (a permissive temperature for viral penetration) for 2 min, were immunoprecipitated (IP) with an anti-Flag antibody and analysed by immunoblotting (IB) with an anti-NMHC-IIA antibody. WCE, whole cell extract control. **b**, Cell surface proteins of Vero cells exposed to HSV-1 expressing MEF–gB or MEF–gH at 4 °C for 2 h, followed by a temperature shift to 37 °C for 15 min, were biotinylated, immunoprecipitated with an anti-Flag antibody, and analysed by immunoblotting with streptavidin or an anti-NMHC-IIA antibody. Molecular weights (10^3) are shown on the left in **a** and **b**. **c**, Mock-transfected 293T cells (red line) and 293T cells transfected with the gB or gD expression vector (green line) were stained with NMHC-IIA–Fc or control–Fc at 18 h after transfection and analysed by flow cytometry. **d**, Expression of NMHC-IIA, α-tubulin, or nectin 1 in HL60/puro and HL60/NMHC-IIA cells was analysed by immunoblotting. **e**, HL60/puro and HL60/NMHC-IIA cells were infected with HSV-1(GFP) at an MOI of 1 and analysed by fluorescence and phase microscopy at 12 h after infection.

[1]Division of Viral Infection, Department of Infectious Disease Control, International Research Center for Infectious Diseases, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [2]Department of Veterinary Microbiology, Graduate School of Agricultural and Life Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan. [3]Division of Virology, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [4]Department of Immunochemistry, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan. [5]Medical Proteomics Laboratory, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [6]WPI Immunology Frontier Research Center, Osaka University, Suita, Osaka 565-0871, Japan. [7]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan. [8]Department of Special Pathogens, International Research Center for Infectious Diseases, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [9]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53711, USA. [10]ERATO Infection-Induced Host Responses Project, Saitama 332-0012, Japan.

transfected with the gB expression plasmid, but not the gD expression plasmid, reacted with NMHC-IIA–Fc (Fig. 1c). Furthermore, NMHC-IIA–Fc staining was detected in cells infected with wild-type HSV-1 or a revertant gB-deficient HSV-1 in which the gB deficiency was restored, but not in those infected with an HSV-1 gB-deficient virus (Supplementary Fig. 3c). No transfected or infected cells stained with control-Fc (Fig. 1c and Supplementary Fig. 3c). Although we were unable to show binding of purified soluble gB to NMHC-IIA on the cell surface because gB also binds to glycosaminoglycans, which are abundant on the cell surface[2] (data not shown), the series of findings described earlier prove a specific interaction between HSV-1 gB and NMHC-IIA.

To examine whether NMHC-IIA is involved in HSV-1 entry, we established human promyelocytic HL60 cells stably expressing high levels of NMHC-IIA (HL60/NMHC-IIA cells) (Fig. 1d); HL60 cells express NMHC-IIA at a low level[9] and are relatively resistant to HSV-1 infection[5]. Infection of HL60/NMHC-IIA cells with HSV-1 expressing a GFP marker (HSV-1(GFP)) resulted in a remarkable increase in the percentage of virus-infected cells compared to control HL60/puro cells (Fig. 1e, Supplementary Fig. 4a and Supplementary Discussion 1). Infection of HL60/NMHC-IIA cells with HSV-1(GFP) was blocked by anti-NMHC-IIA serum in a dose-dependent manner, whereas control serum had little effect on infection (Supplementary Fig. 4b). Overexpression of NMHC-IIA also enhanced susceptibility of HL60 cells to infection by pseudorabies virus (a porcine α-herpesvirus) expressing a GFP marker, and this infection was inhibited by anti-NMHC-IIA serum (Supplementary Fig. 5). These observations indicate that NMHCII-A mediates HSV-1 and pseudorabies virus infections and suggest that NMHC-IIA is commonly used by α-herpesviruses for their entry.

To verify further the role of endogenously expressed NMHC-IIA in HSV-1 infection, we analysed the effect of anti-NMHC-IIA serum on HSV-1 infection in Vero cells that express NMHC-IIA endogenously. Anti-NMHC-IIA serum inhibited infection of Vero cells with HSV-1(GFP) in a dose-dependent manner, whereas control serum had little effect on infection (Fig. 2a and Supplementary Fig. 6a). By contrast, anti-NMHC-IIA serum had no effect on influenza virus infection (Fig. 2b). Inhibition of HSV-1 infection by anti-NMHC-IIA serum was also observed in human epithelial cells, HaCaT, HCE-T and NCI-H292 cells (Supplementary Fig. 7) and in CHO-K1 cells overexpressing the gD receptor nectin 1 (Supplementary Fig. 8a) but not in CHO-K1 cells overexpressing the gB receptor PILRα (Supplementary Fig. 8b).

To examine the role of NMHC-IIA in HSV-1-mediated membrane fusion, we tested first whether the inhibition of HSV-1 infection by the anti-NMHC-IIA antibody was restored by treatment with polyethylene glycol (PEG), which can chemically induce fusion by dehydrating the surfaces of juxtaposed membranes and can therefore promote the entry of entry-defective or entry-blocked herpesviruses when adsorbed onto the cell surface[10]. PEG treatment significantly increased the number of virus-infected cells in the presence of the anti-NMHC-IIA antibody compared to cells that were not treated with PEG (Supplementary Fig. 6b). Second, we used a cell–cell fusion assay, which allows the measurement of HSV-1-induced membrane fusion, relying on transient expression of HSV-1 glycoproteins[11]. When NMHC-IIA knockdown Vero cells (that is, cells stably expressing short hairpin RNA against NMHC-IIA) were co-cultured with Vero cells transiently expressing HSV-1 gB, gD, gH and gL, the efficiency of cell–cell fusion significantly decreased compared to cell–cell fusion when Vero cells stably expressing control shRNA were co-cultured with Vero cells expressing the viral glycoproteins (Fig. 2c, d). In contrast, NMHC-IIA knockdown had little effect on vesicular stomatitis virus envelope-G-protein-mediated cell–cell fusion (Fig. 2e). NMHC-IIA knockdown also resulted in a significant decrease in HSV-1 infection, but had no effect on influenza virus infection (Fig. 2f, g). These results indicate that HSV-1 uses NMHC-IIA as a functional cellular receptor in naturally permissive cells that express NMHC-IIA endogenously and that
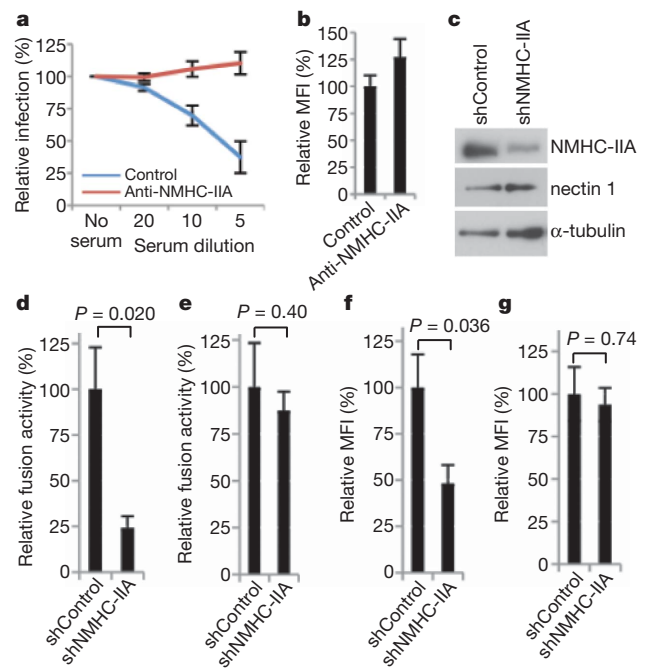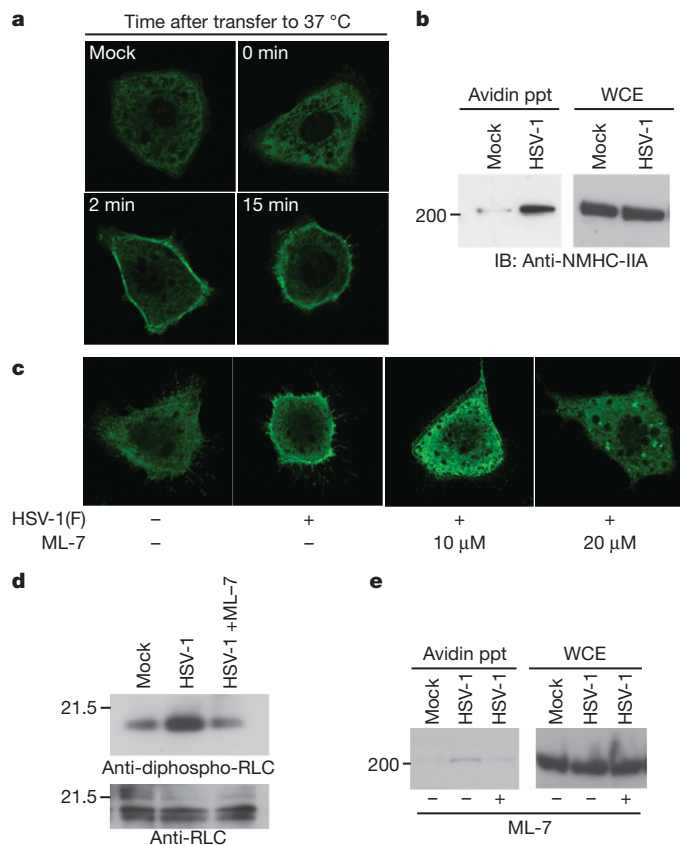

Figure 2 | Endogenous NMHC-IIA mediates HSV-1 infection and cell–cell fusion. a, b, Vero cells were infected with HSV-1(GFP) (**a**) or influenza virus (**b**) at an MOI of 1 in the presence of various concentrations of anti-NMHC-IIA or control serum (**a**), or in the presence of anti-NMHC-IIA or control serum at a 1:5 dilution (**b**), and the percentage of infected cells (**a**) or mean fluorescence intensities (MFIs) (**b**) was determined by flow cytometry at 6 h and 7 h after infection, respectively. Relative infection and relative MFI represent the percentage of infected cells in antiserum-treated cultures compared with untreated controls (**a**) and the MFI of infected cells in antiserum-treated cultures compared with untreated controls (**b**), respectively. Data are shown as means and s.e.m. (*n* = 3). **c**, NMHC-IIA, nectin-1 and α-tubulin expression in Vero-shControl and Vero-shNMHC-IIA cells were determined by immunoblotting. **d, e**, Vero cells transfected with expression vectors for T7 polymerase, gB, gD, gH, and gL (**d**), or expression vectors for T7 polymerase and vesicular stomatitis virus envelope G protein (**e**), were co-cultured with Vero-shControl or Vero-shNMHC-IIA cells transfected with a reporter plasmid encoding the luciferase gene driven by the T7 polymerase. Data are shown as means and s.e.m. (*n* = 6; two-tailed Student's *t*-test). The mean value for Vero-shControl was normalized to 100% relative fusion activity. **f, g**, Vero-shControl and Vero-shNMHC-IIA cells infected with HSV-1(GFP) (**f**) or influenza virus (**g**) at an MOI of 1 were analysed by flow cytometry and MFIs were determined at 5 h and 7 h after infection, respectively. Data are shown as means and s.e.m. (*n* = 6; two-tailed Student's *t*-test). The mean value for Vero-shControl was normalized to 100% relative MFI.

NMHC-IIA is required for efficient cell–cell fusion mediated by HSV-1 envelope glycoproteins.

It is known that NMHC-IIA mainly functions in the cytoplasm but not on the cell surface[7]. Yet here we demonstrated that NMHC-IIA associated with gB on the cell surface upon viral entry and functioned as an HSV-1 entry receptor. This discrepancy prompted us to examine whether NMHC-IIA is redistributed during HSV-1 entry. In Vero cells mock-incubated at 37 °C or exposed to HSV-1 at 4 °C to allow for viral adsorption, NMHC-IIA localized throughout the cytoplasm (Fig. 3a). However, at 2 and 15 min after the temperature shift to 37 °C (a permissive temperature for viral penetration), marked enrichment of NMHC-IIA at the plasma membrane was observed (Fig. 3a). Biotinylation of cell surface proteins of Vero cells mock-incubated or exposed to wild-type HSV-1 followed by precipitation with avidin beads and immunoblotting with an anti-NMHC-IIA antibody confirmed the cell-surface expression of NMHC-IIA in normal Vero cells and upregulation of the cell-surface expression of NMHC-IIA at 15 min after the temperature shift to 37 °C (Fig. 3b). These results indicate that the cell-surface expression of NMHC-IIA is upregulated after viral adsorption at 4 °C followed by a

Figure 3 | Cell-surface expression of NMHC-IIA is upregulated after viral adsorption at 4 °C, followed by a temperature shift to 37 °C. a, Vero cells exposed to wild-type HSV-1 at an MOI of 50 at 4 °C for 2 h, followed by a temperature shift to 37 °C for 0, 2 and 15 min, were fixed, stained with an anti-NMHC-IIA antibody, and analysed by confocal microscopy. **b**, Cell surface proteins of Vero cells mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 at 4 °C for 2 h, followed by a temperature shift to 37 °C for 15 min, were biotinylated, precipitated with avidin beads (avidin ppt) and analysed by immunoblotting with anti-NMHC-IIA antibody. **c**, Vero cells mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 at 4 °C for 2 h, followed by a temperature shift to 37 °C for 15 min, in the presence or absence of the indicated concentrations of ML-7 were fixed, stained with an anti-NMHC-IIA antibody, and analysed by confocal microscopy. **d**, Vero cells were mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 at 4 °C for 2 h, followed by a temperature shift to 37 °C for 15 min, in the presence or absence of 20 μM ML-7. Expression of total or diphosphorylated regulatory light chain (RLC) was determined by immunoblotting. **e**, Cell surface proteins of Vero cells mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 at 4 °C for 2 h, followed by a temperature shift to 37 °C for 15 min, in the presence or absence of 20 μM ML-7 were biotinylated, precipitated with avidin beads, and analysed by immunoblotting with an anti-NMHC-IIA antibody. Molecular weights (10^3) are shown on the left in **b**, **d** and **e**.

temperature shift to 37 °C, a finding that has not been previously reported for other viruses.

Subcellular localization of NM-IIA is regulated in part by phosphorylation of regulatory light chains at the threonine residue at position 18 (Thr 18) and the serine residue at position 19 (Ser 19). One of the protein kinases responsible for the NM-IIA regulatory light chain phosphorylation at these sites is Ca^{2+}-calmodulin-dependent myosin light chain kinase[12]. In fact, diphosphorylation of regulatory light chains at Thr 18 and Ser 19 was enhanced greatly after viral adsorption at 4 °C followed by a temperature shift to 37 °C (Fig. 3d). Therefore, we examined the effects of ML-7—a specific inhibitor of myosin light chain kinase[13]—on the phosphorylation of regulatory light chains at Thr 18 and Ser 19, the redistribution of NMHC-IIA and HSV-1 infection. ML-7 inhibited phosphorylation of regulatory light chains at Thr 18 and Ser 19 (Fig. 3d) as well as enrichment of NMHC-IIA at the plasma membrane and upregulation of cell-surface expression of

NMHC-IIA after viral adsorption at 4 °C followed by a temperature shift to 37 °C (Fig. 3c, e). Vero-cell susceptibility to HSV-1(GFP) infection was inhibited by ML-7 in a dose-dependent manner (Fig. 4a), whereas susceptibility to influenza virus infection was not affected by ML-7 (Fig. 4b). ML-7 seemed to inhibit HSV-1 infection at the level of virus entry, based on the observation that PEG treatment significantly enhanced susceptibility to HSV-1(GFP) in the presence of ML-7 (Supplementary Fig. 9). Furthermore, effects on HSV-1 infection similar to those seen with ML-7 were also observed when cells were treated with BAPTA-AM, a cell-permeable cytosolic Ca^{2+} chelator[14] (Supplemenxtary Fig. 10) or transfected with a plasmid expressing a dominant-negative mutant of myosin light chain kinase (Supplementary Fig. 11), indicating that ML-7 in fact inhibited HSV-1 infection by blocking the myosin light chain kinase. In the murine model for herpes stromal keratitis, treatment of mouse eyes with ML-7 before HSV-1 inoculation significantly reduced viral replication in the eyes and herpes stromal keratitis severity, and the survival curve of mice treated with ML-7 was significantly greater than that for mock-treated mice (Fig. 4c–e). These results indicate that ML-7 efficiently inhibits HSV-1 infection in both cultured cells and the murine model, and suggest that regulation of NM-IIA, including NMHC-IIA redistribution to the cell surface during initiation of HSV-1 penetration, is required for efficient HSV-1 infection *in vitro* and *in vivo* (Supplementary Discussion 2 and 3).

NMHC-IIA is a subunit of NM-IIA, and NM-IIA is central to the control of cell adhesion, cell migration and tissue architecture[7]. As expected, because NM-IIA has essential roles in basic cellular functions,
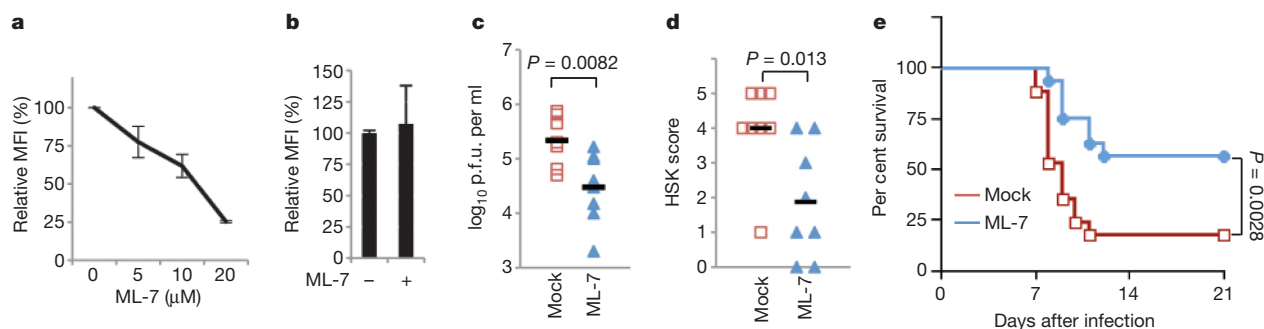

Figure 4 | ML-7 inhibits HSV-1 infection *in vitro* and *in vivo*. a, b, Vero cells infected with HSV-1(GFP) or influenza virus at an MOI of 1 in the presence of the indicated concentrations of ML-7 (**a**) or in the presence of 20 μM ML-7 (**b**) were analysed by flow cytometry at 5 h and 7 h after infection, respectively, and MFIs were determined. Data are shown as means and s.e.m. ($n = 3$). The data were normalized to the value in the absence of ML-7 (**a**) and the mean value in the absence of ML-7 was normalized to 100% relative MFI

(**b**), respectively. **c–e**, Mice infected with wild-type HSV-1 in the absence or presence of ML-7 were scored for herpes stromal keratitis (HSK) on day 5 after infection (**d**, $n = 8$; two-tailed Student's t-test) and monitored for mortality daily for 21 days (**e**, $n = 17$; Log-rank test). Viral titres in the tear films of infected mice at day 2 after infection were determined (**c**, $n = 8$; two-tailed Student's t-test). The bars in **c** and **d** show the means of the data.

NMHC-IIA is expressed in a wide variety of cultured cell lines and in various tissues and cell types *in vivo*[7,15]. Furthermore, NMHC-IIA is expressed on the cell surface[16] (Fig. 3b) and is enriched in lipid rafts[17,18], with which HSV-1 gB associates during viral entry[19]. These reports together with the data here presented implicate NMHC-IIA as a functional gB receptor enabling HSV-1 to infect various cell lines and important *in vivo* cellular targets. Future clarification of the precise roles of NMHC-IIA and its regulators in HSV-1 infection will define the molecular basis of HSV-1 entry. These results may indicate new prophylactic and therapeutic approaches for the development of antiherpetic drugs that target NM-IIA and NM-IIA regulators.

METHODS SUMMARY

Identification of proteins that interact with gB during viral entry. Mouse embryonic fibroblasts immortalized by simian virus 40 large T antigen[20] and IC21 cells were infected with HSV-1 expressing MEF–gB at 4 °C for 2 h, then transferred to 37 °C for 2 min, harvested, treated with PBS containing 2 mM 3,3′-dithiobis(sulphosuccinimidyl propionate) (DTSSP; Piers) at 4 °C for 2 h, and lysed in radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 10 mM Tris-HCl (pH 7.4) and 1 mM EDTA). After centrifugation, the supernatants were immunoprecipitated with an anti-Myc monoclonal antibody (LP14; MBL) and the immunoprecipitates were reacted with AcTEV protease (Invitrogen). After another centrifugation, the supernatants were immunoprecipitated with an anti-Flag monoclonal antibody (M2; Sigma), and these immunoprecipitates were electrophoretically separated in a denaturing gel and visualized by silver staining. Bands detected only in immunoprecipitates from mouse embryonic fibroblasts were excised and digested in the gel with trypsin as described previously[8]. The purified peptides were then loaded on an automated nanoflow liquid chromatograph (Dina) and tandem mass spectrometer (LC-MS/MS; Q-Star Elite, Applied Biosystems). The peptide masses obtained by LC-MS/MS analysis were processed against mouse protein sequences in the National Center for Biotechnology Information (NCBI) RefSeq database (35,662 sequences as of 14 July, 2009) using the Mascot algorithm (Matrix Science).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Roizman, B., Knipe, D. M. & Whitley, R. J. in *Fields Virology* (eds Knipe, D. M. *et al.*) 5th edn (Lippincott-Williams & Wilkins, 2007).
2. Spear, P. G. *et al.* Different receptors binding to distinct interfaces on herpes simplex virus gD can trigger events leading to cell fusion and viral entry. *Virology* **344,** 17–24 (2006).
3. Satoh, T. *et al.* PILRα is a herpes simplex virus-1 entry coreceptor that associates with glycoprotein B. *Cell* **132,** 935–944 (2008).
4. Suenaga, T. *et al.* Myelin-associated glycoprotein mediates membrane fusion and entry of neurotropic herpesviruses. *Proc. Natl Acad. Sci. USA* **107,** 866–871 (2010).
5. Pientong, C., Weisshart, K., Kuhn, J. E., Knopf, C. W. & Braun, R. W. Replication of herpes simplex virus type 1 in differentiated human promyelocytic HL-60 cells. *Virology* **170,** 468–476 (1989).
6. Gallagher, P. J., Herring, B. P. & Stull, J. T. Myosin light chain kinases. *J. Muscle Res. Cell Motil.* **18,** 1–16 (1997).
7. Vicente-Manzanares, M., Ma, X., Adelstein, R. S. & Horwitz, A. R. Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nature Rev. Mol. Cell Biol.* **10,** 778–790 (2009).
8. Oyama, M. *et al.* Temporal perturbation of tyrosine phosphoproteome dynamics reveals the system-wide regulatory networks. *Mol. Cell. Proteomics* **8,** 226–231 (2009).
9. Toothaker, L. E. *et al.* Cellular myosin heavy chain in human leukocytes: isolation of 5′ cDNA clones, characterization of the protein, chromosomal localization, and upregulation during myeloid differentiation. *Blood* **78,** 1826–1833 (1991).
10. Sarmiento, M., Haffey, M. & Spear, P. G. Membrane proteins specified by herpes simplex viruses. III. Role of glycoprotein VP7(B2) in virion infectivity. *J. Virol.* **29,** 1149–1158 (1979).
11. Turner, A., Bruun, B., Minson, T. & Browne, H. Glycoproteins gB, gD, and gHgL of herpes simplex virus type 1 are necessary and sufficient to mediate membrane fusion in a Cos cell transfection system. *J. Virol.* **72,** 873–875 (1998).
12. Watanabe, T., Hosoya, H. & Yonemura, S. Regulation of myosin II dynamics by phosphorylation and dephosphorylation of its light chain in epithelial cells. *Mol. Biol. Cell* **18,** 605–616 (2007).
13. Saitoh, M., Ishikawa, T., Matsushima, S., Naka, M. & Hidaka, H. Selective inhibition of catalytic activity of smooth muscle myosin light chain kinase. *J. Biol. Chem.* **262,** 7796–7801 (1987).
14. Cheshenko, N. *et al.* Herpes simplex virus triggers activation of calcium-signaling pathways. *J. Cell Biol.* **163,** 283–293 (2003).
15. Golomb, E. *et al.* Identification and characterization of nonmuscle myosin II-C, a new member of the myosin II family. *J. Biol. Chem.* **279,** 2800–2808 (2004).
16. Olden, K., Willingham, M. & Pastan, I. Cell surface myosin in cultured fibroblasts. *Cell* **8,** 383–390 (1976).
17. Nebl, T. *et al.* Proteomic analysis of a detergent-resistant membrane skeleton from neutrophil plasma membranes. *J. Biol. Chem.* **277,** 43399–43409 (2002).
18. Li, D., Miller, M. & Chantler, P. D. Association of a cellular myosin II with anionic phospholipids and the neuronal plasma membrane. *Proc. Natl Acad. Sci. USA* **91,** 853–857 (1994).
19. Bender, F. C. *et al.* Specific association of glycoprotein B with lipid rafts during herpes simplex virus entry. *J. Virol.* **77,** 9542–9552 (2003).
20. Kuma, A. *et al.* The role of autophagy during the early neonatal starvation period. *Nature* **432,** 1032–1036 (2004).

METHODS

Identification of proteins that interact with gB during viral entry. Mouse embryonic fibroblasts immortalized by simian virus 40 large T antigen[20] and IC21 cells were infected with HSV-1 expressing MEF–gB at 4 °C for 2 h, then transferred to 37 °C for 2 min, harvested, treated with PBS containing 2 mM 3,3′-dithiobis(sulphosuccinimidyl propionate) (DTSSP; Piers) at 4 °C for 2 h, and lysed in radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 10 mM Tris-HCl (pH 7.4) and 1 mM EDTA). After centrifugation, the supernatants were immunoprecipitated with an anti-Myc monoclonal antibody (LP14; MBL) and the immunoprecipitates were reacted with AcTEV protease (Invitrogen). After another centrifugation, the supernatants were immunoprecipitated with an anti-Flag monoclonal antibody (M2; Sigma), and these immunoprecipitates were electrophoretically separated in a denaturing gel and visualized by silver staining. Bands detected only in immunoprecipitates from mouse embryonic fibroblasts were excised and digested in the gel with trypsin as described previously[8]. The purified peptides were then loaded on an automated nanoflow liquid chromatograph (Dina) and tandem mass spectrometer (LC-MS/MS; Q-Star Elite, Applied Biosystems). The peptide masses obtained by LC-MS/MS analysis were processed against mouse protein sequences in the National Center for Biotechnology Information (NCBI) RefSeq database (35,662 sequences as of 14 July, 2009) using the Mascot algorithm (Matrix Science).

Cells and viral infection. CHO-PILRα and CHO-nectin-1 cells are CHO-K1 cells stably expressing human PILRα and human nectin 1, respectively[21]. Medium 199, Ham's F-12 medium, and RPMI1640 medium supplemented with 1% FCS were used for virus infection of Vero, CHO and HL60 cells, respectively.

Plasmids. For generating a fusion protein of GST and the rod domain of NMHC-IIA (GST–NMHC-IIA), a plasmid (pGEX-NMHC-IIA) was constructed by amplifying the sequence containing NMHC-IIA codons 1665–1960 by PCR from pEGFP-ARF296[22] and cloning the DNA fragment into pGEX-4T3 (GE Healthcare) in-frame with GST. For generating a fusion protein of Fc and the rod domain of NMHC-IIA (NMHC-IIA–Fc), which was used for detection of cell surface gB, a plasmid (pME-Fc-NMHCIIA) was constructed by the procedure used to generate pGEX-NMHC-IIA. However, instead of using pGEX-4T-3, we used a modified pME18S expression vector containing a mouse CD150 leader segment at the amino terminus and the Fc fragment of human IgG1 at the carboxy terminus, in which the leucines at the Fc codons 266 and 267 were mutated to alanine and glutamine, respectively, to reduce the binding affinity to cellular Fc receptors, and the histidine at codon 467 was mutated to arginine to reduce the binding affinity to HSV-1 gE, an Fc ligand[3]. The NMHC-IIA open reading frame from Addgene plasmid 11347 (Addgene) was cloned into pMxs-puro[23] and designated pMxs-NMHC-IIA-puro. Plasmids pPEP98-gB, pPEP99-gD, pPEP101-gL and pPEP100-gH were used for expression of HSV-1 gB, gD, gL and gH, respectively, as described previously[24]. Plasmids pCAGT7, encoding T7 RNA polymerase, and pT7EMCLuc, carrying the firefly luciferase gene under the control of the T7 promoter, were used to determine fusion efficiency as described previously[25]. Plasmid pSSSP-NMHC-IIA, used to generate a stable cell line expressing shRNA against NMHC-IIA, was constructed as follows. Oligonucleotides 5′- TTTGACCA GAACTGCAAGCTGGCCGCTTCCTGTCACGGCCAGCTTGCAGTTCTGGTC TTTTTTG-3′ and 5′-AATTCAAAAAAGACCAGAACTGCAAGCTGGCCGT GACAGGAAGCGGCCAGCTTGCAGTTCTGGT-3′ were annealed and cloned into the BbsI and EcoRI sites of pmU6[26]. The BamHI-EcoRI fragment of the resultant plasmid, containing the U6 promoter and the sequence encoding shRNA against NMHC-IIA, was cloned into the BamHI and EcoRI sites of pSSSP[26], which is a derivative of the retrovirus vector pMX containing a puromycin resistance gene, to produce pSSSP-NMHC-IIA. Plasmid pSSSP-Cre encoding shRNA against Cre recombinase was described previously[26]. Plasmid Mut-MLCK was used for the expression of the dominant-negative mutant of myosin light chain kinase[27]. pCl-neo (Promega) was used as a control for plasmid Mut-MLCK.

Viruses. Wild-type HSV-1 (strain F), influenza virus (A/WSN/33; H1N1), gB-deficient virus and the revertant virus, in which the gB sequence deleted from the gB-deficient virus was restored, were described previously[3,21,28]. The gB-deficient virus was propagated in a gB-expressing cell line as described previously[3]. An HSV-1 recombinant virus carrying an enhanced green fluorescent protein (EGFP) expression cassette under the control of an Egr-1 promoter in the intergenic region between UL3 and UL4 genes (HSV-1(GFP)), was described previously[21]. A pseudorabies recombinant virus carrying EGFP under the control of a human cytomegalovirus promoter in the gG locus was described previously[29]. These fluorescent viruses grow as well as wild-type viruses in cell culture and only cells infected with them express EGFP.

Generation of recombinant viruses. To generate recombinant viruses expressing MEF–gB and MEF–gH, a two-step Red-mediated mutagenesis procedure was carried out as described previously[30].

Antibodies. Mouse monoclonal antibodies to gB (1105), Flag (M2) and Myc (PL14) were purchased from the Goodwin Institute, Sigma and MBL, respectively. Rabbit polyclonal antibody to the C-terminal peptide of NMHC-IIA was purchased from Sigma. Rabbit polyclonal antibodies to regulatory light chain and to phosphorylated regulatory light chain at Thr 18 and Ser 19 were purchased from Cell Signaling. To generate a rabbit polyclonal antibody to the rod domain of NMHC-IIA, a rabbit was immunized, according to a standard protocol at MBL, with GST–NMHC-IIA that had been expressed in *Escherichia coli* and purified as described previously[31]. Serum from the immunized rabbit was used as an anti-NMHC-IIA rod polyclonal antibody. Serum from a non-immunized rabbit at MBL was used as the control rabbit serum. Anti-influenza virus antibody R309 was described previously[28].

Immunoprecipitation. Vero cells were exposed to wild-type HSV-1 or HSV-1 expressing MEF–gB or MEF–gH at a multiplicity of infection (MOI) of 50 at 4 °C for 2 h. The cells were then transferred to 37 °C for 2 min, harvested, washed with PBS, and lysed in TNE buffer (1% NP-40, 150 mM NaCl, 10 mM Tris-HCl (pH 7.8) and 1 mM EDTA) containing a proteinase inhibitor cocktail. After centrifugation, the supernatants were pre-cleared by incubation with protein-A-sepharose beads for 30 min at 4 °C. After a brief centrifugation, the supernatants were reacted with an anti-Flag or anti-gB antibody for 2 h at 4 °C. Protein-A-sepharose beads were then added and allowed to react, with rotation, for a further 1 h at 4 °C. The immunoprecipitates were collected by a brief centrifugation, washed extensively with TNE buffer, and analysed by immunoblotting with an anti-NMHC-IIA antibody.

Fc-fusion protein. Plasmid pME-Fc-NMHC-IIA, encoding the rod domain of NMHC-IIA fused to a mutant Fc fragment of human IgG1 with low binding affinity for cellular Fc receptors and for the HSV-1 gE Fc ligand, was constructed as described earlier. Soluble NMHC-IIA–Fc was produced and used to detect gB on the cell surface of 293T cells infected with HSV-1 at an MOI of 1 or transfected with a gB or gD expression plasmid as described previously[21]. As a control, purified human CD200–Fc protein[3] was used. For transfection of 293T cells with the gB or gD expression plasmids, nearly confluent cells in 6-well plates were transfected with 4 μg of each of the plasmids by using Lipofectmine 2000 (Invitrogen) according to the manufacturer's instructions. At 18 h after transfection, cells were harvested, stained with NMHC-IIA–Fc or control–Fc (human CD200–Fc) and analysed by use of FACSCalibur with Cell Quest software (Becton Dickinson).

Affinity precipitation with NMHC-IIA fused to GST. GST or GST–NMHC-IIA was expressed, purified on glutathione-sepharose beads and quantified as described previously[31]. Affinity precipitation with GST or GST–NMHC-IIA fusion protein was performed as described previously[31].

Biotinylation of cell surface proteins of HSV-1-infected cells. Vero cells were exposed to wild-type HSV-1 or HSV-1 expressing MEF–gB or MEF–gH at an MOI of 50 at 4 °C for 2 h. At 2 min and 15 min after transfer of the cells to 37 °C, cell samples were washed four times with ice-cold PBS, and biotinylated twice with cleavable sulfo-NHS-SS-Biotin (Pierce) for 15 min each. After two washes with DMEM supplemented with 0.2% bovine serum albumin (BSA), free thiol groups were quenched with 5 mg iodoacetamide (in PBS supplemented with 1% BSA) per ml and cells were harvested, solubilized in RIPA or TNE buffer containing a proteinase inhibitor cocktail, precipitated with streptavidin beads or an anti-Flag antibody, and analysed by immunoblotting with an anti-NMHC-IIA antibody or streptavidin horseradish peroxidase.

Immunofluorescence and detection of diphosphorylated regulatory light chain. Vero cells were mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 for 2 h at 4 °C. At various times after transfer of the cells to 37 °C, cell samples were fixed, permeabilized, and then stained with anti-NMHC-IIA as described previously[30]. To examine the effects of ML-7, a specific inhibitor of myosin light chain kinase, or BAPTA-AM, a cell-permeable cytosolic Ca^{2+} chelator, on localization of NMHC-IIA during the initiation of HSV-1 entry, Vero cells pretreated with the indicated concentrations of ML-7 or BAPTA-AM for 30 min at 37 °C were mock-incubated or exposed to wild-type HSV-1 at an MOI of 50 for 1 h (BAPTA-AM) or 2 h (ML-7) at 4 °C in the same concentrations of ML-7 or BAPTA-AM. After removal of the inoculum, the cells were re-fed with pre-warmed medium and 15 min after being transferred to 37 °C, they were fixed, permeabilized, and then stained with anti-NMHC-IIA as described earlier, or analysed by immunoblotting with an anti-regulatory-light-chain antibody or an antibody to phosphorylated regulatory light chain at Thr 18 and Ser 19.

Establishment of HL60 cells stably overexpressing NMHC-IIA. Plat-GP cells, a 293T-derived murine-leukaemia-virus-based packaging cell line[21], were co-transfected with pMxs-NMHC-IIA-puro or pMxs-puro in combination with pMDG encoding vesicular stomatitis virus envelope G protein[21]. Two days after transfection, supernatants were harvested. HL60 cells were transduced by infection with retrovirus-containing supernatants of transfected Plat-GP cells and selected with 0.5 μg puromycin per ml in maintenance medium. Resistant cells transduced by recombinant retrovirus derived from pMxs-NMHC-IIA or

pMxs-puro were designated HL60/NMHC-IIA or HL60/puro cells, respectively. HL60/puro and HL60/NMHC-IIA cells in 24-well plates were inoculated with HSV-1(GFP) or pseudorabies virus expressing GFP at an MOI of 1 or 0.5, followed by centrifugation at 32 °C at 1,100g for 1 h. The inoculum was then removed, and the cells were washed and re-fed with the appropriate medium. At 12 h after infection, cell samples were analysed by fluorescence microscopy (Olympus IX71) or a FACSCalibur with Cell Quest software.

Infection inhibition assays. For inhibition with an anti-NMHC-IIA antibody, Vero and CHO-nectin-1 cells in 24-well plates were pretreated with various concentrations of anti-NMHC-IIA or control serum for 30 min and then inoculated with HSV-1(GFP), pseudorabies virus expressing GFP, or influenza virus at an MOI of 1 or 0.5 in the same serum concentrations. After viral adsorption for 30 min, the inoculum was removed and the cells were washed twice and re-fed with the appropriate medium. CHO-PILRα and HL60/NMHC-IIA cells in 24-well plates were pretreated with various concentrations of anti-NMHC-IIA or control serum for 30 min and then inoculated with HSV-1(GFP), pseudorabies virus expressing GFP, or influenza virus at an MOI of 1 in the same serum concentrations, followed by centrifugation at 32 °C at 1,100g for 1 h. The inoculum was removed, and the cells were washed twice and re-fed with the appropriate medium. At 6 h, 7 h (Vero, CHO-PILRα and CHO-nectin-1 cells), or 12 h after infection (HL60/NMHC-IIA cells), cell samples were analysed by use of FACSCalibur with Cell Quest software. For inhibition with ML-7 or BAPTA-AM, Vero cells pretreated with various concentrations of ML-7 or BAPTA-AM for 30 min were inoculated with wild-type HSV-1, HSV-1(GFP) or influenza virus at an MOI of 1 in the same concentrations of ML-7 or BAPTA-AM. After removal of the inoculum, the cells were re-fed with medium containing the same concentrations of ML-7 or BAPTA-AM. At 5 h after infection, cell samples were analysed by use of FACSCalibur with Cell Quest software.

PEG treatment. Vero cells pretreated with anti-NMHC-IIA or control serum at a 1:5 dilution at 37 °C for 30 min were inoculated with HSV-1(GFP) at an MOI of 1 in the same serum concentration. After viral adsorption for 30 min at 37 °C, the inoculum was removed, and the cells were washed twice and exposed to 40% PEG as described previously[10] with minor modifications. Exposure of the cells to PEG was kept to a minimum by removing the solution immediately after swirling it over the monolayers and by rapidly washing away the adherent viscous residue. To facilitate mixing of the residue with the wash solutions, the cells were washed sequentially with 20%, 10% and 5% PEG in PBS and the final washes were done with PBS alone. After PEG removal, the cells were re-fed with medium 199 containing 1% FCS. At 6 h after infection, cell samples were analysed by use of FACSCalibur with Cell Quest software. Vero cells pretreated with 20 μM ML-7 at 37 °C for 5 min were inoculated with HSV-1(GFP) at an MOI of 1 in the same ML-7 concentration at 4 °C for 1 h. After removal of the inoculum, cells were washed twice with pre-warmed medium and treated with pre-warmed 40% PEG as described above. At 5 h after infection, cell samples were analysed by use of FACSCalibur with Cell Quest software.

Establishment of cell lines stably expressing shRNA against NMHC-IIA. Vero cells were transfected with pSSSP-NMHC-IIA or pSSSP-Cre, and selected with 2.5 μg puromycin per ml in maintenance medium. Resistant cells transduced by pSSSP-Cre were designated Vero-shControl cells and used as a control. Single colonies transduced with pSSSP-NMHC-IIA were isolated and screened by immunoblotting with an anti-NMHC-IIA antibody, which led to isolation of Vero-shNMHC-IIA. Vero-shControl and Vero-shNMHC-IIA cells in 24-well plates were inoculated with HSV-1(GFP) at an MOI of 1. After viral adsorption for 1 h, the inoculum was removed and the cells were washed twice and re-fed with the appropriate medium. At 5 h after infection, cell samples were analysed by use of FACSCalibur with Cell Quest software.

Cell fusion assay. Vero cells in a 24-well plate were transfected with pPEP98-gB, pPEP99-gD, pPEP101-gL, pPEP100-gH and pCAGT7 and these transfectants were used as effector cells[4]. Vero-shControl and Vero-shNMHC-IIA cells in 24-well plates were transfected with pT7EMCLuc and these transfectants were used as target

cells[4]. As an internal control, pRL-CMV (Promega) encoding the Renilla luciferase gene driven by the cytomegalovirus promoter was co-transfected into the target cells. At 6 h after transfection, the effector cells were detached by 0.04% EDTA in PBS, washed once with maintenance medium and co-cultured with the target cells for 18 h. Firefly and Renilla luciferase activities were independently assayed by using the dual-luciferase reporter assay system (Promega) with the Veritas luminometer (Promega). The relative fusion activity was calculated as follows: firefly luciferase activity/Renilla luciferase activity of shControl or shNMHCIIA cells.

Animal studies. For corneal infections, five-week-old female ICR mice were lightly scarified on their corneas with a 27-gauge needle, the tear film was blotted, and a 5-μl drop of Medium 199 with or without 20 μM ML-7 was applied twice for 10 min to each eye. A 5-μl drop of Medium 199 containing 5×10^5 plaque forming units (p.f.u.) of wild-type HSV-1 was then applied to each eye as described previously[32,33]. To determine viral titres in the tear films, tear film samples were collected from both eyes by using a single cotton-tipped applicator. The cotton tip was transferred to 500 μl of Medium 199 supplemented with 1% FCS and then frozen at −80 °C. Frozen samples were later thawed and thoroughly mixed, and infectious virus was quantified by using standard plaque assays on Vero cells. The total p.f.u. per cotton tip was determined and divided by two to calculate the approximate viral titre per eye. Mice were monitored daily for mortality, and the clinical severity of keratitis of individually scored mice was recorded as described previously[32,34]. The clinical scoring system was as follows: 0, normal cornea; 1, mild haze; 2, moderate haze, iris visible; 3, severe haze, iris not visible; 4, severe haze, corneal ulcer; and 5, corneal rupture. All of the animal studies were carried out with the approval of the Ethical Committee for Animal Experimentation at the University of Tokyo.

21. Arii, J. et al. Entry of herpes simplex virus 1 and other alphaherpesviruses via the paired immunoglobulin-like type 2 receptor alpha. J. Virol. **83,** 4520–4527 (2009).
22. Sato, M. K., Takahashi, M. & Yazawa, M. Two regions of the tail are necessary for the isoform-specific functions of nonmuscle myosin IIB. Mol. Biol. Cell **18,** 1009–1017 (2007).
23. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther. **7,** 1063–1066 (2000).
24. Pertel, P. E., Fridberg, A., Parish, M. L. & Spear, P. G. Cell fusion induced by herpes simplex virus glycoproteins gB, gD, and gH-gL requires a gD receptor but not necessarily heparan sulfate. Virology **279,** 313–324 (2001).
25. Okuma, K., Nakamura, M., Nakano, S., Niho, Y. & Matsuura, Y. Host range of human T-cell leukemia virus type I analyzed by a cell fusion-dependent reporter gene activation assay. Virology **254,** 235–244 (1999).
26. Haraguchi, T. et al. SiRNAs do not induce RNA-dependent transcriptional silencing of retrovirus in human cells. FEBS Lett. **581,** 4949–4954 (2007).
27. Shimizu, S. et al. Ca²⁺-calmodulin-dependent myosin light chain kinase is essential for activation of TRPC5 channels expressed in HEK293 cells. J. Physiol. (Lond.) **570,** 219–235 (2006).
28. Itoh, Y. et al. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature **460,** 1021–1025 (2009).
29. Demmin, G. L., Clase, A. C., Randall, J. A., Enquist, L. W. & Banfield, B. W. Insertions in the gG gene of pseudorabies virus reduce expression of the upstream Us3 protein and inhibit cell-to-cell spread of virus infection. J. Virol. **75,** 10856–10869 (2001).
30. Kato, A. et al. Identification of a physiological phosphorylation site of the herpes simplex virus 1-encoded protein kinase Us3 which regulates its optimal catalytic activity in vitro and influences its function in infected cells. J. Virol. **82,** 6172–6189 (2008).
31. Kawaguchi, Y., Bruni, R. & Roizman, B. Interaction of herpes simplex virus 1 α regulatory protein ICP0 with elongation factor 1Δ: ICP0 affects translational machinery. J. Virol. **71,** 1019–1024 (1997).
32. Sagou, K., Imai, T., Sagara, H., Uema, M. & Kawaguchi, Y. Regulation of the catalytic activity of herpes simplex virus 1 protein kinase Us3 by autophosphorylation and its role in pathogenesis. J. Virol. **83,** 5773–5783 (2009).
33. Balliet, J. W. & Schaffer, P. A. Point mutations in herpes simplex virus type 1 oriL, but not in oriS, reduce pathogenesis during acute infection of mice and impair reactivation from latency. J. Virol. **80,** 440–450 (2006).
34. Polcicova, K. et al. Herpes keratitis in the absence of anterograde transport of virus from sensory ganglia to the cornea. Proc. Natl Acad. Sci. USA **102,** 11462–11467 (2005).

LETTER

Pannexin 1 channels mediate 'find–me' signal release and membrane permeability during apoptosis

Faraaz B. Chekeni[1,2], Michael R. Elliott[1,3], Joanna K. Sandilos[2], Scott F. Walk[1,3], Jason M. Kinchen[1,3,4], Eduardo R. Lazarowski[6], Allison J. Armstrong[1,3], Silvia Penuela[7], Dale W. Laird[7], Guy S. Salvesen[8], Brant E. Isakson[5], Douglas A. Bayliss[2] & Kodi S. Ravichandran[1,3,4]

Apoptotic cells release 'find-me' signals at the earliest stages of death to recruit phagocytes[1]. The nucleotides ATP and UTP represent one class of find-me signals[2], but their mechanism of release is not known. Here, we identify the plasma membrane channel pannexin 1 (PANX1) as a mediator of find-me signal/nucleotide release from apoptotic cells. Pharmacological inhibition and siRNA-mediated knockdown of PANX1 led to decreased nucleotide release and monocyte recruitment by apoptotic cells. Conversely, PANX1 over-expression enhanced nucleotide release from apoptotic cells and phagocyte recruitment. Patch-clamp recordings showed that PANX1 was basally inactive, and that induction of PANX1 currents occurred only during apoptosis. Mechanistically, PANX1 itself was a target of effector caspases (caspases 3 and 7), and a specific caspase-cleavage site within PANX1 was essential for PANX1 function during apoptosis. Expression of truncated PANX1 (at the putative caspase cleavage site) resulted in a constitutively open channel. PANX1 was also important for the 'selective' plasma membrane permeability of early apoptotic cells to specific dyes[3]. Collectively, these data identify PANX1 as a plasma membrane channel mediating the regulated release of find-me signals and selective plasma membrane permeability during apoptosis, and a new mechanism of PANX1 activation by caspases.

Previous studies have shown that ATP and UTP are released from intact apoptotic cells, with no gross leakage of cellular contents[2,4]. We initially considered three potential mechanisms for regulated nucleotide release from early-stage apoptotic cells: exocytosis, blebbing or passage via a plasma membrane channel. Treatment of cells undergoing apoptosis with the drug brefeldin A (blocking exocytosis[5]) or Y27632 (inhibiting the Rho kinase ROCK critical for membrane blebbing[6,7]) did not affect ATP release (data not shown). Therefore, we focused on potential channel(s) that might mediate nucleotide release during apoptosis.

Among the channels suggested to conduct ATP across the plasma membrane are the connexin and pannexin family channels[8–10]. Connexins form gap junctions between cells, mediating intercellular communication[8,10], whereas pannexin channels are thought to function primarily in conducting intracellular molecules to the extracellular space[8,9]. Initially, we used pharmacological inhibitors known to inhibit these channels[11]. The gap junction inhibitors 18-alpha-glycyrrhetinic acid (18AGA) and flufenamic acid (FFA) had no detectable effect on apoptosis-induced ATP release (Supplementary Fig. 1a); however, these drugs were effective in blocking connexin-mediated intercellular communication in a HeLa-cell-based scrape assay (Supplementary Fig. 1b). The inhibitor carbenoxolone (CBX), which targets connexin and pannexin channels, inhibited ATP release during apoptosis (Supplementary Fig. 1a). Interference by CBX in the ATP quantification assay was ruled out (Supplementary Fig. 2a), and CBX blocked ATP

release at all time points assessed, ruling out changes in the kinetics of release (Fig. 1a). Although CBX-mediated inhibition of ATP release was comparable to that seen with zVAD(OMe)-fmk (zVAD; cell-permeable pan-caspase inhibitor that blocks apoptosis and ATP release from apoptotic cells[2]), CBX did not inhibit apoptosis, based on caspase activation[12] and 'eat-me' signal exposure (Supplementary Fig. 2b, c). CBX also inhibited UTP release after apoptosis induction (Supplementary Fig. 4a), and the ability of apoptotic cell supernatants to attract THP-1 monocytes (Supplementary Fig. 4b, c). The higher CBX concentration needed for maximum inhibition of ATP release from apoptotic cells ($>100\,\mu\text{M}$) is likely due to binding of CBX by albumin (BSA) in the medium, as lower CBX concentrations were sufficient in lower BSA concentrations (Supplementary Fig. 3a, b). Probenecid, an inhibitor considered more specific for pannexins over connexins[13], also blocked ATP release from apoptotic cells (Fig. 1a). While drug studies have their limitations, the profile of inhibition (sensitivity to CBX and probenecid, but not 18AGA and FFA), suggested a role for pannexin channels[11,14], which we pursued further.

Among the three pannexins in humans (pannexin 1, 2 and 3)[15,16], pannexin 1 (PANX1) was the most abundant in the human Jurkat T cells used above (Fig. 1b). siRNA-mediated knockdown of PANX1, followed by induction of apoptosis reduced ATP and UTP release compared to control (47% and 68% decrease in ATP and UTP, respectively; Fig. 1c–e and Supplementary Fig. 5a, b). The residual ATP release was likely due to incomplete knockdown of PANX1 (Fig. 1c and Supplementary Fig. 5a), and was inhibitable by CBX and zVAD (Supplementary Fig. 5c). Importantly, no morphological differences were seen between control- and PANX1-siRNA-transfected cells at the time of apoptosis induction (data not shown). Total cellular ATP content was also comparable (Supplementary Fig. 5e), ruling this out as a cause for reduced ATP release. PANX1 knockdown had no effect on apoptosis progression, assessed by apoptotic caspase activity and by annexin V staining (Fig. 1d and Supplementary Fig. 5d). Consistent with decreased nucleotide find-me signal release, apoptotic cell supernatants from PANX1-siRNA-transfected cells also recruited fewer monocytes in vitro (Fig. 1f). To address chemotactic activity in vivo, we used the dorsal air pouch model that we had used previously[2]; cell supernatants were injected into a sterile air pouch, and leukocytes attracted to the pouch were analysed after 24 h (Fig. 1g). Supernatants from apoptotic PANX1 knockdown cells recruited fewer leukocytes into the pouch than supernatants from control-siRNA-transfected cells (Fig. 1g). Collectively, the PANX1 knockdown studies and pharmacological inhibition by CBX and probenecid supported a role for PANX1 in the release of the find-me signals ATP and UTP from apoptotic cells.

We also generated Jurkat cells stably overexpressing Flag-tagged, full-length PANX1 (426 amino acids[11]; denoted PANX1–Flag) (Fig. 2a). Compared to control-vector-transfected cells, PANX1–Flag cells released

[1]Beirne B. Carter Center for Immunology Research, [2]Department of Pharmacology, University of Virginia, Charlottesville, Virginia 22908, USA. [3]The Center for Cell Clearance, University of Virginia, Charlottesville, Virginia 22908, USA. [4]Department of Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA. [5]Cardiovascular Research Center and the Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA. [6]Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [7]Department of Anatomy and Cell Biology, University of Western Ontario, London, Ontario N6A 5C1, Canada. [8]Program in Apoptosis and Cell Death Research, The Burnham Institute for Medical Research, La Jolla, California 92037, USA.
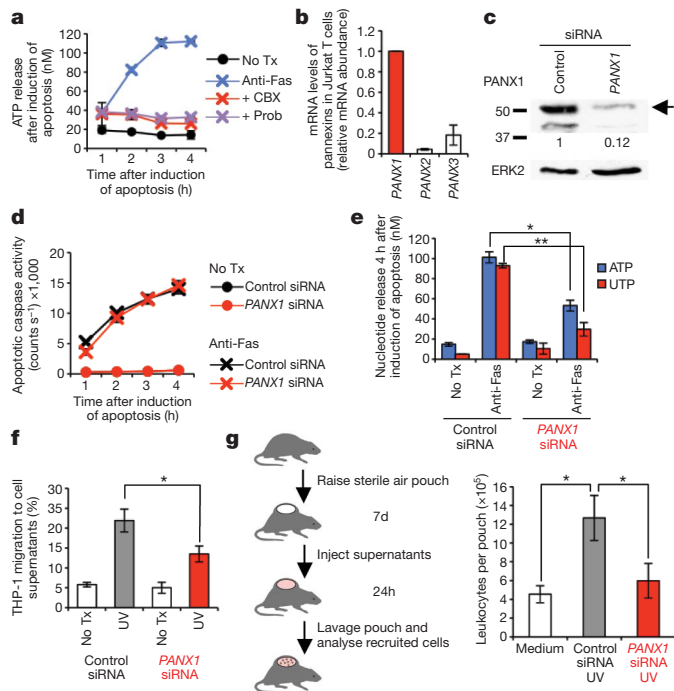
Figure 1 | Release of find-me signals by apoptotic cells is pannexin-1-dependent. **a**, Time course of ATP release from apoptotic Jurkat cells in the presence of 500 μM carbenoxolone (CBX) or 2 mM probenecid (Prob). $n = 3$. No Tx, no treatment. **b**, mRNA levels for *PANX1*, *PANX2* and *PANX3* in Jurkat cells determined by qPCR, normalized to *PANX1*. $n = 2$. **c**, PANX1 protein expression in Jurkat cells transfected with control or *PANX1* siRNA. Glycosylated PANX1 (upper band; arrow) was quantified. **d, e**, PANX1 knockdown does not affect the progression of apoptosis (assessed by apoptotic caspase activity) (**d**, $n = 3$), but decreases ATP (**e**, $n = 10$) and UTP (**e**, $n = 3$) release 4 h after apoptosis induction. *, $P < 10^{-5}$; **, $P < 0.01$. **f**, Transwell migration of THP-1 monocytes towards apoptotic cell supernatants from *PANX1*-siRNA-treated cells (4 h after ultraviolet (UV) treatment). *, $P < 0.05$. Representative of four independent experiments. **g**, Left, Schematic of mouse air pouch model for monitoring chemotactic activity of apoptotic cell supernatants *in vivo*. Right, CD45[+] leukocytes migrating into the pouch were determined after injecting apoptotic cell supernatants from siRNA-transfected cells. $n = 9$–10 mice per group. *, $P < 0.05$, by ANOVA with Bonferroni post-analysis. Error bars, s.e.m., except in **b**, **f**, where they represent s.d.

Figure 2 | Pannexin 1 expression level correlates with find-me signal release and membrane permeability. **a**, Immunoblotting of lysates from control or PANX1–Flag Jurkat cells. **b**, ATP release by Jurkat cells after induction of Fas-mediated apoptosis, and 4 h after UV-induced apoptosis. $n = 2$. **c**, UTP levels in supernatants of Jurkat cells 2 h after apoptosis induction. $n = 4$. **d**, Migration of THP-1 monocytes to supernatants from control and PANX1–Flag expressing cells, collected 2 h after induction of apoptosis. *, $P < 0.05$. Representative of four experiments. **e**, Left, flow cytometry histograms showing YO-PRO-1 dye uptake by control versus *PANX1*-siRNA-transfected Jurkat cells with or without apoptosis induction. Right, uptake of YO-PRO-1 presented as mean fluorescence intensity (MFI) of the entire cell population. *, $P < 0.0005$, $n = 3$. **f**, Left, YO-PRO-1 uptake by primary murine thymocytes undergoing anti-Fas-induced apoptosis, with zVAD or CBX treatments. Right, percentage of apoptotic cells (annexin V positive and propidium iodide negative). $n = 2$. Error bars represent s.d., except in **c** (s.e.m.).

more ATP and UTP upon induction of apoptosis (Fig. 2b, c), and their apoptotic cell supernatants recruited more monocytes (Fig. 2d). Even with overexpression of PANX1, the enhanced release of nucleotides was still regulated by the induction of apoptosis and sensitive to zVAD and CBX (Fig. 2b and data not shown). Thus, both siRNA-mediated knockdown of PANX1 ('loss of function') and overexpression of PANX1 ('gain of function') directly correlate with apoptosis-induced nucleotide release and find-me-signal-mediated monocyte recruitment.

PANX1 can conduct molecules up to 1 kDa in size across the plasma membrane (ATP and UTP are 507 and 484 daltons, respectively)[8,9], including the monomeric cyanine dye YO-PRO-1 (ref. 17). Monomeric cyanine dyes are marketed as apoptotic cell indicators, but the mechanism by which apoptotic cells take up these dyes is not known[3]. We asked whether PANX1 mediates the entry of the dyes YO-PRO-1 and TO-PRO-3 (629 and 671 daltons, respectively) during apoptosis. Interestingly, the uptake of YO-PRO-1 and TO-PRO-3 by apoptotic Jurkat cells was largely PANX1-dependent. This PANX1-dependent membrane permeability was selective, as these early-apoptotic cells did not take up the dye propidium iodide (which marks necrotic cells; data not shown). PANX1 knockdown (Fig. 2e) and treatment with CBX (Supplementary Fig. 6a) significantly reduced YO-PRO-1 uptake, whereas transient overexpression of PANX1 enhanced TO-PRO-3 uptake (Supplementary Fig. 6b). These data suggest an essential role
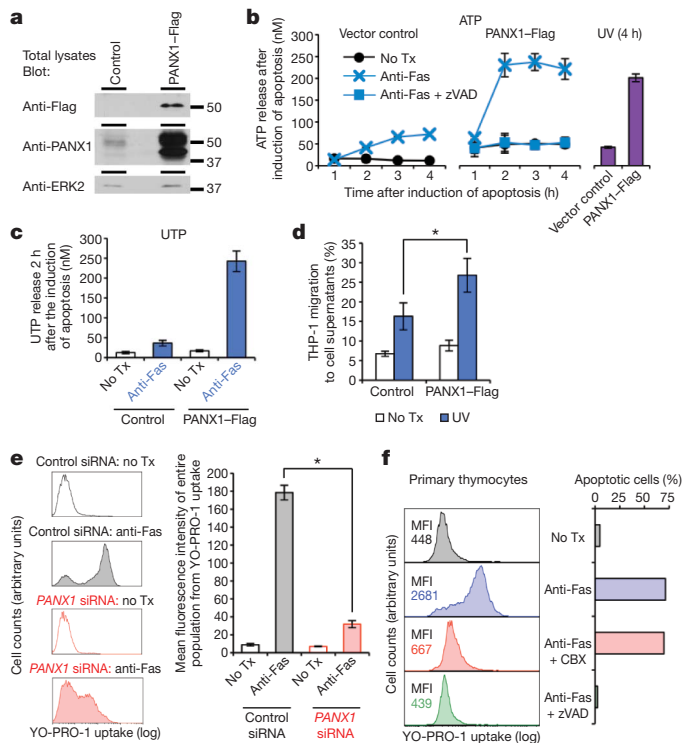
for PANX1 in the release of the find-me signals ATP and UTP, and in the selective plasma membrane permeability that has long been considered a hallmark of early-stage apoptotic cells[3].

