THIS WEEK

EDITORIALS

WORLD VIEW Scientific advisers must know their political place **p.559**

CANCER Tumour growth stalled by tweak to sugar metabolism **p.560**

GNAW NEED Missing molars make worm-eating shrew-rat unique **p.561**

Open for business

If Europe is to achieve the science-investment goals it set for the decade, it must make life easier for researchers coming from abroad.

Europe says that it needs a million more researchers to achieve its aim of boosting research spending to 3% of its gross domestic product by 2020. The continent knows it must make entry easier for foreign scientists, and last week finished taking suggestions on how to do so. The European Commission formally closed a consultation exercise on how to reform its scientific visa, introduced in 2005.

This visa allows researchers to obtain a residence permit without also having to apply for a work permit. Researchers can take advantage of this streamlined process if they sign a 'hosting agreement' with an approved institution. In exchange, institutions vouch that the researchers have the financial means to sustain themselves and sufficient skills for the job. Immigration officials grant researchers a residence permit for at least one year. (The scientific-visa regulations do not apply to the United Kingdom or Denmark, which chose to implement their own rules.)

The scientific visa was a step in the right direction but there is much room for improvement. In 2010, just short of 7,000 researchers, most of them from India, China, the United States and Japan, entered Europe on the scientific visa — a far cry from the continent's 2020 goal.

There are some simple fixes the commission could make. Just obtaining a visa is time-consuming and expensive. The application process often takes more than a month and can take more than three months, according to an assessment published by the commission at the end of last year. This delay has caused top labs, such as the European Molecular Biology Laboratory in Heidelberg, Germany, to miss out on first-choice candidates, who accepted jobs elsewhere during the wait. The Initiative for Science in Europe, a science advocacy group based in Heidelberg, has suggested a sensible fix: research organizations should be able to file the visa applications on behalf of the researchers they wish to hire. These organizations often have in-house legal and administration expertise, and are more likely to be trusted by immigration officials than an individual researcher, all of which will speed up the process.

It costs, on average, €250 (US$314) to apply for a visa — prohibitive for some researchers, especially those from developing countries. And there is large variation in fees. Lower and comparable charges would make the scheme more attractive.

Few professions require individuals to uproot and move as much as science, and the benefits of such freedom can quickly wear off in the face of the mundane realities of finding digs, opening and closing bank accounts, and even just registering for electricity and water time and again. Worse, under the current scientific-visa scheme, researchers are forced to organize their next move at the same stressful time as finishing their existing research project — because their visa lasts only as long as the hosting agreement. The European Union (EU) should grant a grace period of one month, at the end of a research project, say — already done for the J-visa in the United States — to give some breathing space.

Breaking down barriers to entry and movement through the EU is crucial to allow the continent to compete globally for talent and create, by 2014, the long-awaited European Research Area — whereby scientists can collaborate seamlessly across national boarders. The EU is on the right lines, it just needs to give a little more in return. ■

Small steps

Violent opposition to nanotechnology should be countered with public awareness.

In the past two years, Mexican nanotechnology researchers have been subject to a spate of bombings and bomb threats. In the worst of the attacks, two researchers were injured. Police say that if the explosive had gone off properly, a whole building could have collapsed.

Meanwhile, Italian, Swiss and German authorities this summer arrested members of related groups whom they think were responsible for trying to bomb IBM's European flagship nanotechnology lab, and for shooting in the kneecap a nuclear engineer at a firm engaged in nanotechnology and biotechnology research. France has also seen angry protests and attempts to shut down public debates on nanotechnology.

Some policy-makers in Europe and elsewhere have long feared that research on nanotechnology could spark a public backlash — similar to those seen against genetically modified (GM) crops and animal experimentation. Has it arrived — and in a violent fashion?

As the Feature on page 576 highlights, public awareness of nanotechnology remains low and fears of widespread opposition are premature. But how it may pan out is hard to predict. Sympathy for animal-rights violence was always thin on the ground even though opposition to vivisection was quite broad: conversely, few members of the public have ever participated in anti-GM vandalism, yet there is a de facto European moratorium that has all but frozen the industry on the continent.

The more outlandish claims made for nanotechnology stir fear among the public. Opponents know this, so whereas scientists and officials want to talk about environmental and exposure risks, consumer awareness and product regulation, the extremists and some mainstream non-governmental organizations focus on nanometre-scale sensors, cyborgs and swarms of self-replicating robots.

Nanotechnology advocates have an important role here, and one that could help to determine how public awareness of nanotechnology develops. They should continue to work to make public debate informed and accurate, and do more to monitor and test the possible toxicity of novel products. And they should avoid hype. If they paint a true picture of the state of the science, then the distorted version drawn by the extremists will have a greater chance of being recognized as such. ■

What matters for science is who runs the country

A chief scientific adviser is no substitute for a ruling elite that is actually engaged with science and engineering, argues **Colin Macilwain**.

Britain loves its scientific advisers. Almost every government department has one and, earlier this month, Southampton city council became the first local authority to appoint its own: AbuBakr Bahaj, head of the energy and climate-change division at the University of Southampton.

The UK government itself has a new chief scientific adviser in waiting: Mark Walport, director of the Wellcome Trust. Appointed in June, Walport will be the 11th man to fill the role when he takes the post next April. He is a scientific heavyweight whose selection has thrilled Britain's senior scientists.

Walport's appointment is unusual, and heartening. He is voluntarily relinquishing a position that is not only hugely influential — the Wellcome Trust is the biggest biomedical research charity in the world — but is also far better paid than his new role. He is clearly satisfied that his voice will be heard in the corridors of power.

But will it? And what sort of advice does the UK government want, or indeed need, from a chief scientific adviser?

Researchers look to whoever fills the role to strengthen their voice in government and, ultimately, improve their funding prospects. The government, in turn, looks to the scientific adviser to enhance the credibility of its policies.

Scientific advisers in Britain — and the United States, which also has one — have, to an extent, been successful in these duties. But they have failed in a broader task: the mission originally envisaged for them.

The position of scientific adviser wasn't set up to secure science budgets or communicate government policies to the public. Instead, advisers were meant to address competitiveness by bridging the great divide between what UK physicist C. P. Snow called the "two cultures": scientists and engineers on the one hand, and the non-technical elites who govern London and Washington DC, on the other.

It was the launch of Sputnik that led US President Dwight Eisenhower to appoint James Killian as his country's first scientific adviser, in 1957. Seven years later, Harold Wilson was elected UK prime minister after pledging that a new Britain would be "forged in the white heat" of scientific and technological revolution. He appointed his first scientific adviser, Solly Zuckerman, in the same year.

Neither Eisenhower nor Wilson hired a scientific adviser so that their countries' researchers could win more Nobel prizes or publish more papers. What they meant by 'science' was military and industrial competitiveness achieved by harnessing science and technology. What they coveted was Soviet rocketry and German machine tools, not papers in *Nature*.

THE PUBLIC PRESTIGE OF SCIENCE IS HIGHER THAN EVER BUT IT REMAINS REMOVED FROM THE CENTRES OF POWER.

Unfortunately, in both countries, the scientific adviser's role has evolved in ways that marginalize its impact on competitiveness. In the United States, most advisers have little direct access to the president, and busy themselves mainly with interagency mediation on important but arcane matters such as how to coordinate different satellite systems. In the United Kingdom, the adviser is a bigger fish in a smaller pond — but is still seen primarily as a representative of the special interests of academic science, not as a player in economic or industrial policy.

Meanwhile, Germany, Japan, China and France have found no need for a 'chief scientific adviser' — despite frequent cajoling from the English-speaking world. Scientists and engineers already permeate these countries' governing elites.

Contemplating some maddeningly immaculate railway lines in Berlin this summer, I wondered whether this is why the Germans can not only make the trains run on time, but also keep the whole system to a standard that shames the dilapidated, rubble-strewn routes of London or New York.

It seems that whoever runs the railway lines in Germany knows and very much cares about how a railway works — as does their boss. Back in Britain, it is widely accepted that those in charge neither know nor care about how things really work.

Scientific advisers in London and Washington are now part of a political world dominated by 'special advisers', who rarely have any background in science or engineering. In both capitals, the world of finance holds almost limitless sway.

Two years ago, John Beddington, the population biologist who will continue to serve as Britain's chief scientific adviser until Walport takes over, organized a meeting of 200 civil servants with backgrounds in science or engineering, from different government departments, in an effort to forge better links between them.

The exercise brought to mind the fact that no one will ever see fit to organize a meeting of the senior officials who, like UK Prime Minister David Cameron, read philosophy, politics and economics at the University of Oxford. They run the country already — and their networking skills are impeccable.

My point is not that Beddington's efforts were futile; it is rather that in 50 years of trying, the underlying dynamic of London's ruling elites hasn't shifted an inch. The public prestige of science is higher than ever but it remains disturbingly removed from the centres of power.

And under these circumstances, the scientific community's hope that the scientific adviser will exercise meaningful influence is liable always to be frustrated. ■

NATURE.COM
Discuss this article online at:
go.nature.com/f1buhv

Colin Macilwain *writes about science policy from Edinburgh, UK.*
e-mail: cfmworldview@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CANCER

Antitumour metabolism

Changing the way that tumours metabolize sugar can stall their growth.

When cells convert glucose into energy and molecular building blocks, the last step in the metabolic pathway is catalysed by an enzyme called pyruvate kinase. Dividing cells, such as cancer cells, contain a form of this enzyme called PKM2. Matthew Vander Heiden at the Massachusetts Institute of Technology in Cambridge and his colleagues found that small-molecule activators of PKM2 impaired tumour formation in mice bearing human cancer cells. Tumours developed later and remained smaller in treated mice than in untreated controls.

The activators bind to PKM2 at a previously unknown site. The authors suggest that, in so doing, the activators alter tumour metabolism in such a way as to deplete growth-promoting compounds produced earlier in the pathway.

Nature Chem. Biol. **http://dx.doi.org/10.1038/nchembio.1060 (2012)**

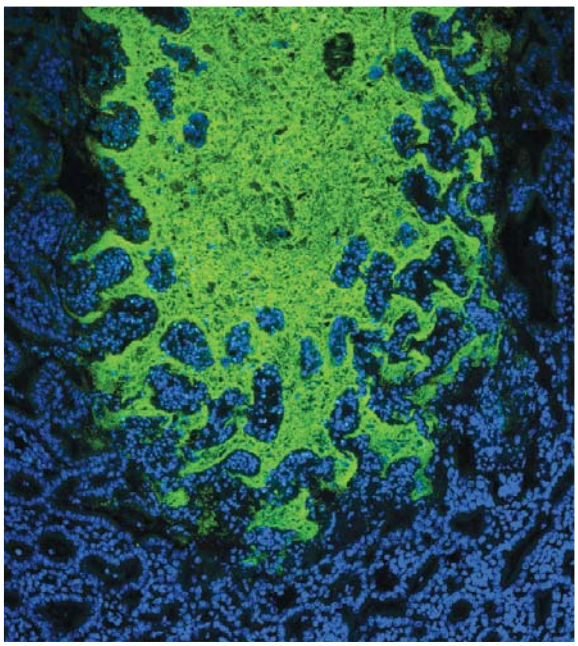
NANOTECHNOLOGY

Strong and stable nanocrystals

A theoretical method could aid in the design of nanocrystalline metal alloys that retain their unusual properties, such as remarkable strength, at high temperatures.

Nanostructured metals, which are made up of individual metal grains less than 100 nanometres in size, are often unstable at high temperatures owing to instabilities arising from the boundaries between grains. Christopher Schuh and his colleagues at the Massachusetts Institute of Technology in Cambridge developed a theoretical framework for predicting the energy associated with the grain boundaries. The researchers designed and produced an alloy of titanium and tungsten that turned out to be more stable at high temperatures than pure nanocrystalline tungsten, and that could be processed in bulk.

Science 337, **951–954 (2012)**



M. MOLLOY/L. KOO/NIAID BIOLOGICAL IMAGING FACILITY

MICROBIOLOGY

Infection breaks truce

Gastrointestinal infections can cause the immune system to attack some of the billions of microbes that are normally tolerated in the mammalian gut.

Yasmine Belkaid at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, and her team demonstrated this immune response when they infected mice with the parasite *Toxoplasma gondii*. The infection caused the bacteria normally resident in the gut to breach the gut wall (pictured), activating the animals' immune systems. This triggered immune cells called T cells to target the gut bacteria. The T cells then developed into long-lasting memory T cells — providing an immunological record of the response.

As a result, a bout of food poisoning could have enduring consequences for the immune system, and could even contribute to the development of inflammatory bowel disease, the researchers suggest.

Science **http://dx.doi.org/10.1126/science.1220961 (2012)**

NEURODEVELOPMENT

Low-flow blood-vessel pruning

The zebrafish brain's complex network of blood vessels develops not only through vessel growth, but also as a result of pruning, which is driven by changes in blood flow.

Jiu-lin Du at the Chinese Academy of Sciences in Shanghai and his team followed the development of transgenic zebrafish larvae in which cells of the blood vessels, blood and brain carry fluorescent tags. Over the course of six days after fertilization, the researchers imaged the creatures' developing midbrain. They found that, as the network of blood vessels expands, pruning occurs — mainly at loop-shaped segments — when cells lining the vessels migrate towards adjacent unpruned segments, resulting in decreased vessel-network complexity. Pruned segments have a lower and more variable blood flow than unpruned ones, and experimental blockade of blood flow triggers the pruning process.

PLoS Biol. 10, **e1001374 (2012)**

EPIGENETICS

Clues from baby blood spots

Blood from the pricked heels of newborn babies can provide insight into the epigenome — chemical changes to DNA that influence gene expression — before diseases set in.

Most babies born in developed countries have drops of their blood blotted onto a slip of paper called a Guthrie card (**pictured**), which is used to screen for certain diseases. Vardhman Rakyan and David Leslie at

SCIENTIFICA/ VISUALS UNLIMITED, INC./SPL

the Queen Mary University of London and their colleagues analysed DNA from the Guthrie cards of three healthy children born in 2000, looking for differences in patterns of DNA methylation, an epigenetic change that involves the addition of methyl groups to DNA. The researchers found about a dozen differences that were still present in samples taken three years later and did not seem to be linked to underlying genetic variation. This suggests that the changes might be the result of differences in conditions encountered in the womb.

Guthrie cards could help researchers to distinguish between epigenetic changes that are driving a disease and those that result from it.

Genome Res. **http://dx.doi.org/10.1101/gr.134304.111 (2012)**

For a longer story on this research, see http://go.nature.com/fgmz6w

MEDICAL DEVICES

Smart way to seal cuts

A 'smart' suture that doubles as a heater and a thermometer could promote wound healing.

J. ROGERS/UNIV. ILLINOIS

John Rogers of the University of Illinois at Urbana-Champaign and his colleagues have developed a prototype silicon-based suture (**pictured**) that is long, thin and flexible enough to be sewn into a wound. The suture contains circuit elements that can both heat tissue and sense its temperature. This enables the device to keep the wound at an ideal temperature for healing and to detect temperature increases associated with infection.

The researchers hope to develop sutures with other functions, such as programmable drug release.

Small **http://dx.doi.org/10.1002/smll.201200933 (2012)**

PALAEONTOLOGY

Jagged jaws identify mollusc

An analysis of the mouthparts of 505-million-year-old fossils suggests that the animals could be among the world's earliest molluscs.

Martin Smith of the University of Toronto, Canada, analysed more than 300 fossils belonging to two early invertebrate groups, *Odontogriphus* and *Wiwaxia*, that have been variously classified as molluscs, annelids or a common ancestor. The fossils are from British Columbia's Burgess Shale, which provides a record of an explosion in new body forms that occurred around the Cambrian era (542 million to 488.3 million years ago).

The mouthparts of many of the fossils — rows of jagged teeth that rest on a grooved tongue — resemble those of modern-day molluscs. The teeth might have worked like a file, allowing the animals to grind soft tissues, the author suggests.

Proc. R. Soc. B **http://dx.doi.org/10.1098/rspb.2012.1577 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

ECOLOGY

Equations in papers = fewer citations

HIGHLY READ on www.pnas.org in July

Biologists have a reputation for avoiding mathematics, and a study of citation rates shows that this is not misplaced.

Tim Fawcett and Andrew Higginson at the University of Bristol, UK, examined whether the inclusion of equations affected the citation rates of papers published in 1998 in three leading ecology and evolution journals. The authors found that papers received 28% fewer citations overall for each additional equation per page of the main text; for citations in experimental papers, this rose to 35%.

The duo recommends that researchers use equations sparingly in their main article text to ensure that their ideas reach a wide audience.

Proc. Natl Acad. Sci. USA 109, **11735–11739 (2012)**

K. ROWE/R. SOC.

EVOLUTION

Rodent that cannot gnaw

A shrew-rat that lacks molars and has oddly shaped incisors is an evolutionary anomaly among more than 2,200 known species of rodent.

Rats, mice and their kin owe their success in part to their ability to gnaw, which enables them to access foods that other mammals cannot.

Jacob Esselstyn of McMaster University in Hamilton, Ontario, and his colleagues report that the shrew-rat *Paucidentomys vermidax* (**pictured**) has a long, pointed face, incisors that differ in shape from those of other rodents, and no molars. These features could be adaptations that help the animal to find its favoured food — soft-bodied earthworms that do not need chewing. The loss of a previously successful evolutionary modification has allowed *P. vermidax* to take advantage of the available food in its forest home.

Biol. Lett. **http://dx.doi.org/10.1098/rsbl.2012.0574 (2012)**

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Scientific integrity

The US defence department has joined President Barack Obama's drive on scientific integrity, introducing a policy that bans its employees from asking scientists or engineers to alter or suppress their professional findings. Details of the policy, dated 26 July, were obtained under the US Freedom of Information Act by the watchdog group Public Employees for Environmental Responsibility (PEER) in Washington DC, and were released by PEER on 21 August. Twenty-two US government agencies have now released draft or final policies on scientific integrity. Four agencies have yet to release anything publicly, including the White House Office of Science and Technology Policy, which is responsible for leading the initiative.

India's nuclear gap

A federal auditing agency has sharply criticized India's nuclear-regulation system, in a report made public on 22 August which may bolster anti-nuclear campaigns in the country. The current regulator — the Atomic Energy Regulatory Board — is merely a subordinate office of central government, with no powers to make rules, enforce compliance or impose effective penalties for safety violations, the report from the Comptroller and Auditor General said. At a time when India is hoping to expand its nuclear-power system, it needs a truly independent regulator, the report urged. See go.nature.com/d4guvd for more.

Stem-cell funding

The US National Institutes of Health is legally allowed to fund research using human embryonic stem cells, a senior US appeals court ruled on 24 August. Two scientists who work on adult stem cells, James Sherley of the Boston Biomedical Research Institute in Watertown, Massachusetts, and Theresa Deisher, chief executive of AVM Biotechnology in Seattle, Washington, had sought to block US government funding for such research and appealed after their lawsuit was thrown out in July 2011 (see *Nature* **476,** 14–15; 2011). Lawyers for the plaintiffs said that they might ask the US Supreme Court, their last recourse, to hear the case. See go.nature.com/mvugzr for more.