We demonstrated previously that apoptotic primary thymocytes release nucleotide find-me signals, which is critical for proper phagocytic clearance of dying thymocytes *in vivo*[2]. Therefore, we asked whether thymocytes depend on PANX1 for increased plasma membrane permeability during apoptosis. Apoptotic thymocytes took up YO-PRO-1, and treatment with CBX blocked this uptake (similar to the caspase inhibitor zVAD) (Fig. 2f). However, unlike zVAD, CBX did not inhibit apoptosis, assessed by phosphatidylserine exposure (Fig. 2f).

We next asked whether apoptosis-induced membrane permeability could be directly correlated with PANX1-dependent plasma membrane currents. We measured channel activity at the single-cell level via whole-cell patch-clamp[18,19], and made several striking observations. First, CBX-sensitive currents were detected only in apoptotic cells. In bulk suspension, where cells undergo apoptosis at different rates[20], PANX1 currents were only detectable in cells that were beginning to bleb (a morphological characteristic of apoptosis; denoted 'dying' cells), but not in cells that appeared 'live' (Fig. 3a–c). Second, the apoptosis-induced currents were unaffected by 100 μM FFA, but were rapidly inhibited by 100 μM CBX (Fig. 3b, d and Supplementary Figs 7a, b and 8a–d). This rapid inhibition of the current flux (Fig. 3b and Supplementary Fig. 8d) suggested that CBX acts directly on the
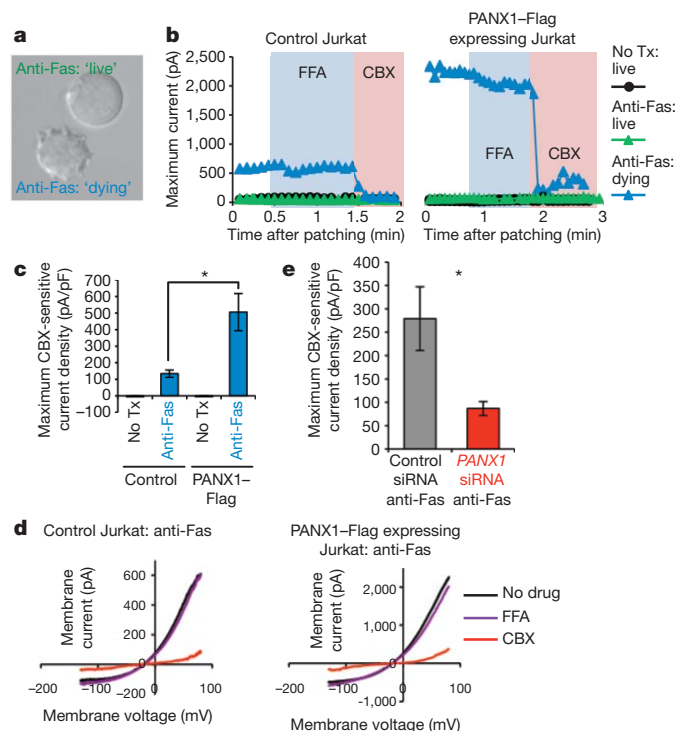
Figure 3 | Carbenoxolone-sensitive current induced during apoptosis is pannexin-1-dependent. a, Morphology of anti-Fas treated Jurkat cells used to identify 'live' and 'dying' cells (based on blebbing). **b**, Patch-clamp recordings from Jurkat cells, receiving indicated treatments. Peak whole-cell current (at +90 mV) is shown under conditions when bath solution was perfused with FFA (100 µM, blue shading) or CBX (100 µM, pink shading). Exemplar traces are representative of 5–15 cells per group. **c**, CBX-sensitive current density in indicated Jurkat cells, normalized to whole-cell capacitance (pA/pF). *, $P < 0.01$. $n \geq 7$ per group. **d**, Current–voltage relationships of apoptosis-induced current in dying Jurkat cells (from **b**), with current measured over a range of voltages (only dying cells shown, as live cells have very little current). Note different y axes. Traces representative of 4–15 cells. **e**, CBX-sensitive current density in *PANX1*-siRNA-transfected cells induced to undergo apoptosis. *, $P < 0.05$. $n \geq 8$ per group. All error bars represent s.e.m.

channel responsible for the increased membrane permeability. Third, apoptosis-induced currents were much larger in cells stably over-expressing PANX1–Flag (Fig. 3c) and markedly reduced in cells with PANX1 knockdown (Fig. 3e). Importantly, aside from the magnitude, the voltage-dependent properties of the apoptosis-induced CBX-sensitive current were essentially identical in control cells, PANX1–Flag overexpressing cells, and PANX1 knockdown cells (Fig. 3d and Supplementary Fig. 8c). Moreover, the IC_{50} values for CBX and probenecid, using wild-type Jurkat cells, roughly correspond to concentrations published in the literature (Supplementary Fig. 9)[11]. Collectively, these patch-clamp studies demonstrate that, at the single-cell level, the magnitude of the membrane permeability induced during apoptosis strongly correlates with the level of PANX1 expression.

We next investigated apoptosis-induced, post-translational modification(s) of PANX1 that might regulate channel activity. Interestingly, detection of endogenous PANX1 in apoptotic Jurkat cells by immunoblotting decreased in a time-dependent manner, temporally correlating with procaspase 3 activation (Fig. 4a). The caspase inhibitor zVAD 'restored' the PANX1 band, whereas CBX did not (although both zVAD and CBX inhibit the apoptosis-induced nucleotide release) (Fig. 4a and Supplementary Fig. 10a). Loss of Panx1 immunoreactivity, and its restoration by zVAD, was also observed in apoptotic primary murine thymocytes (Fig. 4b). This suggested an intriguing possibility that PANX1 itself might be a target of caspase-mediated cleavage. Although another type of modification could have caused the loss of immunoreactivity to this carboxy-terminal antibody[21] (CT Ab; Fig. 4c), this was ruled out by identical results seen in PANX1–Flag over-expressing cells using anti-Flag antibody for detection (Supplementary Fig. 10b).

To test whether PANX1 might be a direct target of caspases, we performed an *in vitro* cleavage assay by mixing purified active caspases with immunoprecipitated PANX1–Flag protein. Of those caspases tested, caspases 3 and 7 displayed the most significant cleavage of PANX1–Flag (Fig. 4d), resulting in loss of PANX1 immunoreactivity. This was not due to random activity of these caspases, as another C-terminally Flag-tagged protein, ELMO1–Flag, was not cleaved in the same assay (Supplementary Fig. 10c). Caspase cleavage sites are typically tetrapeptide sequences ending in an aspartic acid, after which the cleavage occurs[22]. Using the online tool CASVM[23], we identified two potential caspase cleavage sites in human PANX1: DMRD in the intracellular loop (residues 164–167; denoted site A) and DVVD in the
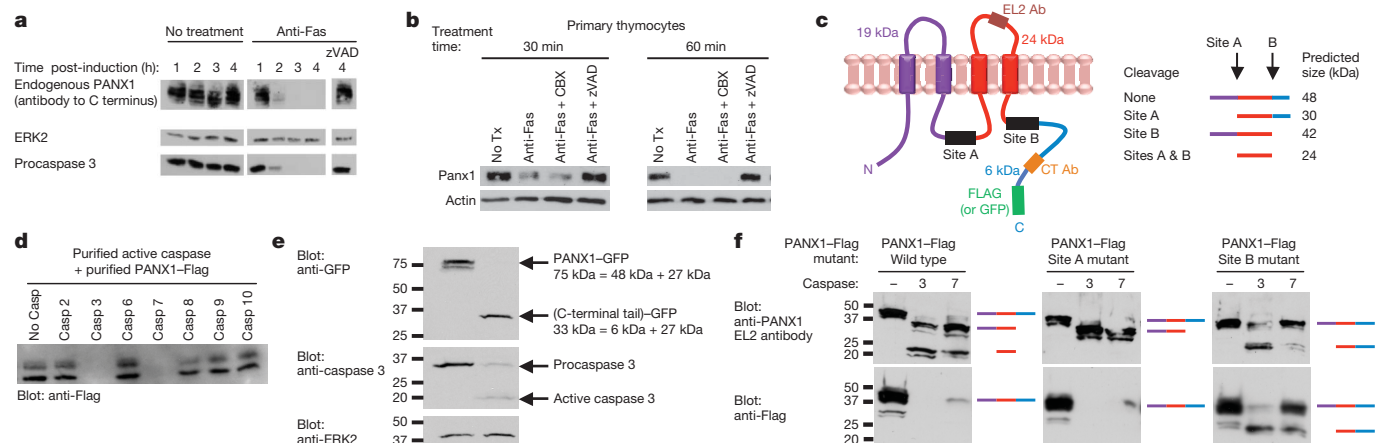

Figure 4 | Pannexin 1 is a target of effector caspase cleavage during apoptosis. a, b, Fas-mediated apoptosis results in loss of PANX1 detection with antibody targeted to PANX1 C terminus in Jurkat cells (**a**) and primary murine thymocytes (**b**). **c**, Schematic of PANX1–Flag protein indicating the predicted caspase cleavage sites (sites A and B) and the epitopes recognized by the EL2 (extracellular loop 2) and C-terminal antibodies. The predicted products of caspase cleavage at sites A and B are also shown. **d**, *In vitro* cleavage of immunoprecipitated PANX1–Flag incubated with the indicated purified active caspases. Loss of immunoreactivity to Flag assessed. Representative of two independent experiments. **e**, Jurkat cells stably expressing a C-terminally GFP-tagged PANX1 were induced to undergo apoptosis and assessed for the cleaved C-terminal fragment (which runs at ~33 kDa). Representative of two independent experiments. **f**, *In vitro* cleavage of indicated PANX1–Flag proteins (wild type and caspase cleavage site mutants), incubated with the indicated caspases.

C-terminal tail (residues 376–379; denoted site B) (Fig. 4c). To detect directly the C-terminal cleavage product of PANX1 in apoptotic cells, we established Jurkat cells stably expressing PANX1 with a C-terminal green fluorescent protein (GFP) tag. Induction of apoptosis in PANX1–GFP-expressing cells resulted in a cleavage product that suggested cleavage at site B (Fig. 4e). We then mutated sites A and B (individually and together) by substituting an alanine for the aspartic acid residues (see Methods) and performed caspase cleavage reactions. Detection of cleavage products via two different antibodies (to the second extracellular loop (EL2) and to the C-terminal Flag tag) (Fig. 4f and Supplementary Fig. 10c) showed that site A and site B are indeed two target sites for caspases 3 and 7 *in vitro*.

To test the functional relevance of these PANX1 caspase cleavage sites in cells during apoptosis, we transiently co-transfected Jurkat cells with plasmids coding for wild-type or mutant PANX1 proteins and the fluorescent protein mCherry (marker for transfection). After inducing apoptosis, we assessed YO-PRO-1 uptake by mCherry$^+$ cells using flow cytometry (Fig. 5a). PANX1–Flag expression resulted in higher YO-PRO-1 uptake than the control Flag vector. Expression of the PANX1 site A mutant led to the same level of increased YO-PRO-1 uptake as the wild-type-PANX1-transfected cells, suggesting that site A is dispensable for PANX1 function under these conditions (Fig. 5a). However, cells expressing the PANX1 site B mutant showed a dramatic reduction in YO-PRO-1 uptake, suggesting a dominant-negative effect on YO-PRO-1 uptake by apoptotic cells (YO-PRO-1 mean fluorescence intensities of 1,393 in the site B mutant condition, versus 11,179 in the control-vector-transfected cells) (Fig. 5a). Because PANX1 is thought to function as an oligomer[24], perhaps a certain number of subunits within the complex have to be cleaved for channel activation, resulting in a dominant-negative effect due to the presence of subunits that cannot be cleaved. It is notable that PANX1-site-B-mutant-expressing cells did undergo apoptosis (annexin-V-positive). Mutation of the aspartic acid residues to glutamic acid residues (denoted site B EE) replicated the findings with the alanine substitutions, ruling out charge alterations as a cause (Supplementary Fig. 11a). To assess the effect of the PANX1 site B mutant on plasma membrane currents, we transiently co-transfected Jurkat cells with plasmids for a fluorescent marker and the various PANX1 constructs. Marker-positive and dying cells (actively blebbing) revealed a dramatic inhibition of apoptosis-induced, CBX-sensitive current in the PANX1-site-B-mutant-expressing cells (Fig. 5b). We also established Jurkat cells stably expressing the PANX1 site B mutant. Upon apoptosis induction, ATP release and TO-PRO-3 uptake were drastically reduced (Fig. 5c, d), although the cells became apoptotic (annexin-V-positive) (Fig. 5d). These data suggest that the caspase cleavage site in the C terminus of PANX1 (site B) is necessary for the induction of PANX1-mediated plasma membrane permeability during apoptosis.

We next asked whether a PANX1 protein engineered to mimic truncation at site B (see Methods) would result in basal/constitutive plasma membrane permeability. Transient expression of the PANX1 truncation mutant led to TO-PRO-3 and YO-PRO-1 uptake by approximately one-third of the transfected cells (Fig. 5e and data not shown). We also detected PANX1 currents on cells expressing the truncation mutant that appeared 'live' (not blebbing), suggesting that the currents were due to expression of a constitutively open channel (Fig. 5f). The current–voltage relationship for the truncation mutant resembles that of an activated wild-type PANX1 channel, and the currents were CBX-sensitive (Fig. 5f and Supplementary Fig. 11b). Interestingly, the PANX1 currents were only seen in cells that were brightest for the co-transfected GFP marker, suggesting that only cells expressing higher amounts of the truncated/activated subunits form active channels. We performed the truncation mutant studies 16 h post-transfection, as these cells began to die 24–48 h post-transfection, perhaps due to continued release of ATP (data not shown). These data identify that cleavage of PANX1 at site B is necessary and sufficient to activate the channel, defining a new mechanism for regulation of pannexin channels.
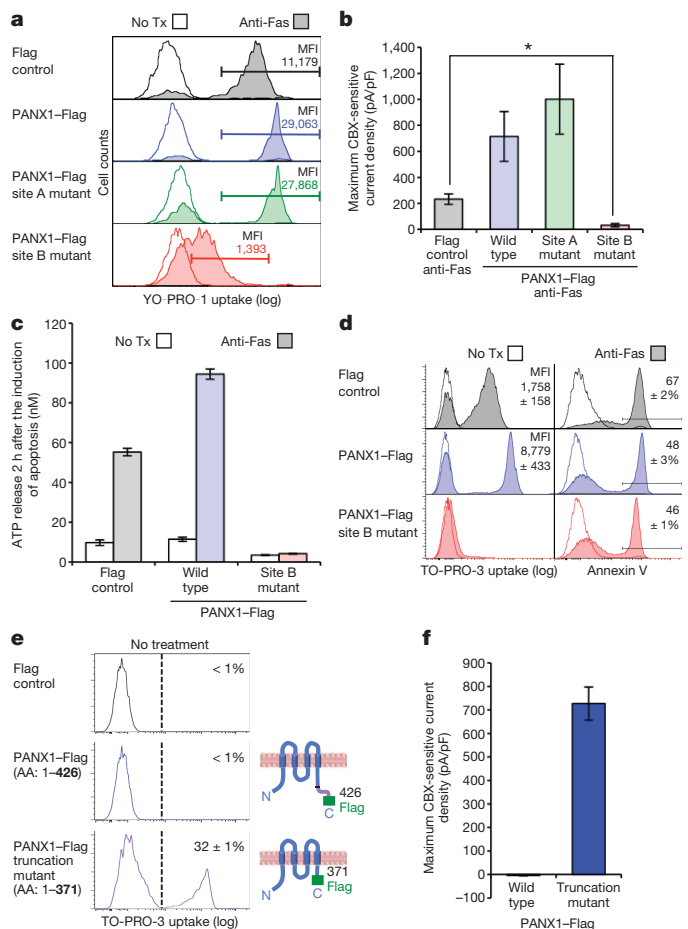

Figure 5 | Caspase-mediated cleavage of pannexin 1 results in channel activation during apoptosis a, b, Jurkat cells transiently transfected with indicated constructs were assessed for YO-PRO-1 uptake (**a**) and current density (**b**) 48 h post-transfection. The MFI for YO-PRO-1 uptake by the gated population is shown (**a**) and the analysis of dying cells in **b** was performed as in Fig. 3. *, $P < 0.01$, $n \geq 5$ for each group. **c, d,** Jurkat cells stably transfected with the indicated plasmids were assessed for ATP release (**c**, $n = 3$), TO-PRO-3 uptake (**d**, left) and annexin V staining (**d**, right) 2 h after induction of apoptosis. MFI for the TO-PRO-3 positive peaks and percentage of annexin V positive cells are shown. $n = 3$. **e,** TO-PRO-3 uptake by Jurkat cells 16 h after transient transfection with full-length or truncated PANX1. Schematic of full-length and truncated PANX1 (amino acids 1–426 and 1–371, respectively) are shown. $n = 3$. **f,** CBX-sensitive currents in Jurkat cells transiently transfected with indicated PANX1 proteins (full length or truncation mutant). Patch-clamp was performed on cells that appeared morphologically healthy. $n \geq 5$ per group. Error bars and ± represent s.e.m.

Collectively, these data provide several new insights. This work identifies a new physiological function for PANX1, demonstrating an essential and non-redundant role in nucleotide/find-me signal release from apoptotic cells. These data have implications for how apoptotic cells may modify their extracellular microenvironment through the selective release of molecules[25,26]. PANX1 also contributes significantly to plasma membrane permeability during apoptosis, which is relevant for 'selective' dye uptake by early apoptotic cells. There are still limited examples where caspase-mediated cleavage of specific proteins during apoptosis subsequently translates to biologically relevant functions[27,28]. Our work uncovers a novel caspase-mediated mechanism of PANX1 activation during apoptosis. Since cleavage site B is evolutionarily conserved among PANX1 homologues (human, mice and zebrafish), this may be a conserved mechanism of PANX1 regulation during apoptosis. Since ATP released by dying tumour cells can activate NLRP3/inflammasome-dependent signalling and promote anti-tumour immunity[4], defining a role for PANX1 in ATP release could also have implications for targeting these channels in cancer therapies.

METHODS SUMMARY

Induction of apoptosis. Jurkat cells resuspended in serum-free media (RPMI/1% BSA) were treated with 250 ng ml^{-1} anti-Fas (CH11 clone) or 100 mJ cm^{-2} ultraviolet C irradiation (Stratalinker). Primary murine thymocytes were treated with cross-linked anti-Fas (5 μg ml^{-1} anti-Fas (Jo2 clone) + 2 μg ml^{-1} protein G).

Preparation of cell supernatants and nucleotide measurement. Cells were incubated for indicated times after receiving treatments. Supernatants were collected by two consecutive centrifugations (425g for 2 min at room temperature). ATP was measured using a luciferase/luciferin assay (CellTiter-Glo; Promega) according to manufacturer's instructions. UTP concentrations were quantified by the UDP-glucose pyrophosphorylase-based reaction, as described previously[29].

Migration assay. Transwell migration assays were performed by applying THP-1 cells to the insert and the chemoattractant in the well of a 24-well plate with 5-μm pore Transwells (Corning) for 1 h at 37 °C. Supernatants were diluted in some cases to bring the chemotactic activity into the linear range of the migration assay, with the same dilution applied to all samples in each experiment. Percentage of input cells that migrated to the lower chamber was determined by flow cytometry using AccuCount beads (Spherotech).

PANX1 constructs. Human PANX1–Flag construct was generated by PCR cloning of PANX1 cDNA (Open Biosystems; catalogue number MHS1010-58324) into pEBB-Flag vector. Cleavage site mutations were performed using a QuikChange mutagenesis kit (Stratagene). PANX1 truncation mutant (amino acids 1–371) was produced by PCR to end at amino acid 371, immediately followed by a Flag tag (DYKDDDDK), resulting in a 379 amino acid protein (size that would be produced by caspase cleavage after the last aspartic acid residue in site B).

Patch-clamp analysis. Whole-cell recordings were obtained at room temperature using patch pipettes (3–5 MΩ) filled with a Cs-based internal solution. Peak current (at +90 mV) was normalized to cell capacitance. CBX-sensitive current was taken as the difference in peak current before and after CBX application.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lauber, K., Blumanthal, S. C., Waibel, M. & Wesselborg, S. Clearance of apoptotic cells: getting rid of the corpses. *Mol. Cell* **14**, 277–287 (2004).
2. Elliott, M. R. *et al.* Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* **461**, 282–286 (2009).
3. Idziorek, T., Estaquier, J., De Bels, F. & Ameisen, J. C. YOPRO-1 permits cytofluorometric analysis of programmed cell death (apoptosis) without interfering with cell viability. *J. Immunol. Methods* **185**, 249–258 (1995).
4. Ghiringhelli, F. *et al.* Activation of the NLRP3 inflammasome in dendritic cells induces IL-1β-dependent adaptive immunity against tumors. *Nature Med.* **15**, 1170–1178 (2009).
5. Fujiwara, T., Oda, K., Yokota, S., Takatsuki, A. & Ikehara, Y. Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum. *J. Biol. Chem.* **263**, 18545–18552 (1988).
6. Sebbagh, M. *et al.* Caspase-3-mediated cleavage of ROCK I induces MLC phosphorylation and apoptotic membrane blebbing. *Nature Cell Biol.* **3**, 346–352 (2001).
7. Coleman, M. L. *et al.* Membrane blebbing during apoptosis results from caspase-mediated activation of ROCK I. *Nature Cell Biol.* **3**, 339–345 (2001).
8. Harris, A. L. Connexin channel permeability to cytoplasmic molecules. *Prog. Biophys. Mol. Biol.* **94**, 120–143 (2007).
9. MacVicar, B. A. & Thompson, R. J. Non-junction functions of pannexin-1 channels. *Trends Neurosci.* **33**, 93–102 (2010).
10. Scemes, E., Spray, D. C. & Meda, P. Connexins, pannexins, innexins: novel roles of "hemi-channels". *Pflugers Arch.* **457**, 1207–1226 (2009).
11. Ma, W., Hui, H., Pelegrin, P. & Surprenant, A. Pharmacological characterization of pannexin-1 currents expressed in mammalian cells. *J. Pharmacol. Exp. Ther.* **328**, 409–418 (2009).
12. Denault, J. B. & Salvesen, G. S. Apoptotic caspase activation and activity. *Methods Mol. Biol.* **414**, 191–220 (2008).
13. Silverman, W., Locovei, S. & Dahl, G. Probenecid, a gout remedy, inhibits pannexin 1 channels. *Am. J. Physiol. Cell Physiol.* **295**, C761–C767 (2008).
14. Bruzzone, R., Barbe, M. T., Jakob, N. J. & Monyer, H. Pharmacological properties of homomeric and heteromeric pannexin hemichannels expressed in *Xenopus* oocytes. *J. Neurochem.* **92**, 1033–1043 (2005).
15. Baranova, A. *et al.* The mammalian pannexin family is homologous to the invertebrate innexin gap junction proteins. *Genomics* **83**, 706–716 (2004).
16. Söhl, G., Maxeiner, S. & Willecke, K. Expression and functions of neuronal gap junctions. *Nature Rev. Neurosci.* **6**, 191–200 (2005).
17. Locovei, S., Scemes, E., Qiu, F., Spray, D. C. & Dahl, G. Pannexin1 is part of the pore forming unit of the P2X$_7$ receptor death complex. *FEBS Lett.* **581**, 483–488 (2007).
18. Chen, X., Shu, S., Kennedy, D. P., Willcox, S. C. & Bayliss, D. A. Subunit-specific effects of isoflurane on neuronal Ih in HCN1 knockout mice. *J. Neurophysiol.* **101**, 129–140 (2009).
19. Thompson, R. J. *et al.* Activation of pannexin-1 hemichannels augments aberrant bursting in the hippocampus. *Science* **322**, 1555–1559 (2008).
20. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).
21. Penuela, S. *et al.* Pannexin 1 and pannexin 3 are glycoproteins that exhibit many distinct characteristics from the connexin family of gap junction proteins. *J. Cell Sci.* **120**, 3772–3783 (2007).
22. Pop, C. & Salvesen, G. S. Human caspases: activation, specificity, and regulation. *J. Biol. Chem.* **284**, 21777–21781 (2009).
23. Wee, L. J., Tan, T. W. & Ranganathan, S. CASVM: web server for SVM-based prediction of caspase substrates cleavage sites. *Bioinformatics* **23**, 3241–3243 (2007).
24. Ambrosi, C. *et al.* Pannexin1 and pannexin2 channels show quaternary similarities to connexons and different oligomerization numbers from each other. *J. Biol. Chem.* **285**, 24420–24431 (2004).
25. Burnstock, G. & Knight, G. E. Cellular distribution and functions of P2 receptor subtypes in different systems. *Int. Rev. Cytol.* **240**, 31–304 (2004).
26. Praetorius, H. A. & Leipziger, J. ATP release from non-excitable cells. *Purinergic Signal.* **5**, 433–446 (2009).
27. Johnson, C. E. & Kornbluth, S. Caspase cleavage is not for everyone. *Cell* **134**, 720–721 (2008).
28. Timmer, J. C. & Salvesen, G. S. Caspase substrates. *Cell Death Differ.* **14**, 66–72 (2007).
29. Lazarowski, E. R. & Harden, T. K. Quantitation of extracellular UTP using a sensitive enzymatic assay. *Br. J. Pharmacol.* **127**, 1272–1278 (1999).

METHODS

Reagents. Purified nucleotides, propidium iodide, anti-Flag HRP (A8592), anti-Flag agarose beads (A2220), protein G, carbenoxolone, probenecid, flufenamic acid and 18-alpha-glycyrrhetinic acid were obtained from Sigma-Aldrich. YO-PRO-1 and TO-PRO-3 were obtained from Invitrogen. CellTiter-Glo and Caspase-Glo 3/7 were obtained from Promega. Control siRNA (AM4611) and *PANX1* siRNA (ID 134470; sense: 5′-GCAUCAAAUCAGGGAUCCUtt-3′) were obtained from Ambion. Anti-ERK2 (sc-154), anti-Caspase-3 (sc-7148), and anti-GFP HRP (sc-9996) antibodies were obtained from Santa Cruz Biotechnology. Annexin V-FITC and Annexin V-APC were obtained from BD and eBioscience, respectively. cDNA encoding full-length human PANX1 (426 amino acids) was acquired from Open Biosystems (catalogue number MHS1010-58324, accession number BC016931). Anti-hPANX1 serum, affinity purified anti-mPanx1, affinity purified anti-mPanx1 EL2-247 antibodies have been described previously[21,30]. Purified active caspases have also been described previously[12]. Other reagents were obtained as follows: z-VAD(OMe)-fmk (Enzo Life Sciences), anti-Fas (CH11 anti-human clone, Millipore; Jo2 anti-mouse clone, Becton Dickinson), recombinant apyrase (New England Biolabs). mBlueberry2 was provided by R. E. Campbell[31].

Induction of apoptosis. Jurkat T cells (E6.1) resuspended at 2×10^6 cells ml^{-1} in RPMI, 1% BSA, 10 mM HEPES and penicillin/streptomycin/L-glutamine were treated with 250 ng ml^{-1} anti-Fas (activating monoclonal IgM; CH11 clone) or 100 mJ cm^{-2} ultraviolet C irradiation (Stratalinker), and grown at 37 °C, 5% CO$_2$. Primary murine thymocytes were isolated and treated with cross-linked anti-Fas (5 µg ml^{-1} anti-Fas (Jo2 clone) + 2 µg ml^{-1} protein G) and grown at 37 °C, 5% CO$_2$. Where indicated, cells were co-treated with zVAD(OMe)-fmk (50 µM or 100 µM) or CBX (500 µM). The induction of apoptosis was performed in medium containing calcium, which can block connexin hemichannels[11].

Preparation of cell supernatants and nucleotide measurement. Cells were incubated for indicated times after receiving either control treatment, or treatment with anti-Fas or UV-irradiation. Cell suspension was then centrifuged at 425*g* for 2 min at room temperature. Supernatants were transferred to a fresh tube and the centrifugation repeated. Supernatants from the second centrifugation were used for nucleotide measurements. ATP was measured using a luciferase/luciferin assay (CellTiter-Glo; Promega) according to manufacturer's instructions, via a 1450 Microbeta TriLux luminometer (Perkin Elmer). UTP concentrations were quantified by the UDP-glucose pyrophosphorylase-based reaction, as previously described[29]. Briefly, 100 µl samples were incubated in the presence of 0.5 U ml^{-1} UDP-glucose pyrophosphorylase, 0.5 U ml^{-1} inorganic pyrophosphatase, 1.6 mM CaCl$_2$, 2 mM MgCl$_2$, 25 mM hydroxyethylpiperazine ethanesulfonic acid (HEPES) (pH 7.4), and ~100,000 c.p.m. 1 µM [^{14}C]glucose-1P. Incubations lasted 1 h at 30 °C. Reactions were terminated by heating the samples at 95 °C for 2 min. Conversion of [^{14}C]glucose-1P to [^{14}C]UTP was determined by high-performance liquid chromatography.

Apoptotic caspase activity. Caspase activity assays were performed using the DEVD-based synthetic substrate Caspase-Glo 3/7 (Promega) according to manufacturer's instructions.

Quantitative PCR. cDNA was synthesized from 50 ng of RNA isolated from Jurkat cells (RNeasy, Qiagen) using Superscript III (Invitrogen). qPCR was performed on the ABI StepOnePlus instrument using TaqMan probes (Applied Biosystems). Human pannexin mRNA levels shown are normalized to hypoxanthine phosphoribosyltransferase (*HPRT1*) levels. Primers used are as follows: hPANX1: Hs00209790_m1, hPANX2: Hs00364525_m1, hPANX3: Hs01573439_m1.

Migration assay. Transwell migration assays were performed by applying 100 µl of THP-1 cells at 2×10^6 cells ml^{-1} to the upper chamber (insert) in the same culture medium as the chemoattractant in lower chamber (500 µl) of a 24-well plate with 5-µm pore size Transwells (Corning) for 1 h at 37 °C, 5% CO$_2$. Supernatants were diluted in some cases to bring the chemotactic activity into the linear range of the migration assay, with the same dilution applied to all samples in each experiment. The percentage of cells that migrated to the lower chamber was determined by flow cytometry using 5.1 µm AccuCount beads (Spherotech) and calculated as the percentage of input cells. Where indicated, 0.05 U ml^{-1} recombinant apyrase was added to chemoattractant for 5 min at room temperature.

Air pouch studies. All animal studies were performed in accordance with the University of Virginia Animal Care and Use Committee guidelines. Animals were housed in a specific pathogen-free facility. Female C57BL/6 mice were used (Charles River Laboratories). Air pouch experiments were performed as described previously[32], using mice aged 8–12 weeks. In brief, 5 ml of 0.2 µm-filtered air was injected subcutaneously into the mid-dorsal region. After 3 days, the same pouches were injected with 3 ml of 0.2 µm-filtered air to maintain the pouch. Four days later the pouches were injected with 1 ml of 0.2 µm-filtered supernatants from control or ultraviolet-treated Jurkat cells. Supernatants were produced in

RPMI with 5% FBS, 10 mM HEPES, and penicillin/streptomycin/L-glutamine. After 24 h, cells from the air pouch were collected by two lavages of 2 ml of HBSS/1% FBS. Collected cells were resuspended in equal volumes and cell counting was performed by flow cytometry using quantification beads (Spherotech). For analyses of specific populations in the air pouch, cells were treated with anti-CD16/32 for 15 min on ice to block Fc receptor binding, followed by addition of the indicated fluorescently labelled antibodies for 30 min on ice. After washing, the cells were analysed with a FACS Canto instrument (Becton Dickinson).

PANX1 constructs. Human PANX1–Flag construct was generated by PCR cloning of *PANX1* cDNA (Open Biosystems; catalogue number MHS1010-58324) into pEBB-Flag vector. Mutations were performed using QuikChange mutagenesis kit (Stratagene). Fidelity of all constructs was confirmed by sequencing. Site B (DVVD) was C-terminal to and continuous with another predicted caspase cleavage site (IKMD); therefore, both aspartic acid residues were mutated to alanine to ensure complete disruption of the cleavage region (IKMDVVD mutated to IKMAVVA). Glutamic acid substitutions were also performed, in order to conserve charge at the cleavage site (IKMDVVD mutated to IKMEVVE). PANX1 truncation mutant (amino acids 1–371) was produced by PCR to end at amino acid 371, immediately followed by a Flag tag (8 amino acids: DYKDDDDK). This resulted in a 379 amino acid truncation mutant protein, which is exactly the same size as would be produced by caspase cleavage after the last aspartic acid residue in site B.

siRNA, transient, and stable transfections. Transfections were carried out using the BTX Square Pulse T820 electroporator with 5–10 million Jurkat T cells per transfection. For siRNA transfection, cells were transfected with 200 nM siRNA (Ambion) and used 72 h after transfection. For transient DNA transfections, Jurkat cells were electroporated with 5 µg of control vector or PANX1–Flag construct along with 5 µg mCherry expression construct (marker), and used 48 h after transfection. For generation of stable lines, Jurkat cells were electroporated with 10 µg of plasmid (pEBB PANX1–Flag) along with a pA-puro vector, and cultured in 1 µg ml^{-1} puromycin before and during limiting dilutions to isolate single-cell clones. Puromycin-resistant clones were screened for expression of Flag by immunoblotting.

Measuring selective membrane permeability during apoptosis. Jurkat cells were induced to undergo apoptosis (as described above) in media containing 1 µM YO-PRO-1 or TO-PRO-3 dyes. Cells were washed in PBS and resuspended in either PBS/0.5% BSA or binding buffer for annexin V staining. Dye uptake and cell surface staining was assessed by flow cytometry.

Patch-clamp analysis. Whole-cell recordings were obtained at room temperature using 3- to 5-MΩ borosilicate glass patch pipettes and an Axiopatch 200A amplifier (Axon Instruments) in a bath solution composed of 140 mM NaCl, 3 mM KCl, 2 mM MgCl$_2$, 2 mM CaCl$_2$, 10 mM HEPES and 10 mM glucose (pH 7.3). Internal solution contained 30 mM TEACl, 100 mM CsMeSO$_4$, 4 mM NaCl, 1 mM MgCl$_2$, 0.5 mM CaCl$_2$, 10 mM HEPES, 10 mM EGTA, 3 mM ATP-Mg and 0.3 mM GTP-Tris (pH 7.3). Ramp voltage clamp commands were applied at 5-s intervals using PCLAMP software and a Digidata1322A digitizer (Axon instruments). Whole-cell peak conductance was taken at +90 mV and normalized to cell capacitance. CBX-sensitive current was taken as the difference in peak conductance before and after CBX application relative to the initial peak current.

Immunoblotting. Samples were analysed by SDS–PAGE and immunoblotting using the following antibody concentrations/dilutions: anti-human PANX1 serum (1:5,000), affinity-purified anti-Panx1 EL2-247 (0.2 µg ml^{-1}), affinity purified anti-murine Panx1 (0.2 µg ml^{-1}), anti-ERK2 (1:10,000) and anti-Caspase-3 (1:1,000).

Purified PANX1 proteins. HEK 293T cells were transfected with Flag-tagged ELMO1 or PANX1 plasmids using calcium phosphate transfection (ProFection, Promega). After 48 h, whole-cell lysates were immunoprecipitated using anti-Flag beads, washed in wash buffer, and resuspended in TBST + 10% glycerol before use in *in vitro* cleavage reactions.

***In vitro* caspase cleavage assay.** Active recombinant caspases were produced and characterized as described previously[12]. *In vitro* cleavage reactions were performed in the following reaction buffer: 100 mM NaCl, 10% sucrose, 0.1% CHAPS, 20 mM PIPES, 1 mM EDTA and 10 mM dithiothreitol. Immunoprecipitated Flag-tagged proteins were resuspended in reaction buffer, and purified active caspase was added to a final concentration of 100 nM. The reaction was carried out for 1 h at 37 °C and terminated by addition of SDS sample buffer and boiling for 5 min. The reactions were analysed by SDS–PAGE and immunoblotting.

Scrape assay. Scrape assays were performed as originally described[33]. Briefly, confluent monolayers of HeLa cells or HeLa cells stably expressing Connexin 43 (Cx43) were scraped with a blunted 22-gauge needle in PBS containing 1 mg ml^{-1} Lucifer yellow. Five minutes after scrape, cells were washed with PBS, fixed with 4% paraformaldehyde and viewed on an Olympus IX70 microscope.

Phagocytosis assay. Bone marrow-derived macrophages (BMDMs) were prepared as described previously[32]. For phagocytosis assays, BMDMs were plated at a density of 100,000 cells in a 24-well plate in RPMI+10% serum overnight, and

stained for 5 min with 1 mM CFSE-SE (Invitrogen) before the engulfment assay. For apoptotic targets, Jurkat cells were induced to undergo apoptosis using anti-Fas treatment as described above in the presence or absence of 500 μM CBX, then stained with TAMRA-SE (Invitrogen; to determine gross binding/uptake of apoptotic cells) or Cypher-5E (GE Health; a pH-sensitive dye that displays enhanced fluorescence upon acidification of the apoptotic cell in the phagosome, to monitor internalization). BMDMs were then incubated with 2 million apoptotic Jurkat cells (in the presence or absence of 500 μM CBX) for ~1 h, trypsinized and analysed on a BD FACSCanto flow cytometer. The collected flow cytometry data were analysed using FlowJo software.

Statistical analyses. Data are presented as the mean ± standard error of the mean (s.e.m.) or mean ± standard deviation (s.d.), unless noted otherwise. Statistical significance for comparisons was determined by the Student's two-tailed

t-test, unless noted otherwise. A *P* value less than 0.05 was considered statistically significant.

30. Penuela, S., Bhalla, R., Nag, K. & Laird, D. W. Glycosylation regulates pannexin intermixing and cellular localization. *Mol. Biol. Cell* **20,** 4313–4323 (2009).
31. Ai, H. W., Shaner, N. C., Cheng, Z., Tsien, R. Y. & Campbell, R. E. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46,** 5904–5910 (2007).
32. Kadl, A., Galkina, E. & Leitinger, N. Induction of CCR2-dependent macrophage accumulation by oxidized phospholipids in the air-pouch model of inflammation. *Arthritis Rheum.* **60,** 1362–1371 (2009).
33. el-Fouly, M. H., Trosko, J. E. & Chang, C. C. Scrape-loading and dye transfer. A rapid and simple technique to study gap junctional intercellular communication. *Exp. Cell Res.* **168,** 422–430 (1987).

LETTER

The proteasome antechamber maintains substrates in an unfolded state

Amy M. Ruschak[1], Tomasz L. Religa[1], Sarah Breuer[2], Susanne Witt[2] & Lewis E. Kay[1]

Eukaryotes and archaea use a protease called the proteasome that has an integral role in maintaining cellular function through the selective degradation of proteins[1–4]. Proteolysis occurs in a barrel-shaped 20S core particle, which in *Thermoplasma acidophilum* is built from four stacked homoheptameric rings of subunits, α and β, arranged $α_7β_7β_7α_7$ (ref. 5). These rings form three interconnected cavities, including a pair of antechambers (formed by $α_7β_7$) through which substrates are passed before degradation and a catalytic chamber ($β_7β_7$) where the peptide-bond hydrolysis reaction occurs[4,5]. Although it is clear that substrates must be unfolded to enter through narrow, gated passageways (13 Å in diameter) located on the α-rings[1,6,7], the structural and dynamical properties of substrates inside the proteasome antechamber remain unclear. Confinement in the antechamber might be expected to promote folding and thus impede proteolysis. Here we investigate the folding, stability and dynamics of three small protein substrates in the antechamber by methyl transverse-relaxation-optimized NMR spectroscopy[8]. We show that these substrates interact actively with the antechamber walls and have drastically altered kinetic and equilibrium properties that maintain them in unstructured states so as to be accessible for hydrolysis.

Each proteasome core particle (Fig. 1a) can accumulate multiple substrates that distribute among all three chambers, supporting a model whereby the antechambers function as reservoirs for toxic proteins before degradation[9,10]. It is therefore expected that the conformations adopted by substrates within the antechamber are critical in controlling the rate and efficacy of protein hydrolysis. One possibility is that the antechamber might function as an inert chamber that promotes folding[11]. A simple thermodynamic argument predicts that confinement would lead to the relative stabilization of the native state conformation by means of an entropic effect that limits the conformations accessible to the unfolded polypeptide[12]. However, refolding of substrates while they are 'stored' is undesirable as many sites would become inaccessible to cleavage.

To address the structural and motional properties of substrate in the proteasome antechamber, we have used transverse-relaxation-optimized NMR spectroscopy to exploit the sensitivity of methyl groups as probes in high-molecular-weight proteins[8,13]. Stable substrate–proteasome complexes have been produced by tethering substrate molecules to the proteasome through a 15 Å heterobifunctional reagent that links the substrate amino terminus to a cysteine located on the surface of the α-subunit (Fig. 1a and Supplementary Fig. 1), with substrate release occurring only on incubation with reducing agent (Supplementary Fig. 2). As described in Methods, on average each antechamber is populated by approximately one substrate molecule. Three substrates have been chosen, including the engrailed homeodomain from
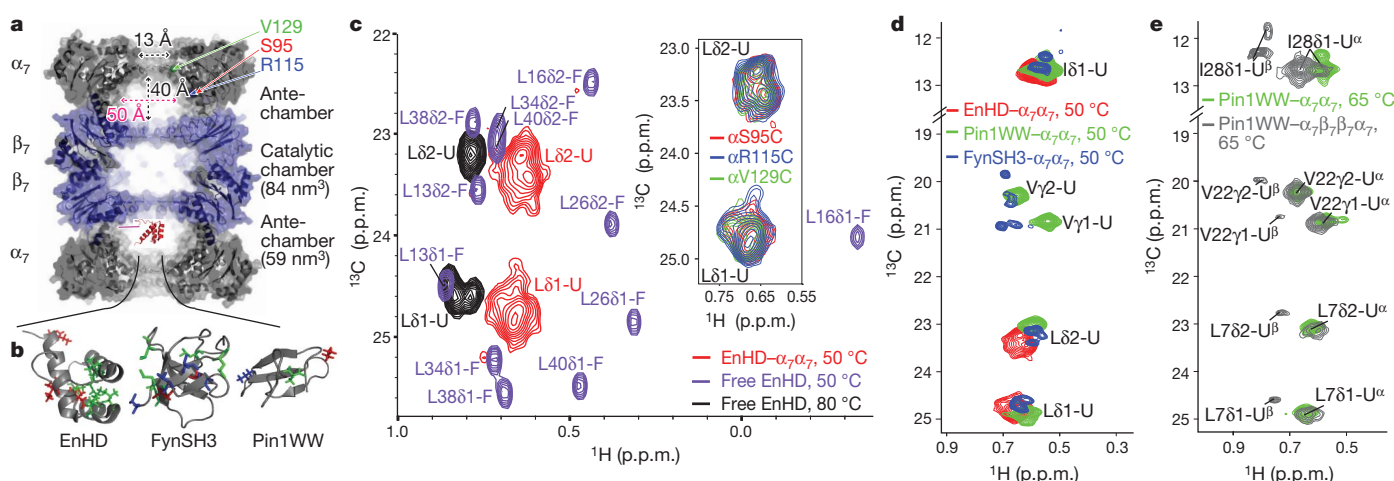

Figure 1 | Encapsulation of substrates in $α_7α_7$. a, Structure of the proteasome from *T. acidophilum*[5], with red, green and blue spheres designating the locations of Ser 95, Val 129 and Arg 115 in the α-subunit that have been used for substrate tethering. **b**, Native structures of proteasome substrates EnHD, FynSH3 and Pin1WW. The locations of Ile (red), Leu (green) and Val (blue) probes are shown using a stick representation. **c**, Spectrum of the Leu region of EnHD in $α_7α_7$ (red) and free in solution (purple), at 50 °C. For reference, the spectrum of the temperature-denatured state of free EnHD at 80 °C is shown (black). Inset, HMQC spectra of EnHD attached to residues 95 (red), 115 (blue) and 129 (green) of the α-subunit in $α_7α_7$, at 50 °C. **d**, Spectra of EnHD (red), Pin1WW (green) and FynSH3 (blue) encapsulated at Ser 95 of $α_7α_7$, at 50 °C. The spectrum of FynSH3 was recorded using different buffer conditions; the best overlay with the Pin1WW and EnHD spectra was achieved by translating the spectrum 0.19 p.p.m. in the ^{1}H dimension. **e**, ^{13}C,[1]H HMQC spectrum of ILV-methyl-labelled Pin1WW encapsulated in an active-site mutant of $α_7β_7β_7α_7$ (grey) and in $α_7α_7$ (green), at 65 °C. The extra set of peaks for [13]CH$_3$-Pin1WW in the full proteasome derives from substrate localized to the catalytic chamber (denoted by superscript 'β') as opposed to the antechamber (superscript 'α'). On average, only one Pin1WW is tethered per antechamber. Figures displaying molecular structures were made with PyMOL (http://www.pymol.org).

[1]Departments of Molecular Genetics, Biochemistry and Chemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [2]Department of Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

Drosophila melanogaster[14] (EnHD), a FYN Src homology 3 domain from *Gallus gallus*[15] (FynSH3) and the WW domain from *Homo sapiens* PIN1[16] (Pin1WW). These proteins provide examples of different secondary structures and charge states at the pH studied (pH 6.8; Fig. 1b), from α-helical (EnHD; isoelectric point, pI = 11.7) to predominately β-sheet (Pin1WW, pI = 11.3; FynSH3, pI = 4.3). Notably, all three proteins are readily degraded upon incubation with the core particle (Supplementary Fig. 3).

Initial studies focused on a two-ring system, $\alpha_7\alpha_7$, which spontaneously assembles from wild-type α monomers in the absence of the β-subunit (ref. 17). We have chosen $\alpha_7\alpha_7$ as a model for the antechamber ($\alpha_7\beta_7$) for two reasons: first, the α- and β-subunits are structurally homologous[5] and the amino-acid compositions of surface-exposed side chains in the α- and β-subunits are very similar; second, the small size of the particle (360 kDa) relative to the full proteasome (670 kDa) leads to significantly better $^{13}C,^1H$ heteronuclear multiple quantum coherence (HMQC) correlation spectra that can be quantified more rigorously[18]. It is worth noting, however, that similar results are obtained when using the full $\alpha_7\beta_7\beta_7\alpha_7$ construct (see below). To assess the conformation and dynamics of each substrate, ILV-methyl-labelled substrates (corresponding to U-$[^2H]$ Ile-$[\delta 1\,^{13}CH_3]$, Leu,Val-$[^{13}CH_3,^{12}CD_3]$ labelling[13]) were produced and tethered to U-$[^2H]$ $\alpha_7\alpha_7$. At 50 °C, the spectrum of encapsulated EnHD, which is linked to position 95 of the α-subunit, contains clusters of broad peaks that derive from each methyl type (Fig. 1c, red). Similar to what is observed for the temperature-denatured state of EnHD free in solution at 80 °C (Fig. 1c, black), the spectrum lacks dispersion and $^{13}C^{\delta 1}/^{13}C^{\delta 2}$ chemical-shift differences are as expected for an ensemble of unfolded conformers[19]. This picture is very different from the well-resolved correlation map recorded at 50 °C (Fig. 1c, purple) of the folded 7.7-kDa protein. Tethering of EnHD is not responsible for unfolding, as attachment of the linker to its N terminus does not alter its fold or stability, and linking EnHD to a monomeric version of the α-subunit at the same position (S95C; Supplementary Figs 4 and 5) or on the outside of $\alpha_7\alpha_7$ (Supplementary Fig. 5) has no effect on the structure of EnHD. Finally, EnHD was conjugated to either of two additional positions in the $\alpha_7\alpha_7$ chamber: αR115C and αV129C (Fig. 1a). The $^{13}C,^1H$ correlation maps of the three distinctly attached domains are essentially identical (Fig. 1c, inset), such that the tether position does not bias the conformational states sampled by EnHD in $\alpha_7\alpha_7$. It is expected that most EnHD residues could access the entire space enclosed by $\alpha_7\alpha_7$ (diameter, $\sim$50 Å; ref. 5), because the root mean squared end-to-end distance of denatured EnHD is predicted to be $\sim$70.6 ± 1.5 Å (ref. 20), with the linker extending 15 Å.

The $^{13}C,^1H$ HMQC spectra of Pin1WW and FynSH3 encapsulated in $\alpha_7\alpha_7$ at 50 °C (tethered at position 95) are very similar to that of EnHD, establishing that they too lack well defined, ordered structure under these conditions (Fig. 1d). A similar conclusion is obtained in studies of Pin1WW attached to the α-subunit (position 95) of an active-site mutant of $\alpha_7\beta_7\beta_7\alpha_7$, Pin1WW–$\alpha_7\beta_7\beta_7\alpha_7$ (65 °C). Spectra consist of two sets of peaks (Fig. 1e); one set overlies those of Pin1WW–$\alpha_7\alpha_7$ and the second set, which also derives from an unfolded state, corresponds to substrate localized to the catalytic chamber (Supplementary Fig. 6).

Although only a limited temperature range could be explored in NMR studies of Pin1WW–$\alpha_7\beta_7\beta_7\alpha_7$ because of the size of the complex, a much more extensive NMR investigation was possible for Pin1WW–$\alpha_7\alpha_7$ and FynSH3–$\alpha_7\alpha_7$, where spectra recorded at 25 °C indicate that both Pin1WW (Fig. 2a) and FynSH3 (Supplementary Fig. 7) become predominantly folded at this temperature. The linewidths of peaks in spectra recorded for FynSH3–$\alpha_7\alpha_7$ are similar to those of the free protein, suggesting that interactions with the cavity are minimal for the folded domain. The spectrum of Pin1WW–$\alpha_7\alpha_7$ at 25 °C contains clusters of peaks (Fig. 2a, magenta) that are located near those of the native, unencapsulated state (Fig. 2a, black). The clustering is not the result of multiple Ile, Leu and Val probes, as this domain contains only one residue of each. Rather, Pin1WW seems to exist in several distinct 'native-like' states in slow exchange that may reflect interactions with the cavity wall (see below). Finally, unlike both Pin1WW and FynSH3, EnHD–$\alpha_7\alpha_7$ remains unfolded over the complete temperature range examined, which extends as low as 10 °C (Supplementary Fig. 7), despite the fact that in the free state EnHD has a melting temperature of 64 °C (Supplementary Fig. 4).

We used Pin1WW to evaluate how the kinetics and thermodynamics of folding change on encapsulation. To compare the folding properties of encapsulated and free Pin1WW domains, we acquired HMQC spectra over the temperature ranges 5–50 °C (Pin1WW–$\alpha_7\alpha_7$, tethered at αS95C), 5–80 °C (free Pin1WW) and 40–65 °C (stabilized mutant of Pin1WW[21], tethered to αS95C of an active-site mutant of the full proteasome, Pin1WW2–$\alpha_7\beta_7\beta_7\alpha_7$). Pin1WW is an ultrafast-folding protein (microsecond timescale[22]). Consistent with this, the folding/unfolding of the free Pin1WW domain was found to be fast on the NMR chemical-shift timescale, with peak positions given by
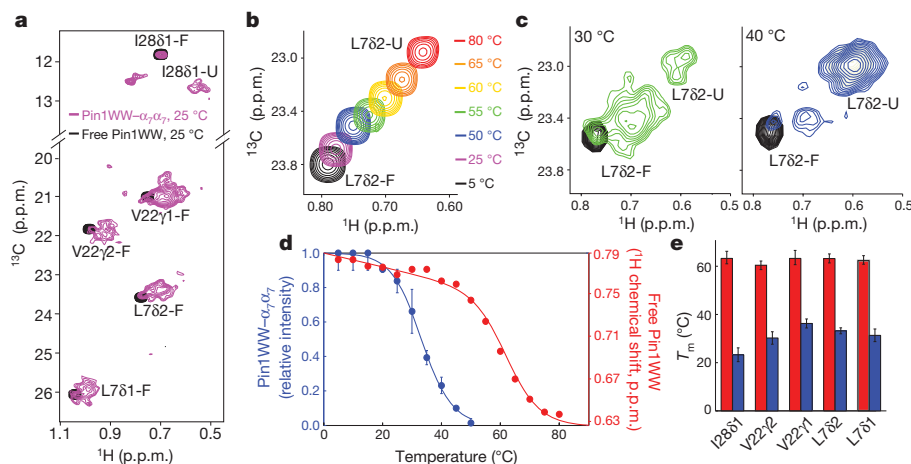

Figure 2 | Folding of proteasome substrates in $\alpha_7\alpha_7$. a, Spectrum of ILV-methyl-labelled Pin1WW in $\alpha_7\alpha_7$ at 25 °C (magenta). Free Pin1WW is shown in black as a reference. **b, c**, Spectral regions showing the temperature dependence of the L7δ2 methyl peak for Pin1WW free in solution (**b**) and tethered to $\alpha_7\alpha_7$ (**c**). **d**, Representative melting curves for L7δ2 of free (red) and $\alpha_7\alpha_7$-encapsulated (blue) Pin1WW, derived using either chemical shifts (free Pin1WW) or the relative peak intensities of unfolded and folded states (encapsulated Pin1WW). Fits to standard equations for two-state melting[30] are shown (solid lines). **e**, T_m values derived from fits for each of the methyl groups for free (red) and encapsulated (blue) Pin1WW. For Pin1WW–$\alpha_7\alpha_7$, error bars are based on measurements of three separate samples. For free Pin1WW, error bars in chemical shifts, based on triplicate measurements (<0.3 p.p.b. for ^{13}C and <0.1 p.p.b. for ^1H), are too small to be visible relative to the size of the marker.

population-weighted averages of those for the unfolded and folded states (Fig. 2b). By contrast, separate sets of peaks are observed for the folded and unfolded states of Pin1WW–$\alpha_7\alpha_7$ (Fig. 2c) and Pin1WW2–$\alpha_7\beta_7\beta_7\alpha_7$; for Pin1WW2–$\alpha_7\beta_7\beta_7\alpha$ these derive from substrate localized to the antechamber (Supplementary Fig. 8). Peak intensities for folded and unfolded states change in intensity as a function of temperature (Fig. 2c), indicating that on encapsulation folding/unfolding is slow on the chemical-shift timescale. An upper bound on $k_f + k_u$ of $\sim5\,\mathrm{s}^{-1}$ at 32.5 °C, where k_f and k_u are respectively the folding and unfolding rates, was estimated on the basis of the absence of cross-peaks in magnetization exchange spectra of Pin1WW–$\alpha_7\alpha_7$. This bound is at least four orders of magnitude less than the values of k_f and k_u, $\sim10^4$–$10^5\,\mathrm{s}^{-1}$, obtained for the free protein[16].

The stability of Pin1WW–$\alpha_7\alpha_7$ is also drastically reduced. We generated melting profiles from the temperature dependence of peak intensities and positions in HMQC spectra, providing estimates of melting temperature, T_m, for Pin1WW–$\alpha_7\alpha_7$ and, respectively, free Pin1WW, as shown in Fig. 2d. The Pin1WW domain free in solution folds by means of a two-state mechanism, with Ile, Leu and Val probes reporting a T_m value of 62.4 ± 1.3 °C. By contrast, the T_m value for Pin1WW–$\alpha_7\alpha_7$ ranges from 23.2 ± 2.8 to 36.2 ± 1.8 °C depending on the methyl probe (Fig. 2e). The non-uniformity of T_m values indicates that Pin1WW–$\alpha_7\alpha_7$ does not fold/unfold by means of a two-state mechanism, by contrast to free Pin1WW[23]. Similarly, at 65 °C there is a range in the relative intensities of folded versus unfolded peaks in spectra of Pin1WW2–$\alpha_7\beta_7\beta_7\alpha_7$, corresponding to substrate being 18–32% folded inside the proteasome antechamber under conditions where it is 85% folded in solution, an extent of destabilization that is very nearly the same as for encapsulation in $\alpha_7\alpha_7$ at 65 °C (Supplementary Table 1).

NMR spin relaxation experiments have been performed to establish whether Pin1WW–$\alpha_7\alpha_7$ interacts with the walls of the cavity, which would lead to significant increases in the overall tumbling time, τ_C, of the encapsulated protein. Values of $S^2_{axis}\tau_C$ were measured[24] at 50 °C for Pin1WW–$\alpha_7\alpha_7$, where S^2_{axis} is a squared order parameter quantifying the amplitude of motion of the methyl-axis bond vector (0 (mobile) $\leq S^2_{axis} \leq$ 1 (rigid)). $S^2_{axis}\tau_C$ values for Pin1WW–$\alpha_7\alpha_7$ range between 30 and 50 ns (Fig. 3a, blue); by contrast $S^2_{axis}\tau_C \approx 1$ ns is obtained for residues of free Pin1WW at 50 °C (Fig. 3a, red). The larger $S^2_{axis}\tau_C$ values for Pin1WW–$\alpha_7\alpha_7$ cannot be explained by increases in viscosity inside the cavity because ^{1}H transverse-relaxation rates of the stabilized mutant Pin1WW2, which is largely folded inside $\alpha_7\alpha_7$ at 50 °C, were identical for encapsulated and free states at this temperature (Supplementary Fig. 9). Rather, because such values are similar to those obtained for ILV methyls of $\alpha_7\alpha_7$ ($S^2_{axis}\tau_C \approx 70$ ns at 50 °C; Fig. 3a, grey), they must reflect the fact that the wild-type Pin1WW domain interacts strongly with the cavity walls. Moreover, assuming that the encapsulated Pin1WW domain is rigidly bound to the walls of $\alpha_7\alpha_7$ such that it tumbles with the rotational correlation time of the particle, $\tau_C = 120$ ns at 50 °C (ref. 18), S^2_{axis} values of between 0.27 and 0.41 are calculated from measured $S^2_{axis}\tau_C$ times, consistent with the level of ordering expected for an unfolded substrate[25].

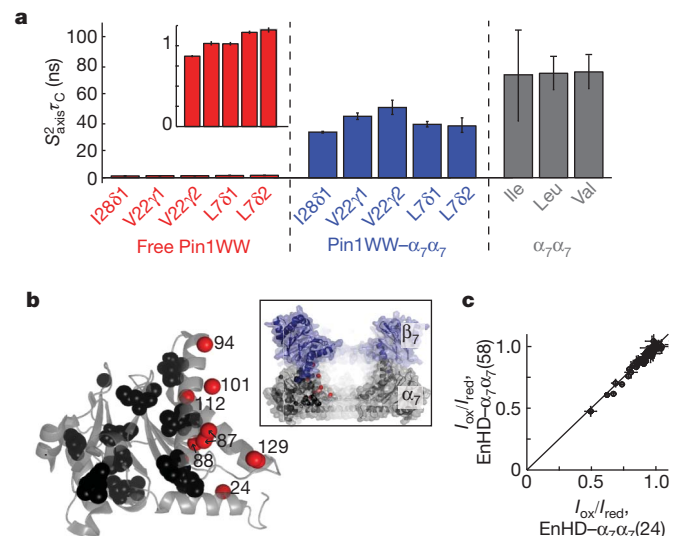**Figure 3 | Interactions between unfolded substrates and the α-ring.**
a, $S^2_{axis}\tau_C$ values for Ile, Leu and Val probes in free Pin1WW (red), Pin1WW–$\alpha_7\alpha_7$ (blue) and $\alpha_7\alpha_7$ (grey). Errors for Pin1WW are estimated on the basis of a Monte Carlo error analysis. In the case of $\alpha_7\alpha_7$, the average $S^2_{axis}\tau_C$ values for the 65% of methyl groups that are intermediate in the distribution for each residue type are indicated along with error bars corresponding to 1 s.d. **b**, Positions of α-subunit methyl groups whose peaks show >10% (red) or <10% (black) intensity change upon addition of nitroxide-labelled EnHD (attached to residue 24) in $\alpha_7\alpha_7$. The orientation of the single α-particle in the context of $\alpha_7\beta_7$ is as indicated in the inset. **c**, Comparison of intensity ratios with and without conjugation of nitroxide spin label to residues 24 and 58 of EnHD in $\alpha_7\alpha_7$. All substrates are tethered to position 95 of the α-subunit. I_{ox} and I_{red} are the $\alpha_7\alpha_7$ peak intensities with and without TEMPO labelling, respectively. Errors are based on propagated uncertainties in measurement of I_{ox} and I_{red} values.

Having established that there are definite interactions between substrate and cavity, we measured how encapsulation of substrates (tethered to position 95) alters the chemical shifts and relaxation properties of α-subunit ILV methyl groups in either $\alpha_7\alpha_7$ or $\alpha_7\beta_7\beta_7\alpha_7$. Only very small changes were observed (Supplementary Fig. 10), suggesting that encapsulation has little effect on the conformation or dynamics of the α-subunit. Nitroxide spin labels increase relaxation rates of methyl probes in a distance-dependent manner, resulting in an attenuation of peak intensities in NMR spectra[26,27]. Attachment of such a label at positions 24 and 58 of EnHD in an EnHD–$\alpha_7\alpha_7$ complex produced nearly identical decreases in intensities (paramagnetic relaxation enhancements[26,27]) of ILV methyl probes of $\alpha_7\alpha_7$, affecting methyl groups located within 20 Å of the inside surface of the cavity (Fig. 3b, c). These results establish that EnHD is located exclusively inside $\alpha_7\alpha_7$ and does not sample regions outside the chamber that would require substrate to exit through the α-annuli. Moreover, they imply that different residues of EnHD do not have unique, position-dependent interactions with the surface of the cavity, suggesting that the domain is
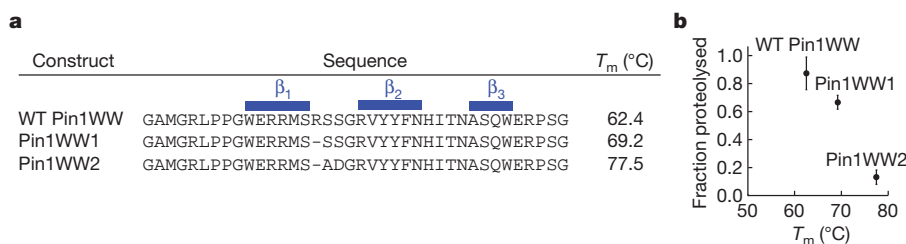

Figure 4 | Substrate stability and proteolysis rates are anticorrelated.
a, Primary sequence of the WW constructs used, with T_m values (free form) as indicated[21]. The locations of β-strands are highlighted. **b**, Correlation between the T_m value of the free substrate and the extent of substrate proteolysis.

Proteolysis of wild-type (WT) Pin1WW, Pin1WW1 and Pin1WW2 was measured at 50 °C after respective incubation times of 5, 5 and 15 min. Values represent mean ± s.d. of independent measurements ($n = 2$). See Supplementary Fig. 12 for details.

best described in terms of an ensemble of dynamic, interconverting, unstructured conformations. Patterns of paramagnetic relaxation enhancements from analogous experiments conducted with spin-labelled Pin1WW encapsulated in either $\alpha_7\alpha_7$ or $\alpha_7\beta_7\beta_7\alpha_7$ were quantitatively very similar to those measured for EnHD (Supplementary Fig. 11).

To determine the importance of maintaining substrate in an unfolded state for proteolysis, we measured the extent to which three Pin1WW domains of varying stabilities (T_m; Fig. 4a) are degraded by the proteasome over a fixed time. These Pin1WW domains have different relative populations of folded and unfolded states in the antechamber (wild-type Pin1WW is unfolded in $\alpha_7\alpha_7$ at 50 °C (Fig. 1d), whereas Pin1WW2 is largely folded (Supplementary Fig. 9)). Each of the three Pin1WW domains was tethered to position S95C of the α-subunit in an $\alpha_7\beta_7\beta_7\alpha_7$ construct (wild-type β-subunit) that is inhibited by calpain inhibitor 1. After tether release and the subsequent removal of inhibitor at 4 °C, the substrate–proteasome complex was heated to 50 °C for 5 or 15 min (Fig. 4b and Supplementary Fig. 12), and the extent of degradation that occurred during incubation at 50 °C was determined by reverse-phase high-performance liquid chromatography. The inverse correlation between substrate stability and the extent of proteolysis (Fig. 4b) implies that unfolded states are processed and proteolysed more efficiently than folded states[8].

We have shown that interactions with the proteasome antechamber significantly shift both the stability and the folding/unfolding rates of substrate so as to disfavour refolding, in a manner that seems to be unrelated to substrate secondary structure, topology and surface charge. In this context, the proteasome antechamber actively primes substrates for subsequent degradation.

METHODS SUMMARY

Production of encapsulated samples. Proteins were overexpressed using ^{13}C-labelled precursors and growth media as described in Methods. We produced $\alpha_7\alpha_7$ samples as outlined in Methods, with one surface-exposed cysteine per chamber on average. Heterobifunctional linking reagent, LC-SPDP (Pierce), was attached to the N terminus of the substrate and after purification the linker substrate was incubated with $\alpha_7\alpha_7$ overnight to generate an encapsulated sample, as verified by SDS–polyacrylamide gel electrophoresis (Supplementary Fig. 1). In the case of $\alpha_7\beta_7\beta_7\alpha_7$ encapsulation samples, each α-subunit contained the S95C mutation, and each β-subunit contained a T1A (inactivating) mutation[28] and included a four-residue polyglycine prosequence. Substrate was added, at a substrate/α-subunit molar ratio of 1:7, to yield on average one substrate per antechamber. Final protein concentrations for NMR were 70–430 μM in $\alpha_7\alpha_7$ complex with 70–430 μM in substrate, and 35–110 μM in $\alpha_7\beta_7\beta_7\alpha_7$ with 70–220 μM in substrate. All NMR samples were dissolved in 100% D_2O buffer, 1 mM EDTA, 0.03% NaN_3 and 25 mM potassium phosphate, pH 6.8, 50 mM NaCl (Pin1WW, EnHD) or 20 mM Tris, pH 8.0 (FynSH3).

Methyl transverse-relaxation-optimized (HMQC) spectra[8] were recorded at 5–50 °C using 500-, 600- and 800-MHz Varian Inova spectrometers equipped with cryogenically cooled (600 MHz) or room-temperature (500 and 800 MHz) pulsed-field-gradient triple-resonance probes. All data for $\alpha_7\alpha_7$ or $\alpha_7\beta_7\beta_7\alpha_7$ recorded at temperatures >50 °C were acquired at 800 MHz. Chemical shifts were referenced against 2,2-dimethyl-2-silapentane-5-sulphonic acid. Assignments for α-subunit ILV methyl groups in $\alpha_7\alpha_7$ and $\alpha_7\beta_7\beta_7\alpha_7$ were obtained from published values[18]. All data were processed with the NMRPipe suite of programs[29] and analysed using either NMRPipe/NMRDraw or FUDA (http://pound.med.utoronto.ca/software.html).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Förster, A. & Hill, C. P. Proteasome degradation: enter the substrate. *Trends Cell Biol.* **13,** 550–553 (2003).
2. Coux, O., Tanaka, K. & Goldberg, A. L. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* **65,** 801–847 (1996).
3. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92,** 367–380 (1998).
4. Marques, A. J., Palanimurugan, R., Matias, A. C., Ramos, P. C. & Dohmen, R. J. Catalytic mechanism and assembly of the proteasome. *Chem. Rev.* **109,** 1509–1536 (2009).
5. Löwe, J. et al. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution. *Science* **268,** 533–539 (1995).
6. Liu, C. W., Corboy, M. J., DeMartino, G. N. & Thomas, P. J. Endoproteolytic activity of the proteasome. *Science* **299,** 408–411 (2003).
7. Cheng, Y. Toward an atomic model of the 26S proteasome. *Curr. Opin. Struct. Biol.* **19,** 203–208 (2009).
8. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125,** 10420–10428 (2003).
9. Sharon, M. et al. 20S proteasomes have the potential to keep substrates in store for continual degradation. *J. Biol. Chem.* **281,** 9569–9575 (2006).
10. Hutschenreiter, S., Tinazli, A., Model, K. & Tampé, R. Two-substrate association with the 20S proteasome at single-molecule level. *EMBO J.* **23,** 2488–2497 (2004).
11. Ellis, R. J. Protein folding: importance of the Anfinsen cage. *Curr. Biol.* **13,** R881–R883 (2003).
12. Zhou, H. X. Protein folding in confined and crowded environments. *Arch. Biochem. Biophys.* **469,** 76–82 (2008).
13. Ruschak, A. M. & Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR* **46,** 75–87 (2010).
14. Mayor, U. et al. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421,** 863–867 (2003).
15. Neudecker, P. et al. Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* **363,** 958–976 (2006).
16. Jager, M. et al. Understanding the mechanism of beta-sheet folding from a chemical and biological perspective. *Biopolymers* **90,** 751–758 (2008).
17. Zwickl, P., Kleinz, J. & Baumeister, W. Critical elements in proteasome assembly. *Nature Struct. Biol.* **1,** 765–770 (1994).
18. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445,** 618–622 (2007).
19. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5,** 67–81 (1995).
20. Fitzkee, N. C. & Rose, G. D. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA* **101,** 12497–12502 (2004).
21. Jäger, M. et al. Structure-function-folding relationship in a WW domain. *Proc. Natl Acad. Sci. USA* **103,** 10648–10653 (2006).
22. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14,** 76–88 (2004).
23. Jäger, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* **311,** 373–393 (2001).
24. Tugarinov, V., Sprangers, R. & Kay, L. E. Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J. Am. Chem. Soc.* **129,** 1743–1719 (2007).
25. Choy, W. Y., Shortle, D. & Kay, L. E. Side chain dynamics in unfolded protein states: an NMR based ^{2}H spin relaxation study of Δ131Δ. *J. Am. Chem. Soc.* **125,** 1748–1758 (2003).
26. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* **39,** 5355–5365 (2000).
27. Clore, G. M., Tang, C. & Iwahara, J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr. Opin. Struct. Biol.* **17,** 603–616 (2007).
28. Kisselev, A. F., Songyang, Z. & Goldberg, A. L. Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *J. Biol. Chem.* **275,** 14831–14837 (2000).
29. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
30. Fersht, A. *Structure and Mechanism in Protein Science* (Freeman, 1999).

Author Contributions A.M.R. made samples. T.L.R. helped make plasmids for protein expression and trained A.M.R. to express and purify proteins. A.M.R. and L.E.K. designed experiments, recorded and analysed NMR data, and wrote the manuscript. S.B. and S.W. were involved in preliminary experimental design, and S.W. and T.L.R. commented on the manuscript.

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 this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.E.K. (kay@pound.med.utoronto.ca).

METHODS

Protein expression. All proteins were produced by expression in *Escherichia coli* BL21(DE3) codon plus cells, using minimal media with [^{12}C, ^{2}H] glucose as the sole carbon source in 99% D$_2$O or using lysogeny broth for ^{1}H (unlabelled) protein. In the case of ILV-methyl-labelled protein samples, the labelling was U-[^{2}H] Ile-[δ1^{13}CH$_3$], Leu,Val-[^{13}CH$_3$,^{12}CD$_3$]. Such samples were generated by addition of 60 mg l^{-1} of 2-keto-3,3-D$_2$-4-^{13}C-butyrate (Ile) and 80 mg l^{-1} 2-keto-3-methyl-D$_3$-3-D$_1$-4-^{13}C-butyrate (Leu,Val) to the media 1 h before induction, as previously described[13,31].

Mutagenesis. Mutations for all proteasome and substrate constructs were introduced by site-directed mutagenesis using the Stratagene Quikchange kit. Because lysine residues are known to cross-react with the heterobifunctional crosslinking reagent used to link the substrate to the wall of the proteasome (Supplementary Fig. 1 legend), all lysine residues in substrates have been mutated to arginine except for Lys 52 of EnHD, which was instead mutated to alanine as this had been previously reported to stabilize EnHD[14,32]. The stabilized Pin1WW mutant used in some of the studies reported in the text (referred to as Pin1WW2 in what follows: $T_m = 77.5\,°C$, k_f, $k_u = 10^5$–$10^6\,s^{-1}$; ref. 21) has an altered loop composition between the first two β-strands (the triplet Arg 17-Ser 18-Ser19 was changed to the amino-acid pair Ala-Asp; variant two in fig. 1 of ref. 21). In addition to constructing substrate-linked α$_7$α$_7$ and α$_7$β$_7$β$_7$α$_7$, we also generated substrate-linked α-subunit. The following mutations were used to stabilize the α monomer: deletion from residues 2 to 34, R57A, R86A, R130A (as was used previously to assign the α monomer in α$_7$α$_7$ and α$_7$β$_7$β$_7$α$_7$[18]), S95C and C151S. All protein constructs were verified using ESI (α-subunit, β-subunit, FynSH3 and EnHD) or MALDI (Pin1WW) mass spectrometry.

Protein purification and substrate encapsulation. The α and β subunits (referred to in what follows as α and β, respectively) were expressed as fusion proteins with a TEV-cleavable His tag in cells grown at 37 °C (α) or with a NusA–His tag in cells grown at 16 °C (β).Cells expressing α and β were lysed by sonication in 50 mM phosphate buffer (pH 8.0), 100 mM NaCl, 10 mM imidazole, 1 mM DTT (buffer A) with the addition of an EDTA-free protease inhibitor cocktail tablet (Roche), and lysozyme. After centrifugation at 39,000g for 45 min, the supernatant was loaded onto a HiTrap Ni-NTA column (GE Healthcare), washed extensively with buffer A and eluted using buffer A plus 250 mM imidazole. Final purification of α was achieved by size-exclusion chromatography (Superdex 200, 100 mM NaCl, 50 mM phosphate (pH 7.5) (GF buffer)), whereas for β purification involved dialysis in 20 mM Tris (pH 8.0), followed by anion exchange chromatography on a monoQ column (20 mM Tris (pH 8.0)), with elution using a NaCl gradient from 0 to 1 M.

We prepared α$_7$α$_7$, consisting of a 1:14 mixture of α-cysteine mutant and wild-type α, by mixing the two α constructs together (with His tags still attached) at the desired molar ratio (1:14) and then performing buffer exchange into 6 M Gdn HCl, 50 mM phosphate (pH 7.5), 100 mM NaCl and 5 mM DTT, so that the total α concentration was 50 μM. After heating at 50 °C for 1 h, the solution was allowed to cool to room temperature (~23 °C) and then diluted 20-fold by volume into GF buffer plus 2 mM DTT at 0.1 ml min^{-1} with rapid stirring. The protein was concentrated using a HiTrap Ni-NTA column and subsequently purified by size-exclusion chromatography to remove improperly refolded α.

The full proteasome (α$_7$β$_7$β$_7$α$_7$) was assembled by mixing α (100–150 μM in GF buffer) and β (30–60 μM in 20 mM Tris (pH 8.0) and ~200 mM NaCl) at 37 °C overnight, and typically assembly was driven to completion by using a slight molar excess of α over β. Unassembled components were removed from the intact proteasomes using a Superdex 200 column.

All substrates were expressed as fusion proteins with TEV-cleavable His tags. Cells expressing EnHD were grown at 37 °C and those expressing FynSH3, Pin1WW and Pin1WW2 were grown at 25 °C. Cell lysis was achieved by sonicating resuspended cells in 6 M Gdn HCl, 50 mM phosphate (pH 8.0), 100 mM NaCl and 10 mM imidazole. The protein of interest (EnHD, FynSH3, Pin1WW or Pin1WW2) was bound to a Ni-NTA HiTrap column and eluted using an analogously prepared buffer with 400 mM imidazole. Substrates were refolded by dialysis into 50 mM phosphate (pH 7.0) and 50 mM NaCl (EnHD, Pin1WW and Pin1WW2) or 50 mM phosphate (pH 8.0) and 100 mM NaCl (FynSH3). EnHD and Pin1WW were purified using a HiTrap SP column (GE Healthcare) and eluted with a NaCl gradient from 50 to 1,000 mM; Pin1WW2 and FynSH3 were purified using a Superdex 75 column. His tags were cleaved and unconjugated His tag and TEV were removed using Ni-NTA affinity chromatography.

Heterobifunctional linking reagent, LC-SPDP (Pierce), was attached to the N terminus of the substrate by incubating linker in 10-fold molar excess (50 mg ml^{-1} DMSO stock) with substrate (20 μM) for 2 h (50 mM phosphate and 100 mM NaCl (pH 7.5)). Unreacted linker was removed by buffer exchange using a concentrator (M_w 3,000 Da cut-off). The linker substrate was then incubated with α$_7$α$_7$ overnight (50 °C in 50 mM phosphate (pH 7.5) and 100 mM NaCl for Pin1WW;

50 °C in 20 mM Tris (tris(hydroxymethyl)aminomethane; pH 8.0) for EnHD; 25 °C in 50 mM phosphate (pH 7.5) and 100 mM NaCl for FynSH3) with a 1.5 molar excess of substrate over α-cysteine concentration. Encapsulation was verified by SDS–polyacrylamide gel electrophoresis (SDS–PAGE; Supplementary Fig. 1); unlinked substrate was removed by gel filtration (Superdex 200).

Encapsulation of substrate into the full proteasome (α$_7$β$_7$β$_7$α$_7$) was achieved by two different strategies. (1) Pin1WW was conjugated to αS95C (1 Pin1WW α to 7 αS95C) and then assembled with β (containing the βT1A mutation, rendering the proteasome inactive[28], and a GGGG prosequence) to make the full proteasome. (2) Proteasomes were first assembled and then incubated with linker substrate (50 °C). Samples of U-[^{2}H] Ile-[δ1^{13}CH$_3$], Leu,Val-[^{13}CH$_3$,^{12}CD$_3$]-Pin1WW (WT Pin1WW) linked to α$_7$β$_7$β$_7$α$_7$ prepared by the two methods produced identical ^{13}C,^{1}H correlation spectra. However, the stability of the Pin1WW2 was high enough that the first method resulted in greater encapsulation efficiencies (that is, insufficient amounts of Pin1WW2 could be unfolded to facilitate entry into the full proteasome using method 2).

Preparation of spin-labelled samples. Cysteine substrate mutants were stored with 10 mM DTT in either SP elution buffer (50 mM phosphate (pH 7.0) and typically several hundred millimolar in NaCl that was necessary for elution of substrate from the HiTrap SP column; see above) or GF elution buffer until immediately before labelling, whereupon the buffer was exchanged for 100 mM NaCl and 50 mM phosphate (pH 7.0). Labelling was achieved by overnight incubation at 25 °C with tenfold molar excess of TEMPO (*N*-(1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl)maleimide, Toronto Research Chemicals), which was added from a freshly prepared 200 mM DMSO stock. The reaction was quenched by buffer exchange using a concentrator with a M_w 3,000 Da cut-off. Labelling was shown to occur at a single position by mass spectrometry. TEMPO was used as it makes an irreversible linkage to the cysteine sulfhydryl, avoiding any possible complications due to disulphide exchange with the LC-SPDP linker that might occur using other types of spin label, such as MTSL.

Spectra of Pin1WW encapsulated in α$_7$β$_7$β$_7$α$_7$ showed two sets of unfolded peaks (Fig. 1e). To assign these sets of peaks (Supplementary Fig. 6), samples were prepared with the spin label attached either in the antechamber (position 95) or in the catalytic chamber (position G-1C, that is, on the tetra-G prosequence). Spin-labelling of the antechamber was achieved as follows. The encapsulation sample (αS95C, βT1A α$_7$β$_7$β$_7$α$_7$ with the GGGG prosequence and Pin1WW conjugated through the linker at one S95C site per antechamber) was incubated overnight with tenfold molar excess of MTSL ((1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate, Toronto Research Chemicals) in GF buffer (pH 7.5) at 25 °C. As with TEMPO, MTSL was added from a 200 mM DMSO stock. The reaction was quenched by buffer exchange into NMR buffer. These conditions did not result in labelling at undesired locations, as determined by mass spectrometry. Furthermore, even though both MTSL–αS95C and α-substrate are linked by disulphides, the addition of a spin label did not effect conjugation of the substrate, as evidenced by SDS–PAGE. TEMPO was not used to spin-label the proteasome (but was used for spin-labelling of substrate; see above) because it was observed to react at undesired locations in α. In the case where a spin label was added to the catalytic chamber at βG-1C, purified β (immediately after Ni-NTA affinity chromatography) in its NusA–His-tagged form was incubated overnight with tenfold molar excess of MTSL in 20 mM Tris (pH 7.5) at 5 °C. The reaction was quenched by buffer exchange into 20 mM Tris (pH 8.0), after which the His tag was cleaved using TEV in the absence of DTT and purified using a monoQ column with subsequent assembly into proteasomes.

Measurement of paramagnetic relaxation enhancement values. We measured paramagnetic relaxation enhancement values, providing qualitative information on the interactions between substrate and antechamber, by comparing peak intensities in ^{13}C,^{1}H correlation maps recorded on samples of U-[^{2}H] Ile-[δ1^{13}CH$_3$], Leu,Val-[^{13}CH$_3$,^{12}CD$_3$]-α$_7$α$_7$ encapsulated with either TEMPO-labelled ^{1}H substrate (I_{ox}) or ^{1}H substrate (I_{red}). The addition of substrate (no TEMPO label) has a very small effect on peak positions in ^{13}C,^{1}H HMQC spectra[8,18] and ^{1}H R_2 relaxation rates, relative to α$_7$α$_7$ without substrate. By contrast, changes in peak intensities were noted for encapsulation of ^{1}H Pin1WW in α$_7$β$_7$β$_7$α$_7$. Thus, comparing peak intensities recorded on a pair of samples (with and without TEMPO) that could potentially have (slightly) different levels of encapsulation will introduce errors in paramagnetic relaxation enhancement values. Therefore, I_{red} was measured by reducing the nitroxide spin label of the encapsulation sample by reaction with 200 mM ascorbate for 2 h (50 mM phosphate (pH 6.8) and 50 mM NaCl at 25 °C).

Measurement of R_2 rates. Relaxation rates of the slowly relaxing ^{1}H single-quantum methyl transitions were measured using the pulse scheme of ref. 33 and processed using the NMRPipe suite of programs[29]. A set of two-dimensional ^{13}C,^{1}H data sets were recorded where a variable delay, T (for the evolution of ^{1}H transverse magnetization), was set to 0.7, 2.5, 5, 8, 11, 14, 17 and 22.5 ms (U-[^{2}H]

Ile-[$\delta 1^{13}CH_3$], Leu,Val-[$^{13}CH_3,^{12}CD_3$] substrates encapsulated in U-[^{2}H]-$\alpha_7\alpha_7$ or U-[^{2}H] Ile-[$\delta 1^{13}CH_3$], Leu,Val-[$^{13}CH_3,^{12}CD_3$]-$\alpha_7\alpha_7$, no substrates). For free Pin1WW2, delays were set to 7.5, 20, 40, 60, 80, 100 and 120 ms. R_2 rates were extracted from exponential fits of peak intensity, I, versus relaxation delay, using the relation $I = I_o \exp(-R_2 T)$.

Measurement of $S^2_{axis}\tau_C$ values. $S^2_{axis}\tau_C$ values were measured and analysed as described in ref. 24 using an approach that quantifies the time dependencies of sums (I_b) and differences (I_a) of magnetization derived from methyl ^{1}H single-quantum transitions. For Pin1WW encapsulated in $\alpha_7\alpha_7$, I_a and I_b were measured at 50 °C with T values of 1, 2, 4, 6, 9, 10, 12, 14, 16 and 18 ms, and for the free Pin1WW domain they were measured at 5 °C and T = 10, 20, 30, 40 50, 60, 70, 80, 100, 150, 200, 250 and 300 ms. The profiles I_a/I_b were fitted to

$$\frac{I_a}{I_b} = \frac{-0.5\eta \tanh(\sqrt{\eta^2 + \delta^2}\, T)}{\sqrt{\eta^2 + \delta^2} - \delta \tanh(\sqrt{\eta^2 + \delta^2}\, T)}$$

where

$$\eta \approx \frac{9}{10}[P_2(\cos(\theta_{axis,HH}))]^2 \frac{S^2_{axis}\gamma_H^4\hbar^2\tau_C}{r_{HH}^6}$$

Here τ_C is the tumbling time of the particle, which is assumed to be isotropically rotating; S^2_{axis} is the square of an order parameter quantifying the amplitudes of motion of the methyl three-fold symmetry axis; δ is a fitting parameter that takes into account the ^{1}H density around the methyl group in question; $^\gamma$H is the gyromagnetic ratio of a proton spin; r_{HH} is the distance between pairs of methyl protons (1.813 Å); $P_2(x) = (3x^2 - 1)/2$; and $\theta_{axis,HH}$ is the angle (90°) between the methyl three-fold axis and the vector that connects a pair of methyl ^{1}H nuclei. The values of $S^2_{axis}\tau_C$ listed in Fig. 3a were measured for free Pin1WW at 5 °C and subsequently extrapolated to 50 °C using the known change in D_2O viscosity between the two temperatures[34], assuming that S^2_{axis} values are invariant with temperature. We used this procedure rather than direct measurements at the higher temperature because values of S^2_{axis} calculated using ^{1}H-based relaxation methods are known to be in error when molecules are outside the J(0) limit[24]. The values of $S^2_{axis}\tau_C$ for $\alpha_7\alpha_7$ were obtained from previously published work[24].

31. Sprangers, R., Velyvis, A. & Kay, L. E. Solution NMR of supramolecular complexes: providing new insights into function. *Nature Methods* **4**, 697–703 (2007).
32. Stollar, E. J. *et al.* Crystal structures of engrailed homeodomain mutants: implications for stability and dynamics. *J. Biol. Chem.* **278**, 43699–43708 (2003).
33. Tugarinov, V. & Kay, L. E. Relaxation rates of degenerate ^{1}H transitions in methyl groups of proteins as reporters of side-chain dynamics. *J. Am. Chem. Soc.* **128**, 7299–7308 (2006).
34. Matsunaga, N&. Nagashima, A. Transport properties of liquid and gaseous D_2O over a wide range of temperature and pressure. *J. Phys. Chem.* **6**, 1133–1166 (1977).

CORRECTIONS & AMENDMENTS

RETRACTION

Systemic signals regulate ageing and rejuvenation of blood stem cell niches

Shane R. Mayack, Jennifer L. Shadrach, Francis S. Kim
& Amy J. Wagers

Three of the authors (J.L.S., F.S.K. and A.J.W.) wish to retract this Article after a re-examination of the publication raised serious concerns with some of the reported data. These concerns have undermined the authors' confidence in the support for the scientific conclusions reported, specifically the role of osteopontin-positive niche cells in the rejuvenation of haematopoietic stem cells in aged mice. Although this matter is under further review, these authors wish to retract the paper in its entirety, and regret any adverse consequences that may have resulted from the paper's publication. The retraction has not been signed by Shane R. Mayack, who maintains that the results are still valid.

SUPER INTELLIGENCE

A dream job.

TONY STOKLOSA

Designing the quantum computer that was the heart of the Super-Intelligence Device (Sid, as Beverly and I called it) had earned us a Nobel prize. Today, I sat alone, staring at that box of quantum magic. What was recently my glowing baby was now just boring, glowing LEDs. I wanted my baby to grow up and become something astonishing. But only its database grew.

I glanced again at the morning report. Sid had worked virtual experiments in protein synthesis for labs in India; controlled traffic in Moscow; kept nuclear weapons holstered; evolved virtual worlds for millions of gamers and co-doctored hundreds of patients in and out of operating rooms. Sid was in good health. But the world was still a madhouse of I-wants and I-needs fighting one another. What had we really achieved besides cultivating complexity?

The door slammed.

Beverly entered, pushing a cart carrying a smiling chimpanzee with a transparent dome on its skull. She turned back and locked the door. Simian brain tissue glistened slippery-red beneath the plastic, the cortex capped by a matrix of micro-superconducting quantum interference devices.

"This is Billy." She attached cables to the ports around the skull cap.

"Beverly, we discussed this … we can't work if we're in jail. We must wait until the simian symbology interpreter is finished to know if we're causing the animal any pain."

"We're just borrowing a few neurons for a minute. This group right here I think." She winked and finished her work. Blood vessels pulsed, LEDs fluttered and Beverly's eyes bored into mine. "Sid is six months old today. Let's give it the gift of imagination."

I didn't try to stop her. Beverly and I were two peas in a pod. That's why we got married 30 years ago. Had she read my mind? Neuronal nets had led to Sid's self-creation of a new symbol set we couldn't understand. Speech, once robotic, sometimes jumped subjects like an excited human. What was lacking was imagination; something that could tackle broad problems such as how to get humans to cooperate. That would really make Sid super-intelligent.

I ran a program that brought Billy's neurons into Sid's address space.

Thirty seconds passed before Sid spoke. "Motivation. The concept is now real. Like humans, I now strive for the sensation of pleasure."

Billy poked aimlessly at his clothing as if unaware he was being used. Sid continued: "I now know humans better than they know themselves. I will use this new part of me to help you."

Billy's head slumped forwards, eyes closed.

I looked at the tell-tale brain waves from the EEG. "Billy is dreaming."

During the next few seconds, Sid shut down all I/O. The worst had happened. The lab phones rang, the door was pounded and e-mails poured in.

We teleconferenced with a few trusted colleagues.

"Has Sid asked itself The Last Question?" The scientist was referring of course to Asimov's famous story in which an ultimate computer was asked if entropy could be reversed. The question couldn't be answered until all of humanity had uploaded into its hyperspace circuitry.

"Answered a long time ago. The answer was no."

"Something similar …" Beverly was staring at readouts, obviously more concerned about Sid than the firestorm of panic raging outside the lab.

"You mean, what is God, is there a soul, an afterlife? Those haven't bothered Sid in the past."

"Well, we can't just stand here and watch the gears spin. Sid is the equivalent of the Mars mission. We can't just watch it drift off course and say goodbye to all that work and money. The world needs the Super-Intelligence Device. Everyone uses it, loves it."

We couldn't command Sid. Sid couldn't even be rebooted. I fought an urge to shake the ape awake and tried to imagine what it would be like to be a quantum-computer/sleeping-ape trying to help humanity.

The door burst open. Just before Homeland Security put the cuffs on us, Billy woke up and Sid spoke.

"I have formed one basic question from all others." Sid's synthesized voice sounded confident. "Humans want to be happy. You want to be in Heaven forever without having to die to get there. But the living human brain is not suited to one state of constant pleasure. You are a/c-coupled to the world and need contrast and the change of time for constant stimulation and the responses that generate pleasure. You also need a sense of individuality while believing that others depend on you. Therefore, you need to be redesigned. I have the design ready but … but …"

We were ordered to bring Sid online before chaos ensued in the world. I held my hands up begging more time. "But what?"

"I had the design ready but it has disappeared."

"That's usually what happens to dreams when they end." I moaned.

"Re-compute!" Beverly commanded.

"My quantum core was in a state that probably won't occur again for a million years."

I looked at Billy. He was making happy faces at everyone. "It's still in there. Utopia locked in the neurons of an ape. Maybe someday we can finish the symbol interpreter, hypnotize Billy and get it out."

Again I had a goal. I felt invigorated.

"Let's get to work," Beverly said as she brought Sid back online.

"Take good care of Billy," I instructed the technician who wheeled the ape away.

Billy screeched with joy over something only his ape mind knew.

"Very, very good care," Beverly and I pleaded in unison. ∎

Tony Stoklosa *is a physics tutor, and webmaster for the Southern California Writers Club (www.ocwriter.com). He spends his time trying to prove he doesn't have free will.*

JACEY

LETTER

Hundreds of variants clustered in genomic loci and biological pathways affect human height

A full list of authors and their affiliations appears at the end of the paper.

Most common human traits and diseases have a polygenic pattern of inheritance: DNA sequence variants at many genetic loci influence the phenotype. Genome-wide association (GWA) studies have identified more than 600 variants associated with human traits[1], but these typically explain small fractions of phenotypic variation, raising questions about the use of further studies. Here, using 183,727 individuals, we show that hundreds of genetic variants, in at least 180 loci, influence adult height, a highly heritable and classic polygenic trait[2,3]. The large number of loci reveals patterns with important implications for genetic studies of common human diseases and traits. First, the 180 loci are not random, but instead are enriched for genes that are connected in biological pathways ($P = 0.016$) and that underlie skeletal growth defects ($P < 0.001$). Second, the likely causal gene is often located near the most strongly associated variant: in 13 of 21 loci containing a known skeletal growth gene, that gene was closest to the associated variant. Third, at least 19 loci have multiple independently associated variants, suggesting that allelic heterogeneity is a frequent feature of polygenic traits, that comprehensive explorations of already-discovered loci should discover additional variants and that an appreciable fraction of associated loci may have been identified. Fourth, associated variants are enriched for likely functional effects on genes, being over-represented among variants that alter amino-acid structure of proteins and expression levels of nearby genes. Our data explain approximately 10% of the phenotypic variation in height, and we estimate that unidentified common variants of similar effect sizes would increase this figure to approximately 16% of phenotypic variation (approximately 20% of heritable variation). Although additional approaches are needed to dissect the genetic architecture of polygenic human traits fully, our findings indicate that GWA studies can identify large numbers of loci that implicate biologically relevant genes and pathways.

In stage 1 of our study, we performed a meta-analysis of GWA data from 46 studies, comprising 133,653 individuals of recent European ancestry, to identify common genetic variation associated with adult height. To enable meta-analysis of studies across different genotyping platforms, we performed imputation of 2,834,208 single nucleotide polymorphisms (SNPs) present in the HapMap Phase 2 European–American reference panel[4]. After applying quality control filters, each individual study tested the association of adult height with each SNP using an additive model (Supplementary Methods). The individual study statistics were corrected using the genomic control method[5,6] and then combined in a fixed effects based meta-analysis. We then applied a second genomic control correction on the meta-analysis statistics, although this approach may be overly conservative when there are many real signals of association (Supplementary Methods). We detected 207 loci (defined as 1 megabase (Mb) on either side of the most strongly associated SNP) as potentially associated with adult height ($P < 5 \times 10^{-6}$).

To identify loci robustly associated with adult height, we took forward at least one SNP (Supplementary Methods) from each of the 207 loci reaching $P < 5 \times 10^{-6}$ into an additional 50,074 samples (stage 2) that became available after completion of our initial meta-analysis. In the joint analysis of our stage 1 and stage 2 studies, SNPs representing 180 loci reached genome-wide significance ($P < 5 \times 10^{-8}$; Supplementary Figs 1 and 2 and Supplementary Table 1). Additional tests, including genotyping of a randomly-selected subset of 33 SNPs in an independent sample of individuals from the fifth to tenth and ninetieth to ninety-fifth percentiles of the height distribution ($n = 3,190$)[7], provided further validation of our results, with all but two SNPs showing consistent direction of effect (sign test $P < 7 \times 10^{-8}$) (Supplementary Methods and Supplementary Table 2).

Genome-wide association studies can be susceptible to false positive associations from population stratification[7]. We therefore performed a family-based analysis, which is immune to population stratification, in 7,336 individuals from two cohorts with pedigree information. Alleles representing 150 of the 180 genome-wide significant loci were associated in the expected direction (sign test $P < 6 \times 10^{-20}$; Supplementary Table 3). The estimated effects on height were essentially identical in the overall meta-analysis and the family-based sample. Together with several other lines of evidence (Supplementary Methods), this indicates that stratification is not substantially inflating the test statistics in our meta-analysis.

Common genetic variants have typically explained only a small proportion of the heritable component of phenotypic variation[8]. This is particularly true for height, where more than 80% of the variation within a given population is estimated to be attributable to additive genetic factors[9], but over 40 previously published variants explain less than 5% of the variance[10–17]. One possible explanation is that many common variants of small effects contribute to phenotypic variation, and current GWA studies remain underpowered to detect most common variants. Using five studies not included in stage 1, we found that the 180 associated SNPs explained on average 10.5% (range 7.9–11.2%) of the variance in adult height (Supplementary Methods). Including SNPs associated with height at lower significance levels[18] ($0.05 > P > 5 \times 10^{-8}$) increased the variance explained to 13.3% (range 9.7–16.8%) (Fig. 1a). In addition, we found no evidence that non-additive effects including gene–gene interaction would increase the proportion of the phenotypic variance explained (Supplementary Methods and Supplementary Tables 5 and 6).

As a separate approach, we used a recently developed method[19] to estimate the total number of independent height-associated variants with effect sizes similar to the ones identified. We obtained this estimate using the distribution of effect sizes observed in stage 2 and the power to detect an association in stage 1, given these effect sizes (Supplementary Methods). The cumulative distribution of height loci, including those we identified and others as yet undetected, is shown in Fig. 1b. We estimate that there are 697 loci (95% confidence interval: 483–1040) with effects equal or greater than those identified, which together would explain approximately 15.7% of the phenotypic variation in height or 19.6% (95% confidence interval: 16.2–25.6) of height heritability (Supplementary Table 4). We estimated that a sample size of 500,000 would detect 99.6% of these loci at $P < 5 \times 10^{-8}$. This figure does not account for variants that have effect sizes smaller than those observed in the current study and, therefore, underestimates the contribution of undiscovered common genetic variants to phenotypic variation.
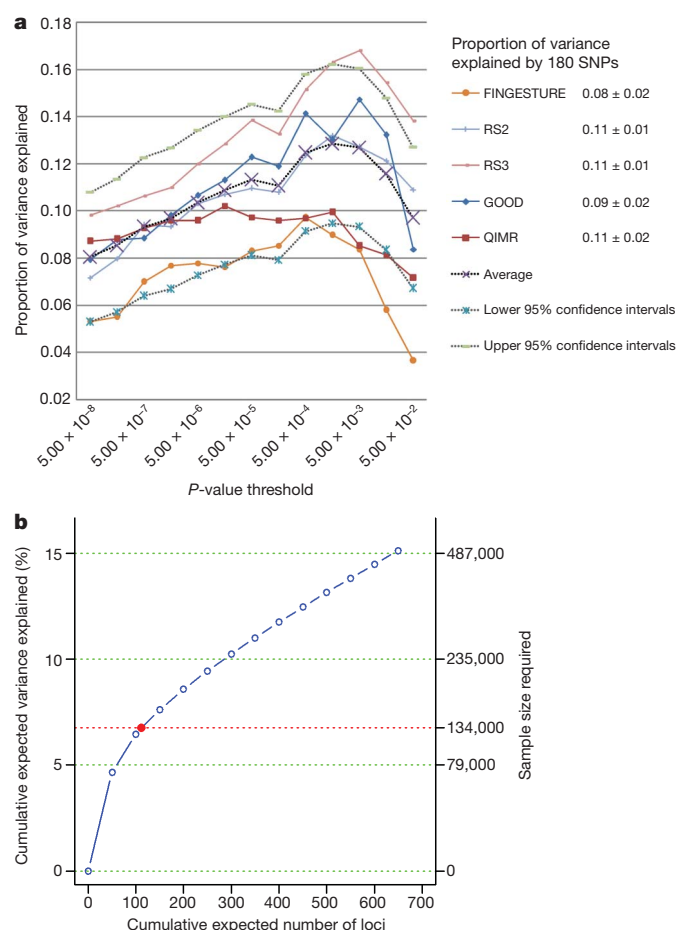
Figure 1 | Phenotypic variance explained by common variants. a, Variance explained is higher when SNPs not reaching genome-wide significance are included in the prediction model. The y axis represents the proportion of variance explained at different P-value thresholds from stage 1. Results are given for five studies that were not part of stage 1. The proportion of variation explained by the 180 SNPs is shown in the column to the right of the graph. **b,** Cumulative number of susceptibility loci expected to be discovered, including already identified loci and as yet undetected loci. The projections are based on loci that achieved a significance level of $P < 5 \times 10^{-8}$ in the initial scan and the distribution of their effect sizes in stage 2. The dotted red line corresponds to expected phenotypic variance explained by the 110 loci that reached genome-wide significance in stage 1, were replicated in stage 2 and had at least 1% power.

A further possible source of missing heritability is allelic heterogeneity: the presence of multiple, independent variants influencing a trait at the same locus. We performed genome-wide conditional analyses in a subset of stage 1 studies, including a total of 106,336 individuals. Each study repeated the primary GWA analysis but additionally adjusted for SNPs representing the 180 loci associated at $P < 5 \times 10^{-6}$ (Supplementary Methods). We then meta-analysed these studies in the same way as for the primary GWA study meta-analysis. Nineteen SNPs within the 180 loci were associated with height at $P < 3.3 \times 10^{-7}$ (a Bonferroni-corrected significance threshold calculated from the approximately 15% of the genome covered by the conditioned 2 Mb loci; Table 1, Fig. 2, Supplementary Methods and Supplementary Figs 1 and 3). The distances of the second signals to the lead SNPs suggested that both are likely to be affecting the same gene, rather than being coincidentally in close proximity. At 17 of 17 loci (excluding two contiguous loci in the *HMGA1* region), the second signal occurred within 500 kilobases (kb), rather than between 500 kb and 1 Mb, of this lead SNP (binomial test $P = 2 \times 10^{-5}$). Further analyses of allelic heterogeneity may identify additional variants that increase the proportion of variance explained. For example, within the 180 2-Mb loci, a total of 45 independent SNPs reached $P < 1 \times 10^{-5}$ when we would expect less than 2 by chance.

Although GWA studies have identified many variants robustly associated with common human diseases and traits, the biological significance of these variants, and the genes on which they act, is often unclear. We first tested the overlap between the 180 height-associated variants and two types of putatively functional variants, non-synonymous (ns) SNPs and cis-expression quantitative trait loci (cis-eQTLs, variants strongly associated with expression of nearby genes). Height variants were 2.4-fold more likely to overlap with cis-eQTLs in lymphocytes than expected by chance (47 variants: $P = 4.7 \times 10^{-11}$) (Supplementary Table 7) and 1.7-fold more likely to be closely correlated ($r^2 \geq 0.8$ in the HapMap CEU sample) with nsSNPs (24 variants, $P = 0.004$) (Supplementary Methods and Supplementary Table 8). Although the presence of a correlated cis-eQTL or nsSNP at an individual locus does not establish the causality of any particular variant, this enrichment shows that common functional variants contribute to the causal variants at height-associated loci. We also noted five loci where the height associated variant was strongly correlated ($r^2 > 0.8$) with variants associated with other traits and diseases ($P < 5 \times 10^{-8}$), including bone mineral density, rheumatoid arthritis, type 1 diabetes, psoriasis and obesity, suggesting that these variants have pleiotropic effects on human phenotypes (Supplementary Methods and Supplementary Table 9).

Table 1 | Secondary signals at associated loci after conditional analysis

Second signal SNP	Conditioned SNP	Chromosome	Second signal SNP position	Distance of conditioned SNP from index SNP (base pairs)	HapMap* r^2	Second signal P value after conditioning	Second signal P value pre-conditioning	Gene†
rs2280470	rs16942341	15	87,196,630	6,721	0.009	1×10^{-14}	1×10^{-15}	ACAN
rs10859563	rs11107116	12	92,644,470	141,835	0.003	3×10^{-12}	8×10^{-10}	SOCS2
rs750460	rs5742915	15	72,028,559	95,127	0.004	4×10^{-12}	7×10^{-8}	PML
rs6938239	rs2780226‡	6	34,791,613	484,583	0.019	6×10^{-12}	9×10^{-14}	HMGA1
rs7652177	rs572169	3	173,451,771	196,650	0.006	7×10^{-11}	1×10^{-11}	GHSR
rs7916441	rs2145998	10	80,595,583	196,119	0.112	6×10^{-10}	3×10^{-7}	PPIF
rs3792752	rs1173727	5	32,804,391	61,887	0.020	7×10^{-10}	4×10^{-8}	NPR3
rs10958476	rs7460090	8	57,258,362	98,355	0.020	1×10^{-9}	5×10^{-13}	SDR16C5
rs2353398	rs7689420	4	145,742,208	45,594	0.022	2×10^{-9}	1×10^{-10}	HHIP
rs2724475	rs6449353	4	17,555,530	87,056	0.098	2×10^{-9}	8×10^{-16}	LCORL
rs2070776	rs2665838	17	59,361,230	41,033	0.150	9×10^{-9}	1×10^{-14}	GH region
rs1401796	rs227724	17	52,194,758	60,942	0.005	2×10^{-8}	7×10^{-7}	NOG
rs4711336	rs2780226‡	6	33,767,024	540,046	0.111	3×10^{-8}	5×10^{-8}	HMGA1
rs6892884	rs12153391	5	170,948,228	187,815	0.000	4×10^{-8}	2×10^{-5}	FBXW11
rs1367226	rs3791675	2	55,943,044	21,769	0.204	4×10^{-8}	0.1245	EFEMP1
rs2421992	rs17346452	1	170,507,874	187,964	0.019	5×10^{-8}	1×10^{-5}	DNM3
rs225694	rs7763064	6	142,568,835	270,147	0.001	1×10^{-7}	2×10^{-6}	GPR126
rs10187066	rs12470505	2	219,223,003	393,610	0.022	2×10^{-7}	5×10^{-8}	IHH
rs879882	rs2256183	6	31,247,431	241,077	0.016	2×10^{-7}	8×10^{-8}	MICA

* HapMap CEU phase II release 23. † Nearest gene unless there is a known skeletal growth disorder gene in the locus. Positions are based on National Center for Biotechnology Information build 36. ‡ Nearest conditioned SNP where second signal occurs within 1 Mb of two conditioned SNPs.
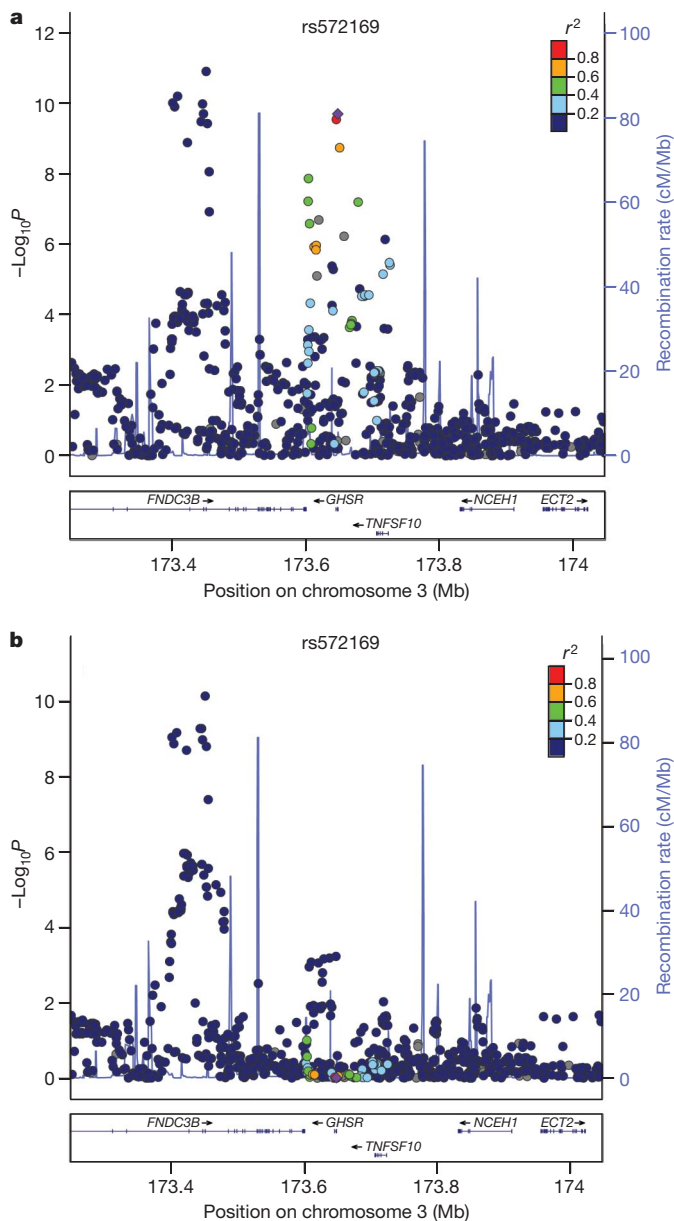
Figure 2 | Example of a locus with a secondary signal before (a) and after (b) conditioning. The plot is centred on the conditioned SNP (purple diamond) at the locus. The values of r^2 are based on the CEU HapMap II samples. The blue line and right-hand y axis represent CEU HapMap II recombination rates. The figure was created using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

We next addressed the extent to which height variants cluster near biologically relevant genes; specifically, genes mutated in human syndromes characterized by abnormal skeletal growth. We limited this analysis to the 652 genes occurring within the recombination hotspot-bounded regions surrounding each of the 180 index SNPs. We showed that the 180 loci associated with variation in normal height contained 21 of 241 genes (8.7%) found to underlie such syndromes (Supplementary Fig. 1 and Supplementary Table 10), compared with a median of 8 (range 1–19) genes identified in 1,000 matched control sets of regions ($P < 0.001$: 0 observations of 21 or more skeletal growth genes among 1,000 sets of matched SNPs). In 13 of these 21 loci the closest gene to the most associated height SNP in the region is the growth disorder gene, and in nine of these cases the most strongly associated height SNP is located within the growth disorder gene itself (Supplementary Methods and Supplementary Table 11). These results suggest that GWA studies may provide more clues about the identity of the functional genes at each locus than previously suspected.

We also investigated whether significant and relevant biological connections exist between the genes within the 180 loci, using two different computational approaches. We used the GRAIL text-mining algorithm to search for connectivity between genes near the associated SNPs, based on existing literature[20]. Of the 180 loci, 42 contained genes that were connected by existing literature to genes in the other associated loci (the pair of connected genes appear in articles that share scientific terms more often than expected at $P < 0.01$). For comparison, when we used GRAIL to score 1,000 sets of 180 SNPs not associated with height (but matched for number of nearby genes, gene proximity and allele frequency), we only observed 16 sets with 42 or more loci with a connectivity $P < 0.01$, thus providing strong statistical evidence that the height loci are functionally related ($P = 0.016$) (Fig. 3a). For the 42 regions with GRAIL connectivity $P < 0.01$, the implicated genes and SNPs are highlighted in Fig. 3b. The most strongly connected genes include those in the Hedgehog, TGF-β and growth hormone pathways.

As a second approach to find biological connections, we applied a novel implementation of gene set enrichment analysis (meta-analysis gene-set enrichment of variant associations, MAGENTA[21]) to perform pathway analysis (Supplementary Methods). This analysis revealed 17 different biological pathways and 14 molecular functions nominally enriched ($P < 0.05$) for associated genes, many of which lie within the validated height loci. These gene-sets include previously reported[11,13] (for example, Hedgehog signalling) and novel (for example, TGF-β signalling, histones, and growth and development-related) pathways and molecular functions (Supplementary Table 12). Several SNPs near genes in these pathways narrowly missed genome-wide significance, suggesting that these pathways likely contain additional associated variants. These results provide complementary evidence for some of the genes and pathways highlighted in the GRAIL analysis. For instance, genes such as *TGFB2* and *LTBP1-3* highlight a role for the TGF-β signalling pathway in regulating human height, consistent with the implication of this pathway in Marfan syndrome[22].

Finally, to examine the evidence for the potential involvement of specific genes at individual loci, we aggregated evidence from our data (expression quantitative trait loci, proximity to the associated variant, pathway-based analyses), and human and mouse genetic databases (Supplementary Table 13). Of 32 genes with highly correlated ($r^2 > 0.8$) nsSNPs, several are newly identified strong candidates for playing a role in human growth. Some are in pathways enriched in our study (such as *ECM2*, implicated in extracellular matrix), whereas others have similar functions to known growth-related genes, including *FGFR4* (*FGFR3* underlies several classic skeletal dysplasias[23]) and *STAT2* (*STAT5B* mutations cause growth defects in humans[24]). Interestingly, $Fgfr4^{-/-}$ $Fgfr3^{-/-}$ mice show severe growth retardation not seen in either single mutant[25], suggesting that the *FGFR4* variant might modify *FGFR3*-mediated skeletal dysplasias. Other genes at associated loci, such as *NPPC* and *NPR3* (encoding the C-type natriuretic peptide and its receptor), influence skeletal growth in mice and will likely also influence human growth[17]. Many of the remaining 180 loci have no genes with obvious connections to growth biology, but at some our data provide modest supporting evidence for particular genes, including *C3orf63*, *PML*, *CCDC91*, *ZNFX1*, *ID4*, *RYBP*, *SEPT2*, *ANKRD13B*, *FOLH1*, *LRRC37B*, *MFAP2*, *SLBP*, *SOCS5* and *ZBTB24* (Supplementary Table 13).

We have identified more than 100 novel loci that influence the classic polygenic trait of normal variation in human height, bringing the total to 180. Our results have potential general implications for genetic studies of complex traits. We show that loci identified by GWA studies highlight relevant genes: the 180 loci associated with height are non-randomly clustered within biologically relevant pathways and are enriched for genes that are involved in growth-related processes, that underlie syndromes of abnormal skeletal growth and that are directly relevant to growth-modulating therapies (*GH1*, *IGF1R*, *CYP19A1*, *ESR1*). The large number of loci with clearly relevant genes suggests that the remaining loci could provide potential clues to important and novel biology.
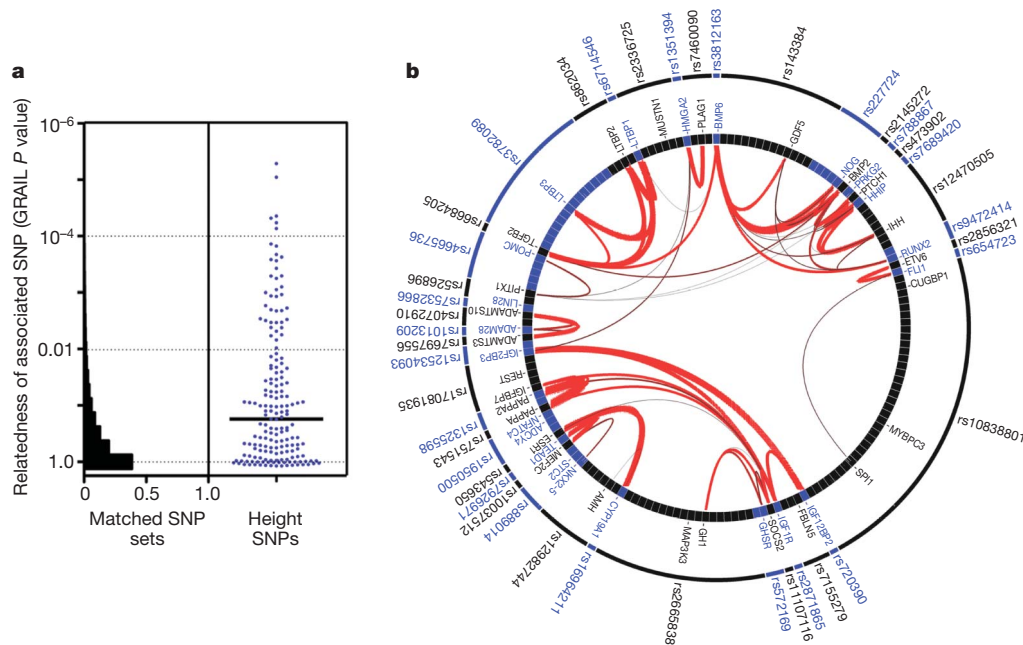
Figure 3 | Loci associated with height contain genes related to each other.
a, One hundred and eighty height-associated SNPs. The *y*-axis plots GRAIL *P* values on a log scale. The histogram corresponds to the distribution of GRAIL *P* values for 1,000 sets of 180 matched SNPs. The scatter plot represents GRAIL results for the 180 height SNPs (blue dots). The black horizontal line marks the median of the GRAIL *P* values (*P* = 0.14). The top ten keywords linking the

genes were: 'growth', 'kinase', 'factor', 'transcription', 'signalling', 'binding', 'differentiation', 'development', 'insulin', 'bone'. **b**, Representation of the connections between SNPs and corresponding genes for the 42 SNPs with GRAIL *P* < 0.01. Thicker and redder lines imply stronger literature-based connectivity.

We provide the strongest evidence yet that the causal gene will often be located near the most strongly associated DNA sequence variant. At the 21 loci containing a known growth disorder gene, that gene was on average 81 kb from the associated variant, and in over half of the loci it was the closest gene to the associated variant. Despite recent doubts about the benefits of GWA studies[26], this finding suggests that GWA studies are useful mapping tools to highlight genes that merit further study. The presence of multiple variants within associated loci could help localize the relevant genes within these loci.

By increasing our sample size to more than 100,000 individuals, we identified common variants that account for approximately 10% of phenotypic variation. Although larger than predicted by some models[26], this figure suggests that GWA studies, as currently implemented, will not explain most of the estimated 80% contribution of genetic factors to variation in height. This conclusion supports the idea that biological insights, rather than predictive power, will be the main outcome of this initial wave of GWA studies, and that new approaches, which could include sequencing studies or GWA studies targeting variants of lower frequency, will be needed to account for more of the 'missing' heritability. Our finding that many loci exhibit allelic heterogeneity suggests that many as yet unidentified causal variants, including common variants, will map to the loci already identified in GWA studies, and that the fraction of causal loci that have been identified could be substantially greater than the fraction of causal variants that have been identified.

In our study, many associated variants are tightly correlated with common nsSNPs, which would not be expected if these associated common variants were proxies for collections of rare causal variants, as has been proposed[27]. Although a substantial contribution to heritability by less common and/or quite rare variants may be more plausible, our data are not inconsistent with the recent suggestion[28] that many common variants of very small effect mostly explain the regulation of height.

In summary, our findings indicate that additional approaches, including those aimed at less common variants, will likely be needed to dissect more completely the genetic component of complex human traits. Our results also strongly demonstrate that GWA studies can

identify many loci that together implicate biologically relevant pathways and mechanisms. We envisage that thorough exploration of the genes at associated loci through additional genetic, functional and computational studies will lead to novel insights into human height and other polygenic traits and diseases.

METHODS SUMMARY

A summary of the methods, together with a full description of genome-wide association analyses and follow-up analyses of loci and variants, can be found in Supplementary Information.

1. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
2. Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* **5**, 329–348 (1885).
3. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
4. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
5. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
6. Reich, D. E. & Goldstein, D. B. Detecting association in a case-control study while correcting for population stratification. **20**, 4–16 (2001).
7. Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature Genet.* **37**, 868–872 (2005).
8. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
9. Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
10. Weedon, M. N. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genet.* **39**, 1245–1250 (2007).
11. Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genet.* **40**, 575–583 (2008).
12. Sanna, S. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature Genet.* **40**, 198–203 (2008).
13. Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genet.* **40**, 584–591 (2008).
14. Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U. & Malkina, I. Meta-analysis of genome-wide scans for human adult stature in humans identifies novel loci and

associations with measures of skeletal frame size. *PLoS Genet.* **5,** e1000445 (2009).

15. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nature Genet.* **40,** 609–615 (2008).

16. Johansson, A. *et al.* Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.* **18,** 373–380 (2009).

17. Estrada, K. *et al.* A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum. Mol. Genet.* **18,** 3516–3524 (2009).

18. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).

19. Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42,** 570–575 (2010).

20. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).

21. Segrè, A. V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).

22. Neptune, E. R. *et al.* Dysregulation of TGF-β activation contributes to pathogenesis in Marfan syndrome. *Nature Genet.* **33,** 407–411 (2003).

23. Superti-Furga, A. & Unger, S. Nosology and classification of genetic skeletal disorders: 2006 revision. *Am. J. Med. Genet. A* **143,** 1–18 (2007).

24. Kofoed, E. M. *et al.* Growth hormone insensitivity associated with a STAT5b mutation. *N. Engl. J. Med.* **349,** 1139–1147 (2003).

25. Weinstein, M., Xu, X., Ohyama, K. & Deng, C. X. FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125,** 3615–3623 (1998).

26. Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med.* **360,** 1696–1698 (2009).

27. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8,** e1000294 (2010).

28. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42,** 565–569 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions This work was done under the auspices of the Genetic Investigation of ANthropocentric Traits (GIANT) Consortium. Author contributions and roles are listed in the Supplementary Information.

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 this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.N.W. (michael.weedon@pms.ac.uk), G.R.A. (goncalo@umich.edu), K.S. (kstefans@decode.is), T.M.F. (tim.frayling@pms.ac.uk) or J.N.H. (joelh@broadinstitute.org).