NASA

Passing of a lunar pioneer

Neil Armstrong, the first man to walk on the Moon, died on 25 August, aged 82. Commemorations poured in last weekend for the modest astronaut, who largely shunned publicity — apart from his recent criticisms of US plans to cut NASA's budget. The landing achieved by Armstrong and his fellow Apollo astronauts in July 1969 (pictured, with Armstrong's shadow) inspired a generation of students to take up scientific research (see *Nature* **460,** 314–315; 2009). For more on the landings, see *Nature*'s Apollo special at nature.com/apollo.

Chimp breeding

The US National Institutes of Health (NIH) has concluded that a major chimpanzee research centre did not violate the agency's breeding moratorium, even though 130 infants were born at the centre to NIH-owned parents between 2000 and 2010 (see *Nature* **479,** 453–454; 2011). The assessment was revealed by *Nature* on 24 August after a freedom-of-information request. The New Iberia Research Center (part of the University of Louisiana at Lafayette) did not charge the NIH for the infants' care, it notes, and the centre's total census decreased in line with other NIH-supported centres. See go.nature.com/1bgbjw for more.

Circumcision boost

The American Academy of Pediatrics (AAP) said on 27 August that boys in the United States should be circumcised because health benefits outweigh the risks. The policy recommendation is a first from the influential group of physicians, and could boost US circumcision rates (which are already higher than in much of the developed world). But the choice is up to parents, the AAP added. See page 568 for more.

RESEARCH

Autism stem cells

The first clinical trial approved by the US Food and Drug Administration to use an experimental stem-cell therapy for autism began recruiting subjects on 21 August. The double-blind study, launched by the Cord Blood Registry in San Bruno, California, and the Sutter Neuroscience Institute in Sacramento, California, aims to enrol 30 children with autism aged 2–7 years. The trial will test whether an infusion of the children's umbilical-cord stem cells can improve their language ability and behavioural symptoms compared with infusions of a saline placebo.

China looks south

Chinese astronomers announced ambitious plans for the country's Antarctic observatory at Dome A — one of the coldest places in the continent — at the International Astronomical

JPL-CALTECH/NASA

Union's 28th General Assembly in Beijing last week. The observatory will include two large telescopes with a combined budget of more than 1 billion yuan (US$157 million). One, KDUST, will search for dark matter and look for Earth-like planets outside the Solar System, and the other, DATE-5, will examine the dust clouds where stars form. See go.nature.com/ftwyqn for more.

PEOPLE

Mathematician dies

American mathematician William Thurston, who won the Fields medal in 1982, died on 21 August, aged 65. Thurston's numerous breakthroughs in the fields of geometry and topology influenced cosmologists' theories about the shape of the Universe. His work also laid the basis for the 2003 solution to the Poincaré conjecture by Russian mathematician Grigory Perelman.

Jailed physicist

Omid Kokabee, an Iranian doctoral student who has been in prison in Iran for 18 months, has lost his appeal against his 10-year jail sentence, it emerged last week. Kokabee, who had been studying laser physics at the University of Texas at Austin, was arrested on charges of conspiring against Iran in February 2011 while leaving Tehran after visiting his family. He was sentenced in May 2012. He has denied all charges, and several science organizations have asserted his innocence and asked for a fair trial. See go.nature.com/7bthm9 for more.

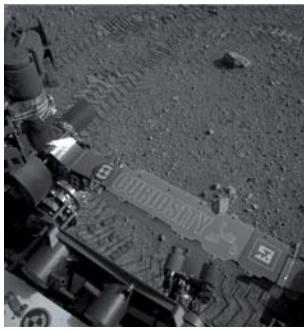
EVENTS

Swine-fever warning

The first-ever detection of African swine fever in Ukraine shows that the disease poses an ongoing risk to the Caucasus and neighbouring areas, the Food and Agriculture Organization of the United Nations warned on 21 August. The disease, caused by an infectious virus, does not affect humans, but can cause large numbers of pig deaths. See page 565 for more on African swine fever's spread across the Caucasus.

Tracks on Mars

NASA's Mars rover Curiosity made its first test drive on 22 August, going forwards, rotating and reversing (tracks, **pictured**) to end up about six metres from its landing site, which has been named 'Bradbury Landing' in honour of the late author Ray Bradbury. Beyond damage to a wind sensor, the mission has been a total success so far, NASA officials said. Curiosity



will now take a few weeks to drive the 400 metres to Glenelg, a spot where three kinds of terrain meet.

Arctic sea-ice low

With a few weeks of the melt season still to go, Arctic sea ice has broken the record for its minimum summer extent. Satellites registered 4.1 million square kilometres of sea ice on 26 August, edging out the record of 4.17 million square kilometres set on 18 September 2007, according to scientists at the US National Snow & Ice Data Center in Boulder, Colorado. Including 2012, all six of the lowest sea-ice minima on record have occurred in the past six years. See go.nature.com/waurw1 for more.

BUSINESS

Alzheimer's setback

A drug aimed at slowing cognitive and behavioural decline in patients with Alzheimer's disease failed to meet those targets in two late-stage clinical trials, the pharmaceutical firm Eli Lilly announced on 24 August. But the antibody, solanezumab, did slow cognitive decline in a subset of patients with milder forms of the condition, the company added. Bapineuzumab, a similar drug being developed by Pfizer and Johnson & Johnson, also failed two large-scale trials recently, and its development has been halted (see *Nature* **488,** 135; 2012). See go.nature.com/nhfbmu for more.

Hepatitis C halt

Bristol-Myers Squibb (BMS) announced on 23 August that it was stopping development of an experimental hepatitis C drug, after nine patients were hospitalized and one died in a clinical trial. The company, based in New York city, had only recently acquired the new drug; it was the main asset of Inhibitex, a firm in Alpharetta, Georgia, for which BMS paid US$2.5 billion in January. The drug was supposed to inhibit a viral RNA polymerase protein and so prevent the hepatitis C virus from copying its genome; other firms have drugs in clinical trials with similar mechanisms of action. See go.nature.com/vet89t for more.

HIV combination

The US Food and Drug Administration on 27 August approved a once-daily pill that combines four drugs in one to treat HIV. The drug was developed by Gilead Sciences of Foster City, California, which hopes that the pill will enable it to maintain profits after a key component of its 3-in-1 pill Atripla comes off patent next year, reducing that therapy to at least 40% of the current price (see *Nature* **488,** 267; 2012). The new pill, called Stribild, contains two previously approved drugs (emtricitabine, tenofovir) and two new ones (elvitegravir, cobicistat).

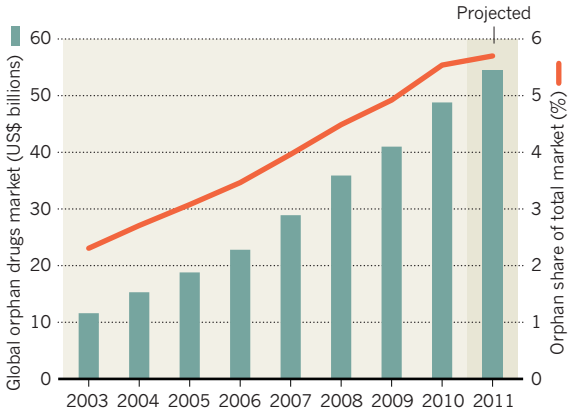
SOURCE: THOMSON REUTERS/KIRAN MEEKINGS/IMS HEALTH

TREND WATCH

Orphan drugs treat rare diseases in a small number of patients, yet still offer good opportunities for pharmaceutical companies to make money, with a global market rising to more than US$50 billion, according to a 22 August report by Thomson Reuters. Such drugs can be priced highly and require smaller clinical trials. They also attract government incentives, enjoy extended exclusivity rights and have high rates of regulatory approval — explaining industry's new interest in the sector, the report notes.

RISE OF ORPHAN DRUGS

Consumer spending on orphan drugs neared 6% of the global pharmaceuticals market by the end of 2011.



NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

K. CHALABOV/RIA NOVOSTI

Mass culls and vigilant hygiene are the main forms of defence against the African swine fever virus.

ANIMAL HEALTH

Pig fever sweeps across Russia

Deadly virus may be poised to spread to neighbouring states.

BY EWEN CALLAWAY

Russian authorities have incinerated tens of thousands of pigs and closed roads in the past few weeks, in an attempt to contain an emerging outbreak of African swine fever, a viral disease so lethal to the animals that it has been likened to Ebola. The spread of the disease comes with a heavy economic toll — last year, the Russian Federation lost 300,000 of the country's 19 million pigs to swine fever, at an estimated cost of about 7.6 billion roubles (US$240 million).

African swine fever was also detected for the first time in Ukraine in late July, and European and Asian countries are on the alert to deal with outbreaks that could cost their pork industries billions of dollars. With no vaccine or cure for the disease, mass culls and vigilant hygiene offer the main defence.

Scientists first encountered African swine fever in the 1920s in domestic pigs in Kenya, where the vicious haemorrhagic fever felled nearly every animal infected. The virus, which is also carried by warthogs and ticks without causing disease, is now endemic in much of sub-Saharan Africa, limiting pig farming there. It does not infect humans.

In 1957, the virus jumped to Portugal after pigs near Lisbon's airport were fed infected human food scraps (the virus particles can survive meat curing processes). It then hit Spain, and import of the region's ham — including the coveted *jamón ibérico* — was banned by many countries, until the disease was eradicated in Spain and Portugal in the mid-1990s.

> ***"If you are a small producer, and you lose all your five pigs, that is devastating to the family."***

The cases now flaring up in Russia, Ukraine and other countries in the Caucasus have their origins in a 2007 outbreak in the former Soviet republic of Georgia, where the virus gained a foothold after being imported from Africa. "It wasn't diagnosed for several months because they weren't really looking for it," says Linda Dixon, an expert on African swine fever at the Institute for Animal Health in Pirbright, UK. The disease quickly jumped to neighbouring Azerbaijan, Armenia and Chechnya, before fanning out across Russia (see 'Pig plague').

The recent spread of the virus means that ▶

the Ukrainian outbreak, now under control after authorities culled 208 pigs and instituted quarantine measures, did not come as a surprise, says Juan Lubroth, the chief veterinary officer at the Food and Agriculture Organization of the United Nations (FAO) in Rome, who is in charge of the organization's response to the outbreak.

Nearby countries, such as Moldova, Belarus and the Baltic states, could be next. To the east, the disease has been detected on the doorstep of Kazakhstan, which shares a long border with China, home to more than 1 billion pigs. China also risks importing the virus through its growing trade with African nations.

Europe's large pig farms are buffered by better biosecurity and hygiene practices. But agencies such as the UK Department of Environment, Food and Rural Affairs in London are nevertheless watching the situation closely. "How this will pan out, we don't really know," Lubroth says. "You're asking me to look into a crystal ball."

The variety of ways in which African swine fever spreads only increases the uncertainty. Pigs can leave virus particles on transport vehicles, for example, exposing whole shipments of uninfected animals. Biosecurity measures, such as scrubbing trucks and decontaminating farmers before they enter and leave pig pens, can help to contain outbreaks. But infected wild boar, whose populations stretch across Russia and Europe, pose a transmission threat that is harder to control. "Boars don't require visas to move across borders," says Lubroth.

The pigs' food can also carry the virus if it includes contaminated pork products. Swill feeding, in which pigs are fed scraps of human food waste, is popular among small-scale farmers. Limiting this practice (which is banned in the European Union) or heat-sterilizing the food scraps can prevent disease transmission, says Dixon. "I remember being taken to a little backyard farm near Nairobi, and that farmer was doing everything correctly. He had a cooker with a big pan of entrails that he was feeding to pigs and he had a little tray of disinfectant outside the pig pens."

SOURCE: C. NETHERTON/OIE

The FAO warns that continued spread of African swine fever could be very costly. Russia does not export its pork, but trade restrictions could prove expensive for other countries where the disease becomes endemic.

Denis Kolbasov, director of the National Research Institute for Veterinary Virology and Microbiology of Russia in Pokrov, says that officials often have little appetite for expensive countermeasures such as widespread culling and quarantine that could disrupt Russia's billion-dollar pork industry. Meanwhile, backyard farmers often do not report suspected cases for fear of losing their livelihood.

"If you are a small producer, and you lose all your five pigs, that is devastating to the family," says Lubroth. "That is the situation I see in many parts of Europe and Africa." African swine fever was especially costly in South Ossetia during a 2008 conflict with Georgia, because many farmers there could not grow crops and relied on livestock for food and income.

While animal-health officials focus on containing the spread of African swine fever, scientists believe that it should be possible to develop a vaccine to eradicate the disease. The lucky few pigs that survive infection are rendered immune, so Dixon's lab and others are working to identify which of the virus's 175 or so genes trigger the immune system. In principle, researchers could engineer these genes into the genome of a harmless virus to create a vaccine. Alternatively, identifying and switching off the disease-causing genes in the virus could lead to an attenuated vaccine. In the longer term, these options offer the best chance of halting the march of the virus, says Lubroth. "I wish I had a vaccine." ■

PHYSICS

US colliders jostle for funds

Heavy-ion facility at risk of premature closure as committee sets budget priorities.

BY EUGENIE SAMUEL REICH

When the Large Hadron Collider (LHC) in Switzerland seized the world record for the highest-energy collisions in 2010, it also sealed the fate of the leading US particle collider. The Tevatron, at Fermi National Accelerator Laboratory in Batavia, Illinois, was closed the following year to save money.

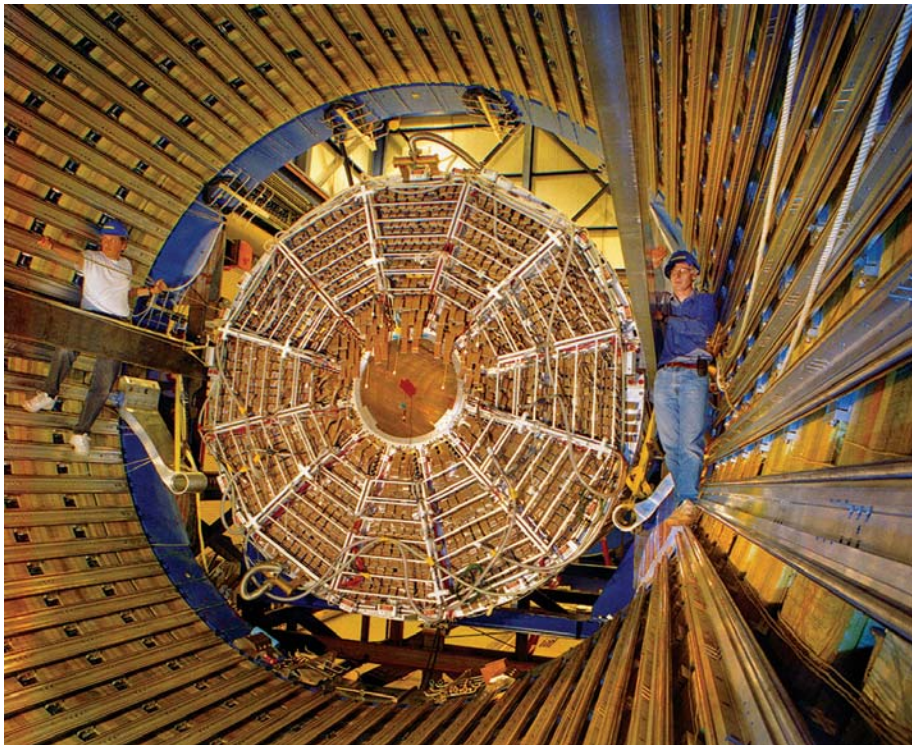
Now, physicists at another US physics facility, the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory in Upton, New York, are trying to avoid a similar end. On 13 August, researchers at the ALICE heavy-ion experiment at the LHC at CERN, Europe's particle-physics lab near Geneva, announced that they had created the hottest-ever man-made plasma of quarks and gluons. This eclipsed the record temperature achieved at RHIC two years earlier by 38%, and raised uncomfortable questions about RHIC's future.

This European achievement comes as a nuclear-science advisory committee to the US Department of Energy and the National Science Foundation prepares to meet on 7–9 September to start a prioritization process. The discussions will pit RHIC against two other US facilities: the Continuous Electron Beam Accelerator Facility (CEBAF) at the Jefferson National Laboratory in Newport News, Virginia, and the Facility for Rare Isotope Beams (FRIB), to be built at Michigan State University in East Lansing (see 'Competing for survival').

The committee has been charged with prioritizing the facilities within various scenarios of flat and slightly rising physics budgets. Nuclear physicist Robert Tribble of Texas A&M University in College Station, who is chairman of the advisory committee, says that in the flat scenario, there is only enough funding to support two of the three facilities. "It doesn't all fit," he says.

Steven Vigdor, who oversees RHIC as associate laboratory director for particle and nuclear physics at Brookhaven, naturally hopes that the committee will decide to scale back plans for another facility. But he admits that the Brookhaven team is already on the back foot because RHIC was ranked lower than the other

BROOKHAVEN NATL LAB./INTERACTIONS.ORG



The STAR detector at the Relativistic Heavy Ion Collider looks for signs of quark–gluon plasmas.

COMPETING FOR SURVIVAL
Flat budgets are forcing US nuclear physicists to plan for a future with just two of three desired facilities.

Name	Cost and location	Status	Purpose
Relativistic Heavy Ion Collider (RHIC)	$649 million; Upton, New York	Operating since 2000.	Explore the properties of quark–gluon plasmas.
Continuous Electron Beam Accelerator Facility (CEBAF)	$310 million; Newport News, Virginia	Upgrade is 60% complete.	Study the arrangement of quarks and gluons in protons and neutrons.
Facility for Rare Isotope Beams (FRIB)	$680 million; East Lansing, Michigan	Construction has not yet begun.	Create new nuclear isotopes and study their properties.

two facilities in a 2007 prioritization effort, also headed by Tribble.

A plan to increase the energy of CEBAF, by contrast, came out on top in the 2007 report, and with the upgrade now 60% complete, the facility looks to be on the safest ground for the latest ranking process. With beams of electrons reaching 12 gigaelectronvolts, the upgraded CEBAF will make further studies of the arrangement of the quarks and gluons that make up protons and neutrons. It will also look for exotic particles, made up of quarks, that are predicted by quantum chromodynamics — the theory that describes how quarks and gluons interact.

Next comes FRIB, which promises to advance the study of nuclear structure by colliding beams of heavy ions to produce hundreds of nuclear isotopes that have never been created before. But the facility is vulnerable because it has not yet been built. Scientists at FRIB have won political support by obtaining US$94.5 million of its $680-million cost from Michigan State University, and by promising to discover isotopes with medical, energy and national-security applications. "The country will get the benefits of the new isotopes," says FRIB director Konrad Gelbke.

To keep CEBAF and FRIB on schedule, however, it is likely that US funding agencies would have to close RHIC prematurely. Vigdor says that would be a mistake. Despite ALICE now generating plasmas at higher temperatures, he says, theorists testing quantum chromodynamics will still want to use RHIC's lower energies and temperatures to map the phase transitions between quark–gluon plasmas and ordinary matter under various conditions.

Vigdor also points out that RHIC's construction is a past investment that is now bearing fruit. Closing it would not necessarily guarantee the funding of a facility such as FRIB. He says that it is "naive" to think that closing RHIC or CEBAF would free up funds for a new machine. "When you terminate a facility, those funds tend to move away from the field."

"When you terminate a facility, those funds tend to move away from the field."

The trade-off between CEBAF and RHIC could have other important implications for their respective host labs. Nuclear physicists hope eventually to build an even higher-energy, polarized electron–ion collider that would be able to image the gluons that hold together quarks in the proton.

Both the Brookhaven and Jefferson labs are eager to host that machine, says Robert McKeown, deputy director for science at Jefferson. But he says that if the effort moves forward, the lab that prevails in the current showdown is likely to win the future collider as well.

Tribble still hopes to avoid having to close any of the three facilities. In 2005, he notes, a similar crisis was averted after an advisory committee laid out the dire consequences of flat funding for the future of US nuclear science. In the end, Congress came through with the budgetary increases required. "What we want to do here is to spell out what will be lost under different budgets," he says. His committee is planning to hold a final meeting in November, in time to influence the budget requests from US funding agencies for the next fiscal year. ■

TIPS/PHOTOSHOT

MORE ONLINE

TOP STORY

It is possible to learn in your sleep go.nature.com/2ubnzd

MORE NEWS

- Marijuana makes kids stupid go.nature.com/rc7l3e
- There are fewer microbes out there than you think go.nature.com/cifrpp
- Innocence Project calls on chemists to support forensics reform go.nature.com/vzalgu

PODCAST

Dieting monkeys don't live longer and Mexican anarchists attack science go.nature.com/eespxa

PUBLIC HEALTH

Doctors back circumcision

US task force finds that public-health benefits of the procedure outweigh risks to individuals.

BY MONYA BAKER

Expectant parents face many anxieties in preparing for a child. For those who have a son, there is an extra complication: deciding whether to keep his foreskin or have it snipped off.

On 27 August, a report by the American Academy of Pediatrics (AAP) concludes for the first time that, overall, boys will be healthier if circumcised[1]. The report says that although the choice is ultimately up to parents, medical insurance should pay for the procedure. The recommendation, coming from such an influential body, could boost US circumcision rates, which, at 55%, are already higher than much of the developed world (see 'Cuts by country'). "This time around, we could say that the medical benefits outweigh the risks of the procedure," says Douglas Diekema, a paediatrician and ethicist at the University of Washington, Seattle, who served on the circumcision task force for the AAP, headquartered in Elk Grove Village, Illinois.

The recommendation is also sure to stir up debate. The practice of circumcision cuts deeper than the body, tapping into religious rituals and cultural identities. What is a harmless snip to some signifies mutilation to others. And in the developing world, many see it as an essential life-saving measure. Condoms are more effective at preventing disease, but are not used consistently.

Diekema says that the task force's reviews of the latest medical evidence allowed it to make stronger policy recommendations than it did in 1999 and 2005. Perhaps the most powerful evidence in favour of circumcision comes from randomized controlled trials in South Africa[2], Kenya[3] and Uganda[4]. These found that, for men who have sex with women, circumcision reduced the risk of infection with HIV. (No protection was observed for men who have sex with men.) The South African and Ugandan trials also found that circumcision reduced infection rates for human papillomavirus (HPV) and herpes. The World Health Organization has already made circumcision part of its HIV-prevention strategy in sub-Saharan Africa, with a goal to circumcise 20 million men by 2015.

The AAP found that, in addition to preventing sexually transmitted infections, circumcision could reduce the rates of urinary tract infections and penile cancer, probably because the foreskin harbours infectious microbes as well as the immune cells targeted by HIV. The most common complications of circumcision — oozing, bleeding and infection — occur in 2% or fewer circumcisions and are easily treated. More serious complications are exceedingly rare, says Diekema. The task force also found no strong evidence that circumcised babies grew up with more urinary difficulties or sexual problems.

Gert van Dijk, an ethicist at the Royal Dutch Medical Association in Utrecht, the Netherlands, thinks that the AAP has underestimated the potential harm of circumcision. He says that it should only be performed when men are old enough to give consent, and disagrees with the AAP that circumcisions are simplest and safest when performed on infants. The very idea of asking whether circumcisions are beneficial is strange to Europeans, van Dijk says. "The integrity of the body is an important thing. We would never amputate a healthy part of a child."

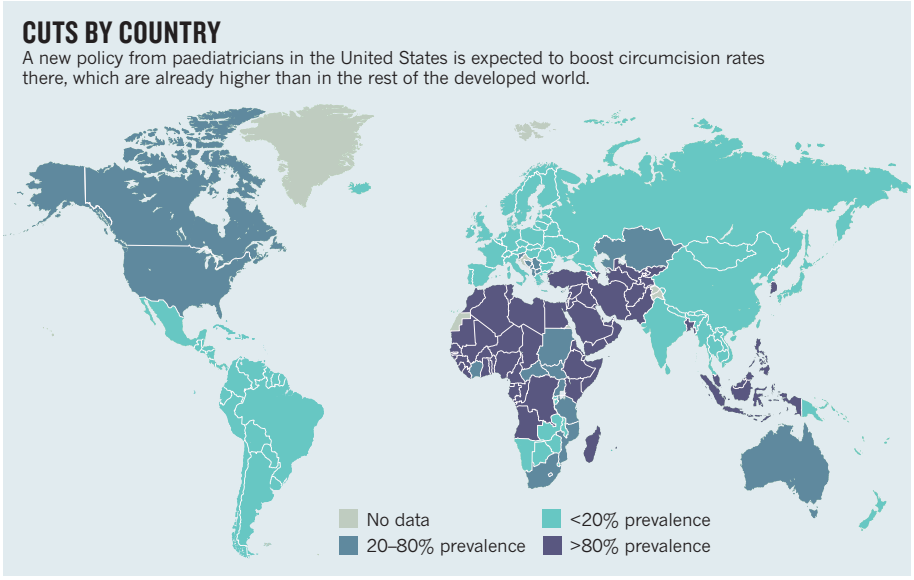
Van Dijk notes that the benefits cited in the African studies do not apply to the Netherlands, where HIV transmission is rare and occurs mainly through sex between men and through needle-sharing among drug users.

Rowena Hitchcock, president of the British Association of Paediatric Urologists, says that she is disappointed with the AAP policy because it recommends an "irreversible, mutilating surgery". She says that her organization is considering a review of its current policy, which recommends circumcision for infants who are at severe risk of urinary tract infections, because the evidence of medical benefit is not definitive.

The number of circumcisions required in order for the wider population to avoid disease may explain some of the national differences in policies and rates. Most men are unlikely to get the infections that circumcisions protect against, so they will not see a direct benefit.

But the cumulative benefits can add up. An analysis published last week by researchers at Johns Hopkins University in Baltimore, Maryland, found that the cost of performing circumcisions and treating complications would be tiny in comparison to the savings from the resulting lower rates of HIV, HPV, herpes and urinary tract infections, as well as from lower rates of bacterial vaginosis and trichomoniasis in women[5]. Each circumcision that is not performed costs the US health-care system US$313, the researchers estimate.

However, national customs may have a bigger role than economic decisions, says David Gollaher, a medical historian at the California Healthcare Institute in La Jolla, who has studied the history of circumcision. Insurance cover sends a signal that a procedure is medically appropriate, he says, which would reinforce existing inclinations in the United States towards action and intervention. "We do a lot more surgeries than anyone else, so circumcision fits into the pattern of doing more," he says. ■



1. Task Force on Circumcision *Pediatrics* **130,** e756–e785 (2012).
2. Auvert, B. *et al. PLoS Med.* **2,** e298 (2005).
3. Bailey, R. C. *et al. Lancet* **369,** 643–656 (2007).
4. Gray, R. H. *et al. Lancet* **369,** 657–666 (2007).
5. Kacker, S., Frick, K. D., Gaydos, C. A. & Tobian, A. A. R. *Arch. Pediatr. Adolesc. Med.* http://dx.doi.org/10.1001/archpediatrics.2012.1440 (2012).

E. BÖMSCH/IMAGEBROKER/FLPA

Rhesus monkeys on calorie-restricted diets age just as quickly as their chubbier counterparts.

AGEING

Calorie restriction falters in the long run

Genetics and healthy diets matter more for longevity.

BY AMY MAXMEN

To those who enjoy the pleasures of the dining table, the news may come as a relief: drastically cutting back on calories does not seem to lengthen lifespan in primates.

The verdict, from a 25-year study in rhesus monkeys fed 30% less than control animals, represents another setback for the notion that a simple, diet-triggered switch can slow ageing. Instead, the findings, published this week in *Nature*[1], suggest that genetics and dietary composition matter more for longevity than a simple calorie count.

"To think that a simple decrease in calories caused such a widespread change, that was remarkable," says Don Ingram, a gerontologist at Louisiana State University in Baton Rouge, who designed the study almost three decades ago while at the National Institute on Aging (NIA) in Bethesda, Maryland.

When the NIA-funded monkey study began, however, studies of caloric restriction in short-lived animals were hinting at a connection. Experiments had showed that starvation made roundworms live longer. Other studies had showed that rats fed fewer calories than their slow and balding brethren maintained their shiny coats and a youthful vigour. And more recently, molecular studies had suggested that caloric restriction — or compounds that mimicked it — might trigger a cascade of changes in gene expression that had the net effect of slowing ageing.

In 2009, another study[2], which began in 1989 at the Wisconsin National Primate Research Center (WNPRC) in Madison, concluded that caloric restriction did extend life in rhesus monkeys. The investigators found that 13% of the dieting group died from age-related causes, compared with 37% of the control group.

One reason for that difference could be that the WNPRC monkeys were fed an unhealthy diet, which made the calorie-restricted monkeys seem healthier by comparison simply because they ate less of it. The WNPRC monkeys' diets contained 28.5% sucrose, compared with 3.9% sucrose at the NIA. Meanwhile, the NIA meals included fish oil and antioxidants, whereas the WNPRC meals did not. Rick Weindruch, a gerontologist at the WNPRC who led the study, admits: "Overall, our diet was probably not as healthy."

NATURE.COM
Read the News & Views article about this research.
go.nature.com/icodos

Further, the WNPRC control group probably ate more overall, because their meals were unlimited, whereas NIA monkeys were fed fixed amounts. As adults, control monkeys in the WNPRC study weighed more than their NIA counterparts. Overall, the WNPRC results might have reflected an unhealthy control group rather than a long-lived treatment group. "When we began these studies, the dogma was that a calorie is a calorie," Ingram says. "I think it's clear that the types of calories the monkeys ate made a profound difference."

Researchers studying caloric restriction in mice have become accustomed to mixed results, which they attribute to genetic diversity among strains. Genetics probably explains part of the variation between the monkey studies, too, as the NIA monkeys were descended from lines from India and China, whereas the Wisconsin monkeys were all from India.

The molecular effects of caloric restriction have also turned out to be complicated. Using compounds such as resveratrol, found in red wine, scientists have triggered the stress response that caloric restriction activates, which shuts down non-vital processes in favour of those that ward off disease. But hopes that ageing could be delayed by targeting a single gene or protein in a single molecular pathway have faded, as researchers have learned that the key pathways vary according to the animal."It may take us a decade to sort out longevity networks," says David Sinclair, a geneticist at Harvard Medical School in Boston, Massachusetts.

Meanwhile, there is a dearth of evidence that caloric restriction slows ageing in humans. Observational studies have found that people of average weight tend to live longest[3]. Nir Barzilai, a gerontologist at Albert Einstein College of Medicine in New York, says that the centenarians he studies have led him to believe that genetics is more important than diet and lifestyle. "They're a chubby bunch," he says.

"When we began these studies, the dogma was that a calorie is a calorie."

A more nuanced picture would suit Ingram, who enjoys an occasional feast of Louisiana crawfish. Ingram says that he looks forward to studies of how diet composition, rather than caloric intake, affects ageing. "Is the human lifespan fixed?" he asks. "I still don't believe that for a minute." ■

1. Mattison, J. A. *et al. Nature* http://dx.doi.org/10.1038/nature11432 (2012).
2. Colman, R. J. *et al.* Science **325,** 201–204 (2009).
3. Berrington de Gonzalez, A. *et al. N. Engl. J. Med.* **363,** 2211–2219 (2010).

W. RUDHART/DPA/CORBIS

Brazilian scientists are trying to understand the effects of logging on carbon emissions in the Amazon.

DEFORESTATION

Brazil unveils tool to track emissions

Carbon releases lag behind Amazon deforestation.

BY JEFF TOLLEFSON

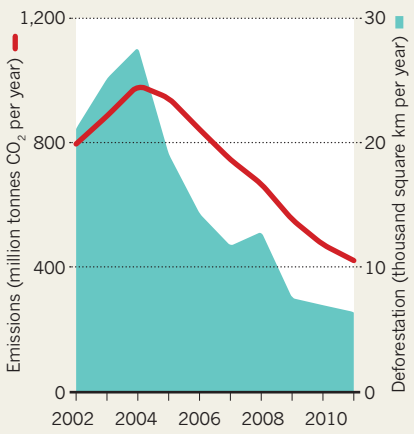
The rate of deforestation in the Brazilian Amazon has plummeted by 77% in the past seven years, but annual carbon emissions associated with deforestation have not fallen nearly as much, says a Brazilian study that combines satellite data and biomass maps to model the change. The difference is in large part due to a natural lag as carbon stocks slowly decay and make their way into the atmosphere as carbon dioxide.

Brazil has long had a sophisticated satellite-based system for monitoring deforestation, but translating forest clearance into emissions has remained a challenge. The picture is more complex than often assumed. The area cleared each year decreased from 27,772 square kilometres — a swathe the size of Sicily — in 2004 to 6,418 square kilometres in 2011. A direct conversion of that lost biomass into carbon would suggest a drop in annual carbon dioxide emissions from more than 1.1 billion tonnes of CO_2 in 2004 to 298 million tonnes of CO_2 — a reduction of nearly 74%.

But numbers calculated by Brazil's National Institute for Space Research (INPE) in Sao Jose dos Campos using the new model, and released on 13 August, suggest a reduction of only 57% in actual emissions, with the difference due to factors such as the time it takes for roots to decay, and the fact that forest debris cut in one year might be burned in another (see 'Carbon lag'). A paper describing the model was published last week (A. P. D. Aguiar *et al. Global Change Biology* http://doi.org/h7g; 2012).

CARBON LAG

Curbing Amazon deforestation has cut greenhouse-gas emissions, but the time it takes for wood and roots to decay delays the pay-off by a few years.



"We don't know how the government is going to use this information at this point, but what we are trying to express is that this is a more correct way to understand carbon emissions," says Jean Ometto, a researcher at INPE who helped to develop the system. He estimates that efforts to forestall deforestation in the Amazon have lowered Brazil's overall carbon emissions by roughly 17% since 2004.

The model is a milepost for scientists and policy-makers working at the intersection of forestry, agriculture and global warming. Many scientists are busy linking satellite measurements to field and aerial studies to create more accurate maps of forest biomass, which can then be used to calculate emissions. INPE's new model carries that work forward, allowing Brazilian scientists to provide annual estimates of both deforestation and carbon emissions, and setting the stage for a deeper analysis of the impacts of logging, agriculture and forest regrowth.

Richard Houghton, a forestry expert at the Woods Hole Research Center in Falmouth, Massachusetts, says that the actual results are not necessarily surprising. But he credits INPE with developing a solid emissions model that will help Brazil build on its forest-monitoring system and tackle emissions in a serious way. "That's where you'd like the rest of the world to be, where Brazil is."

INPE's analysis confirms earlier findings that deforestation is gradually moving into denser forests, which means that emissions per hectare are increasing. The results also suggest that forest regrowth has had little impact on net emissions because many secondary forests are periodically chopped down. INPE scientists say that secondary forests could have a significant impact on net emissions if allowed to regrow.

INPE

One question about the model is how to represent emissions due to logging. INPE scientists assume that some logging has already taken place in clear-cut areas, but they are still working on ways to account for broader timber operations. Earlier research suggests that fully accounting for logging could roughly double overall emissions from the Amazon (G. P. Asner *Science* **310,** 480–482; 2005).

And Sandra Brown, a forestry expert at the non-profit organization Winrock International, based in Arlington, Virginia, asks whether it makes sense to highlight the time lags involved in emissions, which could increase public uncertainty and run the risk of confusing policy-makers. At the same time, she says, scientists do need to better understand the way carbon is cycling through forests. And she says the next major challenge is for INPE to build the effects of widespread logging operations into its emissions model.

"They've got the equipment, and they've got the know-how, and they've got the resources to do it," she says. ■

ASTRONOMY

Recycled dishes form telescope network

Africa refits redundant satellite dishes for radio astronomy.

BY LINDA NORDLING

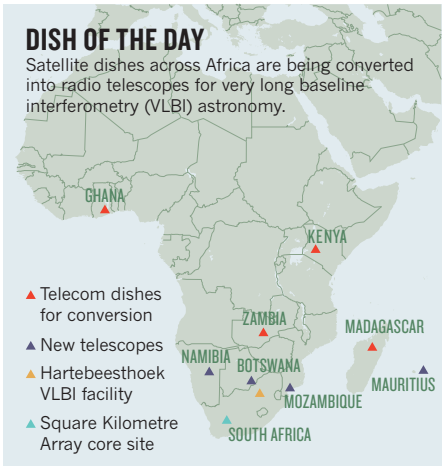
Ageing satellite dishes, once the backbone of Africa's telecommunications system, are being given a new lease of life as radio telescopes.

The thrifty project aims to boost the skills of the continent's scientists as Africa prepares to host the US$2.1 billion Square Kilometre Array (SKA), set to be the world's most powerful radio telescope when it is completed in the mid-2020s. In May, the SKA project leaders decided that Africa, with South Africa at the helm, would split the array with Australia and New Zealand (see *Nature* **485,** 555–556; 2012). The SKA will detect radiation from the early Universe, giving clues to how the first stars and galaxies formed.

South Africa has a nucleus of radio-astronomy expertise, but the rest of the continent has few researchers with the skills to work on the SKA. The African SKA partners hope to train a new generation of radio astronomers using the converted satellite dishes, which have been superseded by the high-speed fibre-optic cables skirting Africa's coast. "We want to make very sure that Africa delivers the best possible instrument, and that African engineers and scientists can maintain and support it," says Anita Loots, associate director of SKA South Africa.

The first dish, in Kuntunse, Ghana, will begin its makeover next month, and the converted dishes will eventually comprise a network of radio telescopes — the African Very Long Baseline Interferometry (VLBI) Network. A VLBI facility can train many geographically distant dishes onto the same astronomical object, and superimposing the signals they collect gives a much higher-resolution image than a solo telescope can obtain. At present, the only member of a VLBI network in Africa is a 26-metre dish at South Africa's radio astronomy hub in Hartebeesthoek, which makes observations with partner telescopes across the world.