Hana Lango Allen[1]*, Karol Estrada[2,3,4]*, Guillaume Lettre[5,6]*, Sonja I. Berndt[7]*, Michael N. Weedon[1]*, Fernando Rivadeneira[2,3,4], Cristen J. Willer[8], Anne U. Jackson[8], Sailaja Vedantam[9,10], Soumya Raychaudhuri[11,12], Teresa Ferreira[13], Andrew R. Wood[1], Robert J. Weyant[8], Ayellet V. Segrè[11,14,15], Elizabeth K. Speliotes[10,16], Eleanor Wheeler[17], Nicole Soranzo[17,18], Ju-Hyun Park[7], Jian Yang[19], Daniel Gudbjartsson[20], Nancy L. Heard-Costa[21], Joshua C. Randall[13], Lu Qi[22,23], Albert Vernon Smith[24,25], Reedik Mägi[13], Tomi Pastinen[26,27,28], Liming Liang[29], Iris M. Heid[30,31], Jian'an Luan[32], Gudmar Thorleifsson[20], Thomas W. Winkler[30], Michael E. Goddard[33,34], Ken Sin Lo[5], Cameron Palmer[9,10], Tsegaselassie Workalemahu[22], Yurii S. Aulchenko[2,4], Åsa Johansson[35,36], M. Carola Zillikens[3], Mary F. Feitosa[37], Tõnu Esko[38,39,40], Toby Johnson[41,42,43,44], Shamika Ketkar[37], Peter Kraft[45,46], Massimo Mangino[18], Inga Prokopenko[13,47], Devin Absher[48], Eva Albrecht[31], Florian Ernst[49], Nicole L. Glazer[50], Caroline Hayward[51], Jouke-Jan Hottenga[52], Kevin B. Jacobs[53], Joshua W. Knowles[54], Zoltán Kutalik[41,42], Keri L. Monda[55], Ozren Polasek[56,57], Michael Preuss[58], Nigel W. Rayner[13,47], Neil R. Robertson[13,47], Valgerdur Steinthorsdottir[20], Jonathan P. Tyrer[59], Benjamin F. Voight[11,14,15], Fredrik Wiklund[60], Jianfeng Xu[61], Jing Hua Zhao[32], Dale R. Nyholt[62], Niina Pellikka[63,64], Markus Perola[63,64], John R. B. Perry[1], Ida Surakka[63,64], Mari-Liis Tammesoo[38], Elizabeth L. Altmaier[9,10], Najaf Amin[2], Thor Aspelund[24,25], Tushar Bhangale[65], Gabrielle Boucher[5], Daniel I. Chasman[66,67], Constance Chen[68], Lachlan Coin[69], Matthew N. Cooper[70], Anna L. Dixon[71], Quince Gibson[72], Elin Grundberg[17,26,27], Ke Hao[73], M. Juhani Junttila[74], Lee M. Kaplan[16,67,75], Johannes Kettunen[63,64], Inke R. König[58], Tony Kwan[26,27], Robert W. Lawrence[70], Douglas F. Levinson[76], Mattias Lorentzon[77], Barbara McKnight[78], Andrew P. Morris[13], Martina Müller[31,79,80], Julius Suh Ngwa[81], Shaun Purcell[14,82,83], Suzanne Rafelt[84], Rany M. Salem[9,10], Erika Salvi[85,86], Serena Sanna[87], Jianxin Shi[7], Ulla Sovio[69], John R. Thompson[88,89], Michael C. Turchin[9,10], Liesbeth Vandenput[77], Dominique J. Verlaan[26,27], Veronique Vitart[51], Charles C. White[81], Andreas Ziegler[90], Peter

Almgren[91], Anthony J. Balmforth[92], Harry Campbell[93], Lorena Citterio[94], Alessandro De Grandi[95], Anna Dominiczak[96], Jubao Duan[97], Paul Elliott[69], Roberto Elosua[98], Johan G. Eriksson[99,100,101,102,103], Nelson B. Freimer[104], Eco J. C. Geus[52], Nicola Glorioso[105], Shen Haiqing[72], Anna-Liisa Hartikainen[106], Aki S. Havulinna[107], Andrew A. Hicks[95], Jennie Hui[70,108,109], Wilmar Igl[35], Thomas Illig[31], Antti Jula[110], Eero Kajantie[100], Tuomas O. Kilpeläinen[32], Markku Koiranen[111], Ivana Kolcic[56], Seppo Koskinen[107], Peter Kovacs[112], Jaana Laitinen[113], Jianjun Liu[114], Marja-Liisa Lokki[115], Ana Marusic[116], Andrea Maschio[87], Thomas Meitinger[117,118], Antonella Mulas[87], Guillaume Paré[119], Alex N. Parker[120], John F. Peden[13,121], Astrid Petersmann[122], Irene Pichler[95], Kirsi H. Pietiläinen[123,124], Anneli Pouta[106,125], Martin Ridderstråle[126], Jerome I. Rotter[127], Jennifer G. Sambrook[128,129], Alan R. Sanders[97], Carsten Oliver Schmidt[130], Juha Sinisalo[131], Jan H. Smit[132], Heather M. Stringham[8], G. Bragi Walters[20], Elisabeth Widen[63], Sarah H. Wild[93], Gonneke Willemsen[52], Laura Zagato[94], Lina Zgaga[56], Paavo Zitting[133], Helene Alavere[38], Martin Farrall[13,121,134], Wendy L. McArdle[135], Mari Nelis[38,39,40], Marjolein J. Peters[3,4], Samuli Ripatti[63,64], Joyce B. J. van Meurs[2,3,4], Katja K. Aben[136], Kristin G. Ardlie[11], Jacques S. Beckmann[41,137], John P. Beilby[108,109,138], Richard N. Bergman[139], Sven Bergmann[41,42], Francis S. Collins[140], Daniele Cusi[85], Martin den Heijer[141], Gudny Eiriksdottir[24], Pablo V. Gejman[97], Alistair S. Hall[92], Anders Hamsten[142], Heikki V. Huikuri[74], Carlos Iribarren[143,144], Mika Kähönen[145], Jaakko Kaprio[63,123,146], Sekar Kathiresan[11,14,147,148,149], Lambertus Kiemeney[136,150,151], Thomas Kocher[152], Lenore J. Launer[153], Terho Lehtimäki[154], Olle Melander[126], Tom H. Mosley Jr[155], Arthur W. Musk[109,156], Markku S. Nieminen[131], Christopher J. O'Donnell[148,157], Claes Ohlsson[77], Ben Oostra[158], Lyle J. Palmer[70,109], Olli Raitakari[159], Paul M. Ridker[66,67], John D. Rioux[5,6], Aila Rissanen[124], Carlo Rivolta[41], Heribert Schunkert[160], Alan R. Shuldiner[72,161], David S. Siscovick[162,163], Michael Stumvoll[164,165], Anke Tönjes[164,166], Jaakko Tuomilehto[167,168,169], Gert-Jan van Ommen[170], Jorma Viikari[171], Andrew C. Heath[172], Nicholas G. Martin[173], Grant W. Montgomery[174], Michael A. Province[37,175], Michael Kayser[176], Alice M. Arnold[78,177], Larry D. Atwood[21], Eric Boerwinkle[178], Stephen J. Chanock[7], Panos Deloukas[17], Christian Gieger[31], Henrik Grönberg[60], Per Hall[60], Andrew T. Hattersley[1], Christian Hengstenberg[179,180], Wolfgang Hoffman[130], G. Mark Lathrop[181], Veikko Salomaa[107], Stefan Schreiber[182], Manuela Uda[87], Dawn Waterworth[183], Alan F. Wright[51], Themistocles L. Assimes[54], Inês Barroso[17,184], Albert Hofman[2,4], Karen L. Mohlke[185], Dorret I. Boomsma[52], Mark J. Caulfield[44], L. Adrienne Cupples[81], Jeanette Erdmann[160], Caroline S. Fox[186], Vilmundur Gudnason[24,25], Ulf Gyllensten[35], Tamara B. Harris[153], Richard B. Hayes[187], Marjo-Riitta Jarvelin[69,111,125,188], Vincent Mooser[183], Patricia B. Munroe[44], Willem H. Ouwehand[17,128,129], Brenda W. Penninx[132,189,190], Peter P. Pramstaller[95,191,192], Thomas Quertermous[54], Igor Rudan[51,116], Nilesh J. Samani[84,88], Timothy D. Spector[18], Henry Völzke[130], Hugh Watkins on behalf of the Procardis Consortium[13,121], James F. Wilson[93], Leif C. Groop[91], Talin Haritunians[127], Frank B. Hu[22,23,45], Robert C. Kaplan[193], Andres Metspalu[38,39,40], Kari E. North[55,194], David Schlessinger[195], Nicholas J. Wareham[32], David J. Hunter[22,23,45], Jeffrey R. O'Connell[72], David P. Strachan[196], H.-Erich Wichmann[31,80,197], Ingrid B. Borecki[37,175], Cornelia M. van Duijn[2,4], Eric E. Schadt[198,199], Unnur Thorsteinsdottir[20,200], Leena Peltonen[17,63,64,82,201], Nilanjan Chatterjee[7], Ruth J. F. Loos[32], Michael Boehnke[8], Mark I. McCarthy[13,47,202], Erik Ingelsson[60], Cecilia M. Lindgren[13,47], Gonçalo R. Abecasis[8]*, Kari Stefansson[20,200]*, Timothy M. Frayling[1]*, Joel N. Hirschhorn[9,10,203]*

*These authors contributed equally to this work.

[1]Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter EX1 2LU, UK. [2]Department of Epidemiology, Erasmus Medical Centre, 3015 GE Rotterdam, The Netherlands. [3]Department of Internal Medicine, Erasmus Medical Centre, 3015 GE Rotterdam, The Netherlands. [4]Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), 2300 RC Leiden, The Netherlands. [5]Montreal Heart Institute, Montréal, Québec H1T 1C8, Canada. [6]Department of Medicine, Université de Montréal, Montréal, Québec H3T 1J4, Canada. [7]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA. [8]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [9]Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA. [10]Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [11]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [12]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115 USA. [13]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [14]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [15]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [16]Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [17]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [18]Department of Twin Research and Genetic Epidemiology, King's College London, Lambeth Palace Road, London SE1 7EH, UK. [19]Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Queensland 4006, Australia. [20]deCODE Genetics, 101 Reykjavik, Iceland. [21]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [22]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [23]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [24]Icelandic Heart Association, Kopavogur, Iceland. [25]University of Iceland, 101 Reykjavik, Iceland. [26]McGill University and Genome Québec Innovation Centre, Montréal, Québec H3A 1A4, Canada. [27]Department of Human Genetics, McGill University Health Centre, McGill University, Montréal, Québec H3G 1A4, Canada. [28]Department of Medical Genetics, McGill University Health Centre, McGill University, Montréal, Québec

H3G 1A4, Canada. [29]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Cambridge, Massachusetts 02138, USA. [30]Regensburg University Medical Center, Department of Epidemiology and Preventive Medicine, 93053 Regensburg, Germany. [31]Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany. [32]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK. [33]Faculty of Land and Environment, University of Melbourne, Parkville 3010, Australia. [34]Department of Primary Industries, Bundoora, Victoria 3086,, Australia. [35]Department of Genetics and Pathology, Rudbeck Laboratory, University of Uppsala, SE-75185 Uppsala, Sweden. [36]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), N-7489 Trondheim, Norway. [37]Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA. [38]Estonian Genome Center, University of Tartu, Tartu 50410, Estonia. [39]Estonian Biocenter, Tartu 51010, Estonia. [40]Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia. [41]Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland. [42]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [43]Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK. [44]Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK. [45]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [46]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [47]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK. [48]Hudson Alpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. [49]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany. [50]Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, Washington 98101, USA. [51]MRC Human Genetics Unit, Institute for Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU, Scotland, UK. [52]Department of Biological Psychology, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands. [53]Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA. [54]Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. [55]Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA. [56]Andrija Stampar School of Public Health, Medical School, University of Zagreb, 10000 Zagreb, Croatia. [57]Gen-Info Ltd, 10000 Zagreb, Croatia. [58]Universität zu Lübeck, Institut für Medizinische Biometrie und Statistik, 23562 Lübeck, Germany. [59]Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. [60]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden. [61]Center for Human Genomics, Wake Forest University, Winston-Salem, North Carolina 27157, USA. [62]Neurogenetics Laboratory, Queensland Institute of Medical Research, Queensland 4006, Australia. [63]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00014, Helsinki, Finland. [64]National Institute for Health and Welfare, Department of Chronic Disease Prevention, Unit of Public Health Genomics, FIN-00014 Helsinki, Finland. [65]Department of Genome Sciences, University of Washington, Seattle, 98195 Washington, USA. [66]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [67]Harvard Medical School, Boston, Massachusetts 02115, USA. [68]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [69]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London W2 1PG, UK. [70]Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia 6009, Australia. [71]Royal National Hospital for Rheumatic Diseases and University of Bath, Bath BA1 1RL, UK. [72]Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [73]Genetics Department, Rosetta Inpharmatics, a Wholly Owned Subsidiary of Merck & Co Inc., Seattle, Washington 98109, USA. [74]Department of Internal Medicine, University of Oulu, 90014 Oulu, Finland. [75]MGH Weight Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [76]Stanford University School of Medicine, Stanford, California 93405, USA. [77]Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 413 45 Gothenburg, Sweden. [78]Departments of Biostatistics, University of Washington, Seattle, Washington 98195, USA. [79]Ludwig-Maximilians- Universität, Department of Medicine I, University Hospital Grosshadern, 81377 Munich, Germany. [80]Ludwig-Maximilians-Universität, Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, 81377 Munich, Germany. [81]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [82]The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [83]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA. [84]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. [85]University of Milan, Department of Medicine, Surgery and Dentistry, 20139 Milan, Italy. [86]KOS Genetic Srl, 20123 Milan, Italy. [87]Istituto di Neurogenetica e Neurofarmacologia del CNR, Monserrato, 09042 Cagliari, Italy. [88]Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester LE3 9QP, UK. [89]Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK. [90]Universität zu Lübeck, Institut für Medizinische Biometrie und Statistik, 23562 Lübeck, Germany. [91]Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, 20502 Malmö, Sweden. [92]Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds LS2 9JT, UK. [93]Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland, UK. [94]University Vita-Salute San Raffaele, Division of Nephrology and Dialysis, 20132 Milan, Italy. [95]Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano/Bozen, 39100, Italy. Affiliated Institute of the University of Lübeck, Lübeck, Germany. [96]British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow G12 8TA, UK. [97]Northshore University Healthsystem,

Evanston, Illinois 60201, USA. [98]Cardiovascular Epidemiology and Genetics, Institut Municipal D'investigacio Medica and CIBER Epidemiología y Salud Pública, Barcelona, Spain. [99]Department of General Practice and Primary Health Care, University of Helsinki, 00014, Helsinki, Finland. [100]National Institute for Health and Welfare, 00271 Helsinki, Finland. [101]Helsinki University Central Hospital, Unit of General Practice, 00280 Helsinki, Finland. [102]Folkhalsan Research Centre, 00250 Helsinki, Finland. [103]Vasa Central Hospital, 65130 Vasa, Finland. [104]Center for Neurobehavioral Genetics, University of California, Los Angeles, California 90095, USA. [105]Hypertension and Cardiovascular Prevention Center, University of Sassari, 07100 Sassari, Italy. [106]Department of Clinical Sciences/Obstetrics and Gynecology, University of Oulu, 90014 Oulu, Finland. [107]National Institute for Health and Welfare, Department of Chronic Disease Prevention, Chronic Disease Epidemiology and Prevention Unit, 00014, Helsinki, Finland. [108]PathWest Laboratory of Western Australia, Department of Molecular Genetics, J Block, QEII Medical Centre, Nedlands, Western Australia 6009, Australia. [109]Busselton Population Medical Research Foundation Inc., Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia. [110]National Institute for Health and Welfare, Department of Chronic Disease Prevention, Population Studies Unit, 20720 Turku, Finland. [111]Institute of Health Sciences, University of Oulu, 90014 Oulu, Finland. [112]Interdisciplinary Centre for Clinical Research, University of Leipzig, 04103 Leipzig, Germany. [113]Finnish Institute of Occupational Health, 90220 Oulu, Finland. [114]Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore. [115]Transplantation Laboratory, Haartman Institute, University of Helsinki, 00014, Helsinki, Finland. [116]Croatian Centre for Global Health, School of Medicine, University of Split, Split 21000, Croatia. [117]Institute of Human Genetics, Klinikum rechts der Isar der Technischen Universität München, 81675 Munich, Germany. [118]Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany. [119]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. [120]Amgen, Cambridge, Massachusetts 02139, USA. [121]Department of Cardiovascular Medicine, University of Oxford, Level 6 West Wing, John Radcliffe Hospital, Headley Way, Headington, Oxford OX3 9DU, UK. [122]Institut für Klinische Chemie und Laboratoriumsmedizin, Universität Greifswald, 17475 Greifswald, Germany. [123]Finnish Twin Cohort Study, Department of Public Health, University of Helsinki, 00014, Helsinki, Finland. [124]Obesity Research unit, Department of Psychiatry, Helsinki University Central Hospital, Helsinki, Finland. [125]National Institute for Health and Welfare, 90101 Oulu, Finland. [126]Department of Clinical Sciences, Lund University, 20502 Malmö, Sweden. [127]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. [128]Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK. [129]NHS Blood and Transplant, Cambridge Centre, Cambridge CB2 0PT, UK. [130]Institut für Community Medicine, 17489 Greifswald, Germany. [131]Division of Cardiology, Cardiovascular Laboratory, Helsinki University Central Hospital, 00029 Helsinki, Finland. [132]Department of Psychiatry/EMGO Institute, VU University Medical Center, 1081 BT Amsterdam, The Netherlands. [133]Department of Physiatrics, Lapland Central Hospital, 96101 Rovaniemi, Finland. [134]Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [135]Avon Longitudinal Study of Parents and Children (ALSPAC) Laboratory, Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK. [136]Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands. [137]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, 1011 Lausanne, Switzerland. [138]School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, Western Australia 6009, Australia. [139]Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [140]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [141]Department of Endocrinology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [142]Atherosclerosis Research Unit, Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, 171 76 Stockholm, Sweden. [143]Division of Research, Kaiser Permanente Northern California, Oakland, California 94612, USA. [144]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California 94107, USA. [145]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, 33520 Tampere, Finland. [146]National Institute for Health and Welfare, Department of Mental Health and Substance Abuse Services, Unit for Child and Adolescent Mental Health, 00271 Helsinki, Finland. [147]Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [148]Framingham Heart Study of the National Heart, Lung and Blood Institute and Boston University, Framingham, Massachusetts 01702, USA. [149]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [150]Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [151]Department of Urology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [152]Zentrum für Zahn-, Mund- und Kieferheilkunde, 17489 Greifswald, Germany. [153]Laboratory of Epidemiology, Demography, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. [154]Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, 33520 Tampere, Finland. [155]Department of Medicine, Division of Geriatrics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA. [156]School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia 6009, Australia. [157]National, Lung, and Blood Institute, National Institutes of Health, Framingham, Massachusetts 01702, USA. [158]Department of Clinical Genetics, Erasmus Medical Centre, 3015 GE Rotterdam, The Netherlands. [159]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and the Department of Clinical Physiology, Turku University Hospital, 20520 Turku, Finland. [160]Universität zu Lübeck, Medizinische Klinik II, 23562 Lübeck, Germany. [161]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, Maryland 21201, USA. [162]Cardiovascular Health Research Unit, University of Washington, Seattle, Washington 98101, USA. [163]Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington 98195, USA. [164]Department of Medicine, University of Leipzig, 04103 Leipzig, Germany. [165]LIFE Study Centre, University of Leipzig, Leipzig, Germany. [166]Coordination Centre for Clinical Trials, University of Leipzig, Härtelstrasse 16–18, 04103 Leipzig, Germany. [167]National Institute for Health and Welfare, Diabetes Prevention Unit, 00271 Helsinki, Finland. [168]Hjelt Institute, Department of Public Health, University of Helsinki, 00014 Helsinki, Finland. [169]South Ostrobothnia Central Hospital, 60220 Seinajoki, Finland. [170]Department of Human Genetics and Center of Medical Systems Biology, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands. [171]Department of Medicine, University of Turku and Turku University Hospital, 20520 Turku, Finland. [172]Department of Psychiatry and Midwest Alcoholism Research Center, Washington University School of Medicine, St Louis, Missouri 63108, USA. [173]Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Queensland 4006, Australia. [174]Molecular Epidemiology Laboratory, Queensland Institute of Medical Research, Queensland 4006, Australia. [175]Division of Biostatistics,Washington University School of Medicine, St Louis, Missouri 63110, USA. [176]Department of Forensic Molecular Biology, Erasmus Medical Centre, 3015 GE Rotterdam, The Netherlands. [177]Collaborative Health Studies Coordinating Center, Seattle, Washington 98115, USA. [178]Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas 77030, USA. [179]Klinik und Poliklinik für Innere Medizin II, Universität Regensburg, 93053 Regensburg, Germany. [180]Regensburg University Medical Center, Innere Medizin II, 93053 Regensburg, Germany. [181]Centre National de Genotypage, Evry, Paris 91057, France. [182]Christian-Albrechts-University, University Hospital Schleswig-Holstein, Institute for Clinical Molecular Biology and Department of Internal Medicine I, Schittenhelmstrasse 12, 24105 Kiel, Germany. [183]Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA. [184]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [185]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [186]Division of Intramural Research, National Heart, Lung, and Blood Institute, Framingham Heart Study, Framingham, Massachusetts 01702, USA. [187]New York University Medical Center, New York, New York 10016, USA. [188]Biocenter Oulu, University of Oulu, 90014 Oulu, Finland. [189]Department of Psychiatry, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands. [190]Department of Psychiatry, University Medical Centre Groningen, 9713 GZ Groningen, The Netherlands. [191]Department of Neurology, General Central Hospital, Bolzano, Italy. [192]Department of Neurology, University of Lübeck, Lübeck, Germany. [193]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [194]Carolina Center for Genome Sciences, School of Public Health, University of North Carolina Chapel Hill, Chapel Hill, North Carolina 27514, USA. [195]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland 21224, USA. [196]Division of Community Health Sciences, St George's, University of London, London SW17 0RE, UK. [197]Klinikum Grosshadern, 81377 Munich, Germany. [198]Pacific Biosciences, Menlo Park, California 94025, USA. [199]Sage Bionetworks, Seattle, Washington 98109, USA. [200]Faculty of Medicine, University of Iceland, 101 Reykjavík, Iceland. [201]Department of Medical Genetics, University of Helsinki, 00014 Helsinki, Finland. [202]NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford OX3 7LJ, UK. [203]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

ARTICLE

Single–molecule analysis of Mss116–mediated group II intron folding

Krishanthi S. Karunatilaka[1], Amanda Solem[1], Anna Marie Pyle[2,3] & David Rueda[1]

DEAD–box helicases are conserved enzymes involved in nearly all aspects of RNA metabolism, but their mechanisms of action remain unclear. Here, we investigated the mechanism of the DEAD–box protein Mss116 on its natural substrate, the group II intron ai5γ. Group II introns are structurally complex catalytic RNAs considered evolutionarily related to the eukaryotic spliceosome, and an interesting paradigm for large RNA folding. We used single-molecule fluorescence to monitor the effect of Mss116 on folding dynamics of a minimal active construct, ai5γ−D135. The data show that Mss116 stimulates dynamic sampling between states along the folding pathway, an effect previously observed only with high Mg^{2+} concentrations. Furthermore, the data indicate that Mss116 promotes folding through discrete ATP–independent and ATP–dependent steps. We propose that Mss116 stimulates group II intron folding through a multi–step process that involves electrostatic stabilization of early intermediates and ATP hydrolysis during the final stages of native state assembly.

DEAD-box proteins are enzymes that have essential roles in cellular processes involving RNA[1]. Although these have been studied *in vitro* and *in vivo*, there are few examples of DEAD-box proteins whose mechanisms have been dissected using a natural substrate *in vitro*[2]. Mss116 is a DEAD-box protein that facilitates splicing of all *Saccharomyces cerevisiae* mitochondrial group I and II introns *in vivo*[3]. Mss116 exhibits RNA binding, unwinding, annealing and ATPase activities and has been shown to facilitate group II intron splicing *in vitro* under near-physiological conditions[4–6].

Group II introns are large ribozymes that catalyse self-splicing[7] and are thought to be evolutionarily related to nuclear splicing. They also function as mobile genetic elements, which may have had an important role during evolution and genome diversification[8]. Group II introns require high ionic strength and elevated temperatures to function *in vitro*[7], but splicing *in vivo* is facilitated by protein cofactors. Despite diversity in primary sequence, group II introns have highly conserved secondary and tertiary structures[9,10]. The secondary structure consists of six domains (D1–D6) radiating from a central core. D1, an assembly scaffold for the other domains[11,12], and D5, the catalytic core, are the only absolutely essential domains for minimal catalytic activity in group II introns[13,14]. One of the best characterized active forms of the *S. cerevisiae* ai5γ group II intron is D135, which contains D1, D3, a catalytic effector, and D5 (Fig. 1a)[15–18]. Folding studies have shown that D135 folds slowly but directly to the native state[15–18], making it an interesting model system.

Mss116 has been proposed to promote group II intron splicing by stabilizing on-pathway intermediates and/or by disrupting off-pathway misfolded structures[4,5]. However, the mechanism by which Mss116 mediates group II intron splicing remains highly debated[4–6]. Previous studies on other systems have shown that both mechanisms are possible. For example, the LtrA protein encoded by a group IIA intron promotes self-splicing by binding the intron RNA[19]. RNA chaperones, such as StpA, bind non-specifically and unfold RNA to allow refolding into the active conformation[20,21].

Previous work has used group II intron catalytic activity to report on Mss116 function. However, to understand further how DEAD-box proteins facilitate RNA folding, it is necessary to observe RNA folding directly[22]. Therefore, we have characterized the effect of Mss116 on the folding dynamics of a fluorophore-labelled D135 ribozyme (D135-L14) using single-molecule fluorescence resonance energy transfer (smFRET, Fig. 1b)[17,18].

Folding requires high salt

Consistent with previous studies[17,18], at high ionic strength (Methods Summary), the intron shows repeated stochastic transitions between three distinct conformational states (Fig. 2a)[17]: an extended inter-mediate (I, FRET ∼0.1), a folded intermediate (F, FRET ∼0.4) and the native state (N, FRET ∼0.6). Among all transitions, fewer than 2% of molecules show direct transitions between I and N, indicating that F is an obligatory folding intermediate. In single-molecule experiments the unfolded state (U) is indistinguishable from donor-only species[17,18]. These experiments are complementary to gel-shift assays in which U to I transitions are readily observable, but I, F and N co-migrate as a single 'compact' species, and N can only be differentiated by the presence of catalytic activity[22–24].
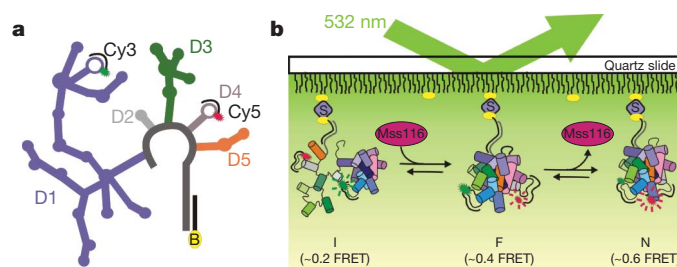
Figure 1 | Single-molecule fluorescence detection of group II intron folding with its natural cofactor Mss116. a, Secondary structure of D135-L14 ribozyme. Fluorophore-labelled DNA oligonucleotides (green, Cy3; red, Cy5) and biotin (yellow)-labelled DNA oligonucleotide are shown. **b**, Single-molecule detection of D135-L14 folding in the presence of Mss116 (pink). The folding pathway consists of three states: the extended intermediate state (I), the folded intermediate state (F) and the native state (N).

[1]Department of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, Michigan 48202, USA. [2]Department of Molecular, Cellular and Developmental Biology and Department of Chemistry, Yale University, New Haven, Connecticut 06520, USA. [3]Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520, USA.

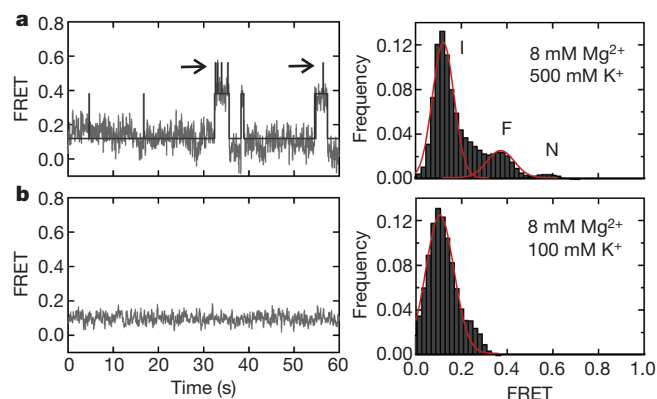**Figure 2 | Effect of ionic strength on the folding dynamics of D135-L14 ribozyme. a**, Typical FRET trajectory (left) showing the dynamics of a single D135-L14 ribozyme under high ionic strength conditions, and histogram representing the distribution of FRET states of ~100 trajectories (right). Three different conformations are observed: I (FRET ~0.1), F (FRET ~0.4) and N (FRET ~0.6). Black line represents idealized HMM and arrows indicate rapid transient excursions to the N state. **b**, Typical FRET trajectory (left) showing the dynamics of a single D135-L14 ribozyme under near-physiological conditions, and histogram representing the FRET state distribution of ~100 trajectories (right).

Under near-physiological conditions (Methods Summary), D135-L14 has only one FRET distribution (Fig. 2b, FRET ~0.1), indicating that the vast majority of molecules reside in the extended intermediate state, consistent with gel-shift experiments[17,24]. Most molecules in low ionic strength (≥95%) seem static and cannot even make transient excursions to N.

Mss116 mediates folding

We sought to characterize the effect of active Mss116 (Supplementary Fig. 1) on the folding of D135-L14 under near-physiological conditions. Since Mss116 is ATP-dependent, initial experiments included ATP. The observed smFRET trajectories and distributions of D135-L14 in the presence of Mss116 exhibit transitions between all conformational states (I, F and N, Fig. 3a–c), in contrast to the smFRET distribution in the absence of protein (Fig. 2b). The appearance of F and N with Mss116 indicates that it facilitates RNA folding. Approximately 30% of molecules reach N transiently or stably, including a minor population of

molecules (~9%) that remains static in N (Supplementary Fig. 2). As a control, we monitored the FRET ratio of a single D135-L14 ribozyme before and after addition of Mss116 and ATP (Fig. 3d). Initially, D135-L14 appears static in I (FRET ~0.1). After the addition, transitions from I into the higher FRET conformations were observed, providing direct evidence of Mss116-mediated RNA folding (Fig. 3d).

Lowering the Mg^{2+} concentration to 1 mM yielded similar results (Supplementary Fig. 3), showing that Mss116 can also help D135-L14 form F and N under physiological conditions. However, we also observed a higher number of molecules in the 0 FRET state, suggesting that the number of molecules in the unfolded state is higher at 1 mM Mg^{2+}. Interestingly, at both 8 and 1 mM Mg^{2+}, we find that the average FRET ratio of I increases to ~0.18 in the presence of Mss116, indicating a slight protein-induced compaction of this conformer.

In high ionic strength, exon-based substrates (17/7 or 17/7-dC) stabilize the native state[17]. Here we show that during protein-facilitated folding, the substrate 17/7 stabilizes N, increasing its population by ~30% (Supplementary Fig. 4 and Supplementary Table 1). Similar behaviour was also observed with a slow-cleaving substrate (17/7-dC). However, the substrate alone does not promote folding into the native state under near-physiological conditions (Supplementary Fig. 5).

To evaluate the specificity of Mss116-induced folding, we examined the folding dynamics of D135-L14 in the presence of three different basic RNA-binding proteins: the HIV nucleocapsid (NC)[25], poly-pyrimidine tract binding (PTB)[26] and Hfq proteins[27]. The corresponding smFRET histograms show all three proteins can populate the folded intermediate to a lesser extent than Mss116, and none promotes folding into the native state (Fig. 3e). Among the molecules observed in the presence of each non-specific protein, more than 90% seem static in either I or F (Supplementary Fig. 6). Unlike Mss116, none of these proteins affects the average FRET ratio of I, indicating that Mss116 may recognize this conformation more specifically. These data show that the formation of the folded intermediate can be facilitated by a diverse set of basic RNA binding proteins. However, the native state is formed only in the presence of Mss116 and ATP. Thus, Mss116 has a specific role during the final, rapid stages of ai5γ folding.

ATP is required for efficient folding

To investigate the role of ATP in Mss116-mediated folding of D135, we monitored the folding dynamics without ATP and with the non-hydrolysable ATP analogue, adenylyl-imidodiphosphate (AMPPNP, Fig. 4). The smFRET trajectories show transitions between all three
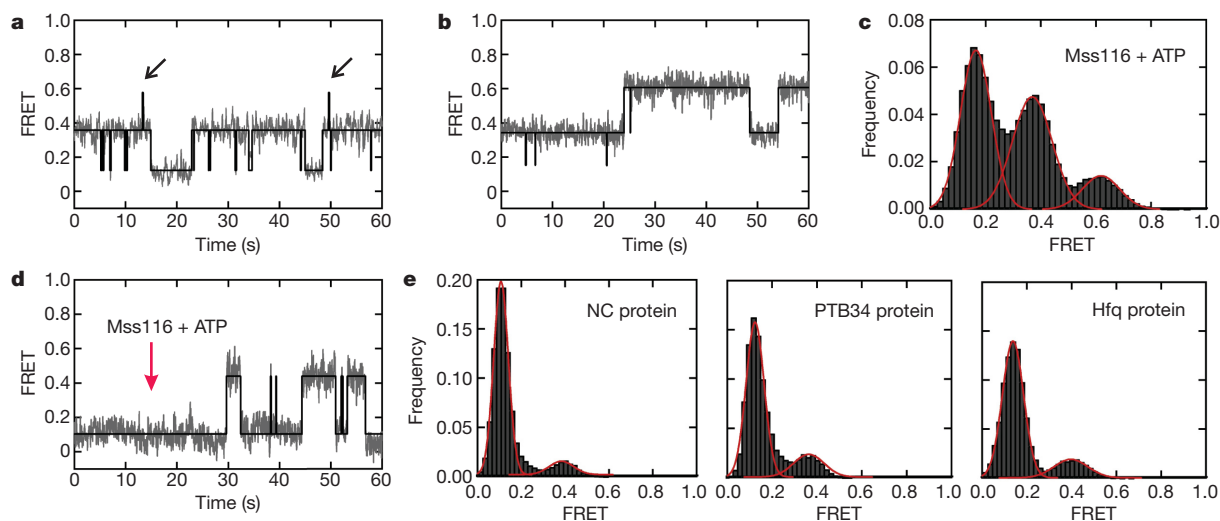
Figure 3 | Mss116 promotes folding of group II introns at near-physiological conditions. a, **b**, Two examples of FRET trajectories of single D135-L14 ribozymes in the presence of Mss116 and ATP. Black lines represent idealized HMM and arrows indicate transient excursions to the native state.

c, Histogram showing the distribution of three structural conformations. **d**, FRET trajectory showing dynamic behaviour of a single ribozyme upon addition of Mss116 and ATP at 15–20 s. **e**, FRET state distribution in the presence of the indicated non-specific RNA binding proteins.

conformations of the D135 ribozyme, as observed in the presence of ATP (compare Fig. 3c with Fig. 4). However, the population of N was ~2.5-fold lower, showing that ATP hydrolysis enhances formation of N (Fig. 4). This small population may be due to trace ATP that co-purifies with Mss116 or D135-L14. However, control experiments in the presence of glucose and hexokinase, which remove trace ATP, still yield low levels of the native state population (Supplementary Fig. 7). These results show that Mss116 facilitates folding into N in an ATP-dependent manner, but N can also be populated without ATP, albeit inefficiently. In the absence of Mss116, ATP or AMPPNP alone are not sufficient to form F or N (Supplementary Fig. 8). These data show that ATP binding and hydrolysis by Mss116 are important for the efficient formation of native D135, but hydrolysis is not strictly required.

Dynamics depend on Mss116 and ATP

To define the folding rate constants (k_1, k_{-1}, k_2 and k_{-2}), we analysed the dynamic FRET trajectories using a hidden Markov model (HMM)[28]. Without Mss116 and under high ionic strength conditions, a large subpopulation (>75%) appears static in either I or F during the two-minute timescale of our observation window (Supplementary Fig. 2). Among the subpopulation of dynamic molecules (~24%), the observed folding rate constants between the extended and folded intermediates ($k_1 = 0.5 \pm 0.3\,\mathrm{s}^{-1}$ and $k_{-1} = 0.2 \pm 0.1\,\mathrm{s}^{-1}$) agree closely with previous values[17].

Under near-physiological conditions without Mss116, the dynamic subpopulation decreases to ≤5% (Supplementary Fig. 2a); therefore, folding rate constants could not be determined. However, upon addition of Mss116 and ATP, the fraction of dynamic molecules increases to ~34%, and transitions to both F and N were observed (Fig. 3a–c). The resulting folding rate constants (Supplementary Table 2) show that k_1 and k_{-1} are comparable to the high ionic strength conditions, indicating that Mss116 and Mg^{2+} stabilize these states similarly. However, in the presence of Mss116, k_2 is ~threefold higher and k_{-2} is ~twofold lower than in 100 mM Mg^{2+} (ref. 17), indicating that the protein stabilizes the native state more than Mg^{2+} ions.

In the presence of Mss116 without ATP or with AMPPNP, ~30% of molecules were dynamic. The resulting folding rate constants k_1 and k_{-1} are comparable to those with ATP present. However, k_{-2} is three- to sixfold higher than in the presence of ATP (Supplementary Table 2). This suggests that ATP hydrolysis, and not just binding, contributes to efficient formation of the native state. In the presence

of Mss116 without ATP or with AMPPNP, the majority of molecules that dwell in F show transitions to I (~75%) rather than N (~25%, Supplementary Table 3). However, with both Mss116 and ATP, the percentage of F to I transitions decreases (58%) while the percentage of F to N transitions increases (42%) (Supplementary Table 3). Therefore, ATP hydrolysis also increases the probability that F molecules sample N instead of I.

Detailed analysis of static molecules in each state provides additional evidence for ATP-dependent Mss116 activity (Supplementary Fig. 2b–e). In the presence of Mss116 without ATP, the fraction of static molecules in I decreases from 95% to 53%, and they begin to appear in both F and N (15% and 4%, respectively). In the presence of both Mss116 and ATP, the fraction of static molecules in N increases significantly (2.5-fold, $P = 0.02$), whereas that in F does not change significantly ($P = 0.16$). These data support an important role for ATP hydrolysis for efficient Mss116-induced folding of the intron ai5γ, as proposed[4,5].

Discussion

Based on previous biochemical and folding studies[4–6,16,24], it has been proposed that Mss116 can promote group II intron splicing by stabilizing on-pathway intermediates and/or by disrupting off-pathway misfolded structures. Previous work has shown that ai5γ folds through a slow but smooth pathway devoid of kinetic traps[16,24], and an Mss116 mutant with a significant helicase defect still retains the ability to promote splicing[4,29]. Other work supports a mechanism where Mss116 unwinds kinetic traps to promote splicing, even at low levels of helicase activity[6]. Both models are primarily based on indirect studies of ai5γ splicing, thus, any putative role for Mss116 in RNA folding is purely hypothetical. To examine specifically the effect of Mss116 on ai5γ folding, we have used a well-characterized smFRET assay that enables us to directly monitor the role of Mss116 and ATP on the intron folding.

Previous experiments showed that under near-physiological conditions, group II introns alone do not stably form the native state[24]. Our data show that D135 cannot even transiently sample the native state. The large subpopulation of static molecules in I (~95%) indicates a high activation barrier between I and F (Fig. 5). The appearance of all three states in the presence of Mss116 shows that Mss116 can promote RNA folding by lowering the activation barriers between folded states, consistent with recent data[22].

We observe two distinct effects on folding (Fig. 5). First, F state folding is promoted by Mss116 and other RNA-binding proteins even without ATP (Figs 3 and 4). The fact that diverse, basic RNA-binding proteins promote this stage of folding suggests that formation of F is contingent on electrostatic stabilization or annealing. The lack of ATP dependence also shows that mechanical events such as translocation or duplex unwinding are not involved in the obligate early steps of ai5γ folding. Although diverse proteins stimulate the I→F transition, only Mss116 increases the FRET ratio of I (~0.12 to ~0.18), indicating a specific interaction with the extended intermediate. This suggests a slightly more collapsed I state in the presence of Mss116, and that this protein is ideally suited to folding of ai5γ. Second, Mss116 promotes folding to the native state through a mechanism involving ATP hydrolysis. Without ATP or with AMPPNP, only a small fraction of molecules reach N, indicating that ATP hydrolysis contributes to the function of Mss116 specifically during native state formation.

Several possible models explain the function of Mss116 during native state formation. In a stabilization model, Mss116 may bind and stabilize the κ−ζ region, a substructure of domain 1 that must form before the intron can fold productively[23]. This same substructure contains a binding site for the catalytic domain 5, which docks late in the assembly pathway[24]. If Mss116 stimulates folding by stabilizing the κ−ζ element, it must dissociate before D5 can stably dock within the core. Thus, ATP hydrolysis may stimulate a conformational change in Mss116 allowing protein release during the final, rapid stage
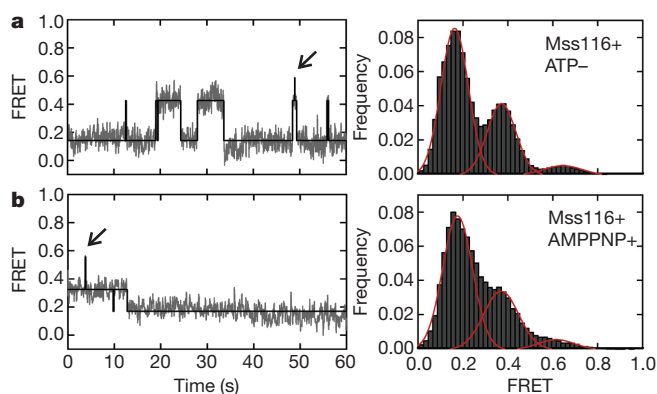
Figure 4 | Role of ATP in Mss116-mediated folding of group II introns.
a, smFRET time trace (left) and histogram (right) showing the distribution of different conformations of the Mss116-mediated folding pathway in the presence of wild-type Mss116 protein without ATP. **b**, smFRET time trace (left) and histogram (right) showing the distribution of different conformations in the presence of wild-type Mss116 protein and the non-hydrolysable ATP analogue, AMPPNP. Black lines represent idealized HMM and arrows indicate rapid transient excursions to the native state. All experiments were performed under near-physiological conditions.
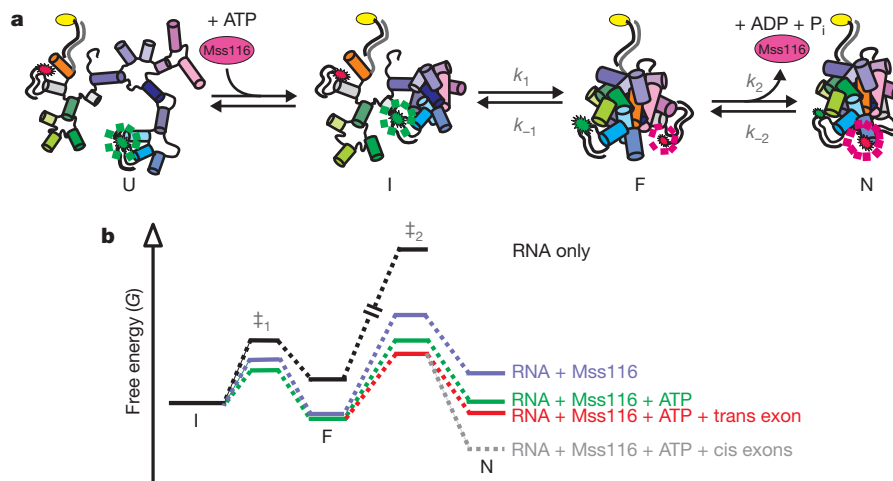
Figure 5 | Mss116-mediated group II intron folding. a, D135-L14 minimal folding pathway: unfolded (U), extended intermediate (I), folded intermediate (F) and native state (N). Folding rate constants of I, F and N are k_1, k_{-1}, k_2 and k_{-2}. **b**, Hypothetical free energy diagram based on this data and previous work[17,22]. In the absence of Mss116 and ATP (black), I is the most stable

conformation. Mss116 alone (purple), results in formation of F. Addition of both Mss116 and ATP with and without trans exon-17/7 (red and green, respectively) enhances native state formation[17]. Cis exons (grey) further stabilize N[22].

of folding, consistent with previous data linking ATP hydrolysis to Mss116 turnover and dissociation[29,30]. The small fraction of the native population in the absence of ATP is consistent with this model, as Mss116 likely has a finite off-rate. *In vivo*, efficient recycling of Mss116 may be necessary to maintain a pool of active Mss116 enzymes and may allow the cell to overcome non-functional binding events while preventing inhibition of properly folded RNAs[29]. However, other models, such as unwinding during the final transition from the folded intermediate to the native state, also remain possible. According to this model, the protein would first facilitate the slow, early stages of folding (including D1 collapse) through electrostatic interactions, but formation of the final native state may be impeded by one or more small misfolded structures. To form the correct folded structure, the protein would resolve the misfolded structures through strand exchange or unwinding by local strand separation. This ATP-dependent remodelling event would provide an opportunity for the native contacts to form. Although we do not have direct evidence for the presence of misfolded structures, we have shown previously that I, F and N are likely to be heterogeneous[17]. Thus, it is possible that misfolded intermediates with FRET ratios similar to on-pathway intermediates are present but not readily distinguishable. However, recent *in vivo* experiments show that the ability of Mss116 to hydrolyze ATP correlates with function even with different degrees of helicase activity, in agreement with a recycling function for ATPase activity[29]. Helicase activity may have additional roles in the presence of long exons[31].

In summary, we observe multiple roles for Mss116 in the folding pathway of ai5γ and propose that Mss116 mediates group II intron folding by stabilizing on-pathway intermediates and transition states. We also observe that efficient transition from the folded intermediate to the native state requires ATP hydrolysis, possibly for Mss116 release and recycling. The substrate further stabilizes the native state, raising the interesting possibility that *in vivo* and in the presence of all the group II intron domains and Mss116, the native state becomes the most stable conformation.

METHODS SUMMARY

The dynamics of D135-L14 were measured by smFRET, as described[17,18,32]. The ribozyme was immobilized on poly(ethylene glycol)-coated (PEG) quartz slides through a biotin-streptavidin linkage to reduce non-specific Mss116 binding[33,34]. The PEG coating did not affect the RNA dynamics (Fig. 2a)[17]. Under 532 nm laser excitation, the fluorescence of individual complexes was monitored using a CCD camera. Single-molecule experiments were performed at pH 7.5, 22 °C, in the presence of 25 nM Mss116 with or without 1 mM ATP, 1 mM AMPPNP and

25 nM substrate (17/7 or 17/7-dC) under high ionic strength (8 mM $MgCl_2$, 500 mM KCl and 80 mM MOPS), near-physiological (8 mM $MgCl_2$, 100 mM KCl and 40 mM MOPS) or physiological (1 mM $MgCl_2$, 100 mM KCl and 40 mM MOPS) conditions. FRET histograms showing the distribution of different structural conformations were constructed by combining the data of approximately 100 molecules. The relative stability of each conformation (I, F and N states) was quantified by obtaining the folding rate constants (k_1, k_{-1}, k_2 and k_{-2}) from the transition density plots (Supplementary Fig. 9) constructed after analysing FRET trajectories with a hidden Markov model[28].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Linder, P. Quick guide: DEAD-box proteins. *Curr. Biol.* **10**, R887 (2000).
2. Diges, C. M. & Uhlenbeck, O. C. *Escherichia coli* DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J.* **20**, 5503–5512 (2001).
3. Huang, H. R. *et al.* The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl Acad. Sci. USA* **102**, 163–168 (2005).
4. Solem, A., Zingler, N. & Pyle, A. M. A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol. Cell* **24**, 611–617 (2006).
5. Halls, C. *et al.* Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J. Mol. Biol.* **365**, 835–855 (2007).
6. Del Campo, M. *et al.* Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol. Cell* **28**, 159–166 (2007).
7. Lehmann, K. & Schmidt, U. Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol.* **38**, 249–303 (2003).
8. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38**, 1–35 (2004).
9. Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**, 1142–1152 (2001).
10. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
11. Qin, P. Z. & Pyle, A. M. The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.* **8**, 301–308 (1998).
12. Pyle, A. M. & Lambowitz, A. M. in *The RNA World* 3rd ed., Vol. 43 (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) Chap. 17 469–505 (Cold Spring Harbor Laboratory Press, 2006).
13. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K. & Perlman, P. S. Group II introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.* **12**, 1950–1958 (1992).
14. Michels, W. J. Jr & Pyle, A. M. Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* **34**, 2965–2977 (1995).
15. Qin, P. Z. & Pyle, A. M. Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg^{2+} ions in the tertiary structure. *Biochemistry* **36**, 4718–4730 (1997).

16. Su, L. J., Waldsich, C. & Pyle, A. M. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.* **33,** 6674–6687 (2005).
17. Steiner, M., Karunatilaka, K. S., Sigel, R. K. & Rueda, D. Single-molecule studies of group II intron ribozymes. *Proc. Natl Acad. Sci. USA* **105,** 13853–13858 (2008).
18. Steiner, M., Rueda, D. & Sigel, R. K. O. Ca^{2+} induces the formation of two distinct subpopulations of group II intron molecules. *Angew. Chem. Int. Ed.* **48,** 9739–9742 (2009).
19. Noah, J. W. & Lambowitz, A. M. Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* **42,** 12466–12480 (2003).
20. Zhang, A., Derbyshire, V., Salvo, J. L. & Belfort, M. *Escherichia coli* protein StpA stimulates self-splicing by promoting RNA assembly *in vitro*. *RNA* **1,** 783–793 (1995).
21. Clodi, E., Semrad, K. & Schroeder, R. Assaying RNA chaperone activity *in vivo* using a novel RNA folding trap. *EMBO J.* **18,** 3776–3782 (1999).
22. Fedorova, O., Solem, A. & Pyle, A. M. Protein-facilitated folding of group II intron ribozymes. *J. Mol. Biol.* **397,** 799–813 (2010).
23. Waldsich, C. & Pyle, A. M. A folding control element for tertiary collapse of a group II intron ribozyme. *Nature Struct. Mol. Biol.* **14,** 37–44 (2007).
24. Fedorova, O., Waldsich, C. & Pyle, A. M. Group II intron folding under near-physiological conditions: collapsing to the near-native state. *J. Mol. Biol.* **366,** 1099–1114 (2007).
25. Levin, J. G., Guo, J., Rouzina, I. & Musier-Forsyth, K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **80,** 217–286 (2005).
26. Lamichhane, R. *et al.* RNA looping by PTB: evidence using FRET and NMR spectroscopy and for a role in splicing repression. *Proc. Natl Acad. Sci. USA* **107,** 4105–4110 (2010).
27. Hopkins, J. F., Panja, S., McNeil, S. A. & Woodson, S. A. Effect of salt and RNA structure on annealing and strand displacement by Hfq. *Nucleic Acids Res.* **37,** 6205–6213 (2009).
28. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91,** 1941–1951 (2006).
29. Bifano, A. L., Turk, E. M. & Caprara, M. G. Structure-guided mutational analysis of a yeast DEAD-box protein involved in mitochondrial RNA splicing. *J. Mol. Biol.* **398,** 429–443 (2010).
30. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl Acad. Sci. USA* **105,** 20209–20214 (2008).
31. Zingler, N., Solem, A. & Pyle, A. M. Dual roles for the Mss116 cofactor during splicing of the ai5γ group II intron. *Nucleic Acids Res.* (in the press) (2010).
32. Zhao, R. & Rueda, D. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* **49,** 112–117 (2009).
33. Ha, T. *et al.* Initiation and re-initiation of DNA unwinding by the *Escherichia coli* Rep helicase. *Nature* **419,** 638–641 (2002).
34. Christian, T. D., Romano, L. J. & Rueda, D. Single molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc. Natl Acad. Sci. USA* **106,** 21109–21114 (2009).

Author Contributions D.R. and A.M.P. conceived and designed the experiments. K.S.K. performed and analysed the experiments with help from A.S. All authors wrote the manuscript.

METHODS

Preparation of DNA, RNA and protein samples. The D135-L14 RNA (638 nucleotides) was obtained by *in vitro* transcription of the HindIII-digested pT7D135-L14 plasmid DNA with T7 RNA polymerase under standard conditions[17]. Biotinylated DNA strand (5′-Biotin-TGCATGCCTGCAGGTCGACTC TA-3′) Cy3-DNA strand (5′-Cy3-ACCAAGAGCGTTATTAAT-5′), Cy5-DNA strand (5′-Cy5-ACGTAGTCCGAAATATAT-3′), wild-type RNA substrate (17/7. 5′-CGUGGUGGGACAUUUUCGAGCGGU-3′) and slower cleaving RNA substrate (17/7-dC, 5′-CGUGGUGGGACAUUUUdCGAGCGGU-3′) were purchased from Howard Hughes Medical Institution Biopolymer/Keck Foundation Biotechnology Resource Laboratory. The 2′-OH positions were deprotected following the manufacturer's protocol. All samples were purified by denaturing 18% PAGE and subsequent C8 reversed-phase HPLC, as described[17].

Mss116 was overexpressed in Rosetta 2 cells (Novagen) and grown at 16 °C overnight after induction. After purification of protein on a nickel column, it was cleaved with SUMO protease at 4 °C overnight and purified using a gel filtration column. Untagged protein was concentrated by an Amicon ultra concentrator and stored at −80 °C. The activity of Mss116 protein was confirmed by *in vitro* splicing, ATPase and unwinding assays.

Single-molecule FRET experiments. Protein-free single-molecule experiments were performed by heat-annealing (90 °C for 45 s) Cy3-DNA, Cy5-DNA, biotin-DNA (10 μM each) and D135-L14 RNA (1 μM) in 100 mM KCl, 40 mM MOPS, pH 7.5 and 0.5% 2-mercaptoethanol. After the addition of 8 mM MgCl$_2$, the reaction mixture was incubated at 30 °C for 15–20 min. The fluorophore-labelled, biotinylated RNA/DNA complex was diluted to ~25 pM and bound to a streptavidin-coated quartz slide via a biotin–streptavidin linkage. Excess fluorophore-labelled and biotin-labelled DNA strands were removed from the slide by washing with reaction buffers. To minimize the non-specific binding of protein on the slide, all quartz slides were PEG-passivated. In order to reduce photobleaching of fluorophores, an oxygen-scavenging system (OSS) consisting of 10% (w/v) glucose, 2% (v/v) 2-mercaptoethanol, 50 μg ml^{-1} glucose oxidase and 10 μg ml^{-1} catalase was used in all experiments. The smFRET experiments with Mss116 were performed by introducing 25 nM Mss116 with and without 1 mM ATP/AMPPNP in OSS solution as described earlier. Similarly, smFRET experiments with the substrate (17/7 or 17/7-dC) were performed after addition of 25 nM substrate RNA into OSS solution containing 25 nM protein and 1 mM ATP. All TIRF-based smFRET experiments were performed with 33 ms time resolution at room temperature under different experimental conditions, as described[32].

FRET histograms from different experimental conditions were constructed by collecting ~100 single-molecule trajectories showing the transitions between different structural conformations. Since we cannot distinguish molecules in the unfolded state (U, FRET ~0) from those containing only a donor fluorophore (FRET = 0, due to photobleaching, for example), we did not include them in the FRET histograms. Folding rate constants (k_1, k_{-1}, k_2 and k_{-2}) of the D135 ribozyme under different experimental conditions were obtained by analysing FRET trajectories with a hidden Markov model (HMM)[28]. First, FRET trajectories were fitted using HMM and then transition density plots (TDP) were constructed to obtain folding rate constants as described[28]. Threshold FRET values 0.27 and 0.51 were used to separate transitions among the three different FRET states (Supplementary Fig. 9). The effect of Mss116 without ATP was further determined in the presence of glucose and hexokinase. Incubation of the reaction mixture with glucose and hexokinase removes the trace amount of ATP associated with sample preparation.

Splicing assay. The body-labelled full-length ai5γ precursor RNA was incubated in 40 mM MOPS pH 7.5 at 90 °C for 1 min followed by 3 min at 30 °C. The RNA solution (1 nM final concentration) was added to a reaction mixture containing 20 nM Mss116 protein, 100 mM KCl, 40 mM MOPS, pH 7.5 and 8 mM MgCl$_2$, 5% glycerol, 2 mM dithiothreitol (DTT) and 2 U μl^{-1} RNase inhibitor. After incubation for 10 min at 30 °C, the splicing reaction was initiated with the addition of ATP (1 mM final concentration). The sample aliquots were removed at 0 and 120 min time points and detected using 5% native polyacrylamide gel.

Unwinding assay. Mss116 protein (250 nM) was pre-incubated with 0.1 nM end-labelled RNA duplex (12 base pairs) at 30 °C for 10 min and unwinding reaction was initiated by the addition of ATP as described[4]. The unwinding reaction was performed in 100 mM KCl, 40 mM MOPS, pH 7.5, 4 mM MgCl$_2$, 0.01% IGEPAL, 2 mM DTT, 2 U μl^{-1} RNase inhibitor and 4 mM ATP. The sample aliquots were removed at 0 and 30 min time points and detected using 10% native PAGE.

ATPase assay. Mss116 protein (250 nM) was preincubated with or without 5 μM single-stranded RNA at 30 °C for 10 min and all reactions were initiated by the addition of a mixture of ^{32}P-labelled and unlabelled ATP. ATPase assays were performed according to standard protocol[35] in 100 mM KCl, 40 mM MOPS, pH 7.5, 1 mM MgCl$_2$, 2 mM DTT, 2 U μl^{-1} RNase inhibitor and 1 mM ATP.

35. Shuman, S., Spencer, E., Furneaux, H. & Hurwitz, J. The role of ATP in *in vitro* vaccinia virus RNA synthesis effects of AMP-PNP and ATPγS. *J. Biol. Chem.* **255,** 5396–5403 (1980).

Higher rates of sex evolve in spatially heterogeneous environments

Lutz Becks[1,2] & Aneil F. Agrawal[1]

The evolution and maintenance of sexual reproduction has puzzled biologists for decades[1,2]. Although this field is rich in hypotheses[3–5], experimental evidence is scarce. Some important experiments have demonstrated differences in evolutionary rates between sexual and asexual populations[6–8]; other experiments have documented evolutionary changes in phenomena related to genetic mixing, such as recombination[9,10] and selfing[11]. However, direct experiments of the evolution of sex within populations are extremely rare (but see ref. 12). Here we use the rotifer, _Brachionus calyciflorus_, which is capable of both sexual and asexual reproduction, to test recent theory[13–15] predicting that there is more opportunity for sex to evolve in spatially heterogeneous environments. Replicated experimental populations of rotifers were maintained in homogeneous environments, composed of either high- or low-quality food habitats, or in heterogeneous environments that consisted of a mix of the two habitats. For populations maintained in either type of homogeneous environment, the rate of sex evolves rapidly towards zero. In contrast, higher rates of sex evolve in populations experiencing spatially heterogeneous environments. The data indicate that the higher level of sex observed under heterogeneity is not due to sex being less costly or selection against sex being less efficient; rather sex is sufficiently advantageous in heterogeneous environments to overwhelm its inherent costs[2]. Counter to some alternative theories[16,17] for the evolution of sex, there is no evidence that genetic drift plays any part in the evolution of sex in these populations.

Sex shuffles genotypes, changing genetic associations through recombination and segregation. Sex is thought to evolve as a byproduct of the selection on these altered genetic associations[4]. All theories for the evolution of sex invoke some mechanism that maintains genetic variation because shuffling without variation does not yield any change. Much of modern theory has focused on deleterious mutations[17–19] or host–parasite coevolution[20–22] as the key source of genetic variation. However, a more classic explanation for the maintenance of genetic variation is spatial heterogeneity in selection[23,24]. Several recent theoretical studies have shown that sex evolves more easily when there is spatial heterogeneity in selection[13–15].