The dish in Ghana was donated to the country's science ministry in May by telecommunications company Vodafone Ghana, and after astronomical instruments are fitted it should be operational by June 2013. The African VLBI Network's organizers want to convert three more dishes — in Kenya, Zambia and Madagascar — and build four radio telescopes from scratch in Namibia, Botswana, Mauritius and Mozambique (see 'Dish of the day'). In July, the project received 120 million rand (US$14.6 million) from the African Renaissance Fund, a development initiative supported by South Africa. A new radio telescope can cost more than 50 million rand, but converting the satellite dishes costs just 15 million to 20 million rand each.



The decades-old equipment can spring some nasty surprises. Last year, an engineer broke the steering mechanism of the dish in Ghana, and its age meant that no spare parts were available, so a completely new control and monitoring system has had to be built.

The project also stands to benefit global VLBI science, says Tom Muxlow of the Jodrell Bank Centre for Astrophysics near Manchester, UK. The most powerful VLBI networks are in the United States and Europe, but equatorial skies are not as well covered — this gap reduces the quality of images available to radio astronomers, potentially missing fine details such as faint jets of plasma from the regions around black holes. "The addition of a dedicated array of African antennas observing the equatorial sky, by itself or in combination with global arrays, has the potential for a truly transformational step in imaging quality," Muxlow says. ■

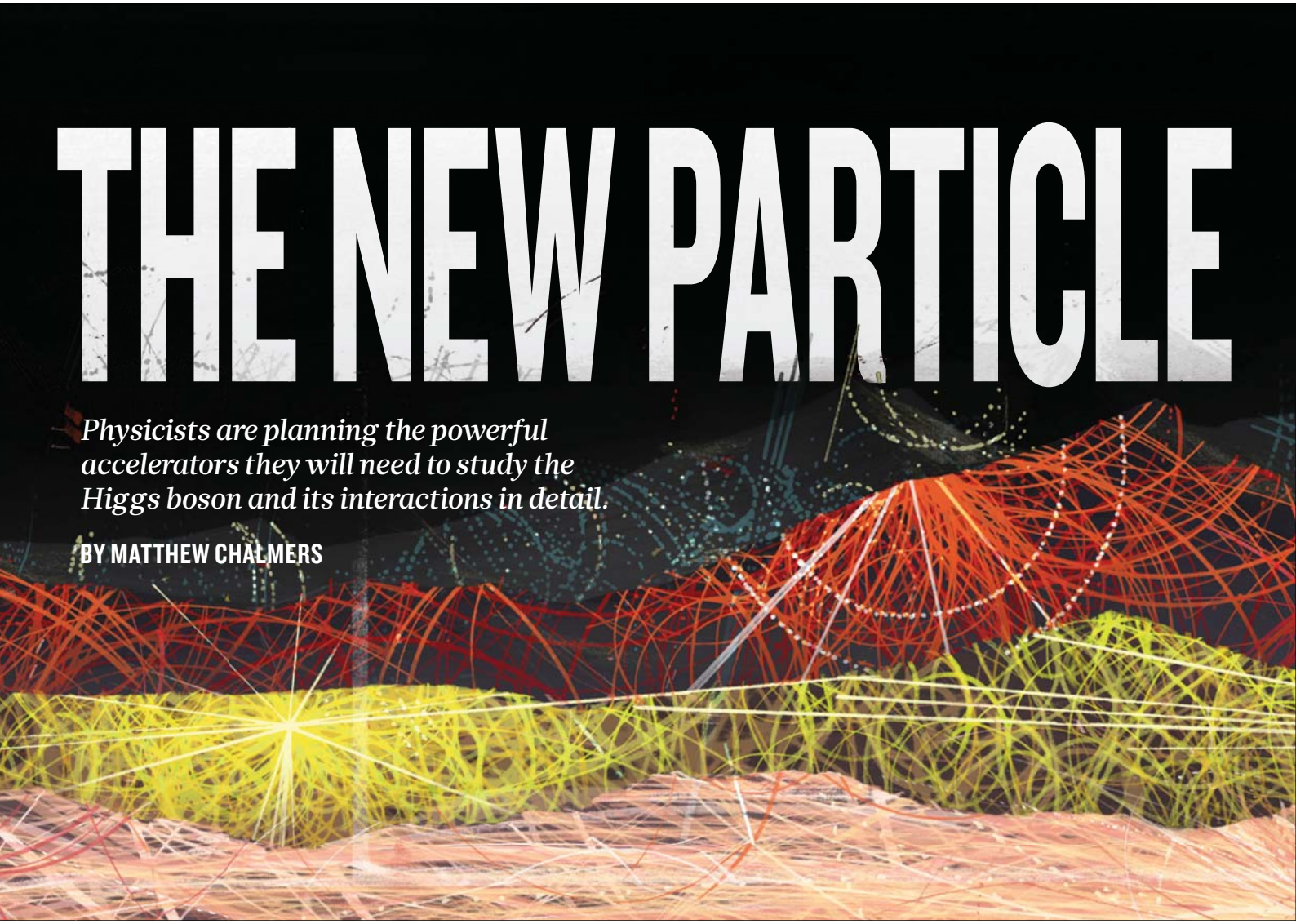
CORRECTION

The News story 'Republican spending plan casts shadow on science' (*Nature* **488,** 441; 2012) should have said that Paul Ryan's plan could cut spending on R&D by 5% below the fiscal-year 2012 budget rather than the President's 2013 request.

THE NEW PARTICLE

Physicists are planning the powerful accelerators they will need to study the Higgs boson and its interactions in detail.

BY MATTHEW CHALMERS



When particle physicists around the world woke up on 5 July, the scenes of joy, relief and tears were still fresh in their minds — along with a huge unanswered question. The memories were of celebrations the previous day, when researchers announced that a new particle very much like the long-sought Higgs boson had at last been found in data from the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory outside Geneva in Switzerland. The question promised to define their discipline's whole future. Is the particle a Higgs boson of maximum simplicity, as predicted by the 40-year-old standard model of particle physics? Or is it something more complex and interesting that will point towards a deeper, more complete theory?

Physicists hope and expect that the LHC will give them some answers over the next few years. But they are already honing their sales pitches for a machine to follow the LHC — a 'Higgs factory' that would illuminate such a theory with measurements far more precise than the LHC can provide.

"We know that there must be new physics beyond the standard model," says Barry Barish, a physicist at the California Institute of Technology in Pasadena. That's guaranteed, he and other physicists argue, by the existence of phenomena that don't easily fit into the model, such as the invisible scaffold of 'dark matter' suspected to comprise a quarter of the mass density of the Universe, or the ability of particles called neutrinos to 'oscillate' from one form to another. Barish heads the global consortium that is designing the International Linear Collider (ILC), one of the candidates for the next big machine. Even if no one yet knows what the new physics will involve, he says, "our strategy is to be ready in the event things fall in place".

The cost, timescales and capabilities of the ILC and other candidate machines will be scrutinized at the European Strategy for Particle Physics workshop in Krakow, Poland, on 10–12 September, which will set out the priorities for this field in Europe for the next five years. American particle physicists are planning a similar exercise at a meeting at Snowmass, Colorado, in June 2013.

But plans are one thing; reality is another. Funding any new machine, particularly in an economic downturn, will be a "daunting task", says Christopher Llewellyn-Smith, director of energy research at the University of Oxford, UK, and director of CERN at the time when the LHC was approved. "It will depend on what other new particles the LHC finds, on whether the new facility is unanimously supported by the community, and on its cost," he explains. "Even if the physics case is as strong as that for the LHC, and the cost is such that it can be done with a constant global high-energy physics budget, it will still be tough."

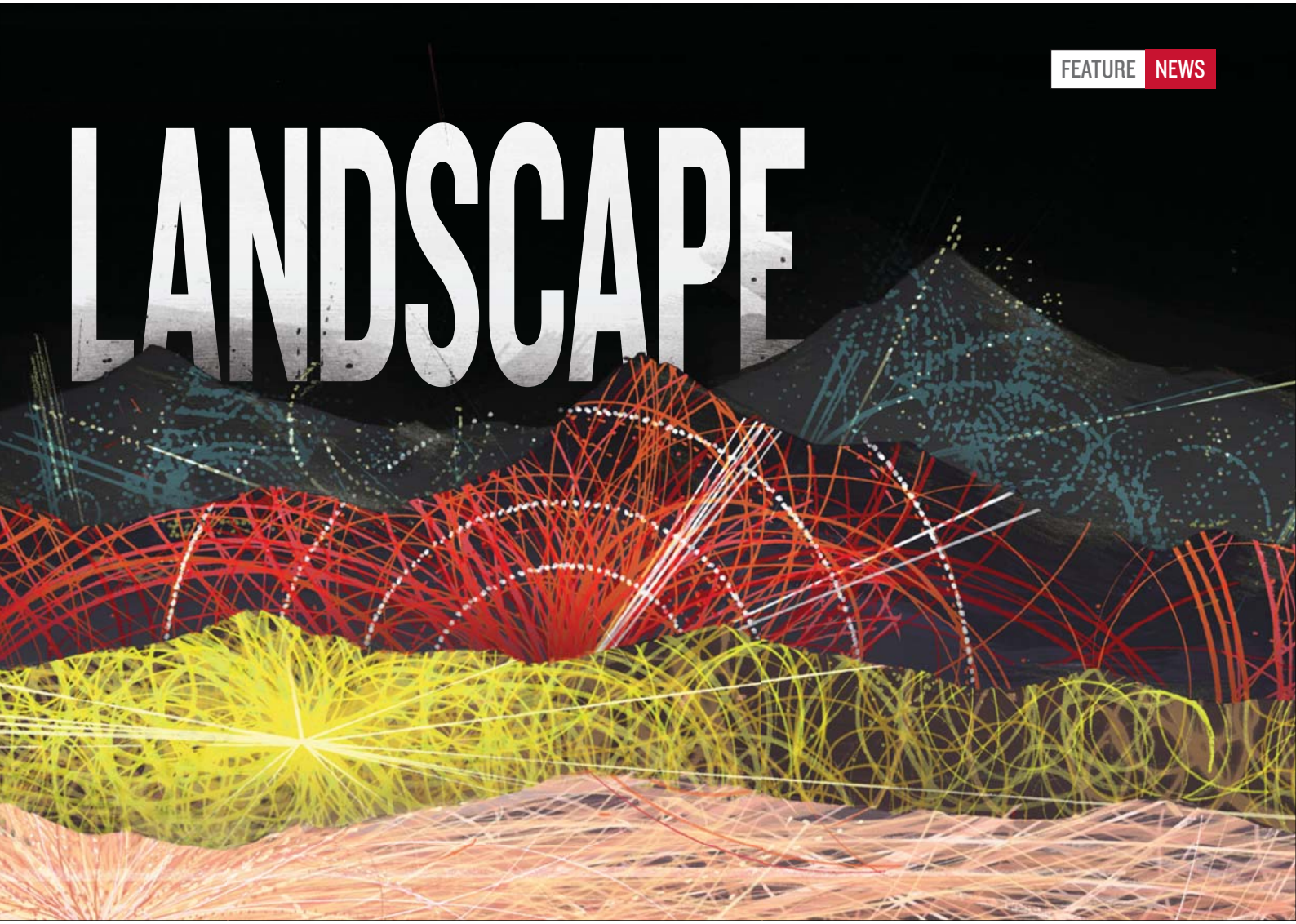
THE LHC LIVES ON

A key issue under discussion at the Krakow workshop will be how far the LHC teams can go in measuring the properties of the new particle. The physicists working there can expect much more data, plus major upgrades over the next ten years.

They already have one piece of good news: the mass of the Higgs-like particle — roughly 125 billion electron volts (GeV) in the energy units favoured by physicists — turns out to lie towards the light end of the range that theorists had estimated. This has two important consequences: it means that a relatively modest new collider would be

ILLUSTRATION BY BRENDAN MONROE

LANDSCAPE



sufficient to produce the Higgs in bulk, and it gives the new particle a rich variety of decay modes that will make it easier for physicists to study its interactions with other standard-model particles.

One priority, for example, is to check the standard model's prediction for how the Higgs interacts with standard-model fermions: entities such as electrons, muons and quarks that have an intrinsic angular momentum, or 'spin', of ½ in quantum units. The probability of an interaction with each particle is supposed to be proportional to its mass — not least because, in the standard model, interaction with the Higgs is what creates the mass.

Another priority is to verify that the new particle's own intrinsic spin has the standard-model value of 0. The LHC physicists can already say that the new particle is a boson — meaning that its spin in quantum units is 0, 1, 2 or some other integer — and that the integer cannot be 1; both conclusions follow from the particle's observed decay into pairs of photons, which are spin-1 bosons. Physicists do not have crazy theories involving bosons with a spin greater than 2, says CERN physicist Albert de Roeck, a scientific coordinator for the team working on the Compact Muon Solenoid detector at the LHC, so their task now is to determine whether it is a spin-2 or a spin-0 'scalar' boson as predicted.

> "WE KNOW THAT THERE MUST BE NEW PHYSICS BEYOND THE STANDARD MODEL."

The LHC will settle the spin question, says CERN's director-general Rolf Heuer, but it is less clear how far the LHC can go in testing the new boson's couplings to other particles — in particular the 'self-interaction' by which the Higgs gives itself mass. At present, all the LHC physicists can say is that the new boson's interactions with other particles are consistent with the standard-model predictions within the present measurement uncertainties of 30–40%. According to de Roeck, the collider should get those uncertainties down to 20% by the end of this year, and conceivably down to "a few per cent" over the next 10–15 years.

But that, for many physicists, is precisely why they need a next-generation machine. A truly stringent test of the standard model, which would reveal tiny deviations that could point the way towards better models, demands that researchers measure the Higgs's interaction with other particles to within 1% uncertainty, possibly as little as 0.1% should the precision of theoretical predictions also improve in the next few years. And that is a level the LHC is unlikely to reach. The machine is like a sledgehammer: it crashes together beams containing hundreds of billions of protons at energies that will eventually reach 7 trillion electron volts (TeV) per beam. This is good for discovering new massive particles, but less so for making precision measurements, because protons are chaotic seas of quarks and gluons that make the collisions messy.

Instead, every proposal for a next-generation machine calls for some form of lepton collider (see 'After the Higgs'). Leptons, a group of light particles that includes electrons, muons and neutrinos, sidestep the

messiness by not participating in the strong quark–gluon interactions that produce it. Leptons are elementary and interact only through the relatively feeble electromagnetic and weak forces. As a result, lepton machines are more like scalpels than sledgehammers: their collisions can be tuned to the mass of a particular particle and the spray of particles created would be comparatively clean and simple to interpret.

MUONS OR ELECTRONS

A relatively cheap option, argue some physicists, would be to place the tubes of a new accelerator alongside the LHC in the existing tunnel, and use them to collide opposing beams of electrons and antimatter electrons (better known as positrons). This proposal, known as LEP3 in honour of the Large Electron–Positron (LEP) collider that occupied the tunnel before the LHC's construction began in 2000, emerged only in the past year as preliminary evidence for the new particle piled up. LEP3 could produce Higgs bosons with just 120 GeV per beam — a total energy of 240 GeV — only a notch up from the original LEP's maximum of 209 GeV. Its production would be boosted further by recent technological advances that would allow for a collision rate, or 'luminosity', some 500 times greater than LEP could have achieved.

Building LEP3 in the LHC tunnel could allow some of the LHC's particle detectors to be reused, as well as making use of CERN's existing infrastructure for power, maintenance and data-taking. Such savings bring LEP3's estimated cost down to between US$1billion and $2 billion, far lower than the LHC's $6-billion price tag. "The idea is there to kill," says LEP3 advocate Alain Blondel at the University of Geneva, who points out that there should be room to build the new lepton collider without removing the LHC: the tunnel was originally intended to have both types of collider running simultaneously.

For all its advantages as a high-output Higgs factory, LEP3 would not be able to study anything much heavier than the Higgs. And that could be a problem if, as many particle physicists hope, the LHC ends up discovering heavier new particles that theorists are predicting from ideas such as supersymmetry, or even finding extra dimensions. Stepping up the energy of LEP3 to study the heavier particles would be virtually impossible because of losses from synchrotron radiation — the stream of photons emitted when any charged particle moves along a curved path. This isn't so much of a problem for the LHC's protons, because energy losses from synchrotron radiation fall off dramatically for particles of higher mass, and protons outweigh electrons by a factor of nearly 2,000. But losses in LEP3 would be severe. The only way to increase the accelerator's energy would be to increase its radius, which would require a new tunnel. Some physicists have talked about drilling a new tunnel stretching out beneath Lake Geneva, and installing an 80-kilometre circular electron–positron machine, although that's not something for the foreseeable future, says Heuer.

Meanwhile, physicists around the world have been exploring concepts for an alternative Higgs factory that would be much smaller than LEP3, perhaps as little as 1.5 km in circumference. By colliding beams of muons, electron-like particles with 207 times the mass of an electron, such a machine has negligible synchrotron-radiation losses and could produce tens of thousands of Higgs bosons from a total collision energy of just 125 GeV, as opposed to LEP3's 240 GeV. It would also be capable of going to much higher energies, to study heavier particles (see *Nature* **462,** 260–261; 2009).

But a muon collider faces major hurdles of its own, not least the fact that muons decay into electrons and neutrinos with a mean lifetime of 2.2 microseconds. That's a very long time in the subatomic realm, where particle lifetimes are often measured in fractions of a trillionth of a nanosecond. But in engineering terms, it is practically instantaneous. Muons for an accelerator would have to be produced by slamming a proton beam into a metal target; then 'cooled', or lined up into an orderly beam; and finally accelerated to the requisite energy, all in a time frame considerably shorter than the blink of an eye. That challenge is being addressed by the muon ionization cooling experiment (MICE) at the Rutherford Appleton Laboratory near Oxford, UK. MICE is expected to conclude its studies by 2016, at which point the cooling technology may be advanced enough for CERN to use it to build a neutrino factory — a stepping stone to a muon collider — that would fire beams of muon neutrinos through Earth to a detector thousands of kilometres away, such as one proposed in Finland.

Nonetheless, many physicists are sceptical. "I doubt I will see a muon collider working in my lifetime," says Brian Foster, a physicist at the University of Oxford. "We've been trying to cool muons for more than ten years, and it is just extremely difficult."

Foster is the European regional director for the rival concept of a linear electron–positron collider. This type of machine would essentially be a long, straight electron accelerator firing right down the barrel of an equally long, straight positron accelerator, with their beams slamming together in the middle. The lack of curvature would eliminate synchrotron radiation losses. And the accelerators could always be bumped up in energy by making them longer on the back end.

Ideas for a high-energy linear collider began to emerge in the 1980s, and eventually converged on two concepts. The ILC, developed by a worldwide consortium of laboratories and universities, would be some 30 kilometres long, and would use proven superconducting accelerator technology to reach energies of 0.5 TeV, with the possibility of upgrading to 1 TeV. The ILC team is soon to publish a technical design report and the cost of the project is currently estimated at $6.7 billion. The Compact Linear Collider (CLIC), championed by CERN, would be almost 50 kilometres long, but would use novel acceleration techniques to reach energies of 3 TeV. CLIC's costs are less clear than the ILC's because only a conceptual design report is available, but its higher energies would open up new realms for discovery as well as for precision measurements.

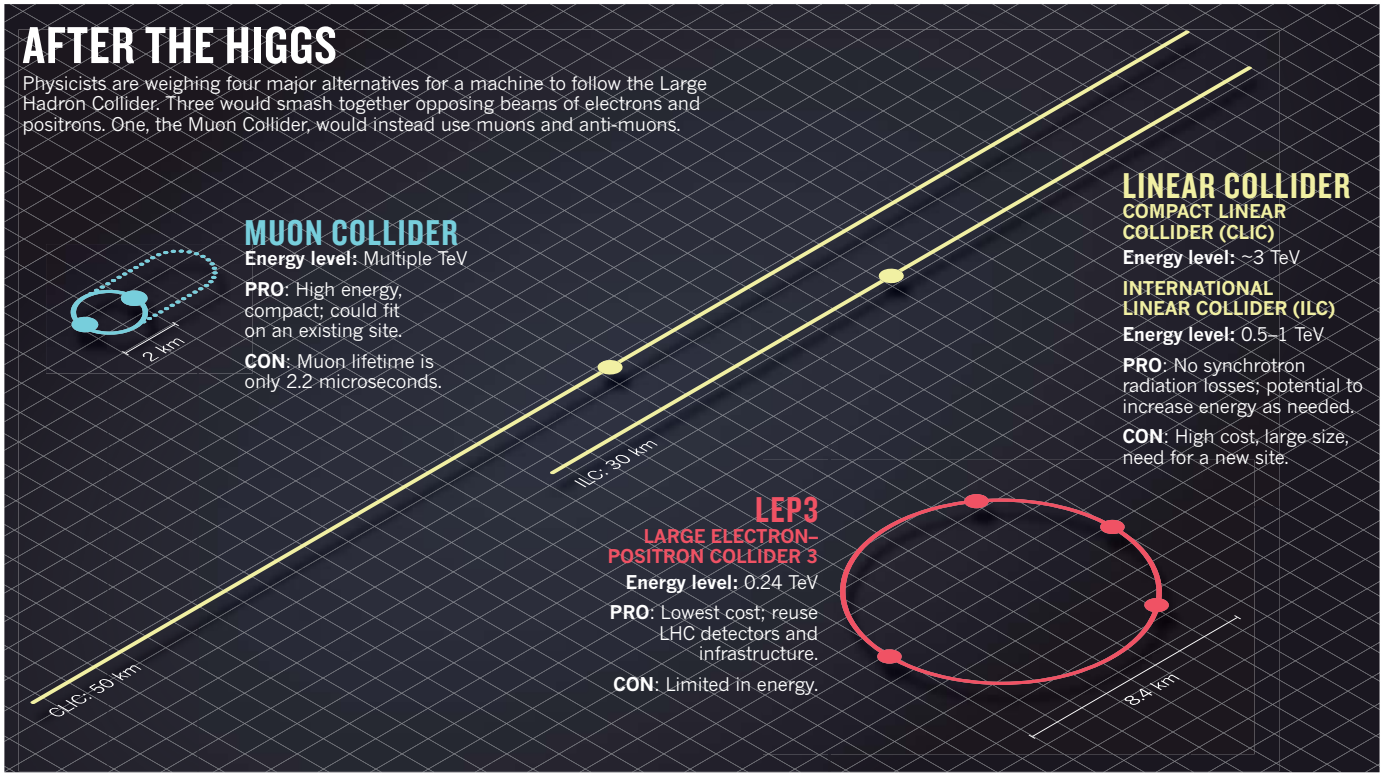
The performance of either design has been extensively studied theoretically, but in practice is a "wide open question" according to Blondel, current spokesperson for MICE. He points to the performance of the Stanford Linear Collider (SLC) at Menlo Park, California, which achieved energies of nearly 100 GeV. "The SLC finally worked very well, but it never quite produced the luminosity that they wanted. It was a very tough machine, and now with the ILC or CLIC we're discussing something that is much more difficult."

Nevertheless, for many, if not most, particle physicists, some form of linear collider seems like the best bet. In June, the International Committee for Future Accelerators, headquartered at Fermilab in Batavia, Illinois, brought the ILC and CLIC together under a single Linear Collider project, headed by former LHC director Lyn Evans. His aim is to deliver a proposal for a single linear collider by the end of 2015.

> "WE SHOULDN'T TAKE IT FOR GRANTED THAT MONEY IS AVAILABLE JUST BECAUSE THE HIGGS HAS BEEN FOUND."

A sensible plan, thinks Evans, is to build a linear collider starting at 250 GeV to probe the Higgs, and then boost its energy in stages until it reaches 500 GeV. At that point it could produce pairs of Higgs bosons and allow researchers to explore how the Higgs couples to itself and also interacts with the heaviest particle of matter, the top quark. Going to higher energies is technically feasible, he says, but requires more electricity — as much as a medium power station's worth. In practice, he says, "I think an upper limit in power [on the hypothetical new site] is the maximum that can be supplied to the CERN site, which is 300 MW."

Technology aside, the multi-billion-dollar question is who would



host the next lepton collider. A rule of thumb is that the host country puts up half the cost in expectation of long-term economic returns, says Foster. But this is not a good economic period to be making that case, especially not for a project that, from a politician's point of view, has no short-term benefit to voters.

GOING GLOBAL

If a linear collider is to be approved in the next few years, says Evans, it will probably not be built at CERN. Despite the European lab's wealth of technical and political infrastructure, it has its hands full with the LHC, which isn't even scheduled to reach its design energy of 7 TeV per beam until 2014 and is also scheduled to undergo a luminosity upgrade around 2022. "I'd bet that the highest priority of the European strategy workshop will be continuing to exploit and upgrade the LHC," says John Womersley, chief executive of Britain's Science and Technology Facilities Council, which controls the country's spending on particle physics.

The United States is also an unlikely site for a new collider, says Fermilab director Pier Oddone, who is chair of the International Committee for Future Accelerators. "Something drastic would have to change," he says. After the closure of Fermilab's 2-TeV Tevatron collider, the energy frontier crossed from the United States to Europe. So the current US strategy is to concentrate on the 'intensity frontier', studying rare particle interactions produced by, for example, intense beams of neutrinos. And yet, says Oddone, "we had a fairly severe budget cut at the beginning of this year and had problems fitting in a facility [a long-baseline neutrino experiment] that costs one-tenth of the ILC". Oddone says that it would also be "very difficult" at this time for the United States to contribute much to a lepton collider built elsewhere.

Many observers think that by far the strongest candidate to host the next project is Japan. After all, notes Evans, Japan made a significant contribution to the LHC in the mid-1990s when the project was under financial strain. "Perhaps it's time for Europe to return the favour," he says. The Japanese premier made positive references to the ILC in December 2011, just after the first preliminary sightings of the new boson were announced. There is a scent of extra funds, because the new accelerator is being discussed as part of a broader economic plan to boost regions devastated by the March 2011 earthquake; the idea is to make it the hub of an 'international city' comprising other research laboratories, industrial zones and education centres. And as Japanese particle physicists update their five-year roadmap this year, the ILC remains at the top of their new-project wish-list. Specifically, explains Atsuto Suzuki, director-general of the KEK laboratory in Tsukuba, the community's recommendation was that "Japan should take leadership of the early realization of an electron–positron linear collider should a particle such as a Higgs boson be confirmed at the LHC".

So is an ILC finally looking like a safe bet? "Good god, no!" says Foster, "but this is the best chance that we've had in a long time." Womersley gives odds of the ILC being built as 50:50 at best. "We shouldn't take it for granted that money is available just because the Higgs has been found," he says, pointing out that there are also strong cases for next-generation neutrino experiments, for example. It would take around ten years from breaking ground to operating an ILC, estimates Oddone, plus the preparatory time. "You're talking 2025 at the earliest, but do you launch such a major project before you know what else the LHC might find? There could be things much wilder than the Higgs."

For many particle physicists, the dream scenario is the LHC exploring the high-energy frontier in Europe; multiple neutrino experiments exploring the intensity frontier in the United States; and a new lepton collider in Japan pinning down the details of all the exotic new particles that so far have not turned up in the LHC's collisions. "I would love to see us going in that direction, if countries put their weight behind the programmes in each region," says Terry Wyatt, a physicist at the University of Manchester, UK, who works on the ATLAS detector at the LHC.

As always in the world of big science, however, making such dreams come true is a question of making the sale to outsiders. "These things would probably be solved outside the particle-physics sphere," says Oddone. "It might be a phone call between a president and a prime minister that decides it." ■ SEE COMMENT P. 581

NATURE.COM
For more about the Higgs boson discovery, see:
go.nature.com/jbnj5l

Matthew Chalmers *is a freelance writer in Bristol, UK.*

Under attack: policemen stand guard outside the Monterrey Institute of Technology and Higher Education after a letter bomb exploded there in August 2011.

ARMED RESISTANCE

Nature assesses the aftermath of a series of nanotechnology-lab bombings in Mexico — and asks how the country became a target of eco-anarchists.

BY LEIGH PHILLIPS

The shoe-box-sized package was addressed to Armando Herrera Corral. It stated that he was the recipient of an award and it was covered in official-looking stamps. Herrera, a computer scientist at the Monterrey Institute of Technology and Higher Education in Mexico City, shook the box a number of times, and something solid jiggled inside. What could it be? He was excited and a little nervous — so much so, that he walked down the hall to the office of a colleague, robotics researcher Alejandro Aceves López, and asked Aceves to open it for him.

Aceves sat down at his desk to tear the box open. So when the 20-centimetre-long pipe bomb inside exploded, on 8 August 2011, Aceves took the full force in his chest. Metal pierced one of his lungs. "He was in intensive care. He was really bad," says Herrera's brother Gerardo, a theoretical physicist at the nearby Centre for Research and Advanced Studies of

A. FRANCO/AP/PRESS ASSOCIATION IMAGES

the National Polytechnic Institute (Cinvestav). Armando Herrera Corral, who was standing nearby when the bomb went off, escaped with a burst eardrum and burns to his legs.

The next day, an eco-anarchist group calling itself Individuals Tending Towards Savagery (ITS) claimed responsibility for the bombing in a 5,500-word diatribe against nanotechnology that it published online. Police found a charred copy of a similar text in the detritus of the explosion. The bombers said that Herrera had been targeted for his role as director of the technology-transfer centre at the Monterrey Institute of Technology and Higher Education (commonly known as Monterrey Tec), "one of the major universities that has staked everything on the development of nanotechnology". The text talked of the potential for the field to cause environmental "nanocontamination", and concluded that technology and civilization as a whole should be held responsible for any environmental catastrophe. Chillingly, the bombers listed another five researchers at Monterrey Tec as presumptive targets, as well as a further six universities.

The incident had precedent. The ITS had already claimed responsibility for bomb attacks in April and May 2011, both targeting Carlos Alberto Camacho Olguín, head of engineering and nanotechnology at the Polytechnic University of the Valley of Mexico in Tultitlán. The first bomb wounded a security guard; the second was identified and disposed of before anyone could be hurt. Last December, the group struck again — this time at the Polytechnic University of Pachuca, where a package containing gunpowder exploded in the hand of a teacher, causing minor burns (see 'A litany of letter bombs'). No other developing country has suffered a comparable string of anti-technology attacks.

CLOSING RANKS

One year on from the bombing at Monterrey Tec, the repercussions are still being felt. Armando Herrera Corral and Aceves will not speak to *Nature* about what happened. "It's too sensitive, you understand?" is all Aceves would say. Herrera has left his job as director of the university's technology park and is now head of postgraduate studies. Other Mexican universities with nanotechnology research programmes have evacuated campuses in response to bomb threats, and universities across the country have introduced stringent security measures. Some researchers are anxious for their own safety; some are furious about being targets. But all the researchers that *Nature* spoke to in Mexico are adamant that the attacks will not discourage them from their research or dissuade students from entering the field.

NATURE.COM
To listen to a podcast on eco-anarchism in Mexico, visit:
go.nature.com/eespxa

So far, there has been little explanation of where the vitriol is coming from. Why are radical environmental groups targeting nanotechnology? Is this field being confronted with the same sort of militant hostility that has dogged genetic-modification research and animal testing? And why Mexico?

Reporting by *Nature* suggests that several broad trends have come together to precipitate the violence. Over the past decade, Mexico has invested heavily in nanotechnology relative to other developing countries, because it sees the field as a route to economic development; mainstream green groups worldwide have grown increasingly concerned about nanotechnology's health and environmental risks; and there has been a shift towards extreme ideas and tactics among radical environmentalists critical of technology. In Mexico, this has been set against a general background of growing violence and political upheaval.

The bombings come at a pivotal moment. Those who study public perception of risk say that the public discourse about nanotechnology is currently fairly moderate but could easily become more polarized. Until the bombings, the radical environmental movement had mostly restricted itself to non-violent actions and property destruction, says Richard Widick, a sociologist at the University of California, Santa Barbara. But, he says, the global economic crisis and the growing perception that ecological catastrophe is imminent could fuel further attacks. "More and more people who have hitherto been able to restrain themselves will just go over the edge," says Widick. "We are going deeper still into an era of deepening and proliferating extremisms. I see a future of environmental struggles marred by violence of every variety."

"WE ARE GOING DEEPER STILL INTO AN ERA OF DEEPENING AND PROLIFERATING EXTREMISMS."

That violence leaves scars. According to Gerardo Herrera Corral, Aceves "still has problems and will do for the rest of his life. There's a piece of shrapnel in his lung they couldn't take out, close to his heart." And only amateurism by the bombers prevented the attack at Monterrey Tec from having more tragic consequences: the police say that only about 8 centimetres of the dynamite in the pipe detonated. The bombers had packed it in such a way that the rest did not burn.

If all the dynamite had gone off, the police say, it could have destroyed the whole building — as well as Herrera, Aceves and dozens of researchers who work alongside them.

Mexico started a concerted nanotechnology push in 2002, when the government identified the field as a strategic sector for development. Dozens of public research institutes signed agreements with foreign institutions, companies and the military, and many opened graduate courses focused on nanotechnology research. Along with other Latin American countries that have invested in the field — Brazil and Argentina, in particular — Mexico views nanotechnology as a pathway to a more powerful research and industrial base. "They see it as a recipe for transition to the knowledge economy. It's less an option than a necessity," says Guillermo Foladori, an anthropologist at the Autonomous University of Zacatecas in Mexico and coordinator of a group of academics studying the regional growth of the field. The most important university in Mexico for nanotechnology, says Foladori, is Monterrey Tec.

TECHNOLOGY BACKLASH

As nanotechnology has been growing in Latin America, a violent eco-anarchist philosophy has taken root among certain radical groups in Mexico. Mexican intelligence services believe that the perpetrators of the bombings last year were mainly young and well educated: their communiqués are littered with references to English-language texts unlikely to have been translated into Spanish. Intelligence services say that the eco-anarchist groups have been around for about a decade. They started off protesting against Mexico's economic and political system by setting off small explosives that destroyed bank machines.

But around 2008, certain groups began to adopt an 'anarcho-primitivist' perspective. (Locally, they are called *primativistas*, says Gerardo Herrera Corral.) This philosophy had won little notice until the past few years, but with increasing media reports of looming global climate disaster, some radical green activists have latched on to it. California-based environmental writer Derrick Jensen — whose popular books call for an underground network of 'Deep Green Resistance' cells — is a highly influential figure in this otherwise leaderless movement, which argues that industrial civilization is responsible for environmental destruction and must be dismantled.

In their writings, anarcho-primitivist groups often express deep anxiety about a range of advanced research subjects, including genetic engineering, cloning, synthetic biology, geoengineering and neurosciences. But it is nanotechnology, a common subject for science-fiction doomsday scenarios, that most clearly symbolizes to them the power of modern science run amok. "Nanotechnology is the furthest advancement that may yet exist in the history of anthropocentric progress," the ITS wrote in its first communiqué, in April 2011.

The same network of 'anti-civilization' anarchists has graduated to violence elsewhere. Attacks include the 2010 attempted bombing of IBM's flagship nanotechnology lab near Zurich, Switzerland, and the non-lethal

shooting in May this year of Roberto Adinolfi, a nuclear engineer for a subsidiary of Italian industrial conglomerate Finmeccanica, which was targeted for its links to nanotechnology (see *Nature* **485,** 561; 2012).

In Mexico, the existing social and political climate may have helped light the fuse, says Miguel Méndez Rojas, coordinator of the department of nanotechnology and molecular engineering at the University of the Americas Puebla in Mexico. He says that the bombings cannot be understood outside the context of what he describes as a dangerous cocktail of poverty and poor education, widespread ignorance of science, ongoing social upheaval and a climate of violence. In July, Mexico City saw some of the country's largest-ever protests, over alleged fraud in this year's presidential election. And since 2006, wars with the major drug gangs have resulted in around 55,000 deaths. Human-rights groups have accused the military and police of illegal arrests, secret and prolonged detention, torture, rape and extrajudicial execution. "I think we are in just the moment for a social explosion," says Méndez Rojas.

Taken together, all these developments made Mexican universities, with their burgeoning nanotechnology industry, a target for violence. In its communiqué from May last year, the ITS warned professors and students: "It would be best for them to walk carefully within and outside the university, that they take warning of every suspicious shape in rooms, buildings, parking areas and campus, because one of these days, we are going to make them pay for everything that they want to do to the Earth with these kinds of nano-scale technologies."

ESCALATING TENSION

The "boom in eco-anarchism" — as CNN Mexico describes it — has had widespread consequences. In the wake of the bombings, officials at Monterrey Tec introduced a slew of security procedures, including sniffer dogs and campus sweeps. Similar procedures have been put in place at the University of the Americas. The institution's Puebla campus was home to the first nanotechnology lab in Mexico, and its site in Monterrey was the first campus in Latin America to offer an undergraduate programme in the field.

"We were very worried that we could be a target," says Méndez Rojas, whose research encompasses the development of nanomaterials for tackling cancer and simple toxicology tests on nanoparticles. After the first attacks last year, he was warned that the ITS was going to target campuses outside Mexico City. On his suggestion, he says, the university formed a task force of professors, security staff and administrators to respond to threats. The campus implemented car checks and a policy that visitors can meet professors only with an appointment; a visitor today undergoes a 15-minute identity check, and is escorted to their meeting by two security guards. Méndez Rojas says that he doesn't receive some visitors as a result, but that, despite the hassles, "I feel safer".

There have been false alarms, including one at Méndez Rojas' university last August. In all, at least ten campuses have received bomb threats, although it is unclear whether they were sent by the ITS or copycats. Greenpeace Mexico, criticized by the ITS for having a soft stance on environmental issues, received an incendiary device from the group last November. Universities in seven states and the capital city have implemented increased security controls, including random bag checks and bomb-evacuation drills, but the Mexican National Association of Universities and Higher Education Institutions warns that only one-third of campuses in the country have taken sufficient action.

The increased security has met with criticism from some quarters. In March, Hugo Aboites, an education specialist at the Autonomous Metropolitan University in Xochimilco, told *La Jornada*, one of the country's leading national daily newspapers, that stringent security precautions could create an environment of "institutionalized fear". The role of universities, he said, is to "train and impart knowledge, not to reproduce police control of the population".