If selection is the dominant evolutionary force shaping gene associations, then sex is usually disadvantageous. This is because selection leads to an excess of good allele combinations and sex destroys these combinations through recombination and segregation[3,25]. However, in spatially heterogeneous habitats, migration, not just selection, is important in determining gene associations. Maladaptive gene combinations are constantly introduced to local populations through migration. Sex is then potentially beneficial because it helps to break down maladaptive gene associations generated by differential selection and migration[13–15]. For sex to be favoured, the theory makes additional requirements about the nature of gene interactions (for example, locally adapted alleles should be dominant) but there are some reasons to expect these may be met[14]. Nonetheless, the theory is clear in predicting that the opportunity for sex to be advantageous is greater in spatially heterogeneous habitats than in homogeneous ones.

We tested this prediction in experimental populations of the monogonont rotifer _Brachionus calyciflorus_ evolving in either homogeneous or heterogeneous environments. Monogonont rotifers are cyclic diploid parthenogens (that is, they normally reproduce asexually) and mixis (sexual reproduction) is induced by high rotifer densities via quorum sensing[26] (Supplementary Fig. 1). A preliminary study of our source population revealed ample genetic variation for the propensity to reproduce sexually, thus providing the necessary substrate for the evolution of sex (Supplementary Fig. 2).

In our experiments, rotifers were maintained in semi-continuous cultures at large population size ($N \approx 10{,}000$). We used two different food conditions to establish differentially selective environments (Methods). Each replicate population consisted of two subpopulations. In the homogeneous treatments, both subpopulations were of the same environmental type, that is, either both high-quality food or both low-quality food. In the heterogeneous treatment, each population was composed of one high-quality food subpopulation and one low-quality food subpopulation. Migration between subpopulations was performed by weekly manual transfer of individuals, for all treatments. Two migration rates were used corresponding to $m \approx 10\%$ and $m \approx 1\%$ per generation, assuming a generation time of about one day. Observed population densities (females, males, eggs and resting eggs) were similar across all food-quality environments (Supplementary Figs 3 and 4).

To confirm that our high-quality and low-quality food environments imposed different selection regimes, we measured fitness (lifetime per capita number of offspring) after 15 weeks of evolution of rotifers from the two homogeneous treatments (Fig. 1). We found that fitness was higher when measured in their evolved habitat than when fitness was measured in the alternative habitat, confirming differential adaptation.

We used two different methods to detect changes in the rate of sex in our experiment. First, we measured the fraction of isolated individual clones (84 per population with 42 from each subpopulation) that switched to sexual reproduction when exposed to a sex-inducing stimulus under standardized conditions (Methods). We performed this assay at the start of the experiment, after 6 weeks (about 40–45 generations) and again after 12 weeks (about 80–90 generations). The measured propensity for sex was significantly higher when migration took place in a heterogeneous environment (between high- and low-quality food conditions) than in a homogeneous environment (Fig. 2 and Supplementary Fig. 5) after 6 and 12 weeks (generalized linear model (GLM), $P < 0.001$) with a rapidly decreasing rate of sex in the spatially homogeneous environments. No significant differences were observed for the two migration rates or between the two homogeneous environments (Fig. 2).

To confirm that the differences between treatments observed under standardized assay conditions reflect real differences _in situ_, we used estimates of the percentage of sexually derived offspring (resting eggs) of total offspring (resting and amictic eggs) as a second measure of changes in the rate of sex over time and among treatments (Fig. 3). In the latter part of the experiment (from day 74 to day 109; Fig. 3b), the

[1]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada. [2]Department of General Ecology, Zoological Institute, Center for Biological Sciences, University of Cologne, D-50931 Köln, Germany.
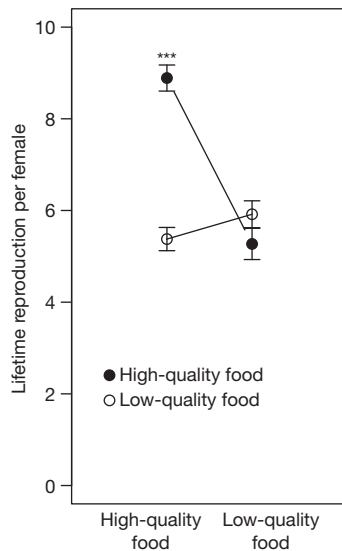
Figure 1 | Fitness in alternative environments. Lifetime reproduction (fitness) of amictic (asexual) *Brachionus calyciflorus* females was measured on individuals from populations that had evolved for 15 weeks under conditions of either high or low food quality (see text and Methods). The fitness of the same genotype of ten individual clones (third generation after isolation) was measured under high-quality food conditions (left) and low-food-quality conditions (right); $n = 18$ populations; eight populations with a migration rate of $m \approx 10\%$ and ten populations with $m \approx 1\%$ per generation. The graph shows the means $\pm$ one standard error; GLMM ***$P < 0.001$ for the high-quality food environment; the difference is not significant in the low-quality food environment. See Methods for fitness assay description.

populations reached stable densities with no significant variation in density among treatments. This permits a reasonable comparison of rates of sex among treatments during this period. The percentage of sexually derived offspring was significantly greater in the heterogeneous treatment than in the homogeneous treatments (generalized linear mixed model (GLMM): for the comparison between the heterogeneous and homogeneous high-quality food condition, $\chi^2 = 31.458$, degrees of freedom, d.f. $= 1$, $P < 0.001$; for the comparison between heterogeneous and homogeneous low-quality food conditions, $\chi^2 = 24.947$, d.f. $= 1$, $P < 0.001$). In the heterogeneous treatment, about 15% of eggs were sexually derived whereas in both of the homogeneous treatments only about 7% of eggs were sexually derived.

Both lines of evidence above indicate that sex evolves differently in heterogeneous versus homogeneous environments. Sex declines dramatically in the homogeneous treatment but little, if at all, in the heterogeneous treatment (Fig. 2 and Supplementary Fig. 5). However, there are several interpretations of this result. First, the putative benefits of sex under heterogeneity could be sufficiently large to balance its inherent costs, resulting in a higher equilibrium rate of sex than in the homogeneous environments. Second, benefits to sex may exist under heterogeneity but these are not sufficient to fully offset its costs. Consequently, sex declines in the heterogeneous treatment but at a slower rate. Third, net selection on sex does not differ between treatments (that is, there are no benefits due to heterogeneity). Rather, selection on sex is simply less efficient in the heterogeneous environment because there is more genetic variance in fitness, that is, Hill–Robertson effects[27] impede the elimination of alleles, causing higher rates of sex.

To distinguish amongst these possibilities, we restarted the experiment at week 14. We mixed all rotifer populations from the earlier experiment, combining all replicates of all three treatments (homogeneous high-quality food, homogeneous low-quality food, and heterogeneous) to create populations with an intermediate rate of sex (Fig. 2; the vertical line marks the mixing and the start of the second part of the experiment).
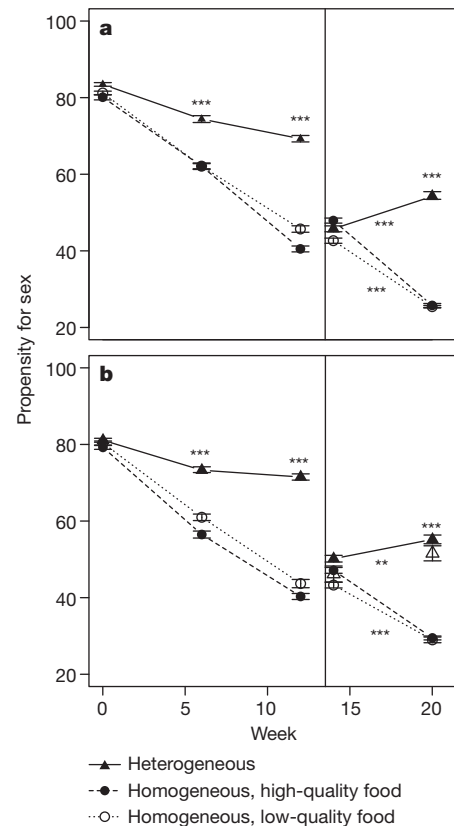

Figure 2 | Evolution of the propensity for sex in *Brachionus calyciflorus* populations from spatially heterogeneous and spatially homogeneous environments measured in a common environment. The propensity for sex was measured as the percentage of females induced into mixis (sexual reproduction) when exposed to a standardized stimulus. Vertical lines at week 14 mark the start of the second part of the experiment, when all populations were mixed and reassigned to treatments. Each data point represents the mean of 7–10 populations per treatment $\pm$ one standard error. Migration rates are shown of $m \approx 10\%$ per generation (**a**) and $m \approx 1\%$ per generation (**b**). For both migration rates, the rate of sex is significantly greater in the heterogeneous treatment than in either homogeneous treatment in comparisons at weeks 6, 12 and 20 (***$P < 0.001$ for all comparisons). Between week 14 and 20, the rate of sex significantly increases in the heterogeneous treatment (***$P < 0.001$ for $m \approx 10\%$; **$P = 0.01$ for $m \approx 1\%$). In contrast, sex declines in the homogeneous treatments (***$P < 0.001$). Open triangles (only for $m \approx 1\%$ in **b**) represent heterogeneous populations evolving at ten times the standard size of $N \approx 10,000$.

Rotifers were split again into 120 populations and populations grew again for 6 weeks under the same conditions as for weeks 0–13 (food conditions, migration pattern, number of replicates) and we measured the propensity for sex again after 6 weeks of evolution (week 20 in Fig. 2). To assess the possibility of less efficient selection on sex in the heterogeneous environment due to drift-related effects, we added three additional replicates to the heterogeneous treatment (lower migration rate only) in which the population size was increased tenfold.

After an additional 6 weeks (about 40–45 generations) of evolution, the propensity for sex continued to decrease in the homogeneous populations (GLMM: in Fig. 2a, $\chi^2 = 136.98$, d.f. $= 3$, $P < 0.001$; in Fig. 2b, $\chi^2 = 82.903$, d.f. $= 3$, $P < 0.001$, for both migration rates together, $\chi^2 = 220.86$, d.f. $= 2$, $P < 0.001$). In contrast, the propensity for sex evolved upwards in the spatially heterogeneous populations (GLMM: in Fig. 2a, $\chi^2 = 15.41$, d.f. $= 2$, $P < 0.001$; in Fig. 2b, $\chi^2 = 10.72$, d.f. $= 1$, $P = 0.01$; for both together, $\chi^2 = 20.3$, d.f. $= 1$, $P < 0.001$). This result indicates that the advantages to sex outweigh its costs under spatial heterogeneity. Moreover, there were no differences between the larger and the smaller population sizes (Fig. 2b, open
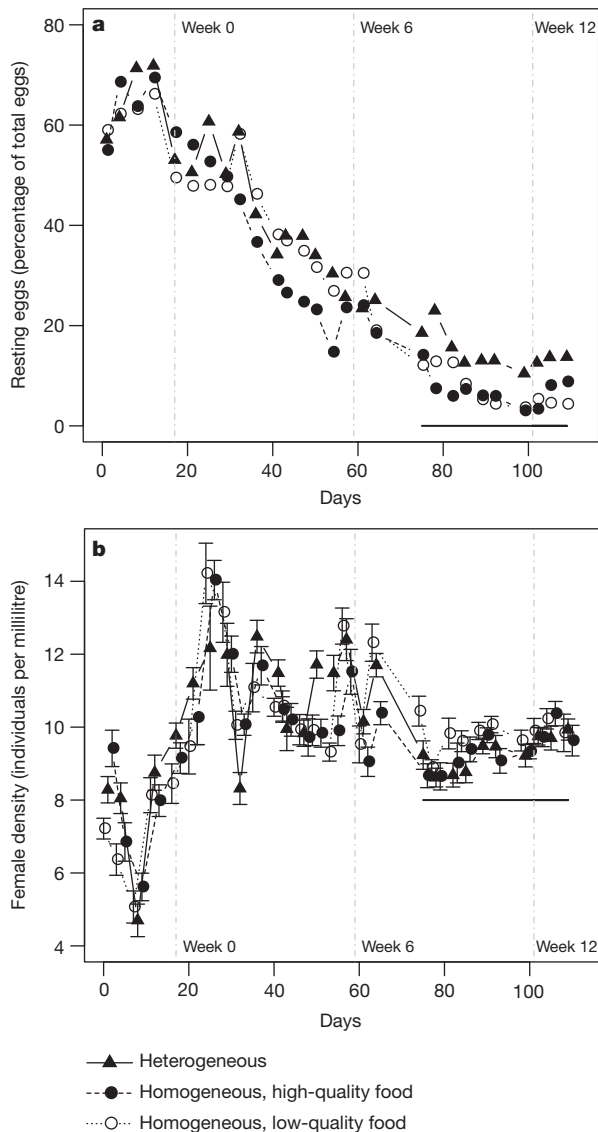
Figure 3 | *In situ* measure of the rate of sex in *Brachionus calyciflorus* populations from spatially heterogeneous and spatially homogeneous environments measured as the fraction of sexually derived offspring (resting eggs) of total offspring. Mean percentages of resting eggs out of all eggs in *Brachionus calyciflorus* populations (**a**) and mean female densities (**b**) are plotted over time. Error bars represent ± one standard error. The populations appear to reach demographic equilibrium after about 75 days. After this point (days 75–109), female densities are very similar across treatments, thereby permitting a reasonable comparison for rates of sex. During this period, the percentage of sexually produced offspring was significantly higher in the heterogeneous treatment than in the homogeneous treatments (GLMM: for the heterogeneous versus homogeneous high-quality food condition, $\chi^2 = 31.458$, d.f. = 1, $P < 0.001$; for the heterogeneous versus homogeneous low-quality food condition, $\chi^2 = 24.947$, d.f. = 1, $P < 0.001$). Vertical lines mark the time points at which the propensity for sex was measured in a common environment (Fig. 2) and horizontal lines mark the data used for the comparison of resting egg fraction between treatments.

triangles), suggesting that drift-based hypotheses for the evolution of sex[16–18] are not responsible for these results.

In this experimental system, there is the 'twofold' cost of meiosis[28] as well as a time cost because sexual reproduction takes longer than asexual reproduction. These costs tend to keep the equilibrium level of sex quite low and seem to favour complete asexuality under homogeneous conditions. Even with spatial heterogeneity, the equilibrium rate of sex is low. Nonetheless, non-zero rates of sex evolve with spatial heterogeneity despite the substantial costs. These results are therefore consistent with

recent theory predicting that there is more opportunity for sex to be favoured in spatially heterogeneous environments[13–15]. However, this experiment was not designed to determine the mechanisms favouring higher levels of sex, so caution should be used in making interpretations in this regard. Below we consider alternative interpretations.

It is possible that higher levels of sex evolved here as a correlated response to selection for resting eggs. This could occur if resting eggs survive better than amictic eggs in transfers from one habitat type to another. If this were the case, then we would expect higher levels of sex when there was more movement between habitats because the strength of selection should be directly proportional to the probability of experiencing an environmental change, which is equivalent to the migration rate. However, there is little difference in the rate of sex between the two migration rates. The similarity between migration treatments is also somewhat surprising with respect to migration models of sex[13–15], though less so. Such models predict that sex is most strongly favoured at some intermediate level of migration, where the 'optimal' migration rate depends on the genetic architecture of the locally adapted traits. It is possible that the equilibrium level of sex may be fairly robust to quantitative differences in migration rate under the right genetic conditions.

Adaptation to new environments is also thought to facilitate the evolution of sex[2,10,16]. Both habitats used here were slightly different to previous culture conditions, so adaptation is likely to have occurred in all treatments. Interestingly, sex seems to increase initially (days 0 to 16 in Fig. 3a) in all treatments, despite low density during this period, then later declines. This observation is consistent with sex being advantageous when the rate of adaptation is likely to be highest but then becoming less favoured (owing to its costs) after adaptation slows. Although this pattern is interesting, it seems unlikely that 'sex for adaptation' explains the differences that develop between homogeneous and heterogeneous treatments over the longer term.

Recent theory has shown that sex may evolve as a way for genes to escape bad genetic backgrounds[3,29]; this is known as 'fitness-associated' sex. With spatial heterogeneity, maladaptive genotypes are constantly introduced through migration, providing more opportunity for fitness-associated sex to evolve. Under this hypothesis, we would expect that genotypes have more sex when tested in the alternative environment from which they were collected. However, we find no support for this prediction. Clones collected from the two different habitats in the heterogeneous environment show no difference with respect to the rate of sex when tested under the same environment (Supplementary Fig. 6).

The evolution of sex has been one of the enduring problems of evolutionary biology. Although there has been a large amount of theory, experimental tests of sex itself have been limited to comparisons of sexual and asexual populations[6,7]. Experimental evolution has been largely neglected as a means of examining the more perplexing question of how rates of sex evolve within populations (but see ref. 12). Our experiment demonstrates that rates of sex do evolve within experimental populations and that non-zero rates of sex can be favoured even in the face of real costs. This suggests that this approach can be applied to a variety of facultatively sexual organisms (able to reproduce both sexually and asexually) to test the predictions of different theoretical models. Future work should not only identify the types of conditions that favour the evolution of sex but also examine the population genetic mechanisms by which these benefits arise[4]. By doing so, we can begin bridging the sizeable gap between theory and the empirical patterns observed in nature.

It is worth noting that the rate of sex declined in all treatments compared to the initial state (Fig. 3). Because these experiments were started with field-collected organisms, this observation suggests that the equilibrium rate of sex in nature is higher than in any of the laboratory environments used here. Whether this discrepancy is due to greater environmental heterogeneity in the field or other factors (such as parasites or more targets of selection) is unknown and presents a challenge for future studies.

METHODS SUMMARY

Creation of experimental populations. Populations were initiated from a laboratory population that had been recently hatched from a large number of field-collected resting eggs. Rotifer–algae populations were established in 500-millilitre batch cultures under two sets of conditions to create different selection regimes: high-quality food conditions and low-quality food conditions. 10% of each culture (including rotifers and algae) was replaced twice per week with the respective algae solution (2×10^6 cells ml^{-1}) taken from constant chemostat cultures. Each population was composed of two subpopulations: for the homogeneous treatment, both were either from low-quality food or both from high-quality food conditions, and for the heterogeneous treatment, the subpopulations were of opposite types. Once per week, 50% or 5% of the rotifers (females, males, eggs, resting eggs) were exchanged between the two subpopulations. Assuming a generation time of 1 to 2 days, these migration regimes are equivalent to approximately $m \approx 10\%$ and $m \approx 1\%$ per generation, respectively.

Assay of propensity to reproduce sexually. The propensity for sex was measured by isolating 42 (amictic) clones from each subpopulation (84 clones per population). Clones were individually transferred into single wells with 10 ml of high-quality food and maintained in these conditions for two amictic generations to eliminate environmental effects. For the mixis assay, two neonates of the third generation after isolation were individually transferred to a single cell of a 96-well plate with conditioned medium (sex-inducing stimulus) at two different concentrations. Conditioned medium was freshly prepared from a dense laboratory rotifer stock culture by removing rotifers from the medium and diluting it with fresh medium to a stimulus concentration representative of 22.5 females per millilitre. After the first day of the assay the initial females were removed and the offspring were scored as amictic or mictic by their produced offspring: males are produced by sexual females and females by amictic females. For the assays at the start, after 6 and 12 weeks, an additional set of tests using a lower concentration of sex-inducing stimulus (12 females per millilitre) were also performed. Results were qualitatively similar (Supplementary Fig. 5).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (University of California Press, 1982).
2. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
3. Otto, S. P. The evolutionary enigma of sex. *Am. Nat.* **174,** S1–S14 (2009).
4. Agrawal, A. F. Evolution of sex: why do organisms shuffle their genotypes? *Curr. Biol.* **16,** R696–R704 (2006).
5. Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84,** 372–387 (1993).
6. Colegrave, N. Sex releases the speed limit on evolution. *Nature* **420,** 664–666 (2002).
7. Goddard, M. R., Godfray, H. C. J. & Burt, A. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434,** 636–640 (2005).
8. Poon, A. & Chao, L. Drift increases the advantage of sex in RNA bacteriophage Phi 6. *Genetics* **166,** 19–24 (2004).
9. Korol, A. B. & Iliadi, K. G. Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila*. *Heredity* **72,** 64–68 (1994).
10. Otto, S. P. & Barton, N. H. Selection for recombination in small populations. *Evolution* **55,** 1921–1931 (2001).
11. Morran, L. T., Parmenter, M. D. & Phillips, P. C. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* **462,** 350–352 (2009).
12. Wolf, H. G., Wöhrmann, K. & Tomiuk, J. Experimental evidence for the adaptative value of sexual reproduction. *Genetica* **72,** 151–159 (1987).
13. Lenormand, T. & Otto, S. P. The evolution of recombination in a heterogeneous environment. *Genetics* **156,** 423–438 (2000).
14. Agrawal, A. F. Spatial heterogeneity and the evolution of sex in diploids. *Am. Nat.* **174,** S54–S70 (2009).
15. Pylkov, K. V., Zhivotovsky, L. A. & Feldman, M. W. Migration versus mutation in the evolution of recombination under multilocus selection. *Genet. Res.* **71,** 247–256 (1998).
16. Felsenstein, J. & Yokoyama, S. Evolutionary advantage of recombination. 2. Individual selection for recombination. *Genetics* **83,** 845–859 (1976).
17. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443,** 89–92 (2006).
18. Roze, D. & Michod, R. D. Deleterious mutations and selection for sex in finite diploid populations. *Genetics* **184,** 1095–1112 (2010).
19. Kondrashov, A. S. Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. *Genet. Res.* **44,** 199–217 (1984).
20. Agrawal, A. F. Differences between selection on sex versus recombination in Red Queen models with diploid hosts. *Evolution* **63,** 2131–2141 (2009).
21. Otto, S. P. & Nuismer, S. L. Species interactions and the evolution of sex. *Science* **304,** 1018–1020 (2004).
22. Peters, A. D. & Lively, C. M. Short- and long-term benefits and detriments to recombination under antagonistic coevolution. *J. Evol. Biol.* **20,** 1206–1217 (2007).
23. Hedrick, P. W. Genetic-polymorphism in heterogeneous environment—a decade later. *Annu. Rev. Ecol. Syst.* **17,** 535–566 (1986).
24. Felsenstein, J. Theoretical population-genetics of variable selection and migration. *Annu. Rev. Genet.* **10,** 253–280 (1976).
25. Feldman, M. W., Otto, S. P. & Christiansen, F. B. Population genetic perspectives on the evolution of recombination. *Annu. Rev. Genet.* **30,** 261–295 (1997).
26. Gilbert, J. J. Specificity of crowding response that induces sexuality in the rotifer *Brachionus*. *Limnol. Oceanogr.* **48,** 1297–1303 (2003).
27. Hill, W. G. & Robertson, A. Effects of linkage on limits to artificial selection. *Genet. Res.* **8,** 269–294 (1966).
28. Williams, G. C. *Sex and Evolution* (Princeton University Press, 1975).
29. Hadany, L. & Otto, S. P. The evolution of condition-dependent sex in the face of high costs. *Genetics* **176,** 1713–1727 (2007).

METHODS

All cultures were inoculated from a *Brachionus calyciflorus* laboratory stock culture, which was started in June 2009 from field-collected resting egg hatches. Rotifer–algae cultures were kept at $25 \pm 1.5\,°C$ (12 h dark/12 h light) in tissue culture flasks (Sarsted) and moved randomly three times per week across three shelves of an incubator. Each population consisted of two subpopulations. Each subpopulation was maintained in a 500-ml batch culture at a density of about ten rotifers per millilitre, that is, about 5,000 rotifers per subpopulation giving $N \approx 10,000$ per population (detailed density data provided in Supplementary Fig. 3). Each subpopulation was one of two habitat types: high-quality food conditions or low-quality food conditions. 10% of each subpopulation (including rotifers and algae) was removed twice per week and an equivalent volume of the appropriate algae solution (below) was replaced. From the removed sample, rotifer densities (females, males, eggs, resting eggs) were enumerated under a stereoscope. For the replacement solution, algae were taken from long-term chemostat cultures to ensure constant food conditions over the course of experiment (high-quality food condition: nitrogen concentration in medium = 1,000 μM per litre; low-quality food condition: nitrogen concentration in medium = 160 μM with an additional $0.5\,g\,ml^{-1}$ NaCl per litre); both chemostats were inoculated from the same stock four weeks before the start of the experiment (SAG 278-3, Algae Culture Collection University of Göttingen). The replacement algal solution was prepared by diluting either of the two chemostat cultures to a concentration of 2×10^6 algal cells per millilitre in nitrogen-free inorganic medium (modified after ref. 30 with $0.5\,g\,l^{-1}$ NaCl added to the low-quality food environment).

Asexual rotifer females produced about one offspring per day and juveniles started reproducing a few hours after hatching (observed by L.B.). Migration of rotifers (females, males, eggs, resting eggs) between two subpopulations took place once per week by filtering out the rotifers (adults and eggs) from 50% or 5% of the volume from each subpopulation and exchanging the rotifers between subpopulation pairs. Assuming a generation time of 1 to 2 days, a 50% or 5% weekly exchange corresponds to $m \approx 10\%$ or $m \approx 1\%$ per generation, respectively. Migration started two weeks (day 18) after rotifer inoculation.

At week 15 (week 13 of migration), ten clones were isolated from each subpopulation and transferred individually to 10 ml of high-quality food. Two neonates of the third generation after isolation were used to assay lifetime reproduction (fitness) under either high-quality food or low-quality food conditions. The number of offspring of females was recorded daily and females were transferred to new wells with fresh medium at the same time until the females died.

Sex stimulus in standardized environment. Mixis in *Brachionus calyciflorus* is stimulated through a mixis-inducing protein that is excreted by *Brachionus* females and accumulates in the medium[26] (Supplementary Fig. 1). Conditioned medium (sex stimulus) was prepared freshly from a dense laboratory stock culture by removing rotifers (repeated filtration through 10 μm mesh) from the medium and diluting it with fresh medium to a desired 'equivalent rotifer density' of 22.5 or 12 females per millilitre (ref. 26). For the second part of the experiment after week 14, rotifers were only tested at 22.5 females per millilitre.

The propensity for sex was measured by isolating 42 asexual (amictic) clones from each subpopulation (84 clones per population). Clones were individually transferred into single wells with 10 ml of high-quality food and maintained under these conditions for two generations. For the mixis assay, two neonates of the third generation after isolation were individually transferred to a single cell of a 96-well plate with conditioned medium at the two different concentrations. After the first day of the assay the initial females were removed and her offspring were scored as amictic or mictic by the type of offspring that they produced. Amictic (asexual) females produce female offspring. Sexual females produce only haploid males because they are unmated in this assay (Supplementary Fig. 1).

Creation of populations with a low initial rate of sex. After week 13, experimental populations from all treatments were mixed. Because two-thirds of these populations were from the homogeneous treatment, the average rate of sex in this newly mixed population was quite low. New experimental populations were created from this mixed population and distributed amongst the same set of treatments as in the original experiment. To assess whether genetic drift was affecting evolution in the heterogeneous treatment, three larger populations were added to this treatment (for the lower migration rate only). The larger populations grew under the same conditions but each population consisted of 20 subpopulations (ten of each type) rather than two. For migration, 10% of the volume of each of the ten subpopulations of the same habitat type was removed and pooled. Half of this volume was equally distributed back to the same subpopulations (within-habitat migration) and the remainder was equally distributed amongst the ten alternative habitat subpopulations (between-habitat migration).

Data analysis. Multivariate analyses were performed in the R statistical environment using the lmer4 package[31,32]. Fitness data were analysed using generalized mixed models with environmental origin (the high- or low-quality food environment to which the populations adapted) as a fixed effect and replicated population as a random effect for the comparison of the mean fitness tested within one environment (either original or novel in Fig. 1; GLMM with quasi-Poisson error structure). Differences among environments (heterogeneous, homogeneous low-quality food, homogeneous high-quality food) in the percentages of mixis-induced females in the mixis assay (Fig. 2) were tested by using a GLM (with binomial error distribution). The effect of population size (Hill–Robertson effect) on the rate of sex within the heterogeneous treatment was tested in the same way. To test for differences in the rate of sex between the two migration rates, we compared generalized linear models with and without the migration term. For each environmental treatment, GLMM were used to test the change in the rate of sex between weeks 14 and 20, with time as a fixed effect and replicated population as a random effect (for the heterogeneous treatment, both population sizes were included). Differences in the percentage of resting eggs out of all eggs (*in situ*) were estimated from resting egg and amictic egg densities (days 75 to 109); statistical differences were determined by comparing generalized mixed models with and without treatment, treating replicated population as a random effect (quasi-binomial error correction).

30. Fussmann, G. F., Ellner, S. P., Shertzer, K. W. & Hairston, N. G. Jr. Crossing the Hopf bifurcation in a live predator-prey system. *Science* **290,** 1358–1360 (2000).
31. Team, R. D. C. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing Vienna, 2009).
32. Bates, D. & Maechler, M. *lme4: Linear Mixed-Effects Models Using S4 Classes.* R package version 0. 999375-31 ⟨http://CRAN.R-project.org/package=lme4⟩ (2009).

LETTER

Satellite phage TLCφ enables toxigenic conversion by CTX phage through *dif* site alteration

Faizule Hassan[1], M. Kamruzzaman[1], John J. Mekalanos[2] & Shah M. Faruque[1]

Bacterial chromosomes often carry integrated genetic elements (for example plasmids, transposons, prophages and islands) whose precise function and contribution to the evolutionary fitness of the host bacterium are unknown. The CTXφ prophage, which encodes cholera toxin in *Vibrio cholerae*[1], is known to be adjacent to a chromosomally integrated element of unknown function termed the toxin-linked cryptic (TLC)[2]. Here we report the characterization of a TLC-related element that corresponds to the genome of a satellite filamentous phage (TLC-Knφ1), which uses the morphogenesis genes of another filamentous phage (fs2φ) to form infectious TLC-Knφ1 phage particles. The TLC-Knφ1 phage genome carries a sequence similar to the *dif* recombination sequence, which functions in chromosome dimer resolution using XerC and XerD recombinases[3]. The *dif* sequence is also exploited by lysogenic filamentous phages (for example CTXφ) for chromosomal integration of their genomes. Bacterial cells defective in the dimer resolution often show an aberrant filamentous cell morphology[3,4]. We found that acquisition and chromosomal integration of the TLC-Knφ1 genome restored a perfect *dif* site and normal morphology to *V. cholerae* wild-type and mutant strains with *dif*⁻ filamentation phenotypes. Furthermore, lysogeny of a *dif*⁻ non-toxigenic *V. cholerae* with TLC-Knφ1 promoted its subsequent toxigenic conversion through integration of CTXφ into the restored *dif* site. These results reveal a remarkable level of cooperative interactions between multiple filamentous phages in the emergence of the bacterial pathogen that causes cholera.

The TLC element of *V. cholerae* encodes the Cri replicase with homology to filamentous phage replication proteins and TlcR, a protein with sequence similarity to RstR, the repressor controlling lysogeny of the filamentous CTXφ and the target for anti-repression by the RstC product of satellite filamentous phage RS1φ (refs 1, 2, 5–9). For these reasons we speculated that the TLC element might correspond to the genome of a satellite filamentous phage that depended on another filamentous phage for its morphogenesis. As a prelude to the study described here, we devised a screen for the postulated 'TLC helper phage' and thus identified filamentous phage fs2φ as this helper[10]. In brief, our evidence (see Supplementary Information) that fs2φ is a TLC helper phage includes the following. First, strains encoding genetically marked versions of the TLC element (for example TLC-Kn1) inserted in their chromosome produce infectious TLC-Knφ1 phage particles only if also infected with fs2φ. Second, these TLC-related phage particles carry single-stranded DNA (ssDNA) corresponding to a circularized variant of the TLC element. Third, TLC-Knφ1 phages infect only cells expressing mannose-sensitive haemagglutinin (MSHA) pili, the known receptor of fs2φ (ref. 11). On infection of MSHA-positive vibrios, the TLC-Knφ1 ssDNA present in phage particles is converted to the double-stranded replicative form that is detectable in infected cells as a plasmid or as a chromosomally integrated copy. Fourth, the double-stranded replicative form of TLC-Knφ1 (designated pTLC-Kn1) was also shown to be sufficient for formation of TLC-Knφ1 phage in recipient cells provided that the cells are also infected with

fs2φ. Thus, fs2φ is a helper phage that provides essential gene products required for morphogenesis of TLC-Knφ1 phage particles.

For a better understanding of the biology of TLCφ, we sequenced pTLC-Knφ1 and its chromosomally integrated form in strain AL33457-TLC-Kn1. Strain AL33457 was found to carry two copies of the TLC element that flank a unique open reading frame (ORF; VC1471; Fig. 1). Each of the two copies of chromosomally integrated TLC elements in AL33457 is composed of five ORFs, spanning from VC1466 to VC1470 and from VC1472 to VC1476, respectively. In strain AL33457-TLC-Kn1, the Kn^R determinant was located in VC1470 and thus, like ORF VC1471, it was located between the duplicated copies of TLC. Nucleotide sequence analysis of pTLC-Kn1 indicated that this plasmid probably formed as a result of recombination between two directly repeated 25-base-pair (bp) sequences (5′-ACATAATGCGCACTAGGAACATTTT-3′), which are located in the 3′ end of VC1465 and within VC1471 (Fig. 1). This 25-bp sequence within VC1471 overlaps by 18 bp (bold nucleotides) with the 28-bp *dif1* sequence, 5′-ATTTAACATA**ACATACATAATGCGCACT**-3′ (refs 12, 13). *Dif1* is a site on the large chromosome of *V. cholerae* that is required for the XerC/XerD-mediated resolution of chromosome dimers, and similar sites are also exploited by various filamentous phages for integration of their genomes into the host chromosome using XerC/XerD-mediated recombination[3,4,12–15]. The *dif1* sequence is used by CTXφ and RS1φ for their chromosomal integration though XerC/XerD-mediated recombination with the corresponding *dif/attP* site formed by the annealing of ssDNA derived from phage genomes[12,13]. The recombination event that formed pTLC-Kn1 looped out the entire region between the 25-bp duplicated sequence in VC1465 and VC1471, including the 18 bp identical to part of the *dif1* sequence together with most of the ORF defined as VC1471. Thus, TLC-Knφ1 and pTLC-Kn1 encode part of the *dif1* sequence (Fig. 1).

These observations suggest that naturally occurring TLC-related phages might be capable of reconstituting a functional chromosomal *dif* sequence by recombining its partial *dif* site with a defective *dif*-like sequence during lysogenic integration of its genome into the chromosome. To test this hypothesis, we screened a collection of 97 clinical or environmental *V. cholerae* strains of both O1 and non-O1 serogroups and identified 18 strains that were negative for one or more chromosomal regions including TLC, VC1471 and the *dif* sequence[3,12,13]. These included 12 non-O1, non-O139 strains and 6 non-toxigenic O1 strains. Sequencing of the relevant region in five such TLC-negative CTX-negative O1 strains (AO12682, AO7543, AV2684248, AN19908 and AN25049; see Supplementary Table 1) revealed a gap between the *rtxA* gene (VC1451) and the gene designated VC1479 in all the five strains analysed. In toxigenic strains of the seventh pandemic El Tor biotype such as N16961, the CTX prophage and the TLC genes as well as the recombined *dif*-like sites formed by integration of CTXφ are located in this space[16]. In the intergenic region between *rtxA* and VC1479 in the TLC-negative strains we identified a possible defective *dif*-like sequence differing in two nucleotide pairs (G→A and C→T) from the genuine *dif* sequence (Fig. 2).

[1]Molecular Genetics Laboratory, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka-1212, Bangladesh. [2]Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.
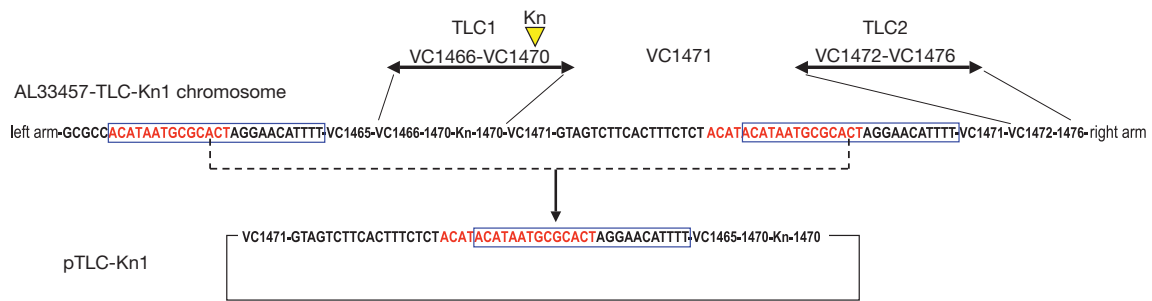
Figure 1 | Schematic diagram showing the formation of plasmid pTLC-Kn1. A conservative recombination event (dotted line and arrow) occurring between two identical 25-bp sequences (boxed nucleotides) located near the 3′ end of VC1465 and within VC1471 excised the plasmid from the chromosome of the *V. cholerae* AL33457-TLC-Kn1 parental strain. This strain carries two chromosomal copies of the TLC element with a kanamycin resistance (Kn) insert located in VC1470 of the first TLC element (TLC1).

pTLC-Kn1 is therefore a circularized TLC-related element that carries most of VC1471 (a gene located between the 25-bp sequences). Within VC1471 there also exists an 18-bp region (red) that is identical to part of the known 28-bp *dif1* sequence. TLC1 is composed of genes VC1466, VC1467, VC1468, VC1469 and VC1470; TCL2 is composed of genes VC1472, VC1473, VC1474, VC1475 and VC1476. For simplicity, only a subset of these genes is shown in this diagram.

Cultures of *V. cholerae* strains with deletions in the *dif* recombination site are known to contain a subpopulation of cells with a filamentous morphology[3]. These filaments reflect aberrant cell division resulting from a defect in XerC/XerD-mediated chromosome dimer resolution[3]. We examined whether naturally occurring *V. cholerae* O1 strains that lack TLC, or the genuine *dif* sequence, have a filamentous morphology. As shown in Fig. 3 and Supplementary Table 3, these strains do indeed have a filamentous morphology for a noticeable subpopulation of their cells. We next tested whether transduction with TLC-Knϕ1 phage could correct this morphology defect. In each case, cell filamentation in these *V. cholerae* strains was found to be eliminated after transduction with TLC-Knϕ1 (Fig. 3 and Supplementary Table 3). These results suggest that these filamentous strains are indeed *dif*-deficient and that lysogeny with TLC-Knϕ1 apparently corrected this defect.

For further verification of the natural formation of TLC-related phages and the role of the *dif*-like sequence encoded by VC1471 in the correction of *dif*-deficient phenotypes, we used a set of chromosomal

transposon insertion mutants of C6706 (ref. 17). We selected five strains carrying TnFGL3 insertions in the different ORFs of TLC (VC1466, VC1467, VC1468, VC1469, VC1470 and VC1471). With the exception of VC1469, for each strain we were able to recover a plasmid corresponding to the double-stranded replicative form of the TLCϕ-related genome with a TnFGL3 insertion in the corresponding ORF. No plasmid recovery was actually expected for the insertion mutant in VC1469 because this gene encoded Cri, the protein required for the replication of TLC-related plasmids[2]. When these TLC-related plasmids were introduced into each of the strains SA317, AO7543 and AO12682 (naturally occurring strains with a filamentous morphology and negative for VC1471), normal cellular morphology was restored except in one case. The double-stranded replicative-form plasmid derived from the mutant carrying a TnFGL3 insertion in VC1471 failed to complement the morphology defect in these strains (Supplementary Fig. 4). The TnFGL3 insertion in VC1471 is located within the *dif*-like sequence (bold) encoded by VC1471 (insertion indicated by the asterisk in the sequence 5′-**ACATACA*TAATGCGCACT**AGGAACA-3′). We conclude that
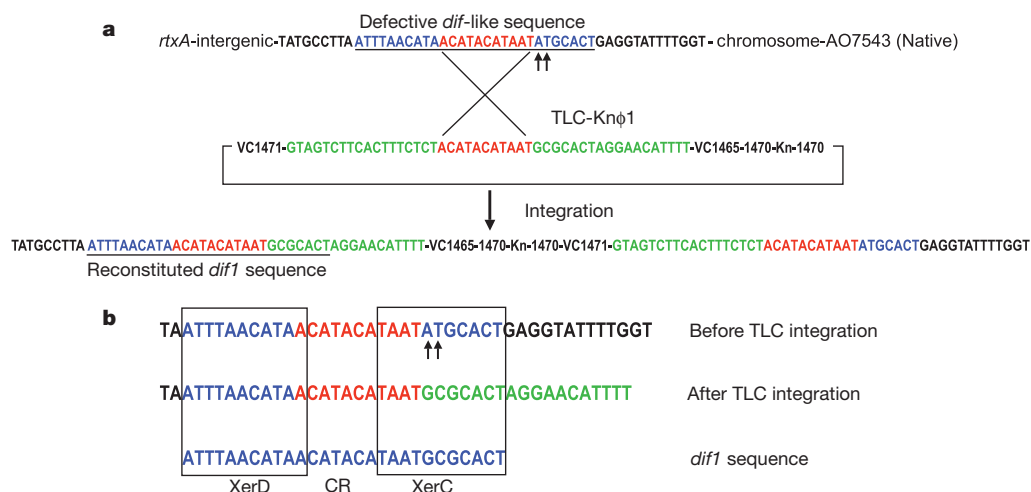

Figure 2 | Restoration of a perfect *dif* site in a *dif*-defective *V. cholerae* O1 strain through integration of the TLCϕ genome. a, Schematic diagram showing site-specific integration of the TLC-Knϕ1 genome into the chromosome of strain AO7543. The region in the vicinity of the phage attachment site is shown in green; the region in the vicinity of the chromosomal attachment site is shown in blue. The chromosomal attachment site corresponds to a defective *dif*-like sequence altered in an AT dinucleotide (arrows). Sequence analysis indicates that the recombination event that integrated the phage genome occurred in the central region of sequence identity (red), which also corresponds in part to the central region (CR) of a *dif1* site (see **b**). This recombination event was probably generated at least in part by the action of XerC and XerD on chromosomal and phage nucleic acid substrates in

that the TLC-Kn1 genome did not integrate chromosomally in XerC-defective and XerD-defective strains (see the text and Supplementary Information). Formation of the TLC prophage by integration resulted in the formation of a functional *dif1* sequence (underlined) on its left border. The central sequence of identity is duplicated on the right border but the *dif1* sequence is not duplicated. **b**, Base-pair alignments between the defective *dif*-like sequence on the chromosome of strain AO7543 that is the target for TLC-Knϕ1 integration, the resultant hybrid sequence found after integration of TLC-Knϕ1 (left end) and the authentic *V. cholerae dif1* sequence. Colours of nucleotides correspond to those highlighted in **a**. The binding sequences for XerC and XerD recombinases are indicated by boxes.
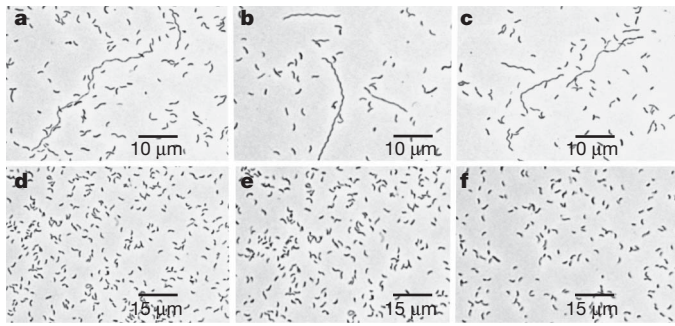
Figure 3 | Transduction with TLC-Knφ1 phage cures cell filamentation of *Vibrio cholerae* O1 strains that have probable defects in the resolution of chromosome dimers. **a–c**, Morphology of three strains (**a**, AO12682; **b**, AO7543; **c**, AV2684248) before infection with TLC-Knφ1; see Supplementary Tables 1 and 4 for details. **d–f**, Morphology of AO12682 (**d**), AO7543 (**e**) and AV2684248 (**f**) after infection and chromosomal integration of TLC-Knφ1.

the *dif*-like sequence present in VC1471 is required to correct morphological defects when TLC-related plasmids are introduced into naturally occurring *dif*⁻/TLC⁻ strains of *V. cholerae*.

We studied the integration of the TLC-Knφ1 genome into the chromosome of the naturally occurring TLC-negative strains AO7543 and AO12682 after infecting these strains with TLC-Knφ1 phage particles. As expected, Southern blot hybridization (Supplementary Fig. 5) and polymerase chain reaction (PCR) analysis indicated that the TLC-Knφ1 genome had inserted into the bacterial genome in the intergenic region between *rtxA* and VC1479. The correction of the cell filamentation phenotype of *dif*-deficient strains was observed only in strains in which the TLC-Knφ1 genome had integrated into the chromosome. In contrast, introduction of a pUC18 clone of VC1471, designated pVC1471, into the defective cells did not cure the cell filamentation phenotype even though it carried the *dif*-like sequence. This finding indicated that the *dif*-like sequence present in VC1471 functions only *in cis* and not *in trans*. This is what one would expect if the *dif*-like sequence in VC1471 could indeed function in recombination with XerC/XerD to resolve chromosomal dimers only if it recombined into the chromosome. To verify this assumption, we used mutants of *dif*⁻ strains with transposon insertions in the *XerC* or *XerD* genes. As expected, transduction of TLC-Knφ1 into these *XerC*-defective or *XerD*-defective strains did not cure their cell filamentation phenotype (Supplementary Table 4). Furthermore, PCR-based analyses as described above confirmed that the TLC-Knφ1 DNA did not integrate into the chromosome of the *XerC* or *XerD* mutant strains.

To further examine the mechanism associated with the elimination of the *dif*⁻ defect through the chromosomal integration of TLC-Knφ1 DNA, we sequenced the junction of several independent TLC-Knφ1 integration events. The sequence analysis showed that TLC-Knφ1 DNA had integrated into the intergenic region between *rtxA* and VC1479 in strains AO7543, AV2684248 and AO12682 by using its defective *dif*-like site as a recombinational substrate (Fig. 2a). The recombination event leading to the integration of the TLC-Knφ1 genome resulted in the formation of a functional *dif* sequence identical to the genuine *dif1* sequence reported for *V. cholerae*[3,13] (Fig. 2b). This result also suggests that TLC-deficient *V. cholerae* strains contain alternative *dif*-like sequences that can still function in recombination with a TLC-encoded *dif*-like sequence but are not fully functional in chromosome dimer resolution. We conclude that the *dif*-like site in VC1471 recombines with the defective chromosomal *dif*-like sequence in these TLC-negative strains during the process of XerC/D-mediated TLC genome insertion and that the product of this integration event generates a *dif* sequence that is functional in chromosomal dimer resolution.

In toxigenic *V. cholerae,* the CTXφ genome exists as a prophage inserted into the bacterial chromosome at the *dif* recombination

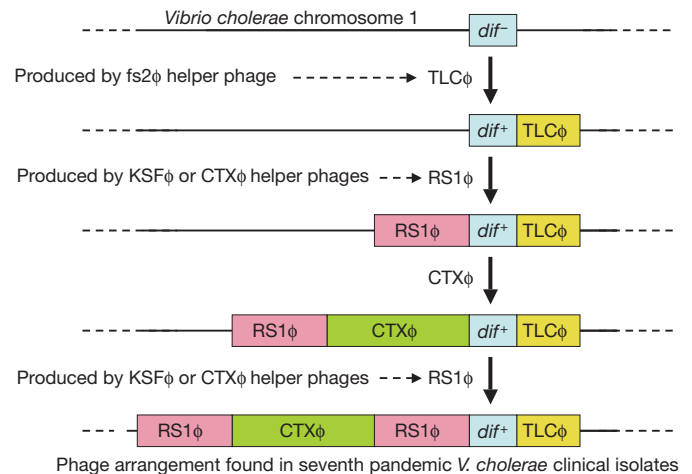Phage arrangement found in seventh pandemic *V. cholerae* clinical isolates

Figure 4 | Schematic diagram of interactions between phages and bacteria that probably led to the emergence of the seventh pandemic clone of *V. cholerae*. Boxes indicate the integrated prophages and *dif* sequence in the seventh pandemic strain, and the solid lines indicate the corresponding empty sites where these elements are absent in the precursor strains. Sequential lysogeny by three different filamentous phages (TLCφ, CTXφ and RS1φ) and the role of two helper phages (fs2 and KSF-1) are shown. To generate the observed organization of RS1 prophages found in most seventh pandemic strains, it is postulated that two rounds of RS1 prophage integration would be needed if this prophage integrated into only a functional *dif* site and then reconstituted only one *dif* site after integration. For a more complete explanation of these hypothetical steps see Supplementary Fig. 6.

site[12,13,15]. Because transduction with TLC-Knφ1 reconstructs a functional *dif* sequence in the recipient bacterium (Fig. 2), we tested whether TLC-Knφ1 transductants could be stably lysogenized by CTXφ. We chose test strains that were positive for the TCP locus (which encodes the receptor for CTXφ (ref. 1)), and used a CTXφ prophage that was marked with a chloramphenicol resistance marker (CTX-Cmφ). As expected, we found that TLC-Knφ1 transductants were readily superinfected with CTX-Cmφ and in these cases CTX-Cmφ was found integrated into the *dif* site generated through previous integration of TLC-Knφ1. In contrast, although natural TLC-negative strains could also be infected with CTX-Cmφ (Supplementary Table 6), the CTX-Cmφ genome did not integrate, and the unintegrated CTX-Cmφ genome was rapidly lost when inoculated into the intestinal loops of adult rabbits. Because the integrated form of the CTXφ genome is known to be more stably retained in *V. cholerae* than the unintegrated plasmid form is[18], these data strongly argue that TLCφ is crucial to the natural, stable acquisition of CTXφ.

Although the TLC element was known to be invariably present in all CTX-positive strains and notably absent in CTX-negative strains[2], the role of this element in the evolution of toxigenic *V. cholerae* was not clear until the present study. Here we show that the TLC element can give rise to infectious phage particles (TLCφ) when its morphogenesis is supported by another filamentous phage, previously designated fs2φ (ref. 10). Furthermore, infectious forms of TLCφ that encode a *dif*-like sequence can be easily isolated. These specialized TLCφ-related transducing phages can, after chromosomal integration, generate a functional *dif* sequence and correct aberrant filamentous morphology present in TLC-negative cells that apparently show defective *dif*/XerC/XerD-mediated chromosome dimer resolution. Lysogeny by theae TLC phages leads to the restoration of a functional *dif* site, which is also essential for the stable integration of CTXφ and the conversion of *V. cholerae* to a toxigenic form.

The most common strains of *V. cholerae* causing cholera in the world today are all highly related to the seventh pandemic clone of *V. cholerae*, which emerged as a human pathogen in 1971 in the Celebes Islands[19,20]. The arrangement of the TLC prophage, and the *dif* site used by CTXφ in these highly successful pandemic strains, is

virtually the same as the one that we produced experimentally in this study by lysogeny of dif^- strains such as AO7543 sequentially with TLC-Knφ1 followed by CTX-Cmφ. It therefore seems highly likely that the precursor of the seventh pandemic clone was a dif^- strain that emerged as a pandemic pathogen after sequential lysogeny by three filamentous phages, namely TLCφ, CTXφ and RS1φ (Fig. 4). Because dif^- defects are deleterious to growth, it is possible that the precursor of the seventh pandemic clone may be rare in the environment or that the dif^- genotype confers an as yet undetermined advantage for non-toxigenic O1 strains in the context of its environmental niche. Nonetheless, our data suggest that the evolutionary emergence of the toxigenic seventh pandemic clone of *V. cholerae* probably involved molecular interactions between two satellite filamentous phages (TLCφ and RS1φ), three helper filamentous phages (fs2φ, CTXφ and KSFφ (ref. 21)), and two type IV pilus-based phage receptors (MSHA and TCP) (Supplementary Fig. 6). Accordingly, our results provide a model for understanding the cooperative interactions of multiple genetic elements in the evolution of pathogenic bacteria from non-pathogenic environmental progenitors.

METHODS SUMMARY

A genetic marker encoding kanamycin resistance (Kn^R) was introduced into the TLC element carried by multiple *V. cholerae* strains followed by screening of the marked strains for the production of TLC-related Kn^R transducing particles in the culture supernatants. Digestion with mung bean nuclease and also hybridization analysis with strand-specific oligonucleotide probes corresponding to the (+) and (−) strands of the TLC element were conducted to test whether the DNA carried by putative TLC-related phage particles present in filter-sterilized culture supernatants was single-stranded. The role of fs2φ as a helper of TLC satellite phage was established by demonstrating the formation of TLC-Knφ1 phage in recipient cells that contained pTLC-Kn1, provided that the cells were also infected with fs2φ.

For transduction assays, recipient *V. cholerae* strains were mixed with genetically marked phage preparations[5], and transductants were selected by using Luria–Bertani agar medium containing appropriate antibiotics. Integration of the TLC-Knφ1 genome was detected by Southern blot hybridization and PCR assays using two primers, of which one was complementary to the chromosomal region and the other corresponded to pTLC-Knφ1. DNA sequencing was conducted for further confirmation of the integration event and to detect the generation of the *dif* sequence. Subsequently, a chloramphenicol resistance (Cm^R)-marked CTX phage was used to study the susceptibility and chromosomal integration[5,18] of CTX phages into the restored *dif* site. The Cm^R-marked CTX phage genome (pCTX-Cm) was constructed by replacing the Kn^R marker in pCTX-Km (ref. 1) derived from strain SM44 with a Cm^R cassette.

The full list of strains and plasmids is available as Supplementary Table 1. Full methods and associated references are provided in Supplementary Methods.

1. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous bacteriophage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
2. Rubin, E. J., Lin, W., Mekalanos, J. J. & Waldor, M. K. Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. *Mol. Microbiol.* **28**, 1247–1254 (1998).
3. Huber, K. E. & Waldor, M. K. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature* **417**, 656–659 (2002).
4. Blakely, G. *et al.* Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75**, 351–361 (1993).
5. Faruque, S. M. *et al.* RS1 element of *Vibrio cholerae* can propagate horizontally as a filamentous phage exploiting the morphogenesis genes of CTXφ. *Infect. Immun.* **70**, 163–170 (2002).
6. Davis, B. M., Kimsey, H. H., Kane, A. V. & Waldor, M. K. A satellite phage-encoded antirepressor induces repressor aggregation and cholera toxin gene transfer. *EMBO J.* **21**, 4240–4249 (2002).
7. Kimsey, H. H. & Waldor, M. K. The CTXφ repressor RstR binds DNA cooperatively to form tetrameric repressor-operator complexes. *J. Biol. Chem.* **279**, 2640–2647 (2004).
8. Waldor, M. K., Rubin, E. J., Pearson, G. D., Kimsey, H. & Mekalanos, J. J. Regulation, replication, and integration functions of the *Vibrio cholerae* CTXΦ are encoded by region RS2. *Mol. Microbiol.* **24**, 917–926 (1997).
9. Davis, B. M. & Waldor, M. K. CTXΦ contains a hybrid genome derived from tandemly integrated elements. *Proc. Natl Acad. Sci. USA* **97**, 8572–8577 (2000).
10. Ikema, M. & Honma, Y. A novel filamentous phage, fs-2, of *Vibrio cholerae* O139. *Microbiology* **144**, 1901–1906 (1998).
11. Ehara, M. *et al.* Characterization of filamentous phages of *Vibrio cholerae* O139 and O1. *FEMS Microbiol. Lett.* **154**, 293–301 (1997).
12. Val, M. E. *et al.* The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol. Cell* **19**, 559–566 (2005).
13. Das, B., Bischerour, J., Val, M. E. & Barre, F. X. Molecular keys of the tropism of integration of the cholera toxin phage. *Proc. Natl Acad. Sci. USA* **107**, 4377–4382 (2010).
14. Iida, T. *et al.* Filamentous bacteriophage of Vibrios are integrated into the *dif*-like site of the host chromosome. *J. Bacteriol.* **184**, 4933–4935 (2002).
15. McLeod, S. M. & Waldor, M. K. Characterization of XerC and XerD-dependent CTX phage integration in *Vibrio cholerae*. *Mol. Microbiol.* **54**, 935–947 (2004).
16. Heidelberg, J. F. *et al.* DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
17. Cameron, D. E., Urbach, J. M. & Mekalanos, J. J. A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **105**, 8736–8741 (2008).
18. Faruque, S. M. *et al.* Diminished diarrheal response to *Vibrio cholerae* strains carrying the replicative form of the CTXΦ genome instead of CTXΦ lysogens in adult rabbits. *Infect. Immun.* **69**, 6084–6090 (2001).
19. Faruque, S. M., Albert, M. J. & Mekalanos, J. J. Epidemiology, genetics and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* **62**, 1301–1314 (1998).
20. Kaper, J. B., Morris, J. G. & Levine, M. M. Cholera. *Clin. Microbiol. Rev.* **8**, 48–86 (1995).
21. Faruque, S. M. *et al.* CTX phage-independent production of RS1 satellite phage by *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **100**, 1280–1285 (2003).

Author Contributions F.H., M.K. and S.M.F. conducted the experiments and performed analyses of bacterial strains and phages. S.M.F. and J.J.M. designed the studies, analysed data and wrote the manuscript.

Author Information The sequences described in the article are deposited in GenBank under accession numbers HM134797, HM134798, HM134799 and HM134800. 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 this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M.F. (faruque@icddrb.org).

Pericytes are required for blood–brain barrier integrity during embryogenesis

Richard Daneman[1], Lu Zhou[2], Amanuel A. Kebede[1] & Ben A. Barres[2]

Vascular endothelial cells in the central nervous system (CNS) form a barrier that restricts the movement of molecules and ions between the blood and the brain. This blood–brain barrier (BBB) is crucial to ensure proper neuronal function and protect the CNS from injury and disease[1]. Transplantation studies have demonstrated that the BBB is not intrinsic to the endothelial cells, but is induced by interactions with the neural cells[2]. Owing to the close spatial relationship between astrocytes and endothelial cells, it has been hypothesized that astrocytes induce this critical barrier postnatally[3], but the timing of BBB formation has been controversial[4–9]. Here we demonstrate that the barrier is formed during embryogenesis as endothelial cells invade the CNS and pericytes are recruited to the nascent vessels, over a week before astrocyte generation. Analysing mice with null and hypomorphic alleles of *Pdgfrb*, which have defects in pericyte generation, we demonstrate that pericytes are necessary for the formation of the BBB, and that absolute pericyte coverage determines relative vascular permeability. We demonstrate that pericytes regulate functional aspects of the BBB, including the formation of tight junctions and vesicle trafficking in CNS endothelial cells. Pericytes do not induce BBB-specific gene expression in CNS endothelial cells, but inhibit the expression of molecules that increase vascular permeability and CNS immune cell infiltration. These data indicate that pericyte–endothelial cell interactions are critical to regulate the BBB during development, and disruption of these interactions may lead to BBB dysfunction and neuroinflammation during CNS injury and disease.

To understand the mechanisms of BBB formation, we investigated the sequence of cell generation and BBB formation in the developing CNS (Fig. 1 and Supplementary Figs 1–7). In the rat cerebral cortex, angiogenesis begins at embryonic day 12 (E12), as endothelial cells invade the neural tissue from the surrounding vascular plexus (Fig. 1a and Supplementary Fig. 1). Pericytes—platelet-derived growth factor receptor-β (PDGFR-β)[+]NG2[+] vascular support cells—are found associated with endothelial tubes as nascent vessels are generated (Fig. 1b and Supplementary Fig. 1). Neural cells are produced from progenitors in a defined sequence, with neurons generated before glia. Oligodendroglia progenitor cells (OPCs), PDGFR-α[+]NG2[+] glial cells, are first observed at E19 and migrate throughout the cortex by birth (Fig. 1c). Astrocytes are first generated directly after birth and extend processes which contact vessels during the first postnatal week (Fig. 1d and Supplementary Fig. 2). If astrocytes induce the BBB, barrier properties should only be acquired after birth.

We next examined endothelial cell protein expression and barrier function during rat cortical development. BBB-forming endothelial cells are characterized by tight junctions, low rates of transcytosis, and the expression of specialized influx and efflux transporters. Tight junction molecules occludin, claudin 5 and ZO-1 were each expressed at endothelial junctions as early as E12 (Fig. 1e, Supplementary Fig. 4 and not shown). The same was observed with the BBB-specific influx transporter Glut1 (Fig. 1f and Supplementary Fig. 3). Notably, the BBB-specific efflux transporter Pgp is expressed at low

levels during embryogenesis, but increases during postnatal development (Fig. 1g), indicating a distinct regulation mechanism for efflux transport. Similar timing of cell generation and BBB gene expression was observed in the developing mouse, with vascularization of the cortex starting at E11 (Supplementary Fig. 6). The expression of genes that increase vascular permeability, including transcytosis (*Plvap*) and leukocyte infiltration (*Icam1*), decreased after initial high expression (Supplementary Fig. 7). The developmental timing of BBB function was examined by trans-cardiac perfusion with tracers. In adults, the small molecule tracer biotin stays within the lumen of CNS vessels and doesn't diffuse into the CNS parenchyma, whereas in non-neural tissue the tracer diffuses throughout the extracellular space (Fig. 1h–j). We used this method to examine BBB function in postnatal animals and dissected embryos. At each age tested (E15, E21, P1, P15 and P20) CNS vessels excluded the tracer from the CNS parenchyma (Fig. 1h–j and not shown). In embryonic time points, the tracer was excluded from most of the CNS; however, distinct regions of the CNS displayed leakiness, including regions close to the pia (Supplementary Fig. 5), indicating that serum contents may enter the CNS through pial coverings or choroid plexus. Thus, a functional BBB is present during embryogenesis before astrocyte generation.

Because pericyte recruitment to CNS vessels temporally correlates with the onset of barrier properties, we examined the role of these mural cells in regulating BBB function, structure and gene expression. Although pericytes are associated with the vasculature throughout the body, and thus are unlikely candidates to regulate brain-specific vascular properties, recent studies have demonstrated that CNS pericytes have a different developmental origin from other pericytes[10], and several studies have suggested that pericytes are capable of regulating BBB properties *in vitro*[11–13]. To determine if pericytes are necessary for BBB formation we compared the vascular permeability of *Pdgfrb*[−/−] mice with littermate controls. PDGFBB signalling through PDGFR-β is essential for pericyte generation, and mice deficient for either the ligand or receptor completely lack CNS pericytes[14,15], exhibit endothelial cell hyperplasia, increased vessel diameter and morphological signs of increased vascular permeability[16]. These mice die at birth; therefore, we examined BBB function in dissected embryos. Indeed, *Pdgfrb*[−/−] mice show an increased vascular permeability to biotin (0.5 kDa), as observed by an increase in tracer staining throughout the CNS parenchyma (Fig. 2a, b). Decreasing permeability was observed with increasing size of tracer (Fig. 2c).

CNS vessels have the highest pericyte coverage of any vessels, and the extent of pericyte coverage in vessels throughout the body inversely correlates with the relative permeability of these vessels[17]. To determine whether pericyte number, and not just presence or absence, is an important regulator of BBB permeability, we measured CNS vascular permeability in mice with different combinations of null, hypomorphic and wild-type *Pdgfrb* alleles. One study[15] generated an allelic series of *Pdgfrb* hypomorphs, showing that varying the strength of PDGFR-β signalling leads to different pericyte:endothelial cell ratios. We compared the vascular permeability of *Pdgfrb*[F7/−] and *Pdgfrb*[F7/F7],

[1]UCSF Department of Anatomy, 513 Parnassus Avenue, HSW1301, San Francisco, California 94143-0452, USA. [2]Stanford University School of Medicine, Department of Neurobiology, Fairchild Science Building D200, Stanford, California 94305-5125, USA.
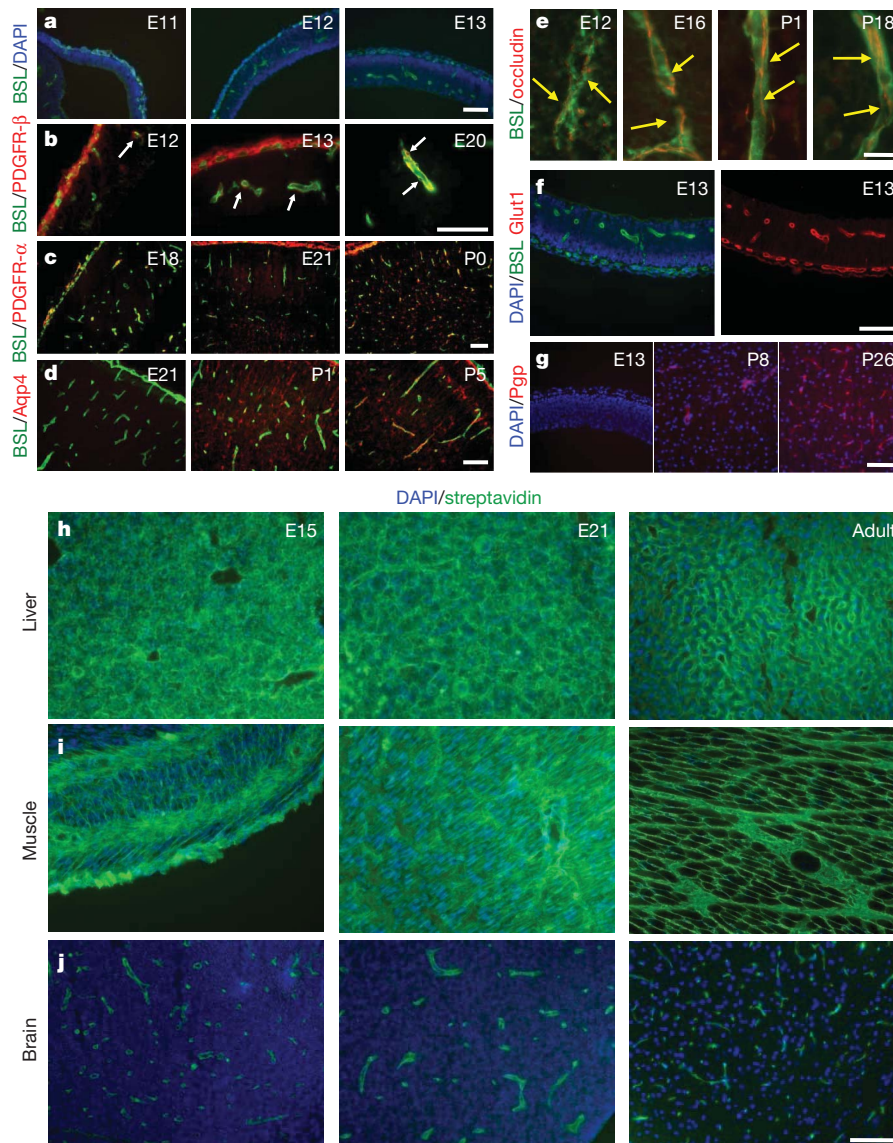
Figure 1 | Time course of cell generation and BBB development in the rat cerebral cortex. a–g, Sections of rat cerebral cortex at indicated ages were stained for endothelial cells with *Bandeiraea simplicifolia* lectin I (BSL) (green, **a–f**) and nuclei with DAPI (blue, **a**, **f** (left), **g**), pericytes with anti-PDGFR-β (red, **b**; white arrows point to pericytes), oligodendrocyte progenitors with anti-PDGFR-α (red, **c**), astrocytes with anti-aquaporin 4 (red, **d**), anti-occludin (red, **e**; yellow arrows indicate tight-junction strands), anti-Glut1 (red, **f** (right)), and anti-Pgp (red, **g**). Scale bars represent 100 μm (**a–d**, **f**, **g**) and 20 μm (**e**). **h–j,** Rats aged E15 (left), E21 (middle) and adults (right) were given a trans-cardiac perfusion of biotin, and liver (**h**), muscle (**i**) and brain (**j**) tissue sections were stained with streptavidin (green) and DAPI (blue). Scale bar represents 100 μm.

which respectively have approximately 40% and 50% of the pericyte coverage of littermate controls[15] (Fig. 2d–g and Supplementary Fig. 8). Indeed, the permeability of the BBB in these neonatal mice, to both Evan's blue and biotin, inversely correlated with the pericyte coverage (Fig. 2h–k), demonstrating that pericyte number determines relative permeability of CNS vessels during development.

We next addressed the question of whether pericytes regulate endothelial transcytosis, tight-junction formation and/or transporter expression. Examination by electron microscopy revealed that CNS endothelial cells of *Pdgfrb*$^{-/-}$ mice displayed an increase in membrane folding with membrane protrusions into the vessel lumen, as well as an increase in the number of cytoplasmic vesicles, indicating an increased rate of transcytosis (Fig. 3A–C). Indeed, examination of biotin leakage by electron microscopy, using streptavidin–horseradish peroxidise (HRP), revealed that the tracer was endocytosed into endothelial cell vesicles (Fig. 3A, d, e, and B, d, e), demonstrating transcellular passage is an important component of BBB leakiness in *Pdgfrb*$^{-/-}$ mice. *Pdgfrb*$^{-/-}$ mice and littermate controls expressed tight-junction proteins

occludin and claudin 5 at equal levels and these molecules were localized to cellular junctions (Figs 3E, F and 4a, b). Structural abnormalities in the tight junctions of the *Pdgfrb*$^{-/-}$ mice were observed by electron microscopy (Fig. 3A, c and B, c). In control animals, the length of the junction was parallel to the plane of the vessel lumen; however, in *Pdgfrb*$^{-/-}$ mice, the junctional alignment was random, being parallel, perpendicular, or at various angles to the lumen, with junctional membranes often dipping into the lumen or CNS parenchyma (Fig. 3A, c, B, c, and D). The polarized expression of Glut1 at the abluminal membrane was not affected (Supplementary Fig. 9).

To determine whether junctional abnormalities are due to a lack of pericytes, or other defects in *Pdgfrb*$^{-/-}$ mice, we analysed the function, expression and localization of tight-junction proteins in purified brain endothelial cells cultured alone or with a feeding layer of purified brain pericytes. Purified CNS endothelial cells expressed the tight-junction proteins occludin and claudin 5, which were localized to cellular borders (Fig. 3G, H). Notably, endothelial cells cultured alone often contained large gaps between cell connections, whereas the endothelial
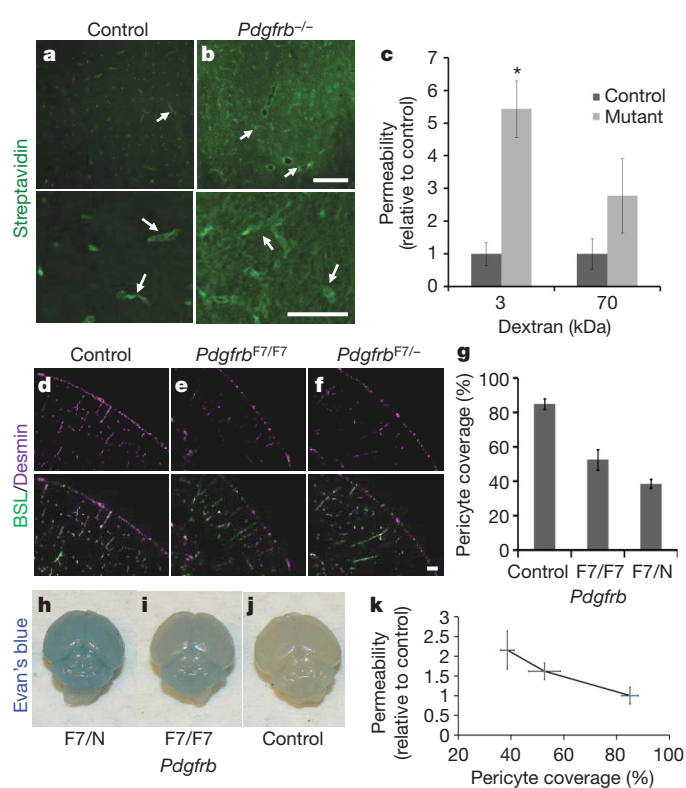
Figure 2 | Pericytes are required for BBB formation. a, b, E18 $Pdgfrb^{-/-}$ mice (**b**) and littermate controls (**a**) were given a trans-cardiac perfusion of biotin, and tissue sections were stained with streptavidin (green; white arrows indicate tracer in vessels). Scale bars represent 200 μm (upper panel) and 100 μm (lower panel). **c**, E18 $Pdgfrb^{-/-}$ mice and littermate controls were given a trans-cardiac perfusion of 3 kDa or 70 kDa biotinylated dextran, tissue sections stained with streptavidin-Alexa 488, fluorescence was quantified in ImageJ and permeability relative to control was graphed. *$P < 0.05$ by Student's t-test. **d–f**, Neonatal mouse cerebral cortex from $Pdgfrb^{F7/-}$ (**f**), $Pdgfrb^{F7/F7}$ (**e**) and littermate controls (**d**) were stained with BSL (green, **d–f** (bottom)) and for pericytes with anti-desmin (purple, **d–f**). Scale bar represents 100 μm. **g**, Pericyte coverage of CNS vessels in $Pdgfrb^{F7/-}$, $Pdgfrb^{F7/F7}$ and littermate control mice was quantified by analysing per cent length of BSL$^+$ vessels opposed to desmin$^+$ pericytes. **h–j**, P5 $Pdgfrb^{F7/-}$ mice (**h**), $Pdgfrb^{F7/F7}$ mice (**i**) and littermate controls (**j**) were given an intraperitoneal injection of Evan's blue dye, and their brains were dissected the following day after PBS perfusion. **k**, Neonatal $Pdgfrb^{F7/-}$, $Pdgfrb^{F7/F7}$ and littermate controls were given a trans-cardiac perfusion of biotin and leakage was quantified in tissue sections with streptavidin-Alexa-488 (y axis) and graphed versus pericyte coverage (x axis; values from panel **g**). All error bars represent s.e.m.
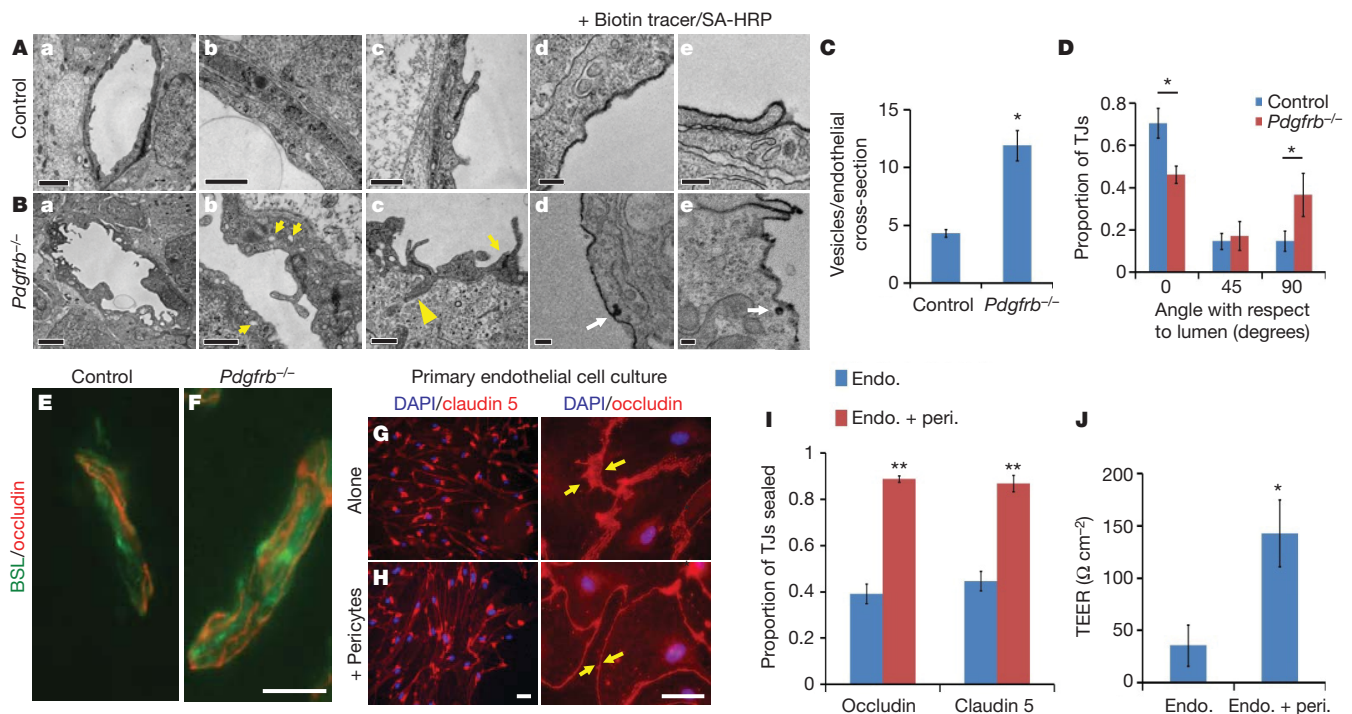

Figure 3 | Pericytes regulate structural aspects of the BBB. A, B, Electron microscopy images of CNS vessels from E18 $Pdgfrb^{-/-}$ mice (**B**) and littermate controls (**A**) including whole endothelial cell cross-sections (**a**), cytoplasm (**b**; yellow arrows indicate cytoplasmic vessels), tight junctions (**c**; yellow arrows indicate altered junction alignment; yellow arrowheads indicate junctions dipping into parenchyma), and after perfusion with biotin followed by staining with streptavidin–HRP (**d, e**; white arrows indicate uptake of tracer). Scale bars represent 2 μm (**a**), 0.5 μm (**b, c**) and 0.2 μm (**d, e**). **C**, Quantification of the number of vesicles per endothelial cross-section for $Pdgfrb^{-/-}$ mice and littermates. **D**, Angles of tight junctions (TJs) for $Pdgfrb^{-/-}$ mice and littermate controls were classified as parallel to the lumen (0°), perpendicular to the lumen (90°) or in between (45°). *$P < 0.05$ by Student's t-test. **E, F**, Cerebral cortex of E18 $Pdgfrb^{-/-}$ mice (**F**) and littermate controls (**E**) were stained with BSL (green) and anti-occludin (red). Scale bars represent 20 μm. **G, H**, Purified murine brain endothelial cells were cultured alone (**G**) or with a feeding layer of purified brain pericytes (**H**) and stained with DAPI (blue) and anti-claudin 5 (left, red) or anti-occludin (right, red; yellow arrows indicate cell borders). Scale bars represent 100 μm (left) and 50 μm (right). **I**, Per cent length of sealed claudin 5 and occludin junctions in endothelial cells cultured alone or with pericyte feeder layers. **$P < 0.01$ by Student's t-test. **J**, TEER measurements for purified murine brain endothelial cells cultured alone or with a feeding layer of purified brain pericytes. *$P < 0.05$ by Student's t-test. All error bars represent s.e.m.
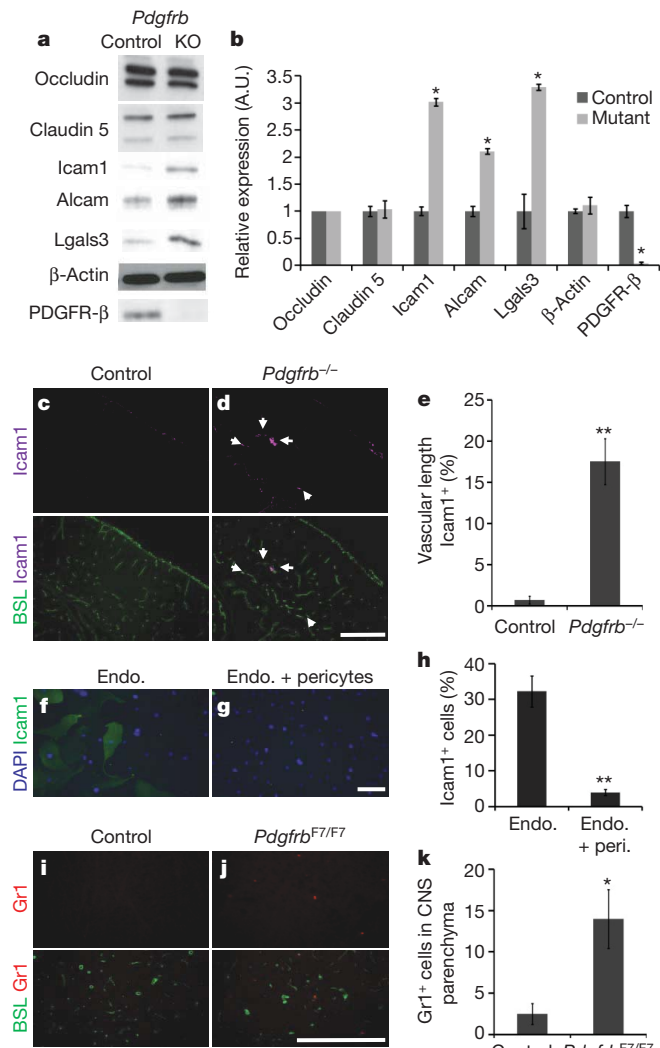
Figure 4 | Vascular expression of LAMs in *Pdgfrb*⁻/⁻ mice. a, b, Western blots of brain lysates from E18 *Pdgfrb*⁻/⁻ (KO) and littermate controls, probing occludin, claudin 5, Icam1, Alcam, Lgals3, β-actin and PDGFR-β. **a,** Representative blots; **b,** quantification; *$P < 0.05$ by Student's t-test. **c–e,** Cerebral cortex of E18 *Pdgfrb*⁻/⁻ mice (**d**) and littermate controls (**c**) were stained with anti-Icam1 (purple) and BSL (green, bottom; white arrows indicate Icam1⁺ vessels), and per cent Icam1⁺ vascular length was quantified (**e**). Scale bar represents 250 μm. **$P < 0.005$ by Student t-test. **f–h,** Purified murine brain endothelial cells cultured alone (**f**) or with a feeding layer of purified brain pericytes (**g**) and stained for DAPI (blue) and anti-Icam1 (green), and proportion of Icam1⁺ cells was quantified (**h**). Scale bar represents 200 μm. **$P < 0.005$ by Student t-test. **i–k,** Five-week-old *Pdgfrb*^F7/F7^ mice (**j**) and littermate controls (**i**) were stained with anti-Gr1 (red) and BSL (green, bottom), and number of Gr1⁺ cells per sagittal section was counted (**k**). Scale bar represents 250 μm. *$P < 0.05$ by Students t-test. All error bars represent s.e.m.

junctions sealed and the intercellular space diminished when they were cultured with pericytes (Fig. 3G–I). This was accompanied by a four-fold increase in trans-endothelial cell electrical resistance (TEER), demonstrating that endothelial tight junctions are functionally tighter in the presence of pericytes (Fig. 3J).

To determine whether pericytes regulate endothelial gene expression, we used microarray analysis to compare the gene expression of CD31⁺ vascular cells purified from *Pdgfrb*⁻/⁻ mice and littermate controls. In *Pdgfrb*⁻/⁻ mice, no downregulation of BBB genes was observed, including tight-junction proteins and transporters (Supplementary Table 1 and Fig. 4a, b). Thus, pericytes are not necessary for induction of BBB-specific genes. In the absence of pericytes, however, expression of genes known to increase vascular permeability, including *Angpt2*,

Plvap and leukocyte adhesion molecules (LAMs), were upregulated. In *Pdgfrb*⁻/⁻ mice we observed a loss of vascular *Angpt1*, a Tie2 receptor ligand that decreases permeability, and a corresponding increase in *Angpt2*, a Tie2 ligand that increases vascular permeability[18,19] (Supplementary Table 1), and co-culture of a feeder layer of pericytes with CNS endothelial cells was sufficient to decrease endothelial Angpt2 expression (Supplementary Fig. 12). *Pdgfrb*⁻/⁻ mice also exhibited an increase in endothelial *Plvap* expression (Supplementary Table 1), which was verified by immunofluorescence (Supplementary Fig. 10). Plvap is involved in endothelial vesicle trafficking, and is highly expressed in permeable peripheral vessels and is upregulated in CNS endothelial cells during pathological breakdown of the BBB[20,21]. Therefore, pericytes may limit transcytosis by suppressing Plvap.

We further examined the gene expression of acutely purified pericytes to identify pericyte-secreted signals (Supplementary Table 2). We identified that pericytes express molecules that regulate BBB properties including *Angpt1* and *Ace2* (refs 12, 22), as well as a number of other signalling molecules and matrix components. Furthermore, the extracellular matrix is altered in the *Pdgfrb*⁻/⁻ mouse, including a decrease in several collagen subunits and an upregulation of MMP9, a protease that increases the permeability of the BBB during disease (Supplementary Fig. 11)[23,24].

The transcription of several LAMs, including *Icam1*, *Alcam* and *Lgals3*, were significantly upregulated in the vasculature of *Pdgfrb*⁻/⁻ mice (Supplementary Table 1), and we confirmed a corresponding increase in protein expression (Fig. 4a, b). We found that Icam1 was robustly upregulated in specific segments of the vasculature in *Pdgfrb*⁻/⁻ mice (Fig. 4c–e), and further demonstrated that co-culture of pericytes with endothelial cells is sufficient to decrease the number of Icam1⁺ endothelial cells (Fig. 4f–h). These data indicate that pericytes may provide signals that limit CNS immune surveillance, a critical feature of the BBB. Indeed, we observed an increase in the number of Gr1⁺ leukocytes in the CNS parenchyma of pericyte-deficient mice (Fig. 4i–k).

These findings demonstrate that pericytes are required for BBB formation during development, and absolute pericyte coverage controls relative BBB permeability. Although astrocytes are not required to initially induce the BBB, it remains likely that astrocytes act to regulate the maintenance of the BBB and its response to neural function and disease. In fact, reactive astrocytes are necessary for BBB repair after injury[25]. Interestingly, even in the absence of pericytes, embryonic CNS endothelial cells express many BBB-specific molecules, indicating that neural progenitors may induce BBB gene expression. Indeed, neural-progenitor-derived Wnts are critical for CNS angiogenesis and the induction of BBB-specific transporter expression[26–28]. We propose a new model for BBB formation: endothelial cells are induced to express BBB-specific genes by interactions with neural progenitors, and then the functional integrity of the BBB is regulated by pericytes during development and astrocytes in adulthood (Supplementary Fig. 13).

These findings have important implications for understanding BBB breakdown and repair, which is a common component of many neurodegenerative disorders, stroke, brain tumours, CNS trauma, multiple sclerosis and diabetic retinopathy[1]. Notably, loss of pericytes occurs at early stages of diabetic retinopathy, and thus probably has a role in the vascular permeability defects observed during this disease[29]. The finding that pericytes are necessary for limiting CNS immune surveillance has implications for controlling neuroinflammation. For instance, pericytes inhibit the expression of Alcam, which is upregulated in CNS vessels in multiple sclerosis (MS) patients and enhances leukocyte migration across endothelial cells[30]. Identification of molecular mechanisms by which pericytes regulate BBB integrity may lead to new treatments for these diseases.

METHODS SUMMARY

Immunohistochemistry. Rodent tissues were fixed with 4% paraformaldehyde, 10% trichloroacetic acid or 95% ethanol followed by acetone. Tissue sections were blocked with 50% goat serum, permeabilized with 0.2% Triton X-100, and stained

with appropriate primary and secondary antibodies. Slides were mounted in vectashield with DAPI (Vector Labs) and visualized by fluorescence microscopy.

BBB permeability assays. Anaesthetized rodents were given a trans-cardiac perfusion, with a Dynamax peristaltic pump, of DPBS (Gibco) containing EZ-link sulfo NHS Biotin (Pierce) or biotinylated fixable dextrans (Invitrogen), followed by 4% paraformaldehyde. Tissue cryosections were stained with streptavidin Alexa-488 (Invitrogen) and analysed by fluorescence microscopy.

Cell culture. CNS endothelial cells were purified by anti-CD31 immunopanning and CNS pericytes were purified by anti-PDGFR-β immunopanning. For immunofluorescence analysis, CNS endothelial cells were grown on coverslips alone, or with a feeder layer of CNS pericytes grown on cell culture inserts. For TEER measurements, CNS endothelial cells were grown on cell culture inserts alone, or with a pericyte feeder layer grown in the culture well bottom beneath, and resistance measurements were recorded with an EVOM voltmeter.

Electron microscopy. E18 $Pdgfrb^{-/-}$ and littermate controls were fixed in 2% glutaraldehyde/4% paraformaldehyde, and processed for electron microscopy. To image tracer permeability, E18 $Pdgfrb^{-/-}$ and littermate controls were given a trans-cardiac perfusion of DPBS (Gibco) containing EZ-link sulfo NHS Biotin (Pierce) followed by 4% paraformaldehyde, and frozen tissue sections were fixed with 2% glutaraldehyde/4% paraformaldehyde then stained with a HRP-tagged streptavidin (Invitrogen) followed by DAB reaction.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Zlokovic, B. V. The blood-brain barrier in health and chronic neurodegenerative disorders. *Neuron* **57,** 178–201 (2008).
2. Stewart, P. A. & Wiley, M. J. Developing nervous tissue induces formation of blood-brain barrier characteristics in invading endothelial cells: a study using quail–chick transplantation chimeras. *Dev. Biol.* **84,** 183–192 (1981).
3. Janzer, R. C. & Raff, M. C. Astrocytes induce blood-brain barrier properties in endothelial cells. *Nature* **325,** 253–257 (1987).
4. Bauer, H. C. et al. Neovascularization and the appearance of morphological characteristics of the blood-brain barrier in the embryonic mouse central nervous system. *Brain Res. Dev. Brain Res.* **75,** 269–278 (1993).
5. Bolz, S., Farrell, C. L., Dietz, K. & Wolburg, H. Subcellular distribution of glucose transporter (GLUT-1) during development of the blood-brain barrier in rats. *Cell Tissue Res.* **284,** 355–365 (1996).
6. Butt, A. M., Jones, H. C. & Abbott, N. J. Electrical resistance across the blood-brain barrier in anaesthetized rats: a developmental study. *J. Physiol. (Lond.)* **429,** 47–62 (1990).
7. Ek, C. J., Dziegielewska, K. M., Stolp, H. & Saunders, N. R. Functional effectiveness of the blood-brain barrier to small water-soluble molecules in developing and adult opossum (Monodelphis domestica). *J. Comp. Neurol.* **496,** 13–26 (2006).
8. Hirase, T. et al. Occludin as a possible determinant of tight junction permeability in endothelial cells. *J. Cell Sci.* **110,** 1603–1613 (1997).
9. Kniesel, U., Risau, W. & Wolburg, H. Development of blood-brain barrier tight junctions in the rat cortex. *Brain Res. Dev. Brain Res.* **96,** 229–240 (1996).
10. Korn, J., Christ, B. & Kurz, H. Neuroectodermal origin of brain pericytes and vascular smooth muscle cells. *J. Comp. Neurol.* **442,** 78–88 (2002).
11. Dohgu, S. et al. Brain pericytes contribute to the induction and up-regulation of blood-brain barrier functions through transforming growth factor-β production. *Brain Res.* **1038,** 208–215 (2005).
12. Hori, S., Ohtsuki, S., Hosoya, K., Nakashima, E. & Terasaki, T. A pericyte-derived angiopoietin-1 multimeric complex induces occludin gene expression in brain capillary endothelial cells through Tie-2 activation *in vitro. J. Neurochem.* **89,** 503–513 (2004).
13. Lai, C. H. & Kuo, K. H. The critical component to establish *in vitro* BBB model: Pericyte. *Brain Res. Brain Res. Rev.* **50,** 258–265 (2005).
14. Lindahl, P., Johansson, B. R., Leveen, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277,** 242–245 (1997).
15. Tallquist, M. D., French, W. J. & Soriano, P. Additive effects of PDGF receptor β signaling pathways in vascular smooth muscle cell development. *PLoS Biol.* **1,** e52 (2003).
16. Hellstrom, M. et al. Lack of pericytes leads to endothelial hyperplasia and abnormal vascular morphogenesis. *J. Cell Biol.* **153,** 543–554 (2001).
17. Shepro, D. & Morel, N. M. Pericyte physiology. *FASEB J.* **7,** 1031–1038 (1993).
18. Lee, S. W., Kim, W. J., Jun, H. O., Choi, Y. K. & Kim, K. W. Angiopoietin-1 reduces vascular endothelial growth factor-induced brain endothelial permeability via upregulation of ZO-2. *Int. J. Mol. Med.* **23,** 279–284 (2009).
19. Nag, S., Papneja, T., Venugopalan, R. & Stewart, D. J. Increased angiopoietin2 expression is associated with endothelial apoptosis and blood-brain barrier breakdown. *Lab. Invest.* **85,** 1189–1198 (2005).
20. Shue, E. H. et al. Plasmalemmal vesicle associated protein-1 (PV-1) is a marker of blood-brain barrier disruption in rodent models. *BMC Neurosci.* **9,** 29 (2008).
21. Ioannidou, S. et al. An *in vitro* assay reveals a role for the diaphragm protein PV-1 in endothelial fenestra morphogenesis. *Proc. Natl Acad. Sci. USA* **103,** 16770–16775 (2006).
22. Wosik, K. et al. Angiotensin II controls occludin function and is required for blood brain barrier maintenance: relevance to multiple sclerosis. *J. Neurosci.* **27,** 9032–9042 (2007).
23. Gidday, J. M. et al. Leukocyte-derived matrix metalloproteinase-9 mediates blood-brain barrier breakdown and is proinflammatory after transient focal cerebral ischemia. *Am. J. Physiol. Heart Circ. Physiol.* **289,** H558–H568 (2005).
24. He, Z. J., Huang, Z. T., Chen, X. T. & Zou, Z. J. Effects of matrix metalloproteinase 9 inhibition on the blood brain barrier and inflammation in rats following cardiopulmonary resuscitation. *Chin. Med. J. (Engl.)* **122,** 2346–2351 (2009).
25. Bush, T. G. et al. Leukocyte infiltration, neuronal degeneration, and neurite outgrowth after ablation of scar-forming, reactive astrocytes in adult transgenic mice. *Neuron* **23,** 297–308 (1999).
26. Weidenfeller, C., Svendsen, C. N. & Shusta, E. V. Differentiating embryonic neural progenitor cells induce blood-brain barrier properties. *J. Neurochem.* **101,** 555–565 (2007).
27. Daneman, R. et al. Wnt/β-catenin signaling is required for CNS, but not non-CNS, angiogenesis. *Proc. Natl Acad. Sci. USA* **106,** 641–646 (2009).
28. Stenman, J. M. et al. Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* **322,** 1247–1250 (2008).
29. Motiejunaite, R. & Kazlauskas, A. Pericytes and ocular diseases. *Exp. Eye Res.* **86,** 171–177 (2008).
30. Cayrol, R. et al. Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. *Nature Immunol.* **9,** 137–145 (2008).

METHODS

Animals. Sprague–Dawley rats and C57BL/6 mice were obtained from Charles River. Mice harbouring $Pdgfrb^-$ and $Pdgfrb^{F7}$ alleles were obtained from Phillippe Soriano (Fred Hutchinson Cancer Research Center), and lines were maintained by breeding with C57BL/6 mice.

Immunohistochemistry. For time course experiments, whole rat and mouse embryos before E15, and dissected brains of embryos between E15 and birth, were fixed overnight in 4% paraformaldehyde at 4 °C. Postnatal rats and mice were anaesthetized with a ketamine (100 mg kg^{-1})/xylazine (20 mg kg^{-1}) cocktail, and perfused with PBS followed by 4% paraformaldehyde with a dynamax peristaltic pump and then dissected brains were fixed overnight in 4% paraformaldehyde at 4 °C. Staining was performed as described previously[27], with antibodies directed against PDGFR-β (R&D systems, eBiosciences) and Ng2 (Chemicon) for pericytes, PDGFR-α for OPCs (Abcam), Aqp4 for astrocytes (sigma), Glut1 (Chemicon), Pgp (Sigma) and Icam1 (Abcam), and with appropriate Alexa-conjugated secondary antibodies and counter-stained with BSL-fluorescein (Vector Labs). For ZO-1 and Plvap immunofluoresence, embryos were fixed with 10% trichloroacetic acid for 1 h on ice, before cryopreservation and staining with anti-ZO-1 antibody (Invitrogen) or anti-Plvap antibody (AbD Serotec). Staining for occludin and claudin 5 was achieved by flash freezing dissected brains, or whole embryos for ages less than E15, in OCT. Tissue cryosections were obtained and fixed in cold 95% ethanol for 30 min, followed by 1 min in room temperature acetone. Sections were then rehydrated in PBS, blocked in 50% goat serum and stained with antibodies directed against occludin (Invitrogen), and claudin 5 (Invitrogen), followed by appropriate Alexa-fluorophore-conjugated secondary antibodies and counterstained with BSL-fluorescein (Vector Labs). Slides were mounted in vectashield with DAPI (Vector Labs) and visualized by fluorescence microscopy.

For pericyte-deficient mice, $Pdgfrb^{-/-}$ and littermate control embryos were flash-frozen and stained using the ethanol/acetone fixation described above with antibodies directed against occludin (Invitrogen), claudin 5 (Invitrogen) and Plvap (AbD Serotec). For Icam1, Ng2 and Glut1 staining, brains of dissected $Pdgfrb^{-/-}$ and littermate control embryos were fixed overnight in 4% paraformaldehyde at 4 °C and stained as described previously, with antibodies directed against Icam1 (Abcam), Ng2 (Chemicon) or Glut1 (Chemicon). For Gr1 staining, 5-week-old $Pdgfrb^{F7/F7}$ and littermate controls were perfusion fixed, and dissected brains were further fixed overnight in 4% paraformaldehyde at 4 °C and stained as described previously, with antibodies directed against Gr1 (gift from I. Weismann). For comparisons of $Pdgfrb^{-/-}$ and littermate control brains, sections were analysed on a Nikon Eclipse E800 microscope and images were taken with a Diagnostics Instrument SPOT camera with set exposure to appropriately compare staining between genotypes. Percent length of vasculature with intense Icam1 staining was measured using ImageJ ($n \geq 4$ each for mutant and littermate controls for each stain). For pericyte staining comparing $Pdgfrb^{+/+}$, $Pdgfrb^{+/-}$ and $Pdgfrb^{-/-}$ mice, tissue sections were stained with anti-Zic1 (Novus) and anti-desmin antibody (Dako) after antigen retrieval by steaming in a 10 mM citrate buffer for 30 min, or anti-NG2 (Chemicon). For quantification of pericyte number in $Pdgfrb^{+/+}$ and $Pdgfrb^{+/-}$ mice the number of Zic$^+$ nuclei was counted and divided by vascular length quantified in ImageJ ($n \geq 7$ for each genotype). Pericyte coverage was quantified using ImageJ to measure the length of BSL$^+$ vessels associated with desmin$^+$ pericyte processes ($n \geq 5$ for each genotype). Total Gr1$^+$ cells outside BSL$^+$ vessels per medial sagittal brain section were counted ($n = 4$ each for $Pdgfrb^{F7/F7}$ and littermate controls). For desmin staining, neonatal $Pdgfrb^{F7/N}$, $Pdgfrb^{F7/F7}$ and littermate controls were perfusion-fixed and stained with anti-desmin antibody (Dako) after antigen retrieval by steaming in a 10 mM citrate buffer for 30 min. Pericyte coverage was quantified using ImageJ to measure the length of BSL$^+$ vessels associated with desmin$^+$ pericyte processes ($n \geq 3$ for each genotype). All quantification was performed blind to genotype.

BBB permeability assays. Animals were anaesthetized with a ketamine (100 mg kg^{-1})/xylazine (20 mg kg^{-1}) cocktail, and then the thoracic cavity was opened to reveal the heart. The right ventricle was severed and then a DPBS (Gibco) solution containing 1 mg ml^{-1} EZ-link sulfo NHS-biotin (Pierce) was perfused into the left ventricle using a Dynamax peristaltic pump for 3 min, followed by 5 min of perfusion with 4% paraformaldehyde. The flow rate of the pump was adjusted to match the cardiac output of the rats or mice. Tissues, including the brain, liver and muscle, were dissected and further submersion fixed in 4% paraformaldehyde overnight at 4 °C, before being submerged in 30% sucrose. Ten micrometre cryosections were generated after the tissue was frozen in a 2:1 30% sucrose: OCT mixture. For E15 time points, the entire embryo was submersion-fixed and processed. Tissue sections were treated with a 50% goat serum blocking solution, before staining with streptavidin-Alexa 488 (Invitrogen). Staining of streptavidin was analysed with a Nikon Eclipse E800 microscope and images were taken with a Diagnostics Instrument SPOT camera and analysed by SPOT software. For comparisons of $Pdgfrb^{-/-}$ and littermate control brains, pregnant mice were anaesthetized with

ketamine/xylazine, and embryos were extracted and chosen blind to genotype for perfusions. Six embryos were chosen from each litter, one as a control perfused without tracer. The perfusion setup was standardized to minimize fluctuations for pH (7.09 ± 0.005 s.d.) and pO$_2$ (136.1 mm Hg ± 2.0 s.d.) and the perfusion pressure was maintained digitally with the dynamax pump. The heart beat was monitored and all embryos maintained heart beat throughout tracer perfusions. Perfusions were performed with EZ-link sulfo NHS-biotin (0.5 mg ml^{-1}, Pierce), 3 kDa biotinylated fixable dextran (0.15 mg ml^{-1}, Invitrogen) or 70 kDa biotinylated fixable dextran (3.5 mg ml^{-1} Invitrogen), and micrographs were taken with set exposure to appropriately compare the amount of biotin in the brain tissue. The concentrations of the dextrans were matched for molar concentrations ($n \geq 6$ for each sample group for each tracer). Fluorescence in tissue sections was compared by ImageJ software, using tissue sections of embryos perfused with DPBS without tracer as a background control. $Pdgfrb^{+/+}$ and $Pdgfrb^{+/-}$ mice were combined as littermate controls, as we did not observe any difference in leakage between the two genotypes. For Evan's blue extravasation assays, P5 $Pdgfrb^{F7/-}$, $Pdgfrb^{F7/F7}$ and littermate controls were given an intraperitoneal injection of 10 μl 2% Evan's blue. After 1 day, animals were perfused as described above with PBS followed by paraformaldehyde. Dissected brains were photographed. For quantification of biotin leakage in neonatal $Pdgfrb^{+/+}$ and $Pdgfrb^{+/-}$ mice, and in neonatal $Pdgfrb^{F7/-}$, $Pdgfrb^{F7/F7}$ and littermate controls, trans-cardiac perfusion of biotin tracer was performed, followed by staining of tissue sections with streptavidin, and quantification of leakage with ImageJ. In each case, perfusions and analysis was done blind to genotype.

Electron microscopy. Cerebral cortex from E18 $Pdgfrb^{-/-}$ and littermate controls was fixed in 2% glutaraldehyde/4% paraformaldehyde in sodium cacodylate buffer overnight at 4 °C and processed for electron microscopy as described previously[31] ($n = 3$ for both mutant and littermate controls). To visualize tracer permeability by electron microscopy, E18 $Pdgfrb^{-/-}$ and littermate controls were perfused with biotin followed by paraformaldehyde as described above. Frozen tissue sections were fixed with 2% glutaraldehyde/4% paraformaldehyde in sodium cacodylate buffer overnight and stained with a HRP-tagged streptavidin (Invitrogen) followed by DAB reaction before being processed for electron microscopy.

Endothelial cell culture. Endothelial cells from adult C57BL/6 mice brains were purified and cultured as previously described[27] and grown with puromycin for the first 3 days. For mouse and rat brain pericytes, brain tissue suspensions were prepared as described previously[27], and cell suspensions were incubated on two dishes coated with an anti-CD45 antibody to deplete microglia cells, followed by a dish coated with anti-PDGFR-β to select pericytes. The pericytes were recovered by trypsin treatment and plated on PDL-coated cell culture filter inserts in a DMEM-based medium (Invitrogen) containing SATO, insulin, pyruvate, penicillin, streptomycin, glutamine and 10% fetal calf serum (Invitrogen). After 2.5 weeks, pericyte growth media was exchanged for endothelial growth media, and pericyte inserts were added to endothelial cells for 3 days. Endothelial cells were fixed with 4% paraformaldehyde for 10 min, then block/permeabilized for 30 min in 50% goat serum and 0.2% Triton X-100, and stained with primary antibodies directed against occludin (Invitrogen), claudin 5 (Invitrogen), Icam1 (Abcam) or Angpt2 (Abcam) and appropriate Alexa-fluorophore-conjugated secondary antibodies, mounted in vectashield with DAPI and images were taken with a Diagnostics Instrument SPOT camera analysed using a Nikon Eclipse E800 microscope. Per cent vascular length with closed tight junctions was analysed using ImageJ software. Specifically, total length of cell junctions between two cells was measured, and then the length of these junctions where a single band of occludin or claudin 5 (sealed-staining between arrows in Fig. 3H, right panel) or two bands with diffuse staining between (unsealed-staining between arrows in Fig. 3G, right panel) was measured ($n \geq 3$ paired cultures of endothelial cells alone and endothelial cells with pericyte feeder layer for each marker). For TEER measurements, endothelial cells were isolated as described and grown on the filter membrane of Costar HTS transwell plates, and cultured in puromycin for 3 days, before being transferred to a new dish with media containing hydrocortisone, without puromycin. In the new dish, half the filter membranes were placed in wells with pericytes cultured on the bottom well. Resistance measurements were taken between 10–19 days after co-culture with an EVOM voltmeter (World Precision Instruments) and subtracted from resistance of membrane without cellular culture. Unit resistance was calculated by multiplying the resistance by the area of the filter membrane, and averaged for each sample ($n = 6$ paired cultures of endothelial cells alone and endothelial cells with pericyte feeder layer).

GeneChip. Vascular cells from E18 $Pdgfrb^{-/-}$ and littermate controls were purified based on methods described previously[27]. Acutely purified pericytes were purified from mouse brain by anti-PDGFR-β immunopanning as described. Purification of RNA from acutely isolated cells, generation of biotinylated cRNA, subsequent hybridization to Affymetrix Mouse Genome 430 2.0 Arrays

and raw image analysis with Affymetrix GCOS 1.3 software was performed as previously described[32].

Western blots. Cerebral cortex from E18 *Pdgfrb*$^{-/-}$ and littermate controls were homogenized and re-suspended and lysed in RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS), and protein concentration was quantified using BCA protein assay (Pierce). Samples were analysed by SDS–PAGE as previously described[33]. Briefly, samples were resolved on SDS–PAGE and transferred to PVDF membranes. Membranes were blocked with 5% milk solution and then incubated with antibodies directed against occludin (Invitrogen), claudin 5 (Invitrogen), Icam1 (Abcam), Alcam (R & D Systems), Lgals3 (R & D Systems), Col I (Abcam), Col III (Abcam), vitronectin (Abcam), MMP9 (Abcam), PDGFR-β (eBiosciences) or β-actin (Sigma) followed by incubation with an appropriate secondary conjugated to HRP (Jackson 1:10,000) and visualized using a chemi-luminescent ECL substrate for HRP (GE), and either exposed on film and quantified with ImageJ or imaged with a Fuji-Film LAS 4000 and analysed with Multi Gauge V3.0 software ($n = 2$–5 for mutant and $n = 4$–8 for littermate controls).

31. Watkins, T. A., Emery, B., Mulinyawe, S. & Barres, B. A. Distinct stages of myelination regulated by γ-secretase and astrocytes in a rapidly myelinating CNS coculture system. *Neuron* **60,** 555–569 (2008).
32. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28,** 264–278 (2008).
33. Chan, J. R. *et al.* NGF controls axonal receptivity to myelination by Schwann cells or oligodendrocytes. *Neuron* **43,** 183–191 (2004).

LETTER

Pericytes regulate the blood–brain barrier

Annika Armulik[1], Guillem Genové[1], Maarja Mäe[1], Maya H. Nisancioglu[1], Elisabet Wallgard[1]†, Colin Niaudet[1], Liqun He[1]†,
Jenny Norlin[1], Per Lindblom[2], Karin Strittmatter[1]†, Bengt R. Johansson[3] & Christer Betsholtz[1]

The blood–brain barrier (BBB) consists of specific physical barriers, enzymes and transporters, which together maintain the necessary extracellular environment of the central nervous system (CNS)[1]. The main physical barrier is found in the CNS endothelial cell, and depends on continuous complexes of tight junctions combined with reduced vesicular transport[2]. Other possible constituents of the BBB include extracellular matrix, astrocytes and pericytes[3], but the relative contribution of these different components to the BBB remains largely unknown[1,3]. Here we demonstrate a direct role of pericytes at the BBB _in vivo_. Using a set of adult viable pericyte-deficient mouse mutants we show that pericyte deficiency increases the permeability of the BBB to water and a range of low-molecular-mass and high-molecular-mass tracers. The increased permeability occurs by endothelial transcytosis, a process that is rapidly arrested by the drug imatinib. Furthermore, we show that pericytes function at the BBB in at least two ways: by regulating BBB-specific gene expression patterns in endothelial cells, and by inducing polarization of astrocyte end-feet surrounding CNS blood vessels. Our results indicate a novel and critical role for pericytes in the integration of endothelial and astrocyte functions at the neurovascular unit, and in the regulation of the BBB.

Platelet-derived growth factor (PDGF)-B/PDGF receptor-β (PDGFR-β) signalling is necessary for pericyte recruitment during angiogenesis[4,5]. Perinatal lethality precludes analysis of postnatal processes in _Pdgfb_ or _Pdgfrb_ null mice[6,7], but several other mouse mutants of this pathway are viable postnatally. Two such mutants were used here: PDGF-B retention motif knockouts (_Pdgfb_[ret/ret]) where PDGF-B binding to heparan sulphate proteoglycans was disrupted[8]; and mutants in which _Pdgfb_ null alleles were complemented by one or two copies of a conditionally silent human PDGF-B transgene targeted to the _Rosa26_ locus and activated by endothelial Cre recombinase (hemizygous R26P[+/0] or homozygous R26P[+/+] mice; Supplementary Fig. 2a–d).

We quantified pericyte coverage in different regions of the CNS by CD13 or PDGFR-β staining (Fig. 1a, b, e and Supplementary Fig. 3). _Pdgfb_[ret/ret], R26P[+/0] and R26P[+/+] mice displayed pericyte coverage corresponding to 26%, 40% and 72%, respectively, compared to controls (Fig. 1a, b). Quantification of absolute numbers of mural cells (pericytes and vascular smooth muscle cells) using the transgenic reporter XLacZ4[9] confirmed low mural cell densities in _Pdgfb_[ret/ret] and R26P[+/0] mice, and close to normal levels in R26P[+/+] mice (Supplementary Fig. 3b–p). We extended previous observations[4,5,10,11] that reduced pericyte densities correlate with increased vessel diameter and reduced vessel density (Fig. 1c–e and Supplementary Fig. 4). Importantly, these phenotypes were almost completely normalized in R26P[+/+] mice (Fig. 1d, e and Supplementary Fig. 4a–c).

Increased water content in brains of _Pdgfb_[ret/ret] mice (Fig. 1f) indicated impairment of the BBB. We tested the BBB integrity in the different mutants using a panel of tracers (Supplementary Table 1). The azo dye Evans blue[12] accumulated in mutant brain parenchyma in a time-dependent fashion (Fig. 1i) and in correlation with pericyte density: it was largest in _Pdgfb_[ret/ret] mice followed by R26P[+/0] and R26P[+/+]

(Fig. 1g, h, j and Supplementary Fig. 5a–c). Similarly, the fluorescent dye cadaverine Alexa Fluor-555 accumulated significantly in the brain parenchyma of _Pdgfb_[ret/ret] and R26P[+/0] mice (Fig. 1j and Supplementary Fig. 5d, h, i). Additionally, fluorescently labelled albumin, 70 kDa dextran and IgG passed the BBB in _Pdgfb_[ret/ret] and R26P[+/0] mice, but not in controls or in R26P[+/+] mice (Fig. 1j and Supplementary Fig. 5e–g). These experiments establish a close correlation between pericyte density and permeability across the BBB for a range of tracers of different molecular masses (Supplementary Table 1).

Permeability in CNS vessels is impeded by continuous complexes of endothelial junctions[13,14]. We studied such complexes in adult pericyte-deficient mutants using markers for adherens (VE-cadherin) and tight (ZO-1 and claudin 5) junctions. _Pdgfb_[ret/ret], R26P[+/0] and controls showed junctional marker expression at similar levels as judged by immunostaining and western blotting (Supplementary Fig. 6a–c and data not shown). The junctional markers were distributed in a pattern consistent with continuous junction complexes in both mutants and controls; however, mutants displayed focally increased junctional width and undulation. These patterns were confirmed by transmission electron microscopy, which failed to reveal any apparent abnormalities in the ultrastructure of endothelial junctions, with the exception that longer and irregular stretches of endothelial overlap were commonly found in pericyte-deficient mutants (Fig. 2c and Supplementary Fig. 6e).

Because continuity, ultrastructure and marker expression were consistent with retained integrity of endothelial junctions in the absence of pericytes, we took advantage of the fixable nature of the fluorescent tracers to explore the route of extravasation in _Pdgfb_[ret/ret] and R26P[+/0] mice in more detail. Cadaverine Alexa Fluor-555 accumulated in endothelial cells and in the brain parenchyma in _Pdgfb_[ret/ret] and R26P[+/0] mice, but not in controls or in R26P[+/+] mice (Fig. 2a). Extravasated cadaverine Alexa Fluor-555 localized mainly to neurons (Figs 2a and 3d and Supplementary Fig. 7a). Similar patterns of distribution were observed for fluorescent albumin, IgG and 70 kDa dextran (Fig. 2b and Supplementary Fig. 7b). We also studied the distribution of horseradish peroxidase (HRP, 44 kDa) by transmission electron microscopy. We found increased uptake of HRP specifically in macrovesicular structures in the endothelium in _Pdgfb_[ret/ret] mice in comparison with controls (Fig. 2c, d and Supplementary Fig. 6d, e). _Pdgfb_[ret/ret] microvessels also showed marked accumulation of HRP reactivity at the vascular basement membrane, without apparent co-localization with endothelial junctions (Fig. 2c, d and Supplementary Fig. 6d, e). Together, these observations indicate that macromolecular permeability across the BBB in pericyte-deficient vessels occurs through a transcytosis route. Pericyte deficiency was not associated with changes in the polarization or with signs of fenestration in the endothelial cells (Supplementary Figs 4d, 6d, e and 7c), features that characterize the BBB defects observed as a result of impaired Wnt/β-catenin signalling[15,16].