But Méndez Rojas says that research activities have not been thrown off course. Despite the attacks, he says, the number of students enrolled in nanotechnology programmes across the country rose to 800 this year, up from 500 in 2011. "Apart from the fear some people may be feeling about the subject, not much will change in the academic community. Researchers in nanoscience and nanotechnology won't switch. They'd lose decades' worth of work and millions in investment," he says.

"THESE KINDS OF ATTACKS — THEY ARE BENEFITING THE DEVELOPMENT OF NANOTECHNOLOGY. IT POLARIZED THE DISCUSSION."

Some researchers in Mexico say that more-moderate groups are stoking fears about nanotechnology. One such body is the Action Group on Erosion, Technology and Concentration (ETC, pronounced et cetera), a small but vocal non-profit organization based in Ottawa, Canada, which was one of the first to raise concerns about nanotechnology and has to a large extent framed the international discussion. Silvia Ribeiro, the group's Latin America director, based in Mexico City, says that the organization has no links to the ITS. The bombings were a "sick development", she says. "These kinds of attacks — they are benefiting the development of nanotechnology," she says. "It polarized the discussion. Do you want nanotech or the bomb?"

ETC wants to see a moratorium on all nanotechnology research, says Ribeiro, who is the lead author on many of the group's reports criticizing nanotechnology research and commercialization. She says that there have not been enough toxicological studies on engineered nanoparticles, and that no government has developed a regulatory regime that explicitly addresses risk at the nanoscale.

However, ETC also infuriates researchers by issuing warnings of a more speculative nature. For example, it has latched on to the concept of 'grey goo' — self-replicating nanorobots run wild — that was raised in the book *Engines of Creation* (Doubleday, 1986) by nanotechnology engineer Eric Drexler. In ETC's primer on nanoscale technologies, it says that the "likely future threat is that the merger of living and non-living matter will result in hybrid organisms and products that are not easy to control and behave in unpredictable ways".

Ribeiro has also criticized genetic modification and vaccination against human papillomavirus in a weekly column in *La Jornada*. Méndez Rojas says that ETC "promotes beliefs, but they are not based on facts, and we need a public discussion of the facts".

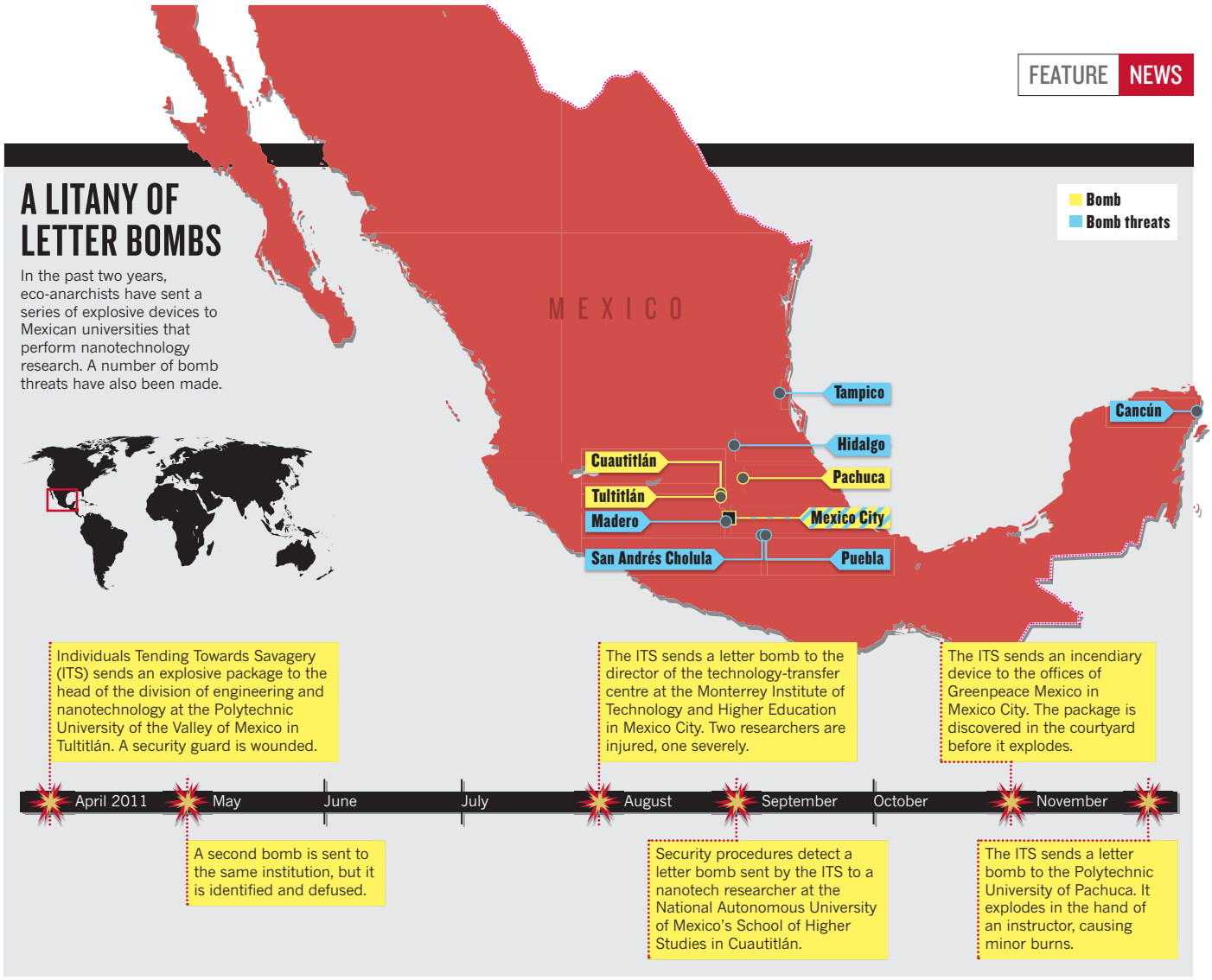
The sentiment is echoed by Beatriz Xoconostle Cázares, a biotechnology researcher at Cinvestav, who is experimenting with transgenic crops resistant to drought and insects — and who regularly debates with ETC in public forums. Last September, Xoconostle arrived at work to find that her lab had been set on fire. A month later, arsonists attacked the lab of a neighbouring researcher.

OPEN DEBATE

Xoconostle does not accuse ETC of responsibility for these acts, but she worries that the organization's communications are helping to spread fears about technology. "These are small groups. But they know how to communicate, and that's a huge advantage. It's becoming a larger group of people who oppose these things." Xoconostle fears that extremist groups might adopt such views and use them to support their acts.

Ribeiro denies that ETC's reports are not based on facts and says that "we have nothing to do with ITS and we strongly and publicly have condemned their violence. Those who exercise violence and those who bluntly and uncritically defend nanotech coincide in hindering a real public open debate on the facts."

The real question now is whether the violence will recur — or spread. The nanotechnology-activist movement does seem to be gaining momentum. For the past four years, nano-critical groups have held an annual International Nanotechnology Activist Summit; the



one last October welcomed 14 environmental and consumer-advocacy groups worldwide, including the European Environmental Bureau — a Brussels-based federation of European green groups, which says it represents a combined membership of 15 million people.

Opposition to nanotechnology has sometimes been hostile outside Mexico. In 2009 and 2010, protesters in France shut down public debates on nanotechnology in Grenoble, Rennes, Lyons and Marseilles. Pièces et Main d'Oeuvre (Parts and Labour), a Grenoble-based group, has organized protests in the city outside Minatec, France's flagship nanotechnology research centre.

But Barbara Harthorn, director of the Center for Nanotechnology in Society at the University of California, Santa Barbara, says that most debate about nanotechnology so far has been measured. She has tracked 125 green groups around the world in an ongoing study of engagement in nanotechnology by non-governmental organizations. She says that most groups restrict themselves to issues of environmental health and safety rather than the more speculative scenarios painted by ETC and the ITS.

At the same time, public awareness of the topic is extremely low, says Harthorn. She collaborated on a meta-analysis of 22 surveys done in the United States, Canada, Europe and Japan between 2002 and 2009, which found that, on average, more than 51% of survey respondents report that they know "nothing at all" about nanotechnology (T. Satterfield *et al. Nature Nanotechnol.* **4,** 752–758; 2009).

"There's a huge public that is undecided, which means that opinion is still highly malleable," says Harthorn. Her own surveys have shown no evidence that the public in general has the same aversion to nanotechnology that has been seen for genetic engineering, because nanotechnology is not viewed as 'messing with nature' in the same way. But subjects' reactions depend on the type of nanotechnology being considered: applications in clean energy are embraced, but uses in food or the far-reaching idea of human 'nano-enhancement' elicit a sharply negative reaction. All this means that there is still a lot to play for in public perception, says Harthorn. If the discourse becomes framed by more speculative notions, the moderate public stance could be lost.

And that creates an opportunity for scientists to tip the debate. Most nanotechnology researchers acknowledge that some areas of their work raise legitimate environmental, health and safety concerns. The most important response, says Gerardo Herrera Corral, is for scientists to engage with the public to address and dispel concerns. Herrera is head of Mexico's only experiment at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, and he points to how CERN dealt with public fears that its Large Hadron Collider could create a black hole that would swallow Earth. "We set up a committee to deal with this. We looked into the real dangers. There were journal articles and we answered all the e-mails we got from people. I mean top-level physicists answering thousands of e-mails."

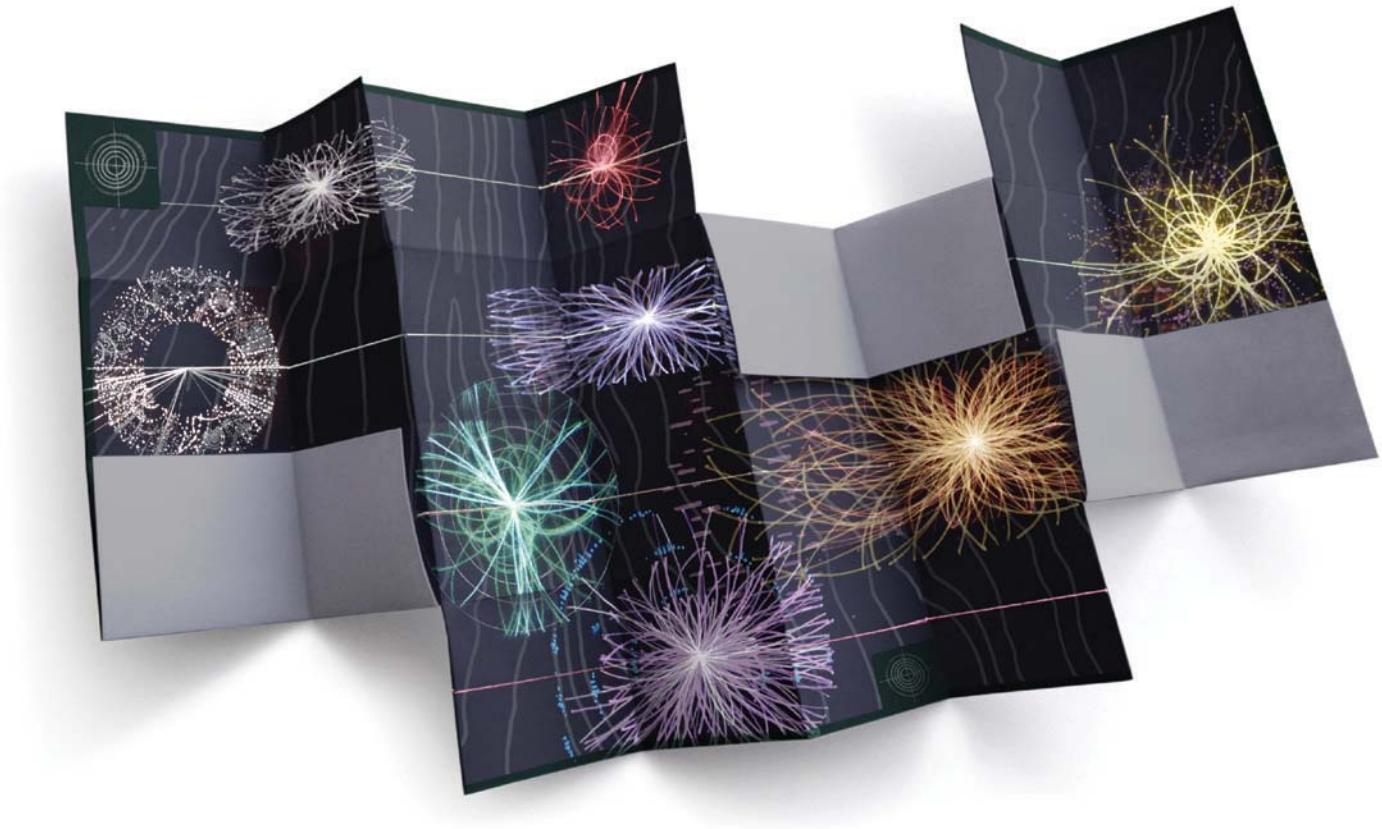
"But this is work we should all be doing," says Herrera. "Even if it's extra work on top of all the other things we have to do. It's just part of our job now."

In Mexico, bomb threats are also becoming part of the job. On 31 May, a hoax threat forced evacuations at the University of Xalapa. The same day, emergency services and military forces descended on the faculty of engineering at the University of Veracruz in Boca del Rio after a suspicious device was found. It turned out to be a professor's forgotten briefcase.

For Xoconostle, the fear is taking its toll. "The fact is I am kind of worried. I'm terrified of these people," the soft-spoken scientist says. "We are in a fight." ■

Leigh Phillips *is an International Development Research Centre fellow at* Nature.

COMMENT



Beyond the Higgs

The Higgs boson is not the end of the story. There is more to map in the new world of extreme physics, says **Jon Butterworth**.

Finding the Higgs boson was on the 'to-do' list for particle physics for so long that I still haven't absorbed the fact that every lecture course, textbook and popular seminar in the field now needs revision.

The Higgs is predicted by the standard model of physics, and because it gives mass to all the fundamental particles, it is required for the model to work. The announcement last month of its discovery at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, was a triumph for one mathematical approach to understanding the world. One could be forgiven for assuming that all particle physicists can now retire to their Geneva chalets or travel the world as pundits. Particle physics: done.

In fact, the discovery of the Higgs marks the dawn of a new era for researchers in extreme-energy physics. After all, we need to understand why the masses of fundamental particles have the values they do, and why some of the patterns in these fundamental particles are as they are.

In the early 1960s, several physicists concluded that, for the standard model to work, the Universe would need to be filled with a quantum field that gives mass to stuff. Peter Higgs said that if this field were there, it would have waves in it — Higgs bosons. Observing them is the only way to know whether the field really exists. It does.

Such vindication alone is strong encouragement to continue — to measure the Higgs boson and see whether it behaves as predicted, and whether it offers clues to other outstanding questions. We have a wild new frontier of physics to explore.

Particle physicists think in terms of energy scales. The strong nuclear force sets a scale at roughly the mass of the proton, equivalent ▶

ILLUSTRATION BY BRENDAN MONROE

to about 1 gigaelectronvolt (GeV). The Large Hadron Collider (LHC) at CERN allows us to study physics up to and beyond the scale at which electroweak symmetry is broken, around 100 GeV. Above this energy, a symmetry exists between the weak nuclear force and electromagnetism that is not present at lower energies. The Higgs is what breaks this electroweak symmetry, and so its mass (around 125 GeV) is on a similar scale.

The Planck scale, on which gravity and quantum forces combine, is the only other fundamental energy scale that has been confirmed. It lies at about 10^{19} GeV, far beyond the reach of current or planned experiments. There must surely be new physics between 100 GeV and 10 billion billion GeV. The LHC will get us partway into this unknown region, but to go beyond that we will need all the clues we can find.

THE GREAT UNKNOWN

First, there are many questions to answer about the nature of the Higgs boson. One is why it is so light. If one assembles the standard model without fine-tuning some parameters, quantum effects mean that the Higgs boson's mass should grow and end up near the Planck scale. This is clearly wrong, and it hints at gaps in the theory.

Supersymmetry and extra space-time dimensions have been proposed as solutions to this problem. Supersymmetry introduces a set of particles that cancel out the quantum effects that would otherwise make the Higgs heavier. Extra dimensions could bring the Planck scale closer to 100 GeV. There is no direct evidence for either theory, but new particles or deviations from the standard model's predictions could turn up at the LHC any time, especially after the machine doubles its energy in 2014.

Reaching ever higher energies requires new developments in technology. The LHC is the biggest project ever in particle physics, yet the CERN budget has not grown in real terms. Research towards making higher-energy beams includes using high-current, low-energy beams to drive higher-energy, lower-current ones; accelerating electrons in the electromagnetic wake of a proton beam; and colliding muons (instead of protons and electrons) to avoid energy sharing between quarks and gluons and energy loss to synchrotron radiation. Not all of these approaches will work, but some might, which would open up new regimes of physics to explore.

Some discoveries can be made without reaching higher energies — the LHC's fast data rate allows for the observation of rare events. Several experiments at the LHC use this feature to address matter asymmetry. Our world is made of matter, not antimatter, yet most fundamental forces do not distinguish between the two. How was this imbalance generated? The LHC 'beauty' experiment, LHCb, is searching for tiny differences in the production of particles and antiparticles in decays involving *b* quarks.

Neutrino experiments might turn out to be our most significant inroad for decades to understanding the dominance of matter. Recent measurements at the RENO experiment in Seoul, South Korea, and the Daya Bay experiment in China suggest that matter–antimatter asymmetry might be especially high among neutrinos, which have a small mass and transform, or 'mix', readily among different types. Proposed experiments in the United States, Europe and Japan could follow up on this possibility.

Neutrinos are unique and mysterious in other ways. Many theorists think that the neutrino is a Majorana fermion — meaning it is its own antiparticle. If so, the neutrino's mass — unlike that of all other particles — does not come from the Higgs boson. Sensitive searches are now under way to find rare isotope decays that could tell us whether or not this is the case.

Other puzzles remain in the standard model, such as dark matter — invisible matter that affects the motions of stars and galaxies but does not seem to be made of any known particle. And we attribute the acceleration of the Universe's expansion to dark energy, but we do not know the physics behind it. A better understanding of how the strong interaction binds quarks and gluons into hadrons would be welcome. Current theory cannot predict how hadrons scatter at very high energies. Quarks and gluons seem to form a liquid at high energy densities, but this form of matter is poorly understood.

The standard model gives us a list of apparently fundamental particles, the only stuff in the Universe that is not made of other stuff. With the discovery of the Higgs, we now have a theory within which these particles can have mass, and this is a huge step forward.

Now we can relish the next steps, at the LHC and beyond. Perhaps researchers can build a linear collider to serve as a 'Higgs factory'. Perhaps another experiment will deliver a surprise that changes everything. Perhaps a breakthrough in the theory will explain the coincidences and patterns of the standard model.

The patterns seen in the periodic table before anyone knew about electrons and nuclei turned out to be a sign of the underlying structure of atoms. Maybe there is another layer of substructure that explains the patterns we see now in particle physics.

■ SEE NEWS FEATURE P. 572

Jon Butterworth *is professor of physics and head of the Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.*
e-mail: j.butterworth@ucl.ac.uk

G. M. B. AKASH/PANOS

Extreme weather, like this flood in 2004 that displaced 30 million people in Bangladesh, is predicted to become more frequent as the world continues to warm.

Time to raft up

Climate scientists should learn from the naysayers and pull together to get their message across, says **Chris Rapley**.

I recently discussed climate change with a British right-wing politician. We both claimed the rational and impartial high ground. But our minds did not meet. He offered stock dismissive arguments: the amplifying processes aren't happening; temperature changes won't be significant; humanity will adapt. He saw climate-change mitigation as a threat to economic progress, and deregulated markets as able to solve any problem. His parting shot was triumphalist: "Among key political power-brokers your case has been lost!"

In the United States, Canada and Australia, climate science has been the focus of right-wing political attack for some time. In the United Kingdom, however, this is a new development. Only four years ago, the UK Climate Change Act passed into law with all-party support. And in 2010, the Conservative Prime Minister, David Cameron, said that he wanted the coalition administration to be the "greenest government ever". Yet recent decisions by the same government will reduce green-energy subsidies, backsliding on previous commitments.

At the international level, the partisan deadlock within the United States has crippled progress towards global emissions reductions. Todd Stern, the US special envoy for climate change, has now suggested that the widely adopted 2 °C limit on global warming may have to be abandoned (see go.nature.com/q7gmvo).

Evidently, the voices of dismissal are trumping the messages of science. A significant factor in their success is an effective communications strategy, which the climate-science community has yet to learn or use. An initiative to redress the balance is crucial if policy-making is to be based on evidence, and if the risks of further prevarication are to be made clear. As political scientists Daniel Sarewitz, Roger Pielke Jr and others have pointed out, from the perspective of policy, "We know enough!".

NATURE.COM
For more on fixing failures in communication:
go.nature.com/mx5fzb

REALITY CHECK

A first step is to understand how the dismissal of climate change is tenable when the evidence to the contrary is so extensive and compelling. Much has been published on this by social scientists and psychologists, but that does not mean that it has been read, understood or assimilated by the climate-science community. As former director of London's Science Museum, the British Antarctic Survey and the International Geosphere–Biosphere Programme, my experience is that it has not, particularly at senior levels.

Some material has been picked up, such as Naomi Oreskes and Erik Conway's book

Merchants of Doubt (Bloomsbury, 2010). Its analysis of the political influence of a network of libertarian activists to stall emissions regulation is understood as a plot perpetuated by "evil" vested interests. This offers a simplistic explanation of an otherwise-perplexing situation. But the insightful writings of social and political scientists, explaining the deeper mechanisms at play, remain largely unknown to most natural scientists. This needs to change.

Part of the problem is that researchers are busy and overwhelmed by information. A lead author on one section of the upcoming report by the Intergovernmental Panel on Climate Change (IPCC) told me that more than 800 papers had been published on her subject in a single year, leaving little time to read more broadly. It is understandable, therefore, that there is a tendency to skip material from unfamiliar fields and authors. But to be of value to society, climate scientists need to master ways to communicate their results effectively (see go.nature.com/euzzf7).

There are also some uncomfortable truths to confront. The unauthorized release of e-mails from the Climatic Research Unit at the University of East Anglia in Norwich, UK, in November 2009 — known as 'Climategate' — has left an aftermath that still needs to be cleared up. The climate-science leadership, fixated on delivering more of the same research and seemingly oblivious to changing realities, has lost its way. In my opinion, the community is in denial about these issues. We climate scientists — from disciplines both natural and social — need to align our purpose, re-establish our legitimacy, identify and understand our target audiences and decide how best to express our message. Above all, we need to develop a new, coherent initiative to engage collectively and actively in the political and public discourse.

WHY DON'T THEY GET IT?

A great deal is known about why people reject the messages of climate science. According to Daniel Kahneman in his 2012 book *Thinking Fast and Slow* (Penguin), the human mind can believe almost anything. Unconscious values, attitudes and beliefs filter our assimilation of evidence[1]. We are influenced by the views of those with whom we identify, and whose esteem we seek.

Once a mindset is established, dissonant facts are met with resistance. As the economist J. K. Galbraith observed, "faced with the choice of changing one's mind and with proving there is no need to do so, almost everyone gets busy on the proof." The more an individual has invested in their position, the more strongly they will seek to justify it[2]. The greater their scientific training, the more effective they will be at doing so.

We know from Oreskes and Conway's book that vested interests have worked to sow doubt. And the media practice of offering 'balanced reporting' has reinforced public perceptions of scientific disagreement, which is a barrier to engagement and belief[3]. But the misinformation campaigns resonate because they tap into deeper human needs.

People go to great lengths to avoid or eliminate anxiety[4]. The implications of climate change are profoundly worrying, so people discount, disavow or deny the discomforting news. Blaming or denigrating the messenger is a common next step.

However, anxiety alone does not account for the fervour of those who are dismissive. An association between climate-dismissive attitudes and people with an individualistic outlook and libertarian politics[5] suggests an ideological root to such beliefs. The implications of climate change challenge fundamental premises about the governance, funding and fuelling of the modern world. Polarized reactions are not surprising. As I have found in my discussions with politicians, far from having been ineffectively communicated, the messages of climate science may have been understood, in a visceral sense, all too clearly — especially by the political right.

NEED TO REGROUP

So what is a climate scientist to do? First, we must recognize and accept that, whether in discussion with decision-makers or the public, we are inextricably embroiled in the policy debate. The science is complex, the projections are uncertain and the social implications are great. We need to respond to questions that go beyond facts, such as 'What does this mean for me?' and 'What are our options?'.

As Roger Pielke Jr discusses in his book *The Honest Broker* (Cambridge Univ. Press, 2007), we need to choose the role that is most appropriate to a given situation and make that choice clear. To draw attention to the risk to food supplies of an increased probability of extreme weather events is to act as an 'issue advocate'. To lay out the climatic consequences on the global food supply of the broadest possible range of mitigation alternatives is to adopt the role of an 'honest broker'. There are dangers. To stray into policy-advocacy or activism is to step beyond the domain of science, and risks undermining legitimacy through the perception — or reality — of a loss of impartiality.

"Once a mindset is established, dissonant facts are met with resistance."

However, as Sarewitz[6] has pointed out, scientists carry authority "in advocating for one particular fact-based interpretation of the world over another". So acting as a 'science arbiter' — explaining the evidence and contesting misinterpretations — is part of the day job.

REBUILDING TRUST

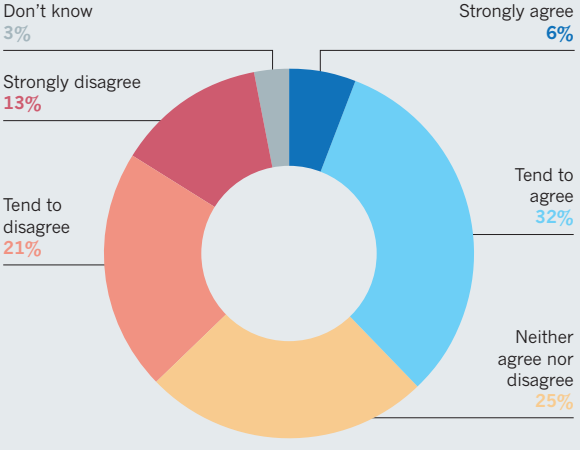
When faced with implacable disagreement, non-experts must decide who to believe. The issue of trust is therefore paramount. And therein lies a problem.

Although levels of trust in scientists are generally high compared to other professions, a study in early 2011 found that only one-third of UK respondents agreed with the statement: "We can trust climate scientists to tell us the truth about climate change"[7]. One-third disagreed (see 'Levels of trust'). Such expressions of distrust have been linked with the allegations that followed Climategate, as well as with the assertion by right-wing politicians that climate scientists are perpetrating a 'hoax'. So the community needs to repair its reputation. But how?

I propose that, as a public statement of our ideals[8], climate scientists should agree and commit to principles of professional conduct — possibly through an equivalent of the medical profession's Hippocratic oath. These principles would cover standards of work, issues of impartiality, transparency of process or accessibility of data, and a willingness to engage positively with non-specialists. The 2010 Singapore statement on research integrity offers a framework on which to build (see www.singaporestatement.org). Weaknesses in the peer-review

LEVELS OF TRUST

A recent UK survey found that about one-third of the public agrees with the statement "We can trust climate scientists to tell us the truth about climate change" and that about one-third disagrees. This is concerning in light of other surveys indicating that more than two-thirds of the UK public trust scientists to tell them the truth generally.



SOURCE: REF. 7

process — concerning conflicts of interest and the degree and effectiveness of critical challenge — need to be addressed jointly by scientists and journal editors, and the solutions made transparent and public.

RULES OF ENGAGEMENT

The climate-dismissive think tanks and organizations have been effective because they have understood and put into practice the insights of social science. They deliver simple messages that are crafted to agree with specific value sets and world views. Their flow of commentary is persistent, consistent and backed up with material that provides deeper arguments. Their narrative is spread and amplified by sympathetic sectors of the media and politics that they have nurtured in person.

In contrast, the climate-science community delivers messages to policy-makers and the public that are often highly technical and detailed. These tend to be fragmented, emphasize uncertainty and are oblivious to the emotions and associations that they trigger. There remains a widespread reliance on the flawed information-deficit model, in which non-experts are viewed by experts as empty vessels who can simply be filled with the 'truth'.

The means of delivering such messages is often through debate, which reinforces partisanship and the impression that established facts remain unresolved. More effective are one-to-one meetings or audience-led dialogue events, such as those developed by the Science Museum's Dana Centre. This centre engages small discussion groups with experts who cover a range of viewpoints, giving people the opportunity to explore controversial issues and allowing them to make up their own minds[9].

Regarding the vast body of evidence on which all climate scientists agree, we need to offer a narrative that is persistent, consistent and underpinned by compelling background material. We need to recognize that this is a necessary but insufficient condition for moving political decision-making towards a practical response. We need to appreciate that the things we climate scientists don't agree on — nuanced disputes at the frontier of our field — are not relevant to policy-making other than to define the current limits of what we know. And we must engage with newspaper editors and politicians in person.

A CALL TO ACTION

Who could organize such an initiative? Climate science extends across many disciplines, represented by no single overarching professional body or voice. One possibility is the Paris-based International Council for Science (ICSU), together with its academic and intergovernmental partners. In March this year, it launched its ten-year strategy, known as Future Earth, which is aimed at revamping its global-change research programmes to better address the needs of society[10].

A. SCHMIDT/AFP/GETTY

Influential US senator James Inhofe (Republican, Oklahoma) still claims that climate change is a hoax.

However, the document disappoints. It provides a vision for a more structured approach with a wider array of partnerships and a closer engagement with decision-makers. But it is written as if Climategate, the rise of the dismissive think tanks and public disengagement had not occurred. These issues are neither acknowledged nor addressed.

Given the need for rapid progress, and with academic leadership focusing elsewhere, a bottom-up approach may provide the way forward. Experience from the International Polar Year (IPY) in 2007–09 shows how effective this can be. Those of us involved in planning the IPY set out to engage the next generation of polar scientists. We had no clear idea of how this might be done. In the end, the young scientists themselves used social networking to establish the Association of Polar Early Career Scientists (APECS) and to agree and pursue their shared purpose. APECS has become an influential force at the highest levels of polar-science organization and planning.

"There remains a widespread reliance on the flawed information-deficit model."

Similarly, I believe that the Internet provides the forum for like-minded and motivated climate scientists from all disciplines to mobilize and transform the impact of climate science on the public and politics. Such an approach could build on the work of existing climate-science websites and initiatives, such as Skeptical Science, Real Climate, Carbon Brief, and of course the IPCC, to develop a more coherent, prioritized and tailored set of messages than are currently available.

The warning signals from the planet are clear. Now is the moment for our community to adopt the rallying cry of sea kayakers confronted with conditions too challenging to handle alone: "Time to raft up!". ■

Chris Rapley *is a professor of climate science in the Department of Earth Sciences at University College London, UK.*
e-mail: christopher.rapley@ucl.ac.uk

1. Kahan, D. M. *et al. Nature Clim. Change* http://dx.doi.org/10.1038/nclimate1547 (2012).
2. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford University Press, 1957).
3. Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C. & Leiserowitz, A. *Nature Clim. Change* **1,** 462–466 (2011).
4. Weintrobe, S. 'The Difficult Problem of Anxiety in Thinking About Climate Change' in *Engaging With Climate Change: Psychoanalytic and Interdisciplinary Perspectives* (ed. Weintrobe, S.) (Routledge, in the press).
5. Poortinga, W., Spence, A., Whitmarsh, L., Capstick, S. & Pidgeon, N. F. *Global Environ. Change* **21,** 1015–1024 (2011).
6. Sarewitz, D. Workshop Paper, AAAS Scientific Responsibility, Human Rights and Law Program (2012); available at http://go.nature.com/7fpus3.
7. Shuckburgh, E., Robison, R. & Pidgeon, N. *Climate Science, the Public and the News Media* (Living With Environmental Change consortium, in the press).
8. Grundmann, R. *Sci. Technol. Hum. Values* http://dx.doi.org/10.1177/0162243911432318 (2012).
9. McCallie, E. *et al. Mus. Soc. Issues* **2,** 165–184 (2007).
10. International Council for Science. *Future Earth: Research for Global Sustainability, A Framework Document* (2012); available at http://go.nature.com/m3r1im.

SSPL/GETTY

Erasmus Darwin was a prominent member of the Lunar Society, which met monthly to discuss science, politics, art and philosophy.

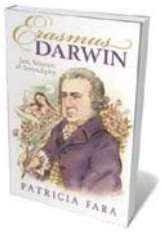
HISTORY

Darwin the eroticist

William Bynum applauds a life of physician and scientific poet Erasmus Darwin, Charles's intriguing grandfather.

Peter Medawar shared a Nobel prize in 1960 for his work on acquired immunological tolerance. But perhaps his best-known work was his 1963 essay 'Is the Scientific Paper a Fraud?' This was not concerned with plagiarism or data falsification: instead, the great biologist was drawing attention to the fact that the finished, published version of scientific research generally presents the paths to the conclusions as simple, logical and smooth. Authors omit the false dawns, failed experiments, discussions with colleagues over coffee or beer, and the many contingencies that almost always accompany a piece of research.

Patricia Fara sets all that complexity centre stage in *Erasmus Darwin: Sex, Science and Serendipity*, her study of the medical doctor, poet, philosopher and abolitionist Erasmus Darwin (1731–1802), grandfather of Charles. Fara leads the reader along the often bumpy trail of her historical research into the life and work of this many-sided man, her attempts to understand his difficult poetry, and her gradual placement of him in the political and social context of his time.



Erasmus Darwin: Sex, Science, and Serendipity
PATRICIA FARA
Oxford University Press: 2012
336 pp. £20.00, $34.95

We follow her as she goes into rare-book rooms, gets stuck in foreign airports, buys facsimile copies of works by Erasmus Darwin and his contemporaries, remembers old trips to museums and tries to make up for her lack of a classical education. (Her undergraduate degree was in physics.) This is not so much a book about Darwin as a book about writing a book about him.

Fara uses this conceit, she tells us, to help readers to appreciate how historians actually work. She uses no impersonal third-person sentence structure of the kind Medawar identified with the scientific paper. So we get sentences such as "I once found myself in a bizarre tourist trap", or "When you get stuck in a research project, one rescue tactic is to let your mind relax and wander about freely". It keeps her narrative flowing and it does portray pretty accurately the ways — casual as well as systematic — in which historical research and writing emerge.

Embedded in Fara's self-conscious account of her project is a convincing study of Erasmus Darwin. When he is not entirely forgotten historically, he often has only a walk-on part in accounts of the life of his remarkable grandson. There is also a tradition, stretching back to Samuel Butler's *Evolution, Old and New* (1879) arguing that all of Charles's major ideas — artificial and

natural selection, sexual selection, adaptation and biological evolution itself — had already been advocated by Erasmus, and that Charles owed much more to his grandfather than he ever admitted.

Fara is less concerned with the relationship of grandfather and grandson than with situating Erasmus Darwin in the history of ideas, and in understanding his poetry. His scientific poems, *The Loves of the Plants*, *The Economy of Vegetation* and *The Temple of Nature*, gave the Lichfield physician a reputation both at home and in Europe.

The first two of these (combined in 1791 to form *The Botanic Garden*) were published during the early years of the French Revolution. Their advanced social and political ideas — such as criticism of slavery — and their perceived erotic content attracted controversy. In Britain at that time, anything smacking of the 'revolutionary' stirred up fears lit by the fires across the Channel: Darwin's friend, the chemist and Unitarian minister Joseph Priestley, published widely on radical issues at the time, and had his house burned down by an anti-French mob. He fled to the United States.

So what of Darwin's poems? In many ways, they are simply ardent celebrations of Britain's industrial achievements, along with fruity expositions of the sexual system on which Linnaeus based his classification of plants. To eighteenth-century sensibilities they seemed arousing — even lines such as "A *hundred* virgins join a *hundred* swains/And fond Adonis leads the sprightly trains". That daring was compounded by Darwin's espousal of biological change through time, his liberal ideas on the evolution of human society, his appreciation of many things French, and his assignment of God to the position of remotely initiating the cosmic process, but taking no further part in its unfolding.

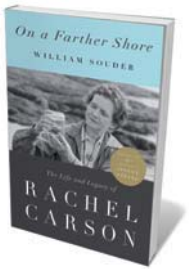
Erasmus Darwin is most commonly seen in association with Priestley, James Watt, Josiah Wedgwood and other members of the Lunar Society. This informal group of industrialists, medical men, inventors and natural philosophers met monthly (by the light of the full Moon) to discuss topics such as the latest discoveries in pneumatic chemistry and ways of building a better steam engine.

Fara explicates both these classical and contemporary reverberations in Erasmus Darwin's writings. She invites modern readers to take him and his ideas seriously, not simply as a relative of the great Charles, but as a substantial figure in his own right. In this she triumphantly succeeds, even while admitting that she doesn't like his poetry. It is not difficult to see why. ■

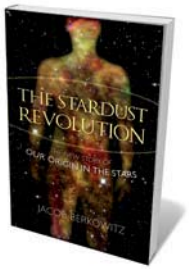
William Bynum *is professor emeritus of medical history at University College London, UK.*
e-mail: w.bynum@ucl.ac.uk

Books in brief

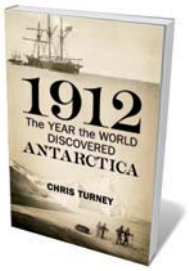
Marie Curie and Her Daughters: The Private Lives of Science's First Family
Shelley Emling PALGRAVE MACMILLAN *256 pp. £16.99 (2012)*
The blazing trajectory of Marie Curie the scientist has left the private woman in the shadows. Shelley Emling delves into the last 20 years of Curie's life, reframing her as mother and humanitarian. The often harrowing tale covers the great physicist's struggle with xenophobia and sexism, her mental and physical breakdowns, and the campaign by American journalist Missy Meloney to supply her with radium. Most compellingly, it bares Curie's relationships with her daughters, the Nobel prize-winning chemist Irène and writer Eve.