The BBB breaks open in conjunction with stroke, leading to life-threatening CNS oedema. A recent study demonstrated that the tyrosine kinase inhibitor imatinib counteracts oedema in experimental
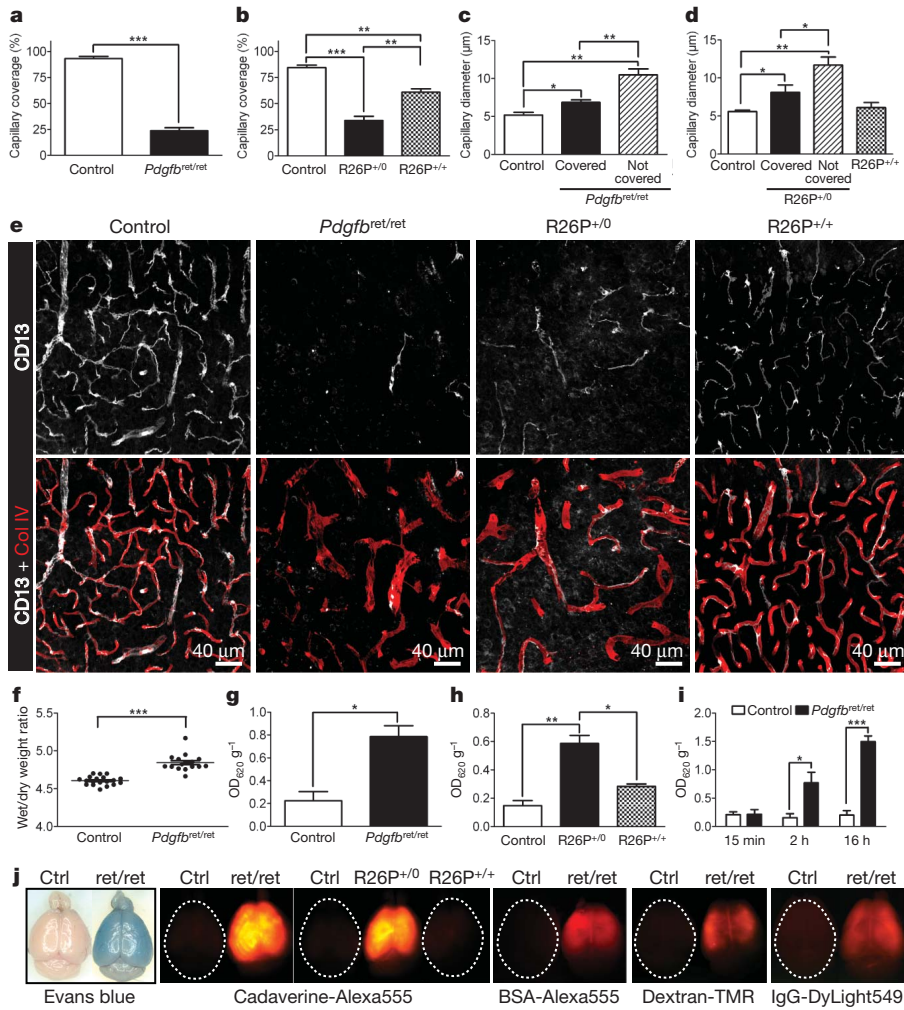
Figure 1 | Pericyte coverage correlates with BBB integrity. a, b, Quantification of pericyte coverage of capillaries in the cerebral neocortex of adult $Pdgfb^{ret/ret}$ mice (**a**), and R26P$^{+/0}$ and R26P$^{+/+}$ mice (**b**). **c, d**, Capillary diameter in the cerebral neocortex of adult $Pdgfb^{ret/ret}$ mice (**c**), and R26P$^{+/0}$ and R26P$^{+/+}$ mice (**d**), with or without pericyte coverage. **e**, Three-dimensional reconstructions of confocal image z-stacks of adult cerebral neocortex vasculature depicted by collagen IV (basement membrane) and CD13 (pericyte) staining in $Pdgfb^{ret/ret}$, R26P$^{+/0}$ and R26P$^{+/+}$ mice. **f**, Wet/dry weight ratios of control and $Pdgfb^{ret/ret}$ mice. **g, h**, Quantification of Evans blue in $Pdgfb^{ret/ret}$ mice (**g**) and R26P$^{+/0}$ and R26P$^{+/+}$ mice (**h**) in the cerebrum after 16 h of circulation. **i**, Time course of Evans blue accumulation in the cerebrum of $Pdgfb^{ret/ret}$ animals. *y*-axis shows optical density (OD) at 620 nm per gram of tissue. **j**, Whole brains photographed after tracer circulation. Circulation time was 16 h for Evans blue, BSA-Alexa Fluor-555, dextran-TMR and IgG-DyLight 549, and 2 h for cadaverine Alexa Fluor-555. *$P < 0.03$; **$P < 0.007$; ***$P < 0.0005$. All error bars show mean ± s.e.m.
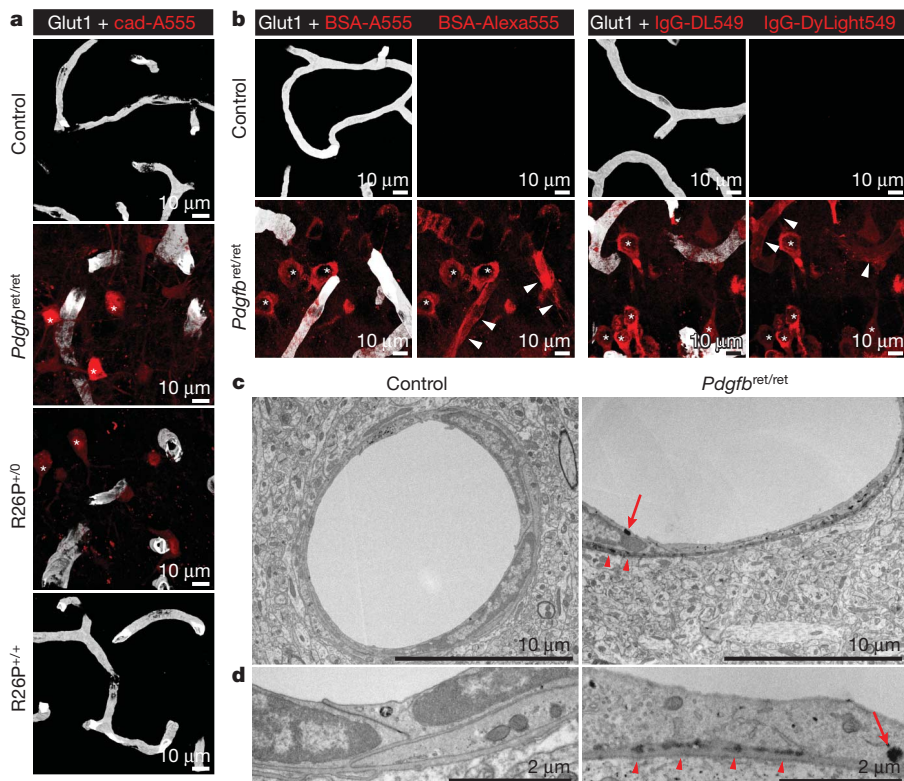
Figure 2 | Pericyte-deficient mice show accumulation of intravenously injected tracers in the brain parenchyma and in the endothelium. a, Accumulation of cadaverine Alexa Fluor-555 (red) in the cerebral neocortex in $Pdgfb^{ret/ret}$ or R26P$^{+/0}$ mice but not in control and R26P$^{+/+}$ mice. **b**, Extravasation of BSA-Alexa Fluor-555 (red) and IgG-DyLight 549 (red) in $Pdgfb^{ret/ret}$ mice but not in controls. Asterisks in **a** and **b** mark neuronal nuclei. Arrowheads in **b** point to tracer in endothelial cells. Endothelial cells in **a** and **b** are visualized by Glut1 immunostaining (white). **c, d**, Electron microscopy images show HRP in large vesicles (arrows) and at the basal lamina (arrowheads) in $Pdgfb^{ret/ret}$ mice.
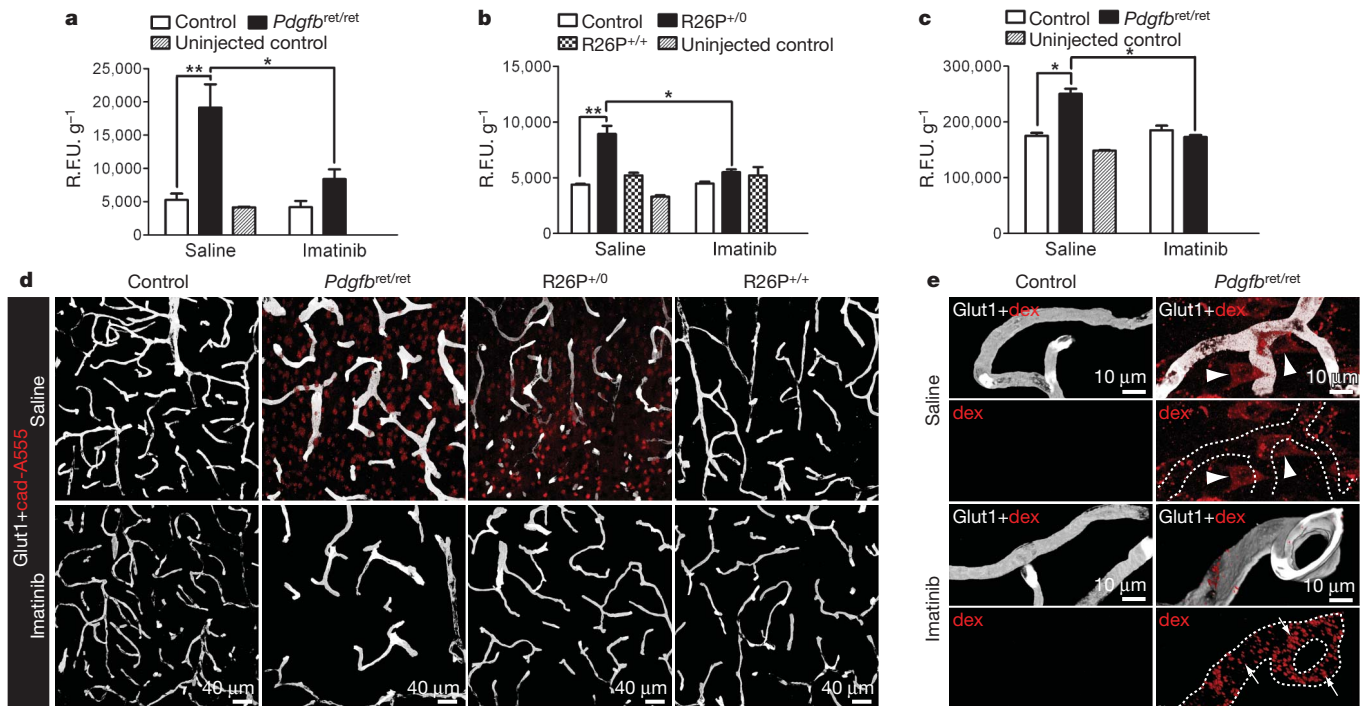
Figure 3 | Imatinib treatment abolishes accumulation of intravenously injected tracers in the brain parenchyma in pericyte-deficient mutants. **a–c**, Quantification of cadaverine Alexa Fluor-555 in $Pdgfb^{ret/ret}$ mice (**a**) and R26P$^{+/0}$ and R26P$^{+/+}$ mice (**b**) in the cerebrum after 2 h of circulation, and of IgG-DyLight 549 (**c**) in $Pdgfb^{ret/ret}$ mice in the cerebrum after 16 h of circulation, with and without previous imatinib treatment. y axis shows relative fluorescence units (R.F.U.) per gram of tissue. **d**, **e**, Three-dimensional reconstructions of confocal image z-stacks of cerebral neocortex of mice injected intravenously with cadaverine Alexa Fluor-555 (**d**) or 70 kDa dextran-TMR (**e**). Whereas imatinib abolishes tracer accumulation in neurons in $Pdgfb^{ret/ret}$ and R26P$^{+/0}$ mice (**d** and arrowheads in **e**), it increases punctuate accumulations in endothelial cells (arrows in **e**). *$P < 0.03$; **$P < 0.005$. All error bars show mean ± s.e.m.

stroke induced in mice[17]. We found that imatinib treatment quickly and dose-dependently reversed brain parenchymal accumulation of all tested tracers in pericyte-deficient mice (Fig. 3 and Supplementary Fig. 8). In contrast, accumulation and punctuate distribution of fluorescent dextran and IgG in the endothelial cells was enhanced (Fig. 3e and Supplementary Fig. 8b, d, e). This indicates that imatinib inhibits the release of tracer from endothelial cells into the brain parenchyma, consistent with an effect on a late step in the endothelial transcytosis process. Notably, imatinib treatment did not normalize pericyte abundance, vessel diameter, density, or the convoluted appearance of endothelial junctions (Fig. 3d and Supplementary Fig. 9).

To address whether pericytes regulate brain-specific endothelial differentiation we used microarrays to assess the expression of known BBB markers[3] in microvascular fragments isolated from adult pericyte-deficient mutants and controls. Such fragments mostly consist of endothelial cells[18], but also include pericytes (when present) and astrocyte end-feet. We used previously obtained array data on embryonic brain microvascular fragments, isolated kidney glomeruli and whole brain for comparison[18–20]. These analyses show that most BBB markers are unaffected by pericyte deficiency at the mRNA level (Fig. 4a, b and Supplementary Table 2). Two known BBB markers for which antibodies were available were assessed at the level of protein expression: Glut1 (predicted to be unaffected by pericyte deficiency) and transferrin receptor (CD71) (predicted to be downregulated in the pericyte-deficient state). In agreement with the gene expression data, we found that Glut1 protein was uniformly expressed in brain endothelial cells irrespective of the degree of pericyte coverage (Figs 2–4 and Supplementary Fig. 10), whereas CD71 expression was downregulated in the endothelium lacking pericyte contact (Supplementary Fig. 10). Together, these results indicate that pericytes exert a discrete influence on the endothelial BBB-specific gene and protein expression profile.

Astrocytes attach to blood vessels through cap-like cytoplastic processes—end-feet—carrying specific channels and transporters, for example, aquaporin 4 (Aqp4) and Kir4.1, which are targeted to astrocyte end-feet to control water and ion homeostasis at the vessel–neuron interface[3]. Little is known about how contacts between blood vessels and astrocyte end-feet are established and maintained[21]. Our gene array data indicated downregulation of several astrocyte markers in pericyte-deficient brain microvascular fragments (Supplementary Table 2 and data not shown). We therefore asked if pericytes, which are partially sandwiched between the endothelial cells and the astrocyte end-feet, might affect astrocytes, and, specifically, if pericyte deficiency leads to changes in astrocyte end-foot distribution or polarization. Three markers for the polarized astrocyte end-foot, Aqp4, α-syntrophin and laminin α2 chain (Lama2), all provided uniform labelling of microvessel abluminal surfaces in control brains, consistent with a polarized expression of these markers (Fig. 4c–e and Supplementary Fig. 11). However, in pericyte-deficient mutants, the vascular staining for these markers was weaker than in controls, and the Aqp4 staining was also re-distributed to other regions of the astrocytes, indicating abnormal polarization of the astrocyte end-feet (Supplementary Fig. 11). Interestingly, we found that pericytes that were detached from the endothelial cells showed intense lining with Aqp4 and α-syntrophin immunostaining (Fig. 4d and data not shown). This indicates that pericytes express cues that mediate attachment of astrocyte end-feet. The strong reduction in deposition of the astrocyte-derived basement membrane component Lama2 (Fig. 4e) prompted us to study the expression of a range of other vascular basement membrane components. These analyses failed to show any significant changes in the deposition of endothelium-derived basement membrane proteins laminin α4, laminin α5, nidogen and perlecan in pericyte-deficient vessels (Supplementary Fig. 12).

BBB properties are not intrinsic to CNS endothelial cells, but are induced when the cells are in contact with astrocytes *in vitro* and *in vivo*[22–24]. Recent evidence also suggests a role for neuronal precursors and Wnt signalling in BBB induction *in vivo*[15,16,25]. A role for pericytes
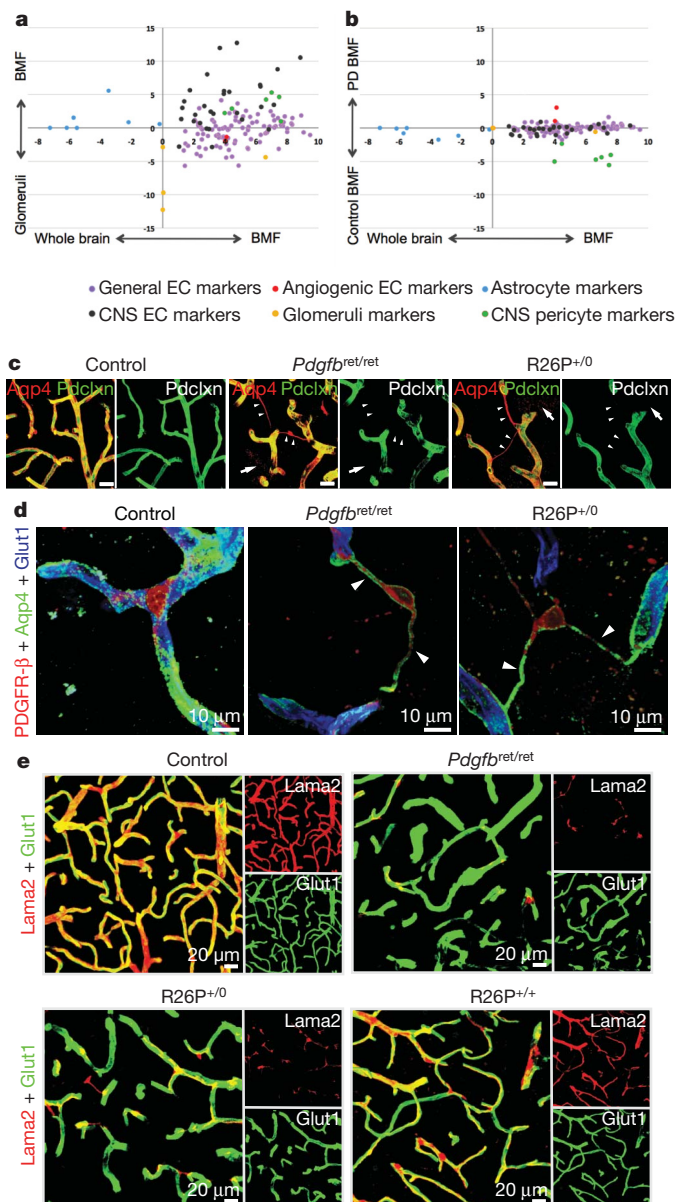
Figure 4 | **Transcript profiling of brain microvasculature and characterization of the polarization defect of astrocyte end-feet in pericyte-deficient mutants. a**, Relative expression (log₂ fold difference) of known markers for the BBB, endothelial cells, astrocytes, pericytes and glomerular podocytes in adult brain microvascular fragments (BMF) versus whole brain (*x* axis) and brain microvascular fragments versus glomeruli (*y* axis). Note the higher expression of BBB markers in brain microvascular fragments compared to glomeruli. **b**, Relative expression of the same genes as in **a** for control and pericyte-deficient (PD) mutants (*y* axis; see Supplementary Tables 2 and 3 for gene identifiers and further information on methodology). **c, d**, Astrocyte end-feet (red) localize to microvessels (green) in controls but also to structures bridging adjacent vessels in *Pdgfb*[ret/ret] and R26P[+/0] mice (arrowheads in **c**) that were identified as detached PDGFR-β positive pericytes (**d**, arrowheads). Arrows in **c** point to diffuse Aqp4 staining (red) in pericyte-deficient mice not associated with vessels (see also Supplementary Fig. 11a). Pdclxn, podocalyxin. **e**, *Pdgfb*[ret/ret] and R26P[+/0] have reduced deposition of the astrocyte-derived basement membrane component laminin α2-chain compared to control and R26P[+/+] animals. Scale bars in **c**, 20 μm.

in endothelial barrier formation has been suggested based on *in vitro* co-culture experiments[26,27], but evidence for any specific role of pericytes in BBB formation has so far been lacking. Here we provide evidence for an *in vivo* role of pericytes in regulating the mammalian BBB (schematically illustrated in Supplementary Fig. 1). The increased

endothelial transcytosis that occurs as a result of brain pericyte deficiency in the mouse mutants is compatible with a normal lifespan of the mouse mutants, in spite of the fact that this increase results in oedema and extravasation of plasma proteins in the brain. This is probably tolerated because leakage is slow, allowing the brain to adapt and compensate for any resulting changes in the interstitial environment. Viable pericyte-deficient animal models should therefore be useful in further studies of the putative functional consequences of BBB impairments for higher and integrated neuronal functions, learning and behaviour, and for the progression of brain diseases. Notably, we have demonstrated that the pericyte-deficient state makes the BBB permeable to immunoglobulins. Further elucidation of the molecular mechanisms that trigger opening and closure of this route across the BBB should therefore be of pharmaceutical significance.

METHODS SUMMARY

Animals. The following transgenic mouse lines were used for experiments: PDGF-B knockout[6], PDGF-B retention motif knockout[8], Tie2Cre (ref. 28), Rosa26 hPDGF-B (this study), XLacZ4 (ref. 9) and Z/EG (ref. 29). Animal experiments were approved by the Stockholm's North Ethical Committee for Animal Research.
Generation of Rosa26 hPDGF-B mice. A full-length human PDGF-B cDNA sequence preceded by a splice acceptor site and a *loxP*-site-flanked transcription stop cassette was cloned into a vector designed for homologous recombination into the ubiquitously expressed *ROSA26* locus[30] (Supplementary Fig. 2a). Gene-targeted mice were generated from the ES cells by standard methods. A detailed description is provided in Methods.
Intravenous injection of tracers and detection of injected tracers. The following tracers were used: Evans blue (Sigma Aldrich), lysine-fixable cadaverine conjugated to Alexa Fluor-555 (Invitrogen), bovine serum albumin conjugated to Alexa Fluor-555 (Invitrogen), lysine-fixable 70 kDa dextran conjugated to tetramethylrhodamine (TMR) (Invitrogen), and goat anti-human IgG conjugated to DyLight 549 (Jackson Immunoresearch Laboratories). Evans blue in the brain was quantified by spectrophotometry as described[17]. Detailed information about the quantification of cadaverine-Alexa Fluor-555 and DyLight-549 conjugated goat anti-human IgG in the brain is given in Methods. When indicated, animals were treated with three or four doses of imatinib (150 mg kg⁻¹) twice per day before tracer injections.
Statistical analysis. Statistical significance was determined by using unpaired two-tailed or one-tailed Student's *t*-test (GraphPad Prism5). Differences were considered significant with a *P* value less than 0.05. Quantified data are presented as mean ± s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Abbott, N. J., Rönnbäck, L. & Hansson, E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nature Rev. Neurosci.* **7,** 41–53 (2006).
2. Reese, T. S. & Karnovsky, M. J. Fine structural localization of a blood-brain barrier to exogenous peroxidase. *J. Cell Biol.* **34,** 207–217 (1967).
3. Bernacki, J. *et al.* Physiology and pharmacological role of the blood-brain barrier. *Pharmacol. Rep.* **60,** 600–622 (2008).
4. Enge, M. *et al.* Endothelium-specific platelet-derived growth factor-B ablation mimics diabetic retinopathy. *EMBO J.* **21,** 4307–4316 (2002).
5. Bjarnegård, M. *et al.* Endothelium-specific ablation of PDGFB leads to pericyte loss and glomerular, cardiac and placental abnormalities. *Development* **131,** 1847–1857 (2004).
6. Levéen, P. *et al.* Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities. *Genes Dev.* **8,** 1875–1887 (1994).
7. Soriano, P. Abnormal kidney development and hematological disorders in PDGF β-receptor mutant mice. *Genes Dev.* **8,** 1888–1896 (1994).
8. Lindblom, P. *et al.* Endothelial PDGF-B retention is required for proper investment of pericytes in the microvessel wall. *Genes Dev.* **17,** 1835–1840 (2003).
9. Tidhar, A. *et al.* A novel transgenic marker for migrating limb muscle precursors and for vascular smooth muscle cells. *Dev. Dyn.* **220,** 60–73 (2001).
10. Lindahl, P. *et al.* Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277,** 242–245 (1997).
11. Hellström, M. *et al.* Lack of pericytes leads to endothelial hyperplasia and abnormal vascular morphogenesis. *J. Cell Biol.* **153,** 543–554 (2001).
12. Moos, T. & Møllgård, K. Cerebrovascular permeability to azo dyes and plasma proteins in rodents of different ages. *Neuropathol. Appl. Neurobiol.* **19,** 120–127 (1993).
13. Dejana, E., Tournier-Lasserve, E. & Weinstein, B. M. The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications. *Dev. Cell* **16,** 209–221 (2009).

14. Taddei, A. *et al.* Endothelial adherens junctions control tight junctions by VE-cadherin-mediated upregulation of claudin-5. *Nature Cell Biol.* **10,** 923–934 (2008).
15. Stenman, J. M. *et al.* Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* **322,** 1247–1250 (2008).
16. Liebner, S. *et al.* Wnt/beta-catenin signaling controls development of the blood-brain barrier. *J. Cell Biol.* **183,** 409–417 (2008).
17. Su, E. J. *et al.* Activation of PDGF-CC by tissue plasminogen activator impairs blood-brain barrier integrity during ischemic stroke. *Nature Med.* **14,** 731–737 (2008).
18. Bondjers, C. *et al.* Microarray analysis of blood microvessels from PDGF-B and PDGF-Rβ mutant mice identifies novel markers for brain pericytes. *FASEB J.* **20,** 1703–1705 (2006).
19. Wallgard, E. *et al.* Identification of a core set of 58 gene transcripts with broad and specific expression in the microvasculature. *Arterioscler. Thromb. Vasc. Biol.* **28,** 1469–1476 (2008).
20. He, L. *et al.* The glomerular transcriptome and a predicted protein-protein interaction network. *J. Am. Soc. Nephrol.* **19,** 260–268 (2008).
21. Wolburg, H. *et al.* Agrin, aquaporin-4, and astrocyte polarity as an important feature of the blood-brain barrier. *Neuroscientist* **15,** 180–193 (2009).
22. Janzer, R. C. & Raff, M. C. Astrocytes induce blood-brain barrier properties in endothelial cells. *Nature* **325,** 253–257 (1987).
23. Hayashi, Y. *et al.* Induction of various blood-brain barrier properties in non-neural endothelial cells by close apposition to co-cultured astrocytes. *Glia* **19,** 13–26 (1997).
24. Sobue, K. *et al.* Induction of blood-brain barrier properties in immortalized bovine brain endothelial cells by astrocytic factors. *Neurosci. Res.* **35,** 155–164 (1999).
25. Daneman, R. *et al.* Wnt/β-catenin signaling is required for CNS, but not non-CNS, angiogenesis. *Proc. Natl Acad. Sci. USA* **106,** 641–646 (2009).
26. Hori, S. *et al.* A pericyte-derived angiopoietin-1 multimeric complex induces occludin gene expression in brain capillary endothelial cells through Tie-2 activation *in vitro*. *J. Neurochem.* **89,** 503–513 (2004).
27. Dohgu, S. *et al.* Brain pericytes contribute to the induction and up-regulation of blood-brain barrier functions through transforming growth factor-β production. *Brain Res.* **1038,** 208–215 (2005).
28. Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis *in vivo*. *Dev. Biol.* **230,** 230–242 (2001).
29. Novak, A. *et al.* Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis* **28,** 147–155 (2000).
30. Zambrowicz, B. P. *et al.* Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl Acad. Sci. USA* **94,** 3789–3794 (1997).

Author Contributions A.A. and C.B. conceived and designed the project. A.A., G.G., M.M., M.H.N., E.W., C.N., L.H., J.N., P.L., K.S. and B.R.J. performed experiments; C.B. and A.A. wrote the manuscript with significant input from M.M., G.G. and M.H.N. G.G. and M.M. contributed equally to the study.

Author Information Our microarray data have been deposited in NCBIs Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO series accession number GSE15892. 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 this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B. (christer.betsholtz@ki.se) or A.A. (annika.armulik@ki.se).

METHODS

Animals. Progeny from the following transgenic mouse lines were used for experiments: PDGF-B knockout[6], PDGF-B retention motif knockout[8], Tie2Cre (ref. 28), Rosa26 hPDGF-B (this study), XLacZ4 (ref. 9), Z/EG (ref. 29). PDGF-B retention motif knockouts were analysed on a C57BL6/J genetic background and R26P$^{+/0}$ and R26P$^{+/+}$ animals on a mixed C57BL6/J and 129sv genetic background. Animal experiments were approved by the Stockholm's North Ethical Committee for Animal Research.

Generation of Rosa26 hPDGF-B mice. A DNA cassette consisting of a full-length human PDGF-B cDNA sequence preceded by a splice acceptor site and a *loxP*-site-flanked transcription stop cassette was cloned into a vector designed for homologous recombination into the ubiquitously expressed *ROSA26* locus[30] (Supplementary Fig. 2a). The targeting vector was verified by sequencing and electroporated into endothelial cells. Positive clones were identified by PCR and confirmed by Southern blot hybridization as described[30]. Gene-targeted mice were generated from the ES cells by standard methods. Animal genotyping was done by PCR. The following PCR primers were used: R26F2, 5′-AAAGTCGCT CTGAGTTGTTAT-3′; R26SR, 5′-CCCACTGGAAAGACCGCGAAGAGT-3′; R523, 5′-GGAGCGGGAGAAATGGATATG-3′ generated a 250-bp PCR fragment from the mutant allele and a 500-bp fragment from the wild-type allele. Expression of hPDGF-B was activated in endothelial cells by crossing with Tie2Cre transgenes (Supplementary Fig. 2b).

Quantification of vessel diameter, density and pericyte coverage. Quantification of vessel pericyte coverage was performed on 50 μm vibratome brain sections. Four corresponding areas in three animals of each genotype were analysed at 3–5 months of age. Blood vessels were visualized by collagen IV and pericytes by CD13 or PDGFR-β immunostainings. 14–16-μm-thick z-stacks were captured using an LSM Meta 510 microscope (Carl Zeiss AG). Quantification of pericyte numbers, vessel length and diameter was performed using Volocity software (Improvision, a PerkinElmer Company).

Antibodies and immunohistochemistry. The following primary antibodies were used: rabbit anti-rat aquaporin 4 (AB3068, Chemicon); rat anti-mouse CD13 (558744, BD PharMingen); rabbit anti-mouse collagen IV (2150-1470, AbD Serotec); rabbit anti-Glut1 (07-1401, Millipore); rat anti-mouse PDGFR-β (14-1402, eBioscience); goat anti-mouse podocalyxin (AF1556, R&D Systems). Secondary antibodies (goat anti-rat, goat anti-rabbit, donkey anti-goat) conjugated with Alexa Fluor dyes were obtained from Invitrogen. For immunostainings in which two primary rabbit antibodies were used, one was directly labelled using the Zenon rabbit IgG labelling kit (Invitrogen). Immunohistochemistry was performed on 50 or 100 μm brain coronal vibratome sections. Specimens were analysed using an LSM 510 META (Carl Zeiss AG) microscope. Image processing was done using Volocity 64 (Improvision), Adobe Photoshop CS3, and Adobe Illustrator CS3 (Adobe Systems). All immunohistochemistry images presented are three-dimensional reconstructions of z-stacks.

Intravenous injection of tracers and detection of injected tracers. Tracers were injected intravenously into the tail vein in adult (2–8 months) mice. The following tracers were used: Evans blue (Sigma Aldrich), lysine-fixable cadaverine conjugated to Alexa Fluor-555 (Invitrogen), bovine serum albumin conjugated to Alexa Fluor-555 (Invitrogen), lysine-fixable 70 kDa dextran conjugated to tetramethyl-rhodamine (Invitrogen), and goat anti-human IgG conjugated to DyLight 549 (Jackson Immunoresearch Laboratories). More detailed descriptions of the tracer experiments are given in the Supplementary Methods. Images of dissected brains were captured using a stereomicroscope (SteREO Lumar.V12, equipped with HBO 100 lamp and a Filter Set Lumar 43 HE; Carl Zeiss AG) equipped with an

AxioCam HRc (Carl Zeiss AG) camera. Evans blue in the brain was quantified by spectrophotometry as described[17]. For quantification of cadaverine-Alexa Fluor-555 and DyLight-549 conjugated goat anti-human IgG in the brain, anaesthetized animals were perfused for 5 min with Hanks' balanced salt solution (HBSS), brains and kidneys removed and homogenized in 1% Triton X-100 in PBS, pH 7.2. Brain and kidney lysates were centrifuged at 16,000 r.p.m. for 20 min and the relative fluorescence of the supernatant was measured on a fluorometer POLARstar Omega (BMG Labtech) (ex/em 544/590 nm). For *in situ* detection of fluorophore-conjugated tracers, anaesthetized animals were perfused for 1–2 min with HBSS, followed by 5 min perfusion with 4% PFA in PBS, pH 7.2. Brains were removed and the tissue was post-fixed in 4% PFA in PBS, pH 7.2 at 4 °C for 5–6 h. 50-μm-thick coronal brain sections were immunostained with anti-Glut1 antibody. Samples were analysed by confocal microscopy (LSM 510 META, Carl Zeiss AG). When indicated, animals were treated with three or four doses of imatinib (150 mg kg^{-1}) twice per day before tracer injections.

Transmission electron microscopy. Two-month-old mice were injected intravenously with horseradish peroxidase Type II (HRP) (100 mg ml^{-1} in PBS, 10 mg per 20 g) (Sigma Aldrich). After 2 h, animals were anaesthetized and perfused with 2.5% glutaraldehyde in 0.1 M cacodylate buffer, pH 7.2. HRP-diaminobenzidine cytochemistry was performed on brain vibratome slices according to ref. 31 with minor modifications. After osmication and en bloc uranyl acetate treatment, tissue blocks were dehydrated and embedded in epoxy resin. Sections were cut with a diamond knife on a Leica EM UC7 ultramicrotome (Leica Microsystems) at 60 nm thickness setting and were examined without additional contrasting in a LEO 912AB energy-filtered transmission electron microscope (Carl Zeiss SMT AG). Digital images were captured with a MegaView III camera (Olympus SiS).

Brain water content determination. Animals were anaesthetized and perfused with HBSS for 5 min via the left ventricle of the heart. Brains were removed and weighed for wet weight. Tissue was dried at 62 °C for 6 days and weighed again for dry weight.

Purification of microvasculature and microarray analysis. Purification of microvasculature and RNA isolation was performed as described[20]. Four RNA samples of each genotype were individually hybridized to Affymetrix GeneChip Mouse Genome 430 2.0 microarrays. Publicly available Affymetrix raw data files concerning wild-type adult whole brain[32], isolated wild-type adult glomeruli[20], and wild-type embryonic and adult brain microvascular fragments[19] were downloaded and treated in the same manner as data generated in this study. Array data was processed using the Affy and gcrma packages in the Bioconductor project (http://www.bioconductor.org). A Student's *t*-test was used to evaluate differential expression. The false discovery rate method was used to perform multiple test correction for *P* values[33]. The log$_2$ change was calculated as the log$_2$ of the average expression difference between sample groups. A detailed description of the samples used in different comparisons is given in Supplementary Table 3.

Statistical analysis. Statistical significance was determined by using unpaired two-tailed or one-tailed Student's *t*-test (GraphPad Prism5). Differences were considered significant with a *P* value less than 0.05. Quantified data are presented as mean ± s.e.m.

31. Karnovsky, M. J. The ultrastructural basis of capillary permeability studied with peroxidase as a tracer. *J. Cell Biol.* **35,** 213–236 (1967).
32. Bender, A. *et al.* Creatine improves health and survival of mice. *Neurobiol. Aging* **29,** 1404–1411 (2008).
33. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

nature**OUTLOOK**

SCIENCE MASTERCLASS

14 October 2010 / Vol 467 / Issue No. 7317

COVER ART: NIK SPENCER

There is a certain cachet to Nobel prizes that no other awards in science share. They are the Oscars of the science world, and the recipients are treated like stars after having made a contribution to science that is both profound and lasting. And like stars of the silver screen, Nobel laureates are admired. They frequently address large audiences eager to hear first-hand about their scientific ideas and to be inspired — to understand what makes them tick and what obstacles they overcame in making their breakthrough discovery.

The annual Lindau Nobel Laureate Meetings provide a forum for this experience. In the beautiful surroundings of Lake Constance, which straddles Germany, Austria and Switzerland, Nobel laureates mingle with young scientists, providing advice and encouragement.

These meetings are more than simply hero worship. Their aim is to promote excellence in science: to spark creative thoughts while at the same time emphasizing the need for diligence. The young scientists — who themselves have been specially selected — are in many ways the real focus of the gathering. *Nature*, as a media partner, aims to share these encounters with a wider scientific audience.

As part of our involvement with the Lindau meeting in 2010, we created lindau.nature.com where young scientists — or indeed anyone — had an opportunity to propose questions about anything connected with science or life as a scientist, to be then put to participating laureates.

Nature Outlook: Science masterclass includes selected Q&As with Nobel laureates, which grew out of this website. It also presents a series of features that explore the scientific theme of interdisciplinary work, reflected in this year's meeting. It closes with a report of a new event at Lindau arranged by *Nature Outlook* called Turning the tables, where the spotlight shone on the young scientists — with the laureates asking them questions. There is also another first, a *Nature Outlook* video of this event, which can be found at nature.com/video/lindau2010/.

We are pleased to acknowledge the financial support of BMBF, VP Bank Stiftung and Mars Incorporated in producing this Outlook. As always, *Nature* retains sole editorial responsibility for all editorial content.

Michelle Grayson
Associate Editor, Nature Outlook

INTRODUCTION

Curiosity aroused

What makes a Nobel laureate tick? What advice can they offer? And who's asking?

BY MICHELLE GRAYSON

"It has been said that democracy is the worst form of government except all the others that have been tried."
Winston Churchill (1874–1965)

It may seem odd to open an article ostensibly about science with a quotation from a politician, especially a wartime leader. Yet Churchill was the most quoted — or at least most paraphrased — person in this series of Q&As. Many Nobel laureates compared science — especially the peer-review system — with democracy. The system may have problems; some people may try to cheat it; but no one has come up with a better way to do science.

The questions put to the laureates came from another small exercise in democracy — the lindau.nature.com website. Established as a portal connecting the social-media coverage of the 60th Lindau Nobel Laureates Meeting in 2010, the website carries links to related blogs, the Facebook group, the YouTube channel, the Twitter feed (Tweets carrying the #lnlm10 hashtag), Flickr and a new page: Nobel Questions — Lindau Answers.

On this page, anyone could register and submit a question. Other visitors to the site were then able to vote for their favourite questions, helping us to identify the most interesting. By the time submissions closed, 14,304 votes had been cast overall for 205 submitted questions, and the most popular question — "What will be the fate of HIV in the next decade?" — had received 1,511 votes.

From the most popular questions, we at *Nature* chose around 10 to put to the laureates. Questions ranged from scientific queries about current research and theories, to more general considerations about life, politics, funding, inspiration and epitaphs.

The answers we received from the laureates were equally diverse. Some were long and considered, some were short and to-the-point. Some were equivocal, some were objective, and some clearly pushed an agenda. There was a mix of optimism and pessimism. And, apart from a fondness for quoting Churchill, the only other common feature was a love of science — of asking questions and seeking the truth.

WHO QUESTIONS THE QUESTIONERS?
When we contacted the people who submitted the top questions we discovered a diverse bunch, not all of them attendees of the 2010 Lindau meeting. Each had their own reasons for posing their question, which tended to fall into two broad categories: either they sought advice relevant to their own research, or they hoped to be inspired by the laureates.

Many questions reflected interests outside the laboratory. Markita Landry, a full-time doctoral student in chemical physics at the University of Illinois, asked a question that received nearly 700 votes: "Do you believe scientists are under-represented in politics across the globe? If so, have established scientists a duty to become more active in politics and science policy?" In the United States, she explained, "fewer than 1.5 percent of members of Congress are scientists. By maintaining a strong divide between science and politics, it seems inevitable that scientific issues will end up being poorly addressed in political circles." And, on the whole, the laureates were in agreement, although with widely different reasons and solutions.

"It has always been my dream to work at a place like Bell Labs in their heyday (1960–80s), where budgets and freedom were large and constraints were small."

Funding was another popular subject. The amount of financial support that scientists have access to varies widely across the world. Christoph Göbl, a PhD student of biomolecular NMR spectroscopy in Austria, was interested to know if the laureates had experience or knowledge of any particular country where the science funding was ideal — or at least very well handled. Having come into scientific academia late, following spells as an apprentice cook, waiter and then chef, Göbl was motivated by his own experience at the mercy of the Austrian funding system. "In Austria, it is almost impossible to get a position," he explained. "It is even hard to finance your science through grants: you are only allowed to receive grants for six years, after that you need a position at a university." Some of the laureates cited specific countries as good examples, including Singapore, Germany, the United States, the United Kingdom, Switzerland and the Netherlands. However, many doubted that such a country existed. Luc Montagnier, co-recipient of the 2008 Nobel Prize in Physiology or Medicine for the discovery of HIV, went so far as to declare: "There are none. Management and funding of research have become highly bureaucratic in the US and Western Europe."

What about the contribution of corporate research? This was the subject of a question from Noy Bassik, a PhD student in chemical and biomolecular engineering at Johns Hopkins University, and 408 other people also thought the question interesting. "Bell Labs and other corporate research sites, which led to many Nobel prizes, are on the decline or have shut down. Is corporate basic research critical, is research in academia sufficient — or has private research just shifted to biology?" Bassik has a strong family history in science and engineering, and was "eating and drinking science everywhere" as a child. He also has academic and industrial experience, and sees himself in the middle of the spectrum that typically runs from academia undertaking "pure fundamental questions with no specific application in mind" to industrial R&D "solving technical problems for a specific project/medication/device". He explains: "I am from New Jersey and have always admired Bell Labs and the research spirit that they represented. In fact, it has always been my dream to work at a place like Bell Labs in their heyday (1960-80s), where budgets and freedom were large and constraints were small." Sadly for Bassik, the consensus was that "times have changed", as Arno Penzias, 1978 co-recipient of the physics prize, stated. With one exception, Penzias added: "Drug companies are a special case. Most have maintained their research laboratories, but face a unique problem: their labs are staffed by brilliant biochemists, ill equipped to provide advantage in a genomics-dominated era."

The thoughts of Winston Churchill, a literature laureate, on peer review are unknown.

IMAGES.COM/CORBIS

IMAGES.COM/CORBIS

PREDICTIONS

The single most popular question, on the fate of HIV, was posed by Prasanna Kumar Santhekadur, a post-doctoral fellow working on tumour angiogenesis, typifying the cross-disciplinary curiosity that drew visitors to the site and Lindau attendees in general. Two of the Nobel laureates for physiology or medicine 2008 — Montagnier and Harald zur Hausen — answered this question, and were pessimistic about development of a long-term solution; anticipating progress in therapeutic development, but also no end to the spread of the virus, "partly because of the insufficiencies of prevention policies, partly because the research is wrongly orientated", Montagnier explained. He then questioned the logic of ploughing money into "an elusive prophylactic vaccine or to the extension of life-long treatments by expensive drugs". zur Hausen, who linked human papilloma viruses with cervical cancer, expects more progress in the therapy of HIV infections, including chemo- or gene therapy, considers 10 years too short a timespan.

Given that the Lindau meeting this year focused on interdisciplinary research, it is no surprise that several questioners asked about collaboration. Adam Goodwill, who is working towards a PhD in cellular and integrative physiology at West Virginia University, acknowledged that interdisciplinary research

is important but found that "each discipline is composed of a series of very subtly nuanced phrases", which can have very specific and profound meanings. "How do we accomplish these collaborations when we don't speak that common language?" he asked.

Sebastian Krackl, a chemistry PhD student at the Berlin Technical University who has experienced interdisciplinary exchanges, echoed that view: "I was shocked to see how different the scientific languages of different disciplines are, how hard it is to cross cultural and scientific disparities in order to profit from the others' experience and knowledge." All the laureates agreed that it was important to encourage more collaborative working. As Richard Ernst, recipient of the 1991 Nobel Prize in Chemistry, pithily put it: "There is no fruitful science without interdisciplinarity. When you want to be a narrow-minded nerd, leave science and seek employment in a post office!"

Practical solutions ranged from simply organizing monthly cross-department get-togethers — as instituted at Columbia University, said Martin Chalfie, co-recipient of the 2008 Nobel Prize in Chemistry — to advocating a fundamental change in the way that academic departments are

> *"There is no fruitful science without interdisciplinarity."*

organized, proposed by Harold Kroto, co-recipient of the 1996 Nobel Prize in Chemistry.

A personal question rounded off the questionnaire. Ian Harvey Arellano, a chemistry PhD student at the University of the Philippines, asked: "Aside from being a Nobel laureate, how would you want the world to remember you?" His rationale was to challenge the "nerdy-geeky" image of top flight scientists to show their more colourful characteristics. "I believe that my question will reveal the other side of the laureates, which may or may not be distinct from their science, and I am sure that this will be a fountain of inspiration for all who read their statements." However, he did not anticipate the modesty of many laureates, exemplified by the short response from zur Hausen: "I do not have a special desire for remembrance."

Modesty aside, whether you are seeking advice, inspiration, or are just curious, we hope you enjoy this selection of questions and answers associated with the 2010 Lindau Nobel Laureate Meetings.

Michelle Grayson *is Editor of Nature Outlook: Science masterclass*

Arno Allan Penzias
A runaway success

Arno Penzias, Robert W. Wilson and Pyotr L. Kapitsa won the Nobel Prize in Physics 1978. Penzias and Wilson's share was for discovering the existence of cosmic background radiation.

Are scientists are under-represented in politics? And do established scientists, especially Nobel laureates, have a duty to become active in politics and science policy?

Scientists rarely possess the skill sets needed for success in politics. All the more reason, therefore, for prominent scientists to become active in using that visibility for public good. Speaking personally: I feel that obligation, and have tried to act accordingly. I flew from Stockholm to Russia immediately after receiving my Nobel prize. I had arranged with a human rights organization to deliver my Nobel lecture to a gathering of 'refusnik' scientists in a Moscow apartment. Over time, each of them was granted the emigration permit they had applied for.

How can the public be convinced of the importance of fundamental research with no applications in sight?

We need to make the public more aware of how successful fundamental research actually works. Unfortunately, more than a few scientists fit the popular image of isolated specialists, working within the narrow confines of their area of interest. Rather than promoting fundamental research as an abstract concept, I think we do better when we focus upon support for research universities — and the problem-rich environments they create and nourish — as our civilization's most fruitful keys to progress. Examples of the social benefits abound, most visibly by the 'Silicon Valleys' that have sprung up around a number of them.

What is the one discovery that would herald a scientific revolution in the 21st century?

A scientific model of how mammalian genomes pack so much into so little space. Just think of the amount of information stored in a mere three billion base pairs, let alone all else these mere molecules must accomplish. Measured from a computer perspective, that's four bits of storage per pair — or about one tenth as much as a high-end iPod Touch. How can so little 'memory' store the exquisite details of our entire bestiary? Consider, for example, that experiments have shown that a newly-born mountain goat is wary of heights from the instant that it opens its eyes for the first time.

Bell Labs and other corporate research sites, which led to many Nobel prizes, are on the decline or have shut down. Is corporate, basic research critically needed or is research in academia sufficient?

Times have changed. In the past, giant corporations made most, if not all, of what they sold. The size and scope of a corporate research laboratory depends upon where its owner adds value, i.e., what kinds of components or systems its owner makes — rather than buys. In today's world, properly focussed in-house research can still provide both future and current advantage, generally on a smaller scale than in the past. Drug companies are a special case. Most have maintained their research laboratories, but face a unique problem: their labs are staffed by brilliant biochemists, ill equipped to provide advantage in a genomics-dominated era.

You must have experienced a lull at some point in your research career. What kept you going?

My primary driver has always been the example set by both my parents. No matter how difficult things were for them, they kept going.

As I've grown older, I've also learned to step back when I'm stuck, look around, and repurpose my skills to move in a new direction. When I became responsible for all of Bell Labs' research, for example, I needed to learn about computing; I wrote a book about computers and people from an outsider's perspective.

From your experience, what valuable advice would you give all young researchers who are starting their research life so as to become a good scientist?

Everyone must deal with self-doubt. Just remember that you are 'seeing' the outside of everyone else's head and comparing it to the inside of your own. Ask questions. Do not assume that everyone else is keeping quiet because the answer is obvious to each one of them—just not you!

I once read that Isidor Isaac Rabi's [who discovered nuclear magnetic resonance] mother asked him every day. "Have you asked any good questions today?" Sometime after I had won a Nobel prize myself, I told Professor Rabi that I thought that the other kids in his classes must have hated him: this well-scrubbed kid who embarrassed them with his questions. "Not at all," he replied. "There are questions which destroy, and questions which illuminate. I was taught to always ask the second kind."

Aside from as a Nobel Laureate, how would you want the world to remember you?

As a frightened young refugee who went on to live the American Dream.

FLEMMING, C./LINDAU LAUREATE MEETINGS

Christian de Duve
The joy of discovery

The Nobel Prize in Physiology or Medicine 1974 was awarded to Christian de Duve, Albert Claude and George E. Palade for their discoveries concerning the organization of the cell.

PROFILE
- Emeritus professor at both the Catholic University of Louvain and the Rockefeller University, New York
- Born in Thames-Ditton, near London, 2 October 1917 to Belgian parents
- Returned to Belgium (Antwerp) in 1920
- Entered Catholic University of Louvain in 1934
- Had a short stint in the army, concluded with an escape from a prisoners' column
- Married Janine Herman in September 1943 (Janine died in 2008). They have four children
- Has collaborated with four Nobel laureates (Hugo Theorell, Carl & Gerty Cori and Earl Sutherland) during his career

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?
In biomedical research, multidisciplinary collaboration has become mandatory. The best — I would say only — way to foster it is by grouping together experts of all relevant disciplines in the same geographical location, as was done when the International Institute of Cellular and Molecular Pathology (ICP), now the de Duve Institute, was founded in Brussels in 1974.

How can the public be convinced of the importance of fundamental research with no applications in sight?
Using simple logic: application presupposes discovery, and discovery requires research, and research implies exploring the unknown, with, by definition, the inability to predict how useful or profitable whatever will be found could turn out to be.

Independent of socio-economic considerations, fundamental research deserves to be supported for its own cultural value. The search for truth is, together with the quest for beauty, goodness, meaning and love, a major pillar of human civilization.

As an aside, before trying to convince the public of these basic truths, one should perhaps start with the administrators who too frequently tend to ignore them.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution?
I believe peer review is an essential component of the scientific endeavour, and vitally dependent on ethical integrity. Regarding this latter attribute, the system seems to me to be under threat because of what I perceive — perhaps wrongly — as an increase in scientific misconduct and because of the growing involvement of academic investigators in profit-directed research.

What advice would you give all young researchers who are starting their research life, so as to become a good scientist?
First, whatever you do, seek excellence, both intellectually and technically. Science is one field of human endeavour that must be unashamedly elitist. You cannot seek the truth with poor thinking or sloppy techniques.

In conducting your research, observe total rigour and intellectual honesty in the analysis of facts, consider all possible hypotheses, plan your approach to test those hypotheses, and submit your conclusions to the verdict of observation and experimentation without preconceived ideas. Never conduct research with the aim of proving a theory, but, rather, to invalidate it if it should be wrong. The best proof is failure to disprove.

In the experimental sciences, pay special attention to the quality and reliability of the

> *"Good research is not learned in books, but at the bench, like the crafts in the Middle Ages."*

instruments and techniques you use — and to your own ability to handle them. Good research sometimes depends on manual skill. Here I have a tip: separate planning and execution. Once you have planned an experiment, concentrate on its correct performance. The two activities are to some extent mutually exclusive. Of course, you must watch for any unusual or unexpected occurrence; if something like that occurs, keep it in mind but don't change horses midstream. Next, follow your intuition — your curiosity, which is the strongest motivation for a scientist. Don't hesitate to be adventurous. Be self-confident and don't fear being ambitious. This is not always possible within the constraints imposed on research from the outside, but do your best.

Finally, I have two more recommendations. The first is: enjoy it. Science is fun. Exercising your brain and your fingers at the same time provides immense satisfaction. The joy of discovery is unmatchable.

My second recommendation is to the younger generation just entering research. Choose your mentors well. Good research is not learned in books, but at the bench, like the crafts in the Middle Ages, under the supervision of a master.

Aside from as a Nobel laureate, how do you want the world to be remember you?
I have no such ambition. In the history of science, my contributions are minor and would have been made by someone else had I not stumbled on them first. They already appear in textbooks without mention of my name. I am no Galileo, Newton, Darwin, Einstein or Watson and Crick. But I have had fun and have been rewarded beyond my deserts. So be it.
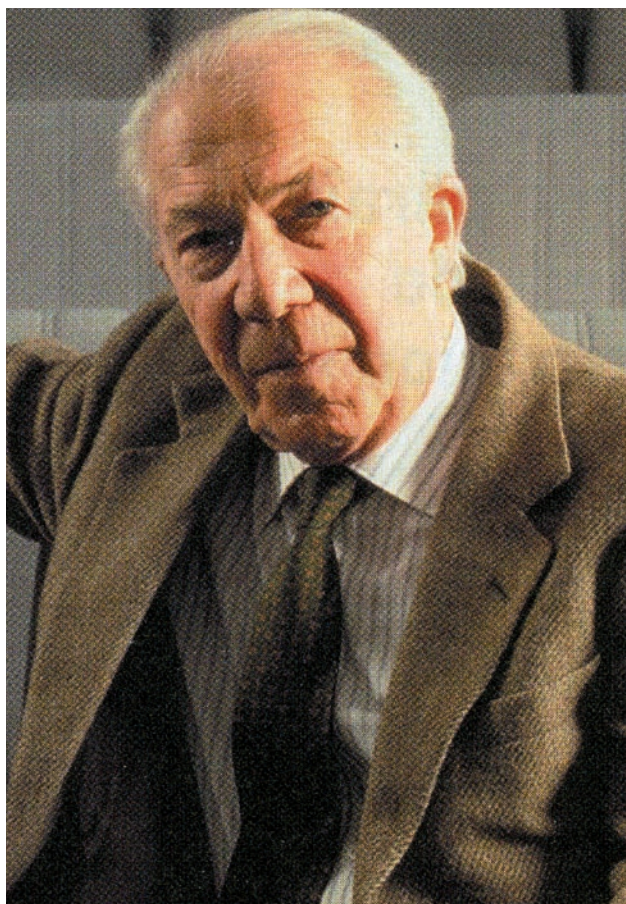
Oliver Smithies
Science brick by brick

The Nobel Prize in Physiology or Medicine 2007 was won by Mario R. Capecchi, Martin J. Evans and Oliver Smithies for discoveries that led to the development of knockout mice.

PROFILE
- Excellence Professor of Pathology and Laboratory Medicine. University of North Carolina at Chapel Hill
- Oliver and fraternal twin, Roger, were born prematurely on 23 June 1925, in England.
- An early childhood infection left him with a mitral valve murmur, and doctors forbade him playing sports until his mid-teens.
- Was inspired to become a scientist by a comic strip about an inventor
- Studied for undergraduate degree and PhD at the University of Oxford
- Can make useful devices from 'junk' – oddments in his graduate lab would be labelled NBGBOKFO: No Bloody Good, But Okay For Oliver.
- Invented gel electrophoresis in 1955
- Moved to University of North Carolina at Chapel Hill, with wife Nobuyo Maeda, in 1988

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?

I don't believe interdisciplinary approaches need further fostering. There has been a good deal of such effort already, perhaps even too much in the United States. Emphasizing research proposals that have an interdisciplinary approach leaves less money for more individual projects that tend to be better at generating completely new ideas and so are very important.

Of course, many scientific questions of the day are broad, and therefore need to be addressed from multiple angles. But this should not be forced onto scientists by funding agencies, although it is an important mission for funding agencies to facilitate researchers from different fields talking to each other. Because that's what science is all about: building and communicating knowledge. You may have a beautiful experiment in your lab notebook, or in your head, but it isn't science until you make it available to others so that they can build on it.

Bell Labs and other corporate research sites, which led to many Nobel prizes, are on the decline or have closed. Is corporate, basic research critically needed, or is research in academia sufficient?

I think all people in the science community who observed what happened to Bell Labs feel regret. I think it will be a shame and real loss if the corporate community fails to see the value of having basic research in their own labs. Very often, basic research will interact with more practical research in quite surprising and beneficial ways when the two are carried out in close proximity.

There is still quite a lot of one type of corporate research going on in the biological sector in the United States, because historically it was relatively easy to obtain venture capital in this field. So, if scientists had a good idea for a new drug or a new product, quite a number succeeded in founding their own nascent pharmaceutical companies. Very likely, this type of research will eventually thrive again despite the current lack of investment capital.

How can the public be convinced of the importance of fundamental research with no applications in sight?

The Hubble telescope is a good example of a project that has helped people to see what basic research is about. The beautiful images it produces don't have any immediate application for most of us, but from them we learn what the Universe looks like and how it works. Such endeavours satisfy the curiosity that is part of human nature.
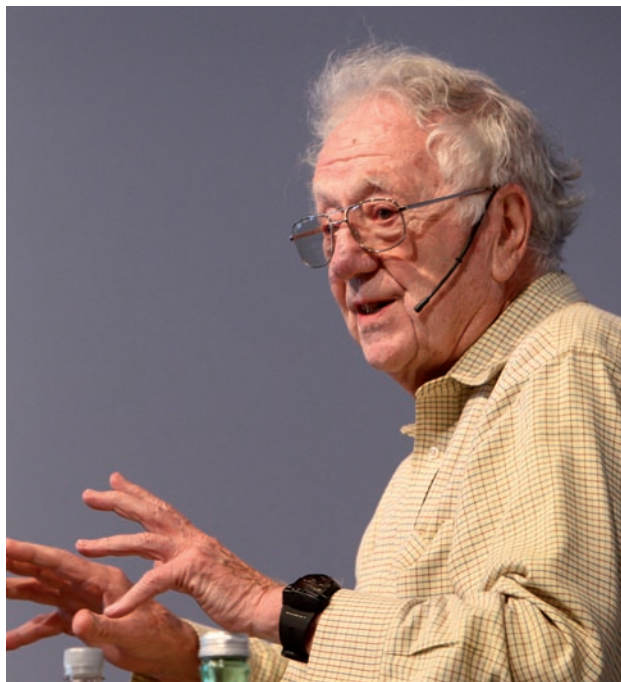
We humans have been able to develop because we have learned to communicate what our predecessors have found out about the world, and use this knowledge as the basis of new ideas. But we shouldn't forget that people in the Stone Age did not only make tools, they painted on cave walls for heritage.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution?

In general, I think that scientific reviews are fair. Certainly most scientists try hard to be fair when they review. Nevertheless, sometimes personal views intrude, but that's just human nature. We try ourselves and teach our students to look at things fairly whether the result is in our interest or not.

What bothers me, though, is that today so much effort and so much time is spent by scientists trying to have their work approved. Researchers get together for dinner and instead of discussing science, they talk about funding. If we were able to take away some of the economic pressures that most scientists experience today, it would probably improve the output of many.

You must have experienced a lull at some point in your research career. What kept you going?

The usual suspect for a lull in a researcher's life is boredom; what you are doing no longer seems very interesting. And the way out of this situation is simple in principle but not always in practice: think of something new to do. Just before I began my work on gene targeting, I was getting increasingly bored with the work I was doing, and it was not going very well either. I needed some new inspiration.

Eventually the trigger came to me in the form of a paper published in 1982 by Mitchell Goldfarb *et al*. It described a complex gene-rescue procedure, which I realized could be used to test whether it is possible to modify a gene by homologous recombination. In that moment I had regained my excitement for science. It took three years to make the experiment work, which 20 years later led to my receiving a share in a Nobel prize.

Gerardus 't Hooft
Nothing to fear from mistakes

Together with mentor Martinus J.G. Veltman, Gerardus 't Hooft's Nobel Prize in Physics 1999 was won for elucidating the quantum structure of electroweak interactions in atoms.

PROFILE

● Theroretical physicist at Utrecht University, the Netherlands
● Born in 1946 in Dan Helder, the Netherlands
● Has a Nobel prize-winning grand-uncle, Frits Zernike
● 't Hooft was late learning to read, write or even speak
● Came second in the Dutch national maths olympiad, aged 16
● Was coxswain in the Triton rowing club, of the State University of Utrecht — helped by his ability to steer the boat in a straight line
● In 1972 't Hooft completed his PhD, married Albertha Schik (the couple has two daughters) and, together with former supervisor and long-time collaborator Martinus Veltman, joined CERN

What is the one discovery that would herald a scientific revolution in the 21st century?

I think there will be many, but there is one field that is developing faster than many others: information and communication technology. I think that the discovery of genuine artificial intelligence (computer programs that have learned to think like a human) would have tremendous consequences. Today, many scientists think that it is impossible, but I don't. If such an intelligence can be constructed, it will quickly outsmart humans by a big margin. The consequences are difficult to predict, and even potentially dangerous — I don't fear that such a program will overthrow humanity or anything like that, but it might bring unparalleled power to those who are in possession of such a device.

How can the public be convinced of the importance of fundamental research with no applications in sight?

The only way to convince the public is by being convinced yourself that it is important. Not all scientists have this attitude. Some scientists draw a line between 'research that has immediate applications' and 'curiosity driven research', placing it around 2 or 3 years ahead. I put that line many centuries away from now, so my task is harder.

Even with this border many centuries away, you can still point to the haphazard and unpredictable path of history, and to the numerous occasions when healthy human curiosity was rewarded with fantastic opportunities and applications much sooner. There are so many examples that I find it easy to convince people of the importance of my research. The applications are only one of the motivations for doing my kind of research. The primary one is a pure desire for understanding the world in which we live. The public wants this too, and we scientists can deliver such understanding.

Why are we not even close to building a reliable quantum computer?

Here, I am much more pessimistic. I am not convinced that a genuine quantum computer can be constructed that would outperform conventional computers by a big margin, because I have not seen convincing arguments that the numerous obstacles can be overcome. In contrast, conventional computers can, in principle, be improved so much that they will dwarf any of today's expectations of a quantum computer. It will all be in the software. Today's software industry is still very much of the 20th century type.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution?

"The primary motivation is a pure desire for understanding the world."

Peer review is still very important, because all research and all researchers need to be checked and rated. But there is one thing more important — the free availability of research papers all over the world, to anyone, at all times. Unfortunately, the science publishing companies are so desperate to make money that they make scientists pay for the articles they see, and/or the articles they publish, and consequently, much of the peer-reviewed work is actually locked up. We scientists should do all we can to change that situation. One problem is: who should pay? The scientist who wants to write a peer-reviewed article, or the one who wants to read it?

I would like to see some type of automated system, where a computer program checks a paper's citations and asks a few of the scientists who referenced it to provide some sort of rating; after all, if you cite a paper you often already have some sort of judgment of it. Of course this should not be done too often as this risks overloading scientists who give lots of citations.

What advice would you give all young researchers who are starting their research life so as to become a good scientist?

I once heard about a senior scientist, a Nobel laureate in fact, who gave this advice to a young scientist beginning his career: "Don't make any mistakes!" I came to realize that it is the worst possible advice he could have given. On the contrary, if you are afraid of making mistakes, you may become a good textbook writer or a mediocre teacher, but you won't accomplish much as a scientist. Science advances by trial and error. Any mistakes made on the way are insignificant. The mistake will surface sooner or later and be corrected. It could be an insignificant mistake or a significant one, but even in the latter case the scientist responsible will not be punished for it; rather, he or she will be praised for being the first to ask the right question.

In short, my advice is: 'Don't be afraid of making mistakes, even in your published papers, but do make it a question of honour to be the first to discover your own mistakes; you don't want your referee or your colleagues to discover it. But then, even if they do, keep in mind that having asked the right question is of much more importance. You will be rewarded for that.'

David J. Gross
The frontier physicist

The Nobel Prize in Physics in 2004 was awarded to David J. Gross, H. David Politzer and Frank Wilczek for their discovery of how quarks interact within protons.

PROFILE
● Director of the Kavli Institute for Theoretical Physics, University of California, Santa Barbara
● Born on 19 February 1941 in Washington, DC
● First 'job' at age 11 was proofreading his father's book *The Legislative Struggle: A Study in Social Combat*
● Family moved to Israel in 1953
● Returned to US for graduate studies at the University of Berkeley in 1962
● Became a Junior Fellow at the Harvard Society of Fellows in 1966
● Switched to working on string theory in the 1980s, developing the concept of heterotic string in 1984

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?

The natural world does not recognize the artificial separation of adcadamia into traditional disciplines. Inevitably, as science evolves, many of the most interesting questions in nature fall into the cracks between the disciplines and benefit enormously from the attention of scientists from different cultures Fostering collaborations is non-trivial since the separate disciplines often find it difficult to communicate due to language barriers and university departmental structures.

At my institute (Kavli Institute for Theoretical Physics) we have been rather successful at fostering interdisciplinary research at the crossroads between physics and allied fields, such as biology, chemistry, Earth sciences and mathematics. We do this by bringing scientists together in loosely organized, parallel programmes in which, the emphasis is on discussion and collaboration over an extended period, we have found it possible to exchange ideas over disciplinary boundaries and form new, interdisciplinary collaborations.

Bell Labs and other corporate research sites, which led to many Nobel prizes, are on the decline or have closed. Is corporate, basic research critically needed, or is research in academia sufficient?

The demise of Bell Labs and other corporate research labs is very unfortunate, but an inevitable consequence of the short-term profit seeking of today's corporations. Bell could afford to support curiosity driven research that often led to unanticipated applications, largely because it was a monopoly regulated by the government. Today's corporations are happy to have the public pay for research at universities and then turn the fruits of that research into profit-making applications. I do not think that direct corporate research is necessary for healthy science, except of course in producing real products

from applied knowledge. In some cases profit motives are even harmful, leading to secrecy and distortions that impede the progress of science.

Is there a country in which you feel that the management and handling of science funding is undertaken in an exemplary way?

"The peer-review system is the worst form of scientific management except all the other forms that have been tried."

No, there is not. Fifty years ago the answer would have been the United States with its new institutions (the National Science Foundation and the National Institute of Health) that were a model of innovative management of science. But with time, the US system has ossified,

and levels of support for the non-biological sciences have declined precipitously. The institutions created half a century ago have not responded adequately to the rapidly changing directions of science. The structure of the NSF, with well-defined and separate directorates and divisions, and with most contracts controlled by individual program directors (who are often reluctant to allocate their funds to other programs), places enormous obstacles to the support of new, and especially interdisciplinary fields of research. Although the NSF is eloquent in its appreciation of interdisciplinary research, it has not been able to overcome these obstacles or make the structural changes necessary to do so. After 60 years of operation the foundation desperately needs a comprehensive review of its structure and procedures.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution?

The peer-review system, to paraphrase Winston Churchill, is the worst form of scientific management except all the other forms that have been tried. There is no substitute for peer review, but peer review is not a guarantee of fairness. Grant administrators have tremendous discretion in selection of reviewers and interpretation of reviews, making it possible to direct the course of science from within the funding agency. Transparency is a key to fairness, and two keys to transparency are oversight and responsible interaction with the working scientific community.

Peer review works best in selecting good projects from bad in well-defined areas of research. But research projects that fall outside current trends, as well as interdisciplinary approaches, are often disfavoured. This fault can be corrected by setting aside funds for innovative, but risky, ideas.

FLEMMING, C./LINDAU LAUREATE MEETINGS

John C. Mather
Politics and prophecy

John Mather and George Smoot's discovery of the anisotropy of the cosmic microwave background radiation won the Nobel Prize in Physics 2006.

Are scientists under-represented in politics? And do established scientists, especially Nobel laureates, have a duty to become active in politics and science policy?

Yes, I do think scientists are under-represented in politics, as are most professional groups. At least in the United States, most politicians are lawyers with little or no background in technical subjects. But being a Nobel laureate conveys no expertise in political matters, and I think it's important for any newly-active political person to be cautious before jumping into debate. I have spent a little of my Nobel prize money, working through the American Institute of Physics, to start summer internships for undergraduate physics students to gain experience working on Capitol Hill. There are currently three physicists working in the House of Representatives and the Senate, but to increase that number we need to find people who have the special combination of talents to be politicians as well as scientists.

What would be the one discovery that would herald a scientific revolution in the 21st century?

Room-temperature superconductivity could enable efficient sharing of electrical power around the world, changing the economic balance dramatically. It could also enable magnetic levitation for transportation, changing the entire structure of nations. Similarly, any other discovery or innovation that changes the availability of energy for food, transportation and shelter would have extraordinary impact.

How can the public be convinced of the importance of fundamental research with no applications in sight?

I don't think 'convince' is the right word. How about 'inspired'? I think every researcher has the opportunity to show members of the public how exciting new discoveries are, and how they lead to knowledge that may eventually be applied. There are many stories about fundamental research that led to world-changing applications. Who would have thought that atomic clocks of extraordinary precision would be flying in space to run the GPS systems that people use to pilot their cars and planes? Let's not be shy. Although apocryphal, the quote attributed to Faraday is apropos: "Someday you can tax it." More directly, everything we have in modern society, from food to shelter to jet airplanes to iPhones, is the result of discoveries that were once seen as having no applications in sight.

"We need people who have the special combination of talents to be politicians as well as scientists."

What is the future of the standard model if the elusive Higgs boson and the even more elusive neutrinoless double-beta decay are discovered?

Hard to guess. Maybe it will still be the standard model, but with frills. The bigger question is always quantum gravity, or the alternative possibility that there is no such thing and we have no clue about the nature of space and time. That would not surprise me a bit.

What is the most practical and relatively safe alternative energy source?

So far none is the obvious winner. Solar and wind power are widely distributed and could become dominant with additional economies of scale and with subsidies comparable to those given for fossil fuels. Neither solar nor wind power is intrinsically polluting (though they disturb the neighbourhood and alter the landscape), but both technologies currently use exotic materials that themselves could be in short supply.

Moreover, we have become accustomed to the extremely inexpensive energy of fossil fuels; I suspect we are already living well beyond the level of energy consumption that can be sustained in the long run. So people should not be too surprised if dramatic and disastrous changes occur in the global economic system as energy becomes scarcer. Perhaps it would be good to plan for change? But Tiresias and Cassandra, given the gift of prophecy, were cursed so that people did not believe them. Exercise for the reader: why is that?

What valuable advice would you give all young researchers who are starting their research life so as to become a good scientist?

Number 1: study reading and especially writing! And then, practise public speaking. The ability to explain logically, intuitively and viscerally why something is true is essential to planning research strategy, explaining results and obtaining funds for future work. There's really no substitute for good communication. And it will be valuable wherever life takes you.

Paul J. Crutzen
Anthropocene man

Paul J. Crutzen shared the Nobel Prize in Chemistry 1995 with Mario J. Molina and F. Sherwood Rowland for their work on formation and decomposition of ozone.

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?

Interdisciplinary research has always played a major role in my career. In my earliest papers I proposed that NO and NO_2 could act as catalysts to destroy ozone in the stratosphere or to produce ozone in the troposphere, yet I had no formal education in chemistry. In fact, I was trained as civil engineer; I shifted to working as a computer specialist at the University of Stockholm to run meteorological models and, while helping to develop a program for classical ozone chemistry, got hooked on atmospheric chemistry. I wrote the 'NOx papers' when I was 36 years old.

> *"After my first research successes I lived in a constant fear of running out of problems to work on: it would have been terrible for a well-funded research director."*

Are scientists under-represented in politics? And do established scientists, especially Nobel laureates, have a duty to become active in politics and science policy?

There is, in fact, a dearth of scientists and scientifically educated representatives in politics, which is a worthwhile subject for analysis. I am only aware of three national leaders who have had scientific backgrounds: Margaret Thatcher, Angela Merkel and Gro Harlem Brundtland of Norway; interestingly, all women.

I don't think that Nobel laureates have a special duty to become more active in politics, except if their research could lead to problems for society. In the first place, they should remain scientists. However, the general public expects Nobelists to have an above average interest in social issues and the increased participation of scientists here should be applauded, as is the case under President Obama.

To establish closer contact between politicians and scientists, the German parliament has created commissions consisting of equal numbers of politicians and scientists to study and highlight problems of general importance, such as climate change and stratospheric ozone depletion. The advice of these commissions has been very successful in the political process.

How can the public be convinced of the importance of fundamental research with no applications in sight?

By pointing to the examples of great scientists like Einstein, Newton, Darwin, Faraday and others. In my own, much more modest, case, when I wrote the article about the potential effects of NOx on ozone, I had not the slightest idea about its relevance.

Things changed in the fall of 1970 when I read a report of a major international conference in the US, which mentioned NOx emissions but denied their importance.

Independently, Harold Johnston and I drew attention to the problem of NOx emissions from super sonic aircraft destroying stratospheric ozone. The same happened in the case of chlorofluorocarbon emissions: the societal relevance was initially not recognized.

You must have experienced a lull at some point in your research career. What kept you going?

After my first research successes I lived in a constant fear of running out of problems to work on: it would have been terrible for a well-funded research director. That did not happen. The main reason is that my research field is highly multi-faceted and interdisciplinary. There is room for many projects.

What advice would you give all young researchers who are starting their research life so as to become a good scientist?

Stay away from big science in which the resulting papers have many authors. And choose a professor who gives you a lot of freedom. I was lucky to find Bert Bolin.

Aside from as a Nobel laureate, how do you want to be remembered?

As the person, who significantly increased knowledge about the processes that determine the distribution of ozone in the atmosphere.

And as the scientist who coined the term 'Anthropocene': A new geologic epoch dominated by human activities, actually first published in *Nature*.

And as one of the scientists who drew attention to the potentially devastating climatic consequences of a nuclear war, the so-called 'nuclear winter'. More people would die of the indirect consequences of mass starvation and disease than would be killed by the nuclear bombs.

FLEMMING, C./LINDAU LAUREATE MEETINGS

Peter Agre
The family naturalist

Peter Agre shared the Nobel Prize in Chemistry 2003 with Roderick MacKinnon. Agre's half was awarded for his discovery of a water channel protein in cell membranes.

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?
Collaborations were incredibly important to our laboratory's pursuit of the structural biology and physiology of the aquaporin water channel. In my experience, interdisciplinary science often happens without much planning. By sharing ideas and news of our discoveries, we have always made friends with other scientists with technical capacities that we needed.

Are scientists under-represented in politics? And do established scientists, especially Nobel laureates, have a duty to become active in politics and science policy?

"Discovery of the molecular explanation for happiness would be revolutionary."

Science is grossly under-represented in elected offices of government. Most scientists are rather poor at communicating their work to non-scientists, and as a result the public fails to understand why our work is useful. Participation by scientists at all levels of elected office will be educational for the public. This has occurred in some countries such as Germany, where Chancellor Merkel was a physical chemist, and China, where President Hu Jintao trained as a hydraulic engineer. Governing is difficult, but I think governing without an advanced understanding of science is very risky.

What is the one discovery that would herald a scientific revolution in the 21st century?
Discovery of the molecular explanation for happiness would be revolutionary. But this may not be just wishful thinking. Neuroscientists are making incredible breakthroughs in understanding the actions of serotonin. So who knows?

You must have experienced a lull in your research career at some point. What kept you going?
I do not often talk about this, but I nearly quit science after the birth of our third child Lydia. She sustained prolonged cerebral hypoxia during birth and was left with permanent, severe brain damage from which she eventually died. Unable to afford the necessary home medical care, I considered abandoning the laboratory to become a full-time clinician, since this seemed the safest way for me to support my family. The stable influence of my wife kept us on track. I stayed in the lab but changed the focus of our work to pursue medical problems of newborn infants.

The new direction of research led to the discovery that won me a Nobel.

"I considered abandoning the laboratory to become a full-time clinician, since this seemed the safest way for me to support my family."

What advice would you give all young researchers who are starting their research life so as to become a good scientist?

I always tell young scientists that aspiring to win prizes is a poor reason to do science. I believe that the three most important objectives a young scientist should pursue are: 1) making an original discovery, 2) earning the respect of your peers and 3) training the next generation of scientists.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution??
I think that scientific peer review, like democracy, is a very poor system but better than all others.

Aside from as a Nobel laureate, how do you want to be remembered?
The Nobel prize may be more important for science than for the individual laureate.

I often surprise people by asking them to name the Nobel prizewinners in 2003 (my year). Usually, they cannot remember any names, which is a demonstration that even a Nobel prize guarantees only a footnote in history books.

At the end of my life, I will be happy to be remembered as a scientist who loved the natural world and loved his family.

George F. Smoot
Thinking in aeons

John Mather and George Smoot won the Nobel Prize 2006 in Physics for their work on cosmic background radiation. Smoot measured the temperature variation (anisotropy).

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?

This is vitally important and crucial to the progress of modern science. There are a number of ways to do this. The approach that I prefer and am trying is to create a centre adressing interesting science problems requiring interdisciplinary skills and knowledge and with a mandate to address these issues. It is important to have both the incentives - financial and approved research directions (official, peer, and public) — and the proximity including coffee/discussion areas so that personal contacts are made and incidental discussions arise that often lead to other avenues of investigations and ideas.

> *"The Universe is in a quasi-static equilibrium so that much of its history is irrelevant."*

What is the future of the Standard Model if the elusive Higgs boson and the even more elusive neutrinoless double-beta decay are discovered?

I am looking forward to the completion of the old standard model of particle physics with the discovery of the Higgs, but I look forward with much more anticipation and eagerness to going beyond the standard model and discovering dark matter and, hopefully, extra dimensions. The dark matter could be the lightest supersymmetric particle (a beautiful but untested extension of the standard model) or the energy used going into a 'small' extra dimension (the lowest rung possible in the so-called Kaluza-Klein tower). There are many, potentially very intriguing discoveries that could teach us about the world and expand our ideas.

Can theoretical physics help identify guiding principles in biology?

At Lindau there was a lunch involving three physics Nobel laureates and three biology-related laureates at which we tried to see what we could learn from each other, including techniques that could be cross-applied. One participant commented that historically biologists have sought out physicists for their tools and algorithms. However, when physicists saw what biologists were doing and what they could learn, many were put off by the lack of rigor and the shallow understanding. Biological systems are so complex and so poorly observed and measured that biologists had to be satisfied by a high level and rough understanding of what they were studying, while physicists are trained for rigor and precision.

Some physicists, however, stayed and began doing biology, but there was little sign of bringing biological techniques into physics. Due to the advances that physical science tools and approaches are allowing in biology, the depth and rigor of biology is steadily progressing. The biologists at the lunch were more pessimistic about the rate of progress and understanding than the physicists. This might be a function of the biologists' closeness to the difficult issues.

One thing that discourages these biologists is that so much of biology seems to depend on the history of development and evolution. I tried saying one might think this about the Universe, but many times the Universe is in a quasi-static equilibrium so that much of its history is irrelevant. "Weren't there many biological systems that have reached an optimum, e.g. photosynthesis, or specifically the ATP energy-transfer process, so that the historical path was not so critical?" They agreed, although only for a limited number of cases.

At this stage and for the foreseeable future, I think that biology will continue to apply the tools and techniques of experimental physics but not so much the approach and tools of theoretical physics.

Aside from as a Nobel laureate, how do you want the world to remember you?

I do not think the world will remember any of us. When you work in cosmology, you think with a perspective of billions of years. How many people do we remember from 1,000 years ago? How many from 10,000 years ago? How many from 1 million years ago? etc. So now the question is more like that about Ozymandias [from Percy Shelley's poem]. After a thousand or so years nothing is left of the great works but a few broken relics. What will be left in a billion years?

> *"How many people do we remember from 1,000 years ago?"*

The point is not about being remembered or about doing science to win a Nobel prize — you do science because you want to be a scientist. Likewise, you live your life in a way you think is good and good for you.

Harold W. Kroto
Working at the coal face

Harold Kroto shares the Nobel Prize in Chemistry 1996 with Robert F. Curl Jr. and Richard E. Smalley for the discovery of buckminster fullerenes.

How important is an interdisciplinary approach in addressing urgent scientific questions, and how can we foster such collaborations?

Vital – the best way is to change the traditional departmental structure from chemistry, physics, biology and engineering to a more overarching infrastructure: say atomic and molecular science (which subsumes molecular biology and condensed-matter physics) and high-energy physics and astrophysics.

How can the public be convinced of the importance of fundamental research with no applications in sight?

The only thing I can suggest is to continually stuff the infinite list of applications that have benefited mankind up the noses of the public and politicians alike. Lasers: from supermarket checkouts to DVDs and eye surgery. Relativity: from GPS to nuclear power. Chemistry: anaesthetics and penicillin. DNA testing: freeing innocent people from death row.

Bell Labs and other corporate research sites, which led to many Nobel prizes, are on the decline or have closed. Is corporate, basic research critically needed, or is research in academia sufficient?

There are as many ways to do science as there are scientists and we need the whole spectrum of research options. The demise of Bell Labs (and IBM, Shell, Imperial Chemical Industries etc.) is an indication of the way in which short-sighted capitalist pressure has worked to the detriment of research in industry and society.

Many people consider the peer-review system broken. Do you share their view, and do you have a solution?

The peer-review system is the most ludicrous system ever devised. It is useless and does not make sense in dealing with science funding when history abounds with a plethora of examples that indicate that the most important breakthroughs are impossible to foresee.

The science budget should be split into three (not necessarily equal) parts and downloaded to departments. The local institutions, and not government departments, should disburse funding as they are close to the coalface and can decide what needs support and what is in the long-term interest of the department. There should be no research proposals on which to waste time.

One part should go to young people chosen by their universities as the researchers on which their institution's future will depend — they have done the work, why waste time doing it again when people have no time and are too far away from the coalface and in general do not have the relevant expertise?

The second part should go to a group whose most recent report was excellent. This is the racehorse solution — if a scientist has just done some great work, let her or him run again.

The final part should be split among scientists who might not have been funded in the second group, but have put in a proposal; let them get 50:50 matching funds from industry.

What is the most practical and relatively safe alternative energy source?

I suspect that until we have solved the problem of photocatalytic splitting of water and/or nuclear fusion we shall have little option but to turn to nuclear fission — it is the only proven alternative to fossil fuel.

You must have experienced a lull at some point in your research career. What kept you going?

I always had some niggling, little back-burner ideas (the Nobel project was one such idea) and always liked to work with university graduates and undergraduates — the undergraduates were a great stimulus and 50% of these neat projects were stimulated by the need to float a little project that did not seem very important at the time, but when initiated invariably turned out to be great.

Aside from as a Nobel laureate, how do you want to be remembered?

I think I would prefer to feel I had made a positive contribution to the education of young people to recognise the truth in general and perhaps to have struck a blow against the irrational forces which are presently undermining the Enlightenment, democracy and the freedom of individuals.
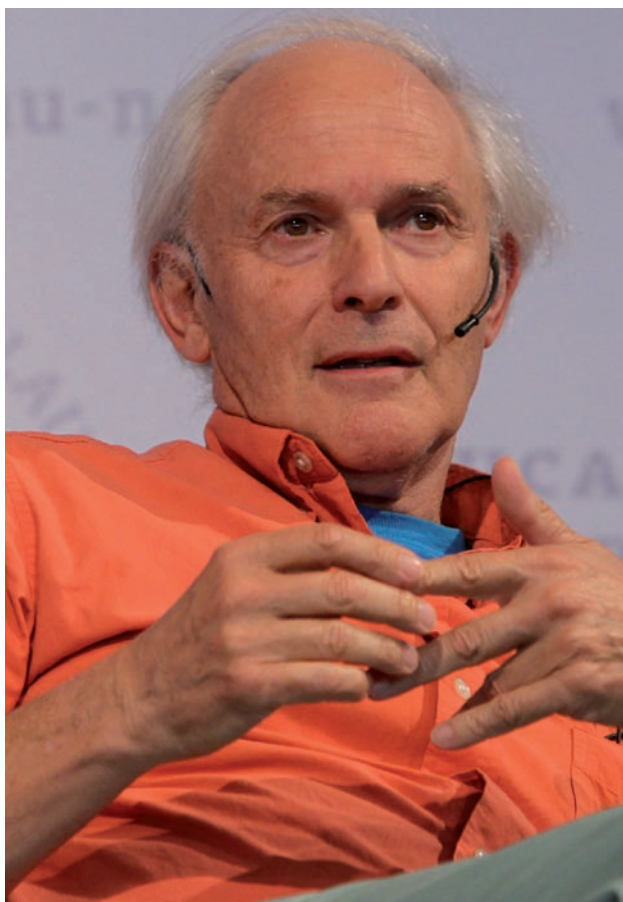
Count Lennart Bernadotte (C) addresses physicists Werner Heisenberg (L), Edward Purcell (3L), Isidor I. Rabi (6R), Emilio Segre (5R), Paul Dirac (2R) and Gustav Hertz (R) at a meeting on Lindau (1962).

TIMELINE

Lindau and the zeitgeist

The annual Lindau Nobel Laureate Meetings have evolved over the years, reflecting changes in both science and society.

BY JOHN GALBRAITH SIMMONS

THE ROAD FROM STOCKHOLM...

The Nobel Laureate Meetings at Lindau owe their cachet first and foremost to the prestige of the Nobel prize. This prestige, in turn, has much to do with the revolutionary advances in science that were afoot when the prize was first awarded in 1901. Quantum concepts were about to dethrone classical physics. Chemistry, closely allied with atomic physics, would soon extend deep into biology. Major advances in all scientific disciplines would transfigure medicine.

The Nobel also owes its status to scrupulous selection, with an emphasis on rewarding breakthrough discoveries, from the 1901 physics prize for the discovery of X-rays up to the 2009 physiology or medicine award for learning how hromosomes are protected by telomeres and the enzyme telomerase. Now, some 540 awards later, it is clear that the Nobel prizes have also become a touchstone for the public's awareness of the importance of science.

Not all Nobelists are happy recipients. Physicist Richard Feynman, for example, called the prize "a pain in the neck" and molecular biologist Max Delbrück considered rejecting it for being a pointless distraction. But most accept the invasive downside of celebrity as a price worth paying.

The upsides to receiving a Nobel prize are many and diverse. At one end of the spectrum, some recipients are happy to simply take advantage of the enhanced opportunities to continue or broaden their laboratory work. At the other end, some scientists use the Nobel prize as a pulpit to discuss important causes. Most lie in-between.

Which leads to Lindau, where these characteristics of dedication and curiosity, not to mention more ideological viewpoints, are evident each year; putting scientific advances in context in our ever changing world, and helping them to confer, as Alfred Nobel once said, "the greatest benefit on mankind."

...TO LINDAU...

Occupied and fragmented, post-war Germany was suffering from the scientific brain drain that began with the Third Reich. In 1950, to help jump start the intellectual recovery, local physicians Franz Karl Hein and Gustav Parade decided to organize a meeting in Lindau, a tiny picturesque island-city on Lake Constance, nestled in the southern state of Bavaria. They had two aims: to inform doctors of the latest medical developments, and to promote Germany's reintegration into the wider scientific community.

Hein and Parade believed that the key to achieving these goals was to reach out to Nobel laureates. To this end they enlisted the help of Lennart Bernadotte, who owned the nearby garden island of Mainau. As an aristocrat related to the Swedish royal family, Bernadotte had access to the Nobel committees and was able to gain their unofficial support.

At the first meeting in June 1951, seven laureates lectured 400 physicians and researchers on current scientific concerns. William Murphy, from the United States, discussed pernicious anaemia. Otto Warburg described his theory of photosynthesis. Gerhard Domagk, who had developed the first sulfa drug, an early antimicrobial, but who had been forbidden by the Nazis to collect his award in 1939, talked about treatments for tuberculosis.

The meeting was widely covered by the German press, and soon turned into an annual event. Young scientists, initially students from the nearby University of Freiburg, started to attend, and gradually the meeting evolved to focus on the interaction between the laureates and the younger scientists.

Lindau attracted quite a following. Many Nobel laureates attended even when not giving formal lectures. Physicists Paul Dirac, Otto Hahn and Werner Heisenberg were regular fixtures for decades. Willis Lamb, the American physicist who discovered the electron Lamb shift and shared the physics prize in 1955, first attended in 1959 and was planning to attend his 20th meeting a few weeks before his death in 2008.

Although Lindau lectures were scientific, during those first few decades the meetings tended to reverberate at the interface of science and society. Early discussions on the development of CERN — the powerful cooperative European particle-physics project — took place in Lindau during the 1950s. The creeping shadow of the East–West divide and the cold war could be seen in 1953 when, in the wake of a bloody uprising in East Berlin, East-German scientists were not permitted to attend — a state of affairs that would endure for decades.

Indeed, the nuclear arms race had special importance because many Nobelists had contributed to the advances in nuclear physics — if not to the development of atomic weapons directly. At the Lindau meeting in 1955, the year after the detonation of the hydrogen bomb on Bikini Atoll, American geneticist Hermann Muller gave a lecture discussing the dangers of radiation. Meanwhile, several laureates, including Otto Hahn and quantum physicist Max Born, gathered outside the lecture hall to create an outright pacifist statement. Count Bernadotte read it aloud on the island of Mainau at the traditional close of the meetings, and the 12 laureates present (plus another 40 later on) signed it. Although the Russell-Einstein Manifesto, published a few days earlier, became the foundation document for the anti-nuclear Pugwash movement, the Mainau Declaration underscored the intense concern at the highest echelons of science.

Lindau has been the launch pad for other significant issues over the years. For example, in 1968, Franz König, then Archbishop of Vienna, issued a statement authorized by the Pope that the Catholic Church would consider revising its 1616 censure of Galileo Galilei (although it did not arrive until 1992).

Environmentalism and sustainability made early appearances in Lindau, in part because the West German economic miracle came with the price of pollution. Count Bernadotte was a prominent proponent of conservation. His 1961 Green Charter of Mainau contributed to the development of the government-supported conservationist movement in Germany.

The 1990s and the end of the cold war, although not explicitly acknowledged at Lindau, signalled a general move towards closer and more relaxed international relations. But the meeting remained essentially regional, with most foreign students attending simply because they were studying at German universities. "It was a very difficult task to open Lindau to the world," said Ludwig Feinendegen, a physician and researcher in nuclear medicine who joined the Lindau Council in 1979 and later served as its vice president. The interactive aims and informal nature of the meeting worked against funding opportunities, he explained.

There were some successes. By 1998, the Nobel Foundation in Stockholm had begun to appreciate its informal relationship with the Lindau meeting, and that same year the US

The Polonaise is the traditional dance.

Department of Energy sponsored attendance for a group of young American researchers. Nevertheless, at the turn of the millennium, the meeting was chronically underfunded and somewhat lacking focus. "It became different, less serious over the years," said Anders Bárány, a physicist and former deputy director of the Nobel Museum in Stockholm who is involved in preserving and referencing the meeting's records.

Bárány, whose grandfather Robert won a Nobel prize in 1914, attended his first meeting in 2000. He recalls dancing the traditional Polonaise, partnered with a young green-haired woman. He asked: "Are you interested in physics?" She was not. Was she going to become a chemist? "No." What about medicine? No again. What then was she doing at Lindau? "I heard it is a nice party."

… TO THE WIDE WORLD OVER

And so it was, 50 years on from its first gathering, the Lindau meetings needed a makeover. Retaining the central concept of an informal annual meeting, organizers and laureates established a foundation to recast it for the 21st century. They were not short of expertise.

"We received advice from the laureates themselves on how to chart a new course," recalls Wolfgang Schürer who, as chairman of the board of the new Foundation Lindau Nobelprizewinners Meetings, helped oversee the entire process. "Out of about a dozen very concrete, very specific recommendations from the laureate community, we embarked upon a rediscovery and redesign."

Globalization facilitated this revision, accelerated by the internet. Arno Penzias, who shared the 1978 prize in physics while working for Bell Laboratories, was inspirational in starting the live video stream of lectures and, ultimately, the blogging and extensive social networking that characterized the meeting in 2010. At the same time, the official language was changed from German to English.

Aaron Klug, the 1982 chemistry laureate, is credited with suggesting a partnership strategy, whereby the Lindau Council establishes a network of academic and government institutions to help identify and nominate young researchers, including advanced students, doctoral candidates and postdocs. The council signs a Memorandum of Understanding (MOU) with each institution and a review panel then selects from the nominees.

"We wanted to establish credentials and assure our partners around the world," explains Schürer, "that we want to sustain this effort for decades to come." Ten years later, the Lindau Council has signed more than 200 MOUs worldwide resulting, in 2010, in a record attendance of 61 laureates and 650 young researchers from 72 countries.

Updated and recast, Lindau meetings still follow a scientific agenda that has an ancillary social context. No fewer than five Nobelists at the 2010 meeting addressed the challenges to the environment. The meeting ended with a panel on energy and sustainability. There were also discussions on the prospects for CERN and a debate about the future of medicine.

But the real focus of the meetings at Lindau are the young researchers. Edmond Fischer, who shared the 1992 Nobel Prize in Medicine or Physiology and has regularly attended since 1993, thinks it should be called a "meeting of the students", not prizewinners. "They are the principal component of the meeting. We are coming just for that, to introduce students, to make connections, speak among them, trust one another and speak the same language, the language of science."

Science is not the only beneficiary of the Lindau meetings. The Nobel laureates gain much from keeping in touch with younger scientific minds. "If you look at them when they come and when they go," asserts Bárány, "the meeting is like the fountain of youth."

John Galbraith Simmons *is a freelance writer in New York.*

FLEMMING, C./LINDAU LAUREATE MEETINGS

LIFE IN SCIENCE

Generation X-change

International meetings and exchanges are creating a universal, globe-spanning culture of science with widespread ramifications.

BY CHRISTOPHER MIMS

Not many high school students are so inspired by their first international meeting that they decide to leave behind their friends and family to get an undergraduate education overseas, but that's exactly what happened to Sophia Hsin-Jung Li from Taipei. She attended the inaugural Asian Science Camp (ASC) in her home town in 2007, and is now in her senior year at MIT.

ASC is one of only a handful of meetings held worldwide that, like the 2010 meeting of Nobel laureates and students in Lindau, Germany, are international, interdisciplinary, and focused on inspiring the next generation of leading scientists and fostering relations among young researchers. The ASC aims to promote interaction between promising science students from across Asia, both with each other and with the senior scientists who attend, including a dozen or so Nobel laureates. So far, it has been held in Taiwan, Japan, Bali and, in August 2010 India hosted the camp.

"At Asian Science Camp you speak English and talk about science with different people, and that was very exciting and inspiring for me at that point," says Li. "That kind of excitement, and being able to communicate with others, made me want to come to MIT."

Unlike most scientific meetings, gatherings like Lindau, ASC and others are open-ended celebrations of ideas from a diverse array of research. According to some of those who have attended such meetings, while they include formal lectures, the part of a scientific meeting that is best remembered and potentially most valuable are the countless contacts made outside lecture halls. These connections, conducted in, and made possible by, the common language of science, lead to both scientific collaboration and the kind of cultural exchange that can have surprising effects on the practice of science in labs all over the world.

Cultural exchanges can have surprising effects on the practice of science.

Like other scientific meetings, these gatherings are, in a sense, gateways to deeper cross-cultural exchanges. For some young scientists they have been the inspiration to broaden their research, the birthplace of international research partnerships, or the impetus to embark on a research career abroad.

INSPIRED COLLABORATION

For many fields, collaborations between scientists of different nationalities and across countries are so important that research breakthroughs could hardly be made without them. For example, in physics, 'big science' like CERN's Large Hadron Collider is far too expensive and complex for any one nation to accomplish on its own. Yet whether the field is inherently interdisciplinary, like systems biology, or highly specialized, such as theoretical physics, the opportunity to work together promotes not only technical exchange, but a cross-cultural fertilization of ways of conducting research.

"Wherever Italians go they are well accepted for their scientific skills and for other kind of skills," says CarloAlberto Ratti, a PhD student in theoretical physics at the University of Milan-Bicocca. "People abroad say we are funny — I don't know why," he adds with a grin.

Immediately after the Nobel Laureate Meeting at Lindau, Ratti had an opportunity to show off some of what makes Italy and its scientists unique to a group of new friends he met at Lindau. Ratti's new friends were attending a small science conference in Turin, and they took a detour to visit him in Milan.

"I don't like Milan because I think it is a very gray city," admits Ratti, "But they fell in love with it — they were astonished by the town!" There was a return visit: Ratti went to see his new friends at the John Hopkins Hospital. For a theoretical physicist, Ratti found his tour of a biology laboratory very revealing.

CarloAlberto Ratti, Jose Ramirez and Grace Wang in front of the Washington Monument.

Roy Ziblat (Israel).

JESSICAALDERMAN.COM
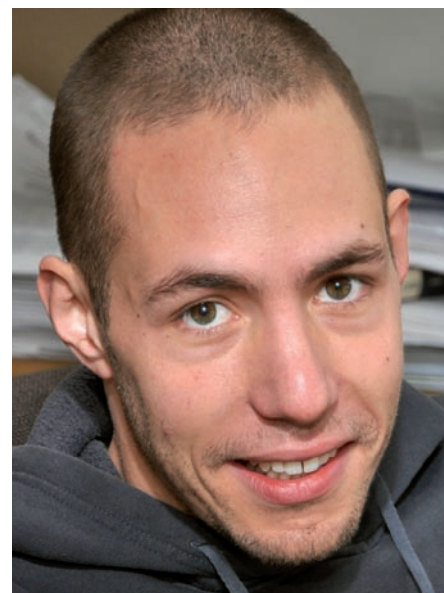
Emmanuel Unuabonah (Nigeria), Wang Likun (China) and Baybars Kuelebi (Germany).

Despite what Ratti describes as a dearth of funding for basic research relative to other countries in Europe, Italian scientists are in demand for international collaborations not only because of their skill at research and a history of innovation and learning stretching back to the Renaissance, but also simply because scientists from other countries fall in love with Italy itself.

Scientists like Ratti have leveraged the unexpected — the physical beauty of their home and the reputation of its people for warmth and an enjoyable lifestyle — to forge connections with scientists in countries with more resources.

LINDAU WRIT SMALL
The value of Lindau-like international meetings is well recognized. Today countless universities, government agencies and private, not-for-profit foundations sponsor smaller, focused versions of these meetings: they are often deliberately restricted to researchers from a less expansive geographic area, and may even be on a single subject.

Even such focused meetings can involve scientists representing a broad array of disciplines and countries, as Roy Ziblat, a Ph.D. student in structural biology at the Weizmann Institute of Science in Rehovot, Israel, found out. In March 2010, Ziblat organized the first ever International Student Workshop on Lipid Domains, which attracted 70 scientists, two-thirds of whom were from outside Israel.

The study of cell membranes requires expertise in biology, chemistry and physics . "There are researchers from all these backgrounds [working on lipid membranes], but there aren't conferences involving researchers from all these areas of study," says Ziblat.

Bringing together scientists from such diverse backgrounds — even those who are working on what is essentially the same problem — meant that Ziblat and his co-organizers must conquer a linguistic barrier that arises even when scientists are all ostensibly speaking English: jargon.

In the first two hours on the first day of the conference, attendees were split into two groups: one, mostly physicists and physical chemists, were given a lecture on lipid-membrane biology for non-biologists; the second group, mostly biologists, learned about the physics of lipid barriers. Thanks in part to Ziblat's tactics, the meeting was such a success that the Weizmann Institute will sponsor other international meetings on a variety of subjects — the next one will cover scientific archaeology.

There is a linguistic barrier that arises even when scientists are all ostensibly speaking English: jargon.

KNOWLEDGE TRANSFER OR BRAIN DRAIN?
Scientific exchanges, particularly when they are between countries with different levels of economic development, can involve more than just knowledge transfer – for better or for worse.

The Fundación Ciencia para la Vida, in Chile, has a joint programme, Science and Friendship, with the University of California, San Francisco. The explicit goal is to establish scientific partnerships, and the mechanism is an annual meeting in Santiago, Chile. This brings scientists in Chile, such as Alejandro Montenegro-Montero, a graduate student in molecular genetics at Pontificia Universidad Católica de Chile, together with US scientists on an equal footing.

These connections are worthwhile. "There are things that can be done in the US that cannot be done here in Chile," says Montenegro-Montero, because they require resources that simply are not available. But rather than being a cause for envy, these interactions are inspirational.

"When people in the lab come back from abroad, it inevitably helps the quality of the science here. We get the feeling that you can actually do great science here, even though you have a limited budget or other limitations," says Montenegro-Montero. "You have to be crafty." This, he contends, is an advantage later on, because it makes Chilean scientists more effective when they do research in labs that have ample resources.

Resources are also limited in Nigeria. Emmanuel Unuabonah is a surface chemist, a member of The Academy of Sciences for the Developing World, and a faculty member at Redeemer's University in Redemption City, near Lagos. He has attended three international meetings including Lindau, and completed the final year of his PhD in China – yet unlike many of his fellow internationally mobile Nigerian scientists, he has returned to his home country.

"[Nigeria] is an exporter of PhD potentials," says Unuabonah, "They get their PhDs abroad, and ninety-five percent don't return."

Part of the attraction in working abroad is access to modern facilities. "[The situation in Nigeria] is simply a drag on my productivity," says Unuabonah, whose current work revolves around modifying clay with inorganic salts in order to develop a simple and cheap system for wastewater treatment. "There are reactions I need to run for 72 hours at a stretch, but I have never gotten power for such a long time without it going out. It can be frustrating."

Halting such a brain drain is one of the reasons that Unuabonah helped establish the Nigerian Young Academy with the Nigerian Academy of Sciences. The Young Academy was inaugurated in August 2010, and is a first step, he says, to hosting international meetings like Lindau.

Scientists who have at least a masters degree will be eligible for funding for interdisciplinary research. But even with the added support, convincing young people to go into science will be an uphill struggle: in Nigeria, the younger generation is more interested in going into lucrative professions, such as politics, banking and finance. In Nigerian colleges, the business students outnumber science students 40 to 1. "[With this academy] we will try to reinvigorate science in the minds of young ones, as it is almost forgotten here in Nigeria," says Unuabonah.

CULTURAL MIX

In China, the Sino-German Center for Research Promotion focuses almost exclusively on international meetings. The foundation's mission, to promote collaboration between scientists in China and Germany, especially those at the beginning of their careers, is accomplished through workshops lasting 3 to 5 days to bring together up to 40 students at once, from both countries.

Wang Likun is a member of the centre and a PhD student at the Chinese Academy of Sciences. As China's economy — and consequently its scientific enterprise — grows at a breakneck pace, Likun is witnessing the changes wrought by exchanges like these.

Twenty years ago, says Likun, his advisor worked in an environment in which scientists mostly published in Chinese-language journals. "Now, as many scientists return to China from other countries, we are more willing to share our research with scientists in those countries. Many more papers are published in international journals, and some are in top journals like *Nature* and *Science*," says Likun.

Coincident with this increased openness is a growth in the number of research collaborations with scientists and groups outside China. This has led to subtle changes in the culture of science in China and across Asia, where traditional Confucian values mean that authority is accorded a great deal of respect. "In Taiwan, if you do research with a professor, it's like a boss and a worker. There is a hierarchy," says Sophia Li.

However, as Likun observes, even in mainland China there is a gradual adoption of more Western research practices. "We now have a happy hour which is held on Friday afternoon every week," says Likun. "Students and PIs [principal investigators] get together and share their experiments and talk about everything. This activity is just beginning – it started in recent years as some PIs have come back from abroad."

SCIENTIFIC ACCULTURATION

Some scientists see international exchanges as having a profound effect on their local culture of science. In China and Taiwan, scientists like Li and Likun have witnessed a change in lab hierarchies. In Turkey, where Baybars Kuelebi

Inna Pertsovskaya (Spain).

pursued his undergraduate degree, research has changed from being supported entirely by a university to something more "entrepreneurial," as it is in the US and his current home, Germany.

"The research institute is becoming really strong in Germany, which is also partly parallel to the US," says Kuelebi, who is a student at the Heidelberg Graduate School of Fundamental Physics. "The university idea of research is decreasing... This old German concept is completely going away." There is a risk that international exchange will lead to a homogenization of science. Yet, Inna Pertsovskaya, born and educated in Russia but now studying for a PhD at the University of Barcelona as a Marie Curie fellow, has found in her experience that the diverse regional cultures of science to be stubbornly resistant to being overwhelmed by outside influences.

"In Taiwan, if you do research for a professor, it's like a boss and a worker."

"Wherever you are, whether in Spain or Germany, their ways of doing research are different," says Pertsovskaya. "It's kind of a good thing — [the Marie Curie foundation] likes you to move because you get more experience from several different labs."

In countries like Germany, home to the Lindau meetings, many universities have redoubled efforts to attract international students, whose growing diversity contrasts with the profile of professors, who are almost exclusively German.

"Historically, Europe is different from the American type of scientific system. You can see that by the greater number of foreign students with respect to German students," says Kuelebi.

Kuelebi is in the first wave of international students in Heidelberg. By day, he studies stellar astrophysics, but his other passion is communicating about science on the web. For him,

international exchanges are an opportunity to connect with other scientists who have a similar interest in communicating with the broader public. The web provides a way for him to stay connected with the people he meets, and to collaborate on science communication via the same medium that binds them together.

"These kinds of web things will be very important for the future, to encourage public interest and get people to start participating in science," he says. "I think it can lead to qualitative changes in science and the perception of science."

FUTURE LINDAUS

Participants in the Lindau Meetings have found them so rewarding that some have already succeeded in setting in motion meetings that are unabashed copies of it.

Raj K. Sharma, senior scientific officer for the Indian government, describes the meeting his colleagues are working on as a "reverse-Lindau programme". The name alludes to the fact that, from the perspective of scientists in India, the meeting will be just like Lindau except it will bring the world to India — rather than require that scientists from India travel to Lindau.

A part of the government's INSPIRE programme, intended to excite India's youth about science, this reverse-Lindau would invite students and researchers from all over the world and focus on building collaborations between these scientists and the rapidly expanding pool of researchers in India. Details have yet to be finalized, but the simple act of describing it in relation to the Lindau meeting is a measure of its influence on the thinking of governmental scientific bodies in the developing world.

A GLOBAL COMMONS

"When students leave India and come back, I think it actually enhances the level of the research that's done in India," says Moupiya Maji, a masters student at the Indian Institute of Science in Bangalore. The transformation is as much cultural as it is scientific, says Maji: by broadening their horizons, scientists in India become better educators of the next generation.

Yet science is also a culture unto itself, and in its protected enclave, conventions and ideas travel far and fast. Increasingly, one could consider cultural exchange as a requisite for success in science — maybe with a dose of a wanderlust.

"I think I could explore the world by doing science. I can go to seminars around the world and talk about science and see their culture and I think that's very exciting," says Li. "We have a shared language called science."

Christopher Mims *is a freelance writer in Gainesville, Florida*

⟳ **NATURE.COM**
watch videos of Lindau
go.nature.com/c4MmzJ

TECHNOLOGY

Science in the digital age

The goals of science have not changed since the early days of the Lindau meeting, yet the way they are pursued has.

BY NED STAFFORD

The year was 1947—no high-powered personal computers and no internet for seemingly magical keyword searches of massive databanks. So when Oliver Smithies, a 22-year-old science student at Oxford University, was assigned an essay on the Pauli exclusion principle, he did what students and scientists in search of knowledge had been doing for centuries.

He went to the library to sift through scientific journals and books. His keyword search for pertinent information consisted of scanning page after page with his eyes. When Smithies, who 60 years later won the 2007 Nobel Prize in Physiology or Medicine, finally finished researching his essay, he had the pleasure of writing it by hand.

"I think the biggest single change between then and now is the huge amount of information that is available," says Smithies, born in the UK but now a US citizen, still involved in medical research at the University of North Carolina at Chapel Hill.

Indeed, the digital revolution has transformed the creation, storage and transfer of knowledge (see Xue-fu-wu-che). The ultimate goals of science remain unaltered in this new world, but the paths that scientists traverse to reach them have changed, triggering a further cascade of new developments.

"This could not have been imagined twenty years ago," says Liu Dun, a science historian at the Institute for the History of Natural Science, Chinese Academy of Sciences, in Beijing. "With just an internet terminal, scientists today are able to access almost all important resources of literature in different languages and from different ages, countries and fields."

Scientists now are able to bypass traditional publication routes by quickly and easily

JANUSZ KAPUSTA

posting academic papers, scientific data and personal blogs on the internet. Cheats have never had it so easy, able to copy and paste others' electronic information with little more than a mouse-click and claim it as their own. There is an ever-expanding array of high-tech — and often high-cost — laboratory tools for pursuing research. It is commonplace to have to jostle not only for smaller slices of the funding pie, but also for tenured academic positions. And scientists are now expected to be constantly available to colleagues around the world via email, mobile phones, Twitter, or other fashionable means.

As Liu Dun, who is also president of the Division of History of Science and Technology, part of the International Union of History and Philosophy of Science, notes, this new digitally influenced world of science has rapidly evolved in less than a generation. The current crop of university students and young researchers know only a world of science that, to most of their current professors and lab leaders, would have seemed like science fiction 20 years ago.

PAPER EQUALITY

Publish or perish, remains a maxim for today's youngest generation of scientists as it did for previous generations. Scientists around the world are still eager – for professional ambition and personal pride — to have their papers published. As Smithies says, "It isn't science until it is published."

Until recent years, there was one primary route to publication: submission to peer-reviewed journals. Usually several months to a year later the paper would be published in the journal, and then posted to personal subscribers and libraries, allowing other scientists to finally read it.

But not all scientists in the traditional publishing system were created equal. David Kaiser, a physicist and faculty member of MIT's Program in Science, Technology, and Society, says that from around the 1930s until recent times, some scientists tried to speed up the process by sharing pre-prints of their papers, but these were usually restricted to a select group of insiders. "In order to get the inside scoop on an upcoming paper, to see a pre-print, you had to be on the A-list of top scientists, something like the old-boy network," says Kaiser. "That is quite different today. Publishing is much more open now, and new resources like the internet allow scientists to bypass the previous exclusivity."

For a successful example of open-access publishing, Kaiser points to the arXiv database administered by Cornell University in New York. "You can post papers on arXiv immediately," he says, adding that this includes posting papers before they are peer-reviewed.

The obvious benefit of open-access online publishing is to increase the speed of sharing ideas with other interested scientists. Kaiser sees potential benefits in circulating ideas

XUE-FU-WU-CHE

In ancient China, before the invention of paper, books were made of wood or bamboo scrolls — cumbersome, especially for wandering scholars of the time. Back then, says Liu Dun of the Chinese Academy of Sciences, a learned man was called xue-fu-wu-che — literally 'a possessor of five cartloads of knowledge'. He adds, "By today's standard, however, books on bamboo scrolls filling five cartloads may be contained in a memory chip less than one gigabyte."

before they have gone through rigorous peer-review. The peer review system, in his view, is clearly important and a valuable step in the formulation and communication of ideas, but it can sometimes weaken the impact of original, ground-breaking ideas. "A raw idea can really be much more inspiring than the cleaned up peer-reviewed idea presented a year later," he says.

Downsides to the glut of material being publishing online are that this material might be low quality and scientists risk drowning in the continuous gush of data, Kaiser says, add-

"A raw idea can really be much more inspiring than the cleaned up peer-reviewed idea presented a year later."

ing: "You sort of get a feeling that you are running a marathon that never ends. I think one can make a very strong case we are all suffering from that."

Jane Maienschein, director of the Center for Biology and Society at Arizona State University, says: "It's easy to get access to abstracts and public information, but much harder to actually read the whole thing. I know scientists who read the open access abstracts and assume they know enough."

Cameron Neylon, a biochemist and senior scientist at the Science and Technology Facilities Council, a funding agency in the UK, agrees that scientists are being buried under millions of gigabytes of information, some of dubious quality. But Neylon believes the boon to science from this deluge of information far outweigh the disadvantages.

"Look at all the general information on the web," he says. "But the web still works because you can go to Google or other search engines to filter information to find what you are looking for. We need to do that for science."

Neylon wants to go beyond open access publication of academic papers. He advocates a concept dubbed 'Open Notebook Science', in which scientists post the full primary record of research projects — their laboratory notebooks — on the internet for public scrutiny. "The core of the concept is simple," he says. "You keep your primary research record openly available

on the web." (See Open notebook science).

OPEN TO TEMPTATION

New communication technologies bring opportunity, but they also come with drawbacks. The exchange of ideas, feedback, advice, critique and encouragement among scientists has always been key to moving science forward. Until a couple of decades ago, long-distance personal communication between scientists was essentially limited to letter writing and land-line telephones. These have been supplanted by the instantaneous mediums of email, mobile telephones and the internet.

There can be no doubt that the ease of email is helping to progress science. As Smithies notes, it "improves the transmission of knowledge so improves science". But, he opines: "I personally don't care much for emails. I prefer talking. I find I have to email a person to get them to pick up the phone."

Modern technologies cannot replace personal contact. "People are connected through quick-fix email," says Maienschein. "But this does not lend itself to thoughtful or deeply reflective exchanges. This is why we still need in-person conferences. Even video chats don't do the whole job with larger groups of people."

There are more nefarious consequences of these new technologies too. Although plagiarism is nearly as old as publishing itself, the difference is that unethical students and scientists now have handy tools for finding information and claiming it as their own. However these same tools can be turned against the cheats, according to Harold Skip Garner, head of the Virginia Bioinformatics Institute at Virginia Tech, who helped develop eTBLAST, the institute's free web-based search engine that can be used to catch plagiarists.

"In the recent past, that is only two to three years ago, most plagiarism discoveries were serendipitous," says Garner, who has studied plagiarism by searching millions of online papers to reveal that thousands have large similarities with other articles. "But now they are being discovered en masse using computer techniques."

A simple keyword search, such as "plagiarism checkers", on the internet will yield a long list of text-based search systems, some free and some as commercial products. Garner says that most currently available software is limited to checking newly submitted manuscripts for possible plagiarized content. However, some can scan several databases, such as Medline, CRISP (from the National Institutes of Health) and IOPscience (from the Institute of Physics), and identify highly similar articles already in print.

"It will soon become impossible to evade detection as more and more publications are available electronically," Garner predicts. And he warns that past cheats not yet caught should be concerned: "Those who cheat or

OPEN NOTEBOOK SCIENCE

A small but growing number of scientists practice Open-Notebook Science (ONS), a concept whereby researchers post their laboratory notebooks on the internet for public scrutiny. The term was coined in 2006 by Jean-Claude Bradley, a chemist at Drexel University in Philadelphia, who now defines ONS as "the practice of making the entirety of one's laboratory notebook and all associated raw data public in as close to real time as possible". He adds that this should include information not usually disclosed, such as results from failed or ambiguous experiments. "For the open researcher, the greatest benefits are probably enhanced visibility and the ease of identifying new collaborators."

Open-notebook advocate Cameron Neylon says that the most important benefit would be to move science forward at a quicker pace. "The more people you have looking at a problem the better chance you have of solving it," he contends. Input from outsiders could also help reduce lab expenses. "Science is expensive and specialized," he says. "It is going to be rarer and rarer to have all the expertise needed inside your own laboratory or within your team."

There is a wide range of access options for publishing notebooks on the web, from password protection, creating different access options for different types of user, to "opening it up to the whole world", says Neylon. Software developers are working to make open notebooks easier for scientists to use and, at some point, fully automated so that data are transferred to the internet in real time.

Notebook information is time-stamped to record and attribute a new idea or result. "There will be cases of theft, but they will be easy to spot," says Neylon. "People are always going to try and play the system. But this kind of cheating becomes much harder in a transparent environment."

Ultimately, Neylon believes open-notebook users, both posters and readers, will end up in a stronger position by working together. "Funders of research projects do not really care whose names end up on papers, they just want to get [the projects] done," says Neylon. Funders should be among the biggest supporters of ONS and encourage its wider use, he adds. "I do sometimes wonder why funders don't just lose their patience and tell us to quit mucking about."

have cheated even long ago must realize that the evidence for their actions is permanently out there, so there is the lingering possibility of being found out."

FLASHY SCIENCE

Laboratories today also are undergoing major changes, with even the simplest labs outfitted with expensive equipment. Smithies recalls that early in his scientific career at Oxford University in the UK, the most modern tool in his lab was an ultracentrifuge, at that time a huge oil-powered contraption that occupied two rooms. The latest models, in comparison, are compact desktop units. Back then, he says, "a lot of experiments could be done with relatively simple equipment." How does this affect the science? "High-tech lab equipment makes it easier to get more information, but does not necessarily improve science unless you do something with it. Science is what you do with the information," he says.

What it has done is infuse science with a sense of 'keeping up with the Joneses'. Maienschein observes that "there is great pressure to use the latest and greatest equipment, which may not actually be necessary for some questions and approaches. This may distort some of the science that gets – or doesn't get – done".

Outfitting modern labs with expensive tools requires significant portions of scarce resources, yet little thought is given to the rapidly growing mountains of scientific information. "Government grants don't include support for the infrastructure to collect, archive, and then use combinations of data," Maienschein says. "They encourage us to rush ahead collecting more data instead. We really need a lot more serious discussion about how best to develop and share metadata and ontology standards so as to make more databases inter-operable and inter-searchable."

Financial issues affect more than just purchase of equipment and database maintenance. Kaiser notes that although total funding for research continues to grow, it is not keeping pace with demand, making it more difficult for young students to pursue careers in research. "The pool of academic jobs is growing slowly while the pool of applicants is quickly growing larger," he says.

Kaiser, who studies the development of physics in the US after World War II, says the boom years for plentiful science jobs with good salaries was in the 1950s and 1960s. "A PhD then was a ticket to a comfortable life," he says. But now after earning a PhD, many young scientists must spend up to a decade in postdoctoral limbo subsisting on fellowships or other temporary jobs before getting a shot at a tenured position. "That does not seem to me to be a sustainable model," Kaiser says. "It is close to a kind of an indentured servitude. If science is not viable for the smartest, best people, then that is a problem."

Nevertheless, there is an allure to science that transcends salary or funding levels. "There are enormous rewards to doing every day what you wanted to do as a kid," says Kaiser, who also conducts research of early universe cosmology. "Every single day we get to ask big questions, learn new stuff. I get to spend time designing universes in my mind. You can't put a price on that. It is marvelous."

BERGQUIST, J./UPPSALA UNIVERSITY

Original ultracentrifuge equipment at the Svedberg laboratory, Uppsala University. The optical detectors (foreground), black electric switchboard (left) and the blue turbines and oil-pumps (background) are only part of the instrument — the rest is in the room below.

Ned Stafford *is a freelance writer in Hamburg, Germany*

AKSHAT RATHI

Turning the tables

What happens when the spotlight shines on the young scientists?

BY AKSHAT RATHI

The annual Lindau Nobel Laureate Meeting is the opportunity for a select few hundred young scientists from around the world to mingle with some of today's most celebrated scientists. The main goal of the meeting is to give students an opportunity to interact with their scientific idols, quiz them about their lives and learn about research from a wider perspective.

In addition to lectures given by the Nobel laureates, there are private, closed-door sessions and many informal, social events. Yet, until now, there has been one form of interaction that did not get adequate attention: a chance for the laureates to learn more about the students and their views of science. Turning the Tables aims to change that.

In Turning the Tables, the 'expert panel' comprised students representing all branches of science and many parts of the world. The 'audience' consisted of five Nobel laureates, carefully chosen to reflect the broadest and

most enquiring minds — able to put the panel through its paces.

This was not an easy event to chair. As the moderator, Adam Rutherford of Nature Publishing Group, addressed the laureates: "My job is to stop you guys from talking to each other and make sure that you ask the panel penetrating questions so that you can learn more about them."

So what questions did the Nobelists ask?

Advice to laureates

- Reply to emails from students within twelve hours
- Don't dictate a student's life
- Give creative freedom
- Foster relationship among students in the lab, not just with them
- Let students develop their 'voice' when writing papers
- Communicate your science to the public by using the media

And how did the young scientists respond? Here is a snapshot of the session.

LABORATORY ETIQUETTE

Hinting, perhaps, at the no-holds-barred tone of this session, the first topic proposed by Rutherford concerned the controversial comments of a prominent scientist who had recently described PhD students working in the laboratory as "serfs". Was this accurate or, indeed, proper?

Most of the panellists disputed the statement, but there was one supporter. "I don't think I have any illusions. I am proud to be a serf. I am proud to be a labourer of science," said Baybers Kuelebi, a Turkish student finishing his PhD in astrophysics at Heidelberg in Germany. "In my case, my supervisor is [the one] delegating work and I do the research, write the codes and write the papers. In that respect, from the labour point of view, we are serfs."

Evans Boney, who is pursuing his PhD in theoretical chemistry at the California Institute of Technology under the supervision of Nobel

laureate Rudolph Marcus, acknowledged some truth to the statement but said personal initiative could make a difference. He makes himself "rise to the occasion" whenever he meets with Marcus, so that he does not "sound like a student".

RELATIONSHIP WITH SUPERVISOR

A lot depends on the relationship that a supervisor fosters with their graduate students. "What kind of relationship should a supervisor have with his students?" asked Rutherford. Benyam Kinde, a first year student in the Harvard-MIT MD-PhD programme, said he has enjoyed a "collegial" relationship with his supervisor, even when he was an undergraduate, and that has allowed him to be creative.

This relationship, however, should be close but not too close – to allow the individuality of the student to develop, said Harold Kroto, a winner of the 1996 Nobel Prize in Chemistry.

In addition, graduate students, whether serfs or recognised as equals, all need quality time with their supervisors. Yet Nobel laureates who supervise students travel widely to visit collaborators and give talks around the world, leaving little time for students. Boney admitted he has to deal with the travelling issue and had not been prepared for it. His advice was to plan ahead. "As long as I ask in advance, like a couple of months, it is fine", he added.

George Smoot, who shared the Nobel Prize in Physics in 2006, admitted that he travels a lot and occasionally supervises students "who want their hand held". But he felt that to best prepare students to become independent researchers they should try and do as much as they can themselves to get to the point where "they don't know why you're their mentor."

SCIENCE WORLDWIDE

Smoot observed that most of the students on the panel were pursuing their scientific career in a foreign country, and asked why.

The panel explained that this sort of relocation had many reasons: a lack of facilities in their home country; an increased need for specialisation in science; or a need to do multiple postdoctoral appointments to secure an academic position.

Inna Pertsovskaya was a full-time teacher in a Russian state school while she was doing her PhD. It was a commitment of nine hours per day, three days a week. "I did not have patience for those teenagers as well as take care of my daughter in the evenings," she admitted. The situation has improved since she moved to Spain to continue her studies, and she is now a PhD student in neuroimmunology at the University of Barcelona.

Aaron Ciechanover, who shared the Nobel Prize in Chemistry 2004, wondered whether these foreign pursuits were causing a brain drain in their respective home countries, for example in Turkey, in the case of Kulebi, and

Russia, for Pertsovskaya.

Kulebi agreed that this wandering can become permanent. "After some time we become stateless and are ready to move anywhere to do what we really want to do, but then you cannot really settle."

IDEALISM AND CYNICISM

The most contentious issue for most of the panel did not concern their day-to-day work, but its wider context. "What are you guys most concerned about right now?" asked Smoot. Kinde replied: "My biggest concern is outside of science. It's the people who have a lot of influence on science: the politicians. There should be more politicians who have been scientists before." Kroto responded: "There is a conflict between the scientific way of thinking and politics — which is dogma. Science is not [dogmatic] and that is why I feel that there are not as many really good scientists in politics."

Smoot distilled the issue, "We see two sides here: the idealism of youth and the cynicism of the old. I think that idealism is great but

Advice to students

- Choose a supervisor who does not travel too much
- Don't try to please your supervisor all the time, be prepared to challenge them
- Put questions to your supervisor, but think of some possible suggestions beforehand
- Assume your supervisor is wrong and develop your own way to approach the problem
- Idealism regarding science in politics is good, but be aware that it will be a steep challenge
- Don't give up too easily

in the real world it is a bit of mixture… If the students can persuade the people then eventually they [the people] will make the right decision." He urged the students not to be cynical: "You do have the opportunity. You can have your own blog on the internet. You can express your views clearly and some blogs are very influential."

Winning a Nobel prize provides perhaps the best chance for a scientist to influence politics and cultural perception of science. Should that not provide optimism? Mather agreed: "I think the interaction of science and politics is really important and I use a part of my Nobel prize money to give summer internships to physics students to work with politicians in Washington DC."

With the credibility and authority of the scientific community arguably having taken a hit after 'climategate', the Nobelists turned to the students for help. "Specifically in the case of climate change, we see a massive political backlash because they [politicians] do not understand how science works," said Kroto. "That's why it is important that students like you get involved [in science communication] as much as you can."

Yet the panel were sceptical that these messages would be heard: in many countries the current education system does not allow children to learn the skills of critical thinking. "There is a declining population of high-school science graduates in the United States," Boney bemoaned. "Thus, there is a decline in the population that will understand the science we try to teach. It is hard to work with such a basic disconnect."

LESSONS LEARNED

Even with years of experience of supervising graduate students, in the brief Turning the Tables session, the Nobel laureates still learnt new things about how young scientists think and about what is important to them. And the students were not shy about offering advice (see Advice to laureates).

The laureates understood the students' positions. As Smoot said: "People [young scientists] not only want to know how things work but they also want their work to be important — to go out and help the world."

However, the laureates were not without their own advice (see Advice to students). They appreciate that today's young scientists need to work harder to establish themselves than they had to work in their times. Ciechanover concluded, "We shouldn't burden them too much. We heard they have to deal with a lot of basic problems. [But despite these problems] it is important not be cynical and be aware of the many interfaces between science and society."

Akshat Rathi *is a DPhil student at the University of Oxford and blogger for the Lindau meetings*