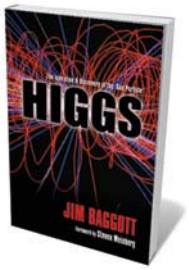
On a Farther Shore: The Life and Legacy of Rachel Carson
William Souder CROWN *544 pp. £19.99 (2012)*
It is 50 years since biologist Rachel Carson's *Silent Spring* appeared in book form, following serialization in *The New Yorker*. This paradigm shift in the way we think about the environment, and the sensitive and searching intelligence behind it, make for a many-layered story eloquently told by William Souder. Carson's pull towards science and development as a writer are explored, and Souder picks his way through the polarized reactions to her book. Acceptance of Carson's warnings on the misuse of pesticides such as DDT was widespread, but protestations from industry and others persist to this day.



The Stardust Revolution: The New Story of Our Origin in the Stars
Jacob Berkowitz PROMETHEUS *312 pp. £23.95 (2012)*
Having set out to write a book on the nexus of evolutionary biology and astronomy, physics writer Jacob Berkowitz ended up at 'extreme genealogy'. We might inherit brown eyes or big bones, but we also carry a heritage from the stars — chemical bonds and cell molecules. Berkowitz is an amiable tour guide: starting with astrochemist Lucy Ziurys's radiotelescope tracking of oceans of molecules pooled across the Milky Way, he finishes in open-ended style at the search for Earth-like planets. In between, he tacks back and forth through the history of astrobiology to entertaining effect.



1912: The Year the World Discovered Antarctica
Chris Turney BODLEY HEAD *368 pp. £20 (2012)*
A century on, the exploits of 'golden age' explorers in Antarctica still grip us. The expeditions of Robert Falcon Scott, Roald Amundsen and Douglas Mawson are oft-told tales, but in this engaging treatment by climate scientist Chris Turney, the forays of Japan's Nobu Shirase and Germany's Wilhelm Filchner get an airing too. The portraits are nuanced. Shirase's crew both bag key geological findings and hunt penguins out of boredom — while Filchner's team, trapped by sea ice on the *Deutschland*, first set up a magnetic observatory in record time, then rapidly succumb to drunken mayhem.



Higgs: The Invention and Discovery of the 'God Particle'
Jim Baggott OXFORD UNIVERSITY PRESS *304 pp. £14.99 (2012)*
Science writer Jim Baggott gives the lowdown on a showdown — one of the biggest in science. The Higgs boson or 'something very much like' it emerged at CERN in Geneva on 4 July, and Baggott chronicles the science leading up to that moment, threaded through with the stories of the main players, from Emmy Noether to Peter Higgs. A foreword by theoretical physicist Steven Weinberg, a useful glossary, and a tendency towards brevity and clarity make for a handy guide to the long hunt for an elusive quarry.

NEUROSCIENCE

The ears have it

Andrew King enjoys a personal account of the impact of sound on life, evolution and the brain.

Why does scraping your fingernails down a blackboard evoke such a strong reaction? This is one of many questions tackled in Seth Horowitz's *The Universal Sense*. Hearing provides us with an incredibly rich source of information about the world and, through speech and music, plays a central part in communication. In this enjoyable book, Horowitz describes the ingenious ways in which different species produce and respond to sound, and how these have shaped the creatures' acoustic environments.

Horowitz begins by asking how noisy Earth might have been before the emergence of the first creatures capable of hearing the primordial sounds of rain, volcanoes and meteorite impacts, and how life changed this soundscape. He then highlights the characteristic features of everyday sounds, such as voices and birdsong, and explains how the acoustic properties of buildings, road surfaces and other environmental objects influence what we actually hear.

Moving on to the auditory abilities of different species, Horowitz looks at the challenge of hearing underwater, where sound travels much faster than it does through air. He also details the specializations that have evolved in frogs and bats for producing and listening to very-low- or very-high-frequency sounds, respectively.

The Universal Sense: How Hearing Shapes the Mind
SETH S HOROWITZ
Bloomsbury: 2012. 320 pp. $25.00, £15.99

After a brief overview of the organization of the hearing regions of the brain, Horowitz describes how effectively sounds can evoke emotional reactions or grab the listener's attention. Of course, any of our senses can achieve this, but acoustic signals, especially music, seem to have a particularly powerful influence on our emotions and memories. This reflection leads neatly into a chapter on the challenge of understanding our response to music. Whether we find a piece of music pleasant or not is much more than a matter of the physics of sound. Horowitz emphasises the need to take account of both cultural and individual differences when attempting to explain our musical preferences, and offers an appropriate critique of studies that claim listening to Mozart can boost your intelligence.

Horowitz looks too at how the use of music and sound effects in entertainment and advertising often relies on principles of auditory perception that were in use long before we understood their biological basis. He discusses the body's reaction to silence and loud sounds, and how certain, often rhythmic, sounds can induce hypnosis and other mind-altering states. As the author points out, a truly silent environment doesn't really exist because sounds that are inaudible to humans can usually be heard easily by other species. Even in the absence of any sound sources, the acoustic properties of the environment influence our perceptions. This is demonstrated by the often unsettling experience of stepping into an anechoic chamber, where the walls suck out the reverberation and echoes that inhabit more typical rooms.

Horowitz's attention to the discomfiting side of the auditory world doesn't stop there. He goes on to consider how the more sinister applications of sound have been exploited in psychological warfare, as illustrated, for example, by the howl of German Stuka aircraft during their dive-bombing raids in the Spanish Civil War and the Second World War.

In the penultimate chapter, Horowitz speculates on the promise of future research, from the challenge of restoring hearing by regrowing the ear's sensory hair cells — which all too easily succumb to age, loud sounds and infection — to listening to the soundscapes of other planets. Finally, he celebrates what he calls the "song of the brain" itself: the distinctive chorus of clicks, signalling neural activity, that can be picked up over an audio monitor when an electrode is advanced into the brain. He speculates that the mind may emerge from this "neuronal orchestra" of activity across the brain much as music consists of — but is also so much more than — a collection of notes.

The Universal Sense is a clear, humorous, personal account, punctuated by numerous anecdotes from Horowitz's own experience as a scientist, electronic-music producer and sound designer. Although these link together the different sections of the book, their inclusion does skew the content towards the author's interests. The book's coverage of the neuroscience of hearing is patchy. For example, Horowitz says little about the possible neural basis for differences in musical or linguistic abilities. But this is not intended to be a textbook. Instead, by rooting his tale in familiar examples — such as that blackboard — Horowitz provides a fascinating insight into the remarkable impact that sound has on the workings of the brain. ■

Sounds can have powerful influences on our emotions and memories.

RUNPHOTO/GETTY

Andrew J. King *is a Wellcome Principal Research Fellow and professor of neurophysiology at the University of Oxford, UK.*
e-mail: andrew.king@dpag.ox.ac.uk

J. MCCORMACK

Jon McCormack's *Morphogenesis Series* explores a computerized evolution of native Australian plants.

Q&A Jon McCormack
Species futurologist

Media artist Jon McCormack uses computer algorithms to imagine the future of native Australian species. As he prepares two new works — Codeform *and* Fifty Sisters *— for the* Ars Electronica Festival *in Linz, Austria, he talks about digital evolution and virtual ecosystems.*

Tell me about your work.
I use evolutionary algorithms to create artificial life forms that would be almost impossible to design directly. I employ a process similar to selective breeding that evolves aesthetic and behavioural traits. The computer is able to find nuances and complexity that I could never imagine. It is a creative partner, a way to make the unimaginable tangible. I've been told that visiting my installations feels like taking a walk in a strange forest.

What influence has existing Australian wildlife had on what you do?
Growing up there, I was exposed to flora and fauna that left an indelible mark on my view of nature. On one school trip to the bush, we spent several days looking at everything we could find in a square metre of wild forest. It was amazing to see the variety of organisms in such a tiny space. Students just a hundred metres apart found entirely different species.

How do your algorithms imagine the future of native species?
I've been inspired by Hungarian biologist Aristid Lindenmayer, who in the 1960s devised a formal mathematical system to model the growth of plants and simple multicellular organisms. A book he co-authored, *The Algorithmic Beauty of Plants* [Springer, 1990], showed how simple replacement rules could give rise to complex forms. I used this idea as a basis to make early animations of imaginary plants. In 2006, I produced work for a large digital billboard on a freeway in Queensland. I made a series of still pictures of imaginary hybrids of plants that were once native to the region, but had been removed to make way for the road. At first people thought I was an incredible painter. They were disappointed to discover that these strange combinations were 'grown' in a computer. But I couldn't have imagined those plants on my own.

N. HIGGINS

Can your imaginary organisms interact?
Codeform presents a virtual, interactive ecosystem. Your museum ticket will be scanned, and its bar code will produce the genetic material for a digital creature. In a warehouse-sized room with eight high-definition stereoscopic projectors, you will see your creature moving and breeding with others. The piece plays with the idea of DNA as a code for life that you are given and cannot change. I hope the scale and strangeness of the surroundings will expand peoples' imagination about real and artificial life.

What inspired your *Fifty Sisters* project?
I've used oil company logos as the basis for a set of 50 exotic-looking artificial plants depicted on large digital prints. The BP logo is floral, a clean veneer hiding something deep. There's no denying the benefits of oil. Cheap and abundant energy is one reason we live in relative safety and comfort. But we need to appreciate that our reliance on fossil fuels has many costs. Oil was once phytoplankton and other prehistoric flora, and has taken hundreds of millions of years to produce. Ironically, the burning of fossil fuels is returning our climate to conditions similar to those that were prevalent when these plants evolved.

How are computers changing the way we think about art?
Computers still leave most of the real creativity to the programmer. We're on the cusp of a very different relationship, however, where the computer won't be a passive tool but will contribute to the creative process. Current devices have yet to take advantage of that potential. A digital camera uses sophisticated algorithms to optimize the shot, sometimes even waiting for the subject to smile before opening the shutter. This puts the user in a very passive role. The mass reduction of creativity to the lowest common denominator is dangerous. Machines should allow for accidents and serendipity, contributing their intelligence without telling us what to do.

Could a computer be creative on its own?
When Charles Babbage built his Analytical Engine, a precursor to the computer, his muse and collaborator Ada Lovelace wrote that it "can do whatever we know how to order it to perform" but "has no pretensions whatever to originate anything". That criticism has endured. But we now have machines that have beaten the world's best chess players and proved theorems that humans couldn't. We haven't built creative machines yet; I suspect we will eventually. A program written by composer David Cope has generated symphonies that sound like Mozart and Rachmaninoff to informed listeners. Visual artist Harold Cohen has spent decades developing a software 'artist' that generates figurative drawings. There is debate about how creative these programs really are, but I think they challenge the assumed superiority of human creativity. As computers continue to develop, our understanding of creativity will change.

JASCHA HOFFMAN

Correspondence

A call to commission more women writers

We have analysed the gender distribution of authors of News & Views articles in *Nature* and of Perspectives in *Science* for 2010 and 2011. Our numbers indicate that both journal sections under-represent women scientists.

We divided the articles into three broad subject categories: biological and chemical sciences (which includes medical sciences); physical sciences; and Earth and environmental sciences. We compared the proportion of women authors with the proportion of women scientists employed in 2006 in the United States in science and engineering in each of the three categories (see go.nature.com/bkechu).

We found that the proportion of women commissioned to write *Nature* News & Views articles was much lower than the proportion of women scientists overall: female authorship was 17.3% for the biological and chemical sciences, 8.1% for physical sciences and 3.8% for Earth and environmental sciences, with the proportion of women authors of Perspectives in *Science* being slightly larger. However, the pool of women scientists in these disciplines was significantly higher than the proportion of female authorship at 32%, 16% and 20%, respectively.

It should be pointed out that a large proportion of invited News & Views authors are full professors, and the percentage of full professors who are women is lower than that for all scientists. Also, the proportion of women full professors is smaller in the European Union than in the United States. However, the present proportion of women authors of News & Views and of Perspectives is very low, and we believe that it is still fair to conclude that fewer women than men are offered the career boost of invitation-only authorship in each of the two leading science journals.

In response to earlier criticism, *Nature* increased the proportion of women authors in its Insight section (D. Conley *Nature* **438,** 1078; 2005). It is time to extend gender parity for commissioned writers across *Nature* and *Science*.

Daniel Conley, Johanna Stadmark *Lund University, Sweden. daniel.conley@geol.lu.se*

Competing interests: expanding rapidly

It is biased reports, not "the taint of scandal", that are the real danger in scientific research (*Nature* **488,** 5; 2012). Without full disclosure of conflicts of interest (COI), universities and journals cannot begin to deal with the problem — but a disclosed conflict is still a conflict.

Declarations of conflicting interests shift to the reader the responsibility that should be borne by the editors to detect whether reported findings could have been warped by bias. The reader, even when forewarned, is not always in a position to judge the extent to which the financial interest of an author or investigator has consciously or unconsciously influenced their assessment of the evidence.

Our courts deal with this problem more effectively. For example, a judge or juror with a financial interest in a case must not only disclose it, but also withdraw from the proceedings.

As conflicts of interest become more common, there is a risk that the proliferation of footnotes disclosing them will desensitize our apprehension of bias.

Arleen B. Rifkind *Weill Cornell Medical College, New York, USA. arifkind@med.cornell.edu*

Competing interests: judged in perpetuity

Full disclosure of competing interests is necessary so that readers can judge the conduct, validity, merit and reporting of research (*Nature* **488,** 5; 2012). But this needs to go beyond simply revealing financial connections, because conflict with non-financial interests can sometimes be more serious (D. F. Horrobin *Br. Med. J.* **318,** 466; 1999).

Arguably, anyone who writes in a scientific journal may be biased in some way or another. Authors who are tempted not to disclose competing interests need to remember that readers continue to act as 'peer reviewers' after publication.

Thomas C. Erren *University Hospital of Cologne; University of Cologne, Germany. tim.erren@uni-koeln.de*

Biomass energy holds big promise

Australia's government strongly supports the production of energy from renewable sources. But it is not yet tapping into biomass — the country's most cost-competitive renewable source.

Australia could generate at least 50 million tonnes of economically available biomass every year, with potentially millions more coming from sustainable forest management and from decomposed waste. With the help of mature bioenergy technologies already in use in other countries, this could provide more than 20% of the country's primary energy for heat, electricity and transportation. Modern biomass-fuelled plants have a fuel-to-energy conversion efficiency of more than 85%, comparing favourably with Australia's present coal-fired condensing plants and gas-fired turbine systems.

We should follow the lead of countries such as Austria, the land area of which is roughly 1% of Australia's (much of it comprising urban regions and alpine reserves). Austria produces more than 20 million tonnes of biomass as wastes and residues for energy production. In 2009, this constituted more than half of the 30% of primary energy that came from renewable sources.

Using sustainably produced biomass instead of fossil fuels can cut greenhouse-gas emissions, depending on effective use of

the heat produced and increased net carbon sequestration by expanded plantings. Integrating farmed forestry plantations and woody energy crops within current farming areas would increase biomass supply without disrupting water run-off or the production of food and fibre. Fewer conventional water-cooled power plants would also mean reduced consumption of potable water.

Andrew Lang *SMARTimbers Cooperative, Ballarat, Australia.*
Heinz Kopetz *Austrian Biomass Association, Vienna, Austria.*
Albert Parker *University of Ballarat, Ballarat, Australia.*
albertparker@y7mail.com

Curb temptation to skip quality control

Daniel MacArthur cautions scientists and journal editors against jumping to "false positive" conclusions (*Nature* **487,** 427–428; 2012). But temptations to leap over the quality-control gap are rife in scientific publishing.

Journals are proliferating, even in the top tier, so they have to compete for eye-catching papers. The quality of these publications necessarily depends on the vanishing time of a small pool of expert reviewers. Those who are competent in statistics are in particularly short supply.

Authors are driven to publish in the leading journals so that they can reap the benefits of academic respect, such as promotion and tenure. Never mind being overly conscientious in underpinning the conclusions — refuted results do not always end in retraction. And if a paper is retracted, it may not always have serious repercussions.

The upshot is that ambitious researchers want to push the limits of credibility in interpreting their results, and the same can be said of journal editors.

This publication 'currency' needs to be brought under tighter control to curb inflationary practices among authors and editors.

Herman Tse *University of Hong Kong, Hong Kong.*
herman@graduate.hku.hk

Speed up reviews of misconduct

We call for greater transparency and speed in investigations of scientific misconduct. For the benefit of the scientific community, swift assessment of the validity of published results should be distinguished from legally convoluted verdicts that concern wilful deceit.

Journal publishers and academic institutions have differed markedly in their response to recent cases of alleged misconduct. For example, the committee investigating publications by the Dutch social psychologist Diederik Stapel (*Nature* **479,** 15; 2011) published an interim report within two months and is releasing its findings in under a year (www.commissielevelt.nl). And a panel investigating the papers of another social psychologist, Dirk Smeesters (*Nature* **487,** 18; 2012), reported its results within a week of his resignation. Such prompt and detailed reporting contrasts with that of other protracted enquiries that have culminated in retractions of papers.

Web-based initiatives such as Retraction Watch (www.retractionwatch.com) can help by offering timely, although informal, alerts about suspect papers. Even if fabrication or falsification cannot be proved beyond reasonable doubt, statistical indications of problems with the data — such as gross errors or remarkably consistent findings — are scientifically relevant.

Jelte M. Wicherts, Marcel A. L. M. van Assen *Tilburg University, the Netherlands.*
j.m.wicherts@uvt.nl

PhDs fit for industry and commerce, too

You continue to lament the shortcomings of PhD training (see, for example, *Nature* **472,** 259–260; 2011 and *Nature* **486,** 304; 2012) but overlook the encouraging results of reforms to PhD programmes in many European countries.

Programmes in the region have modernized, expanding the traditional apprenticeship model in recognition of the fact that most new PhDs are not destined for academia. Industry and commerce are already welcoming this new batch of graduates.

Training has been broadened to develop such skills as scientific and lay presentation, teaching, grant application, time management, linguistic abilities and networking. These complement the rigour acquired in setting up and completing a three-year research project, and are valuable in any job that demands creative synthesis and the use of initiative, whether inside or outside academia. Flying in the face of convention, students can delegate some of their PhD work and learn how to become managers.

These reforms are being promoted by the European Commission, the European Universities Association Council for Doctoral Education (see, for example, *Nature* **468,** 125; 2010 and *Nature* **482,** 557–559; 2012) and by ORPHEUS, an independent organization that represents more than 100 European biomedical and medical faculties (www.orpheus-med.org).

Michael Mulvany *ORPHEUS, Aarhus University, Denmark.*
mm@farm.au.dk
Zdravko Lackovic *ORPHEUS, University of Zagreb School of Medicine, Croatia.*

Institute to continue climate monitoring

Contrary to what you imply in your online News article, staff cuts are not sounding a "death knell" for the Lauder research station run by the National Institute of Water and Atmospheric Research (NIWA) in New Zealand (http://doi.org/h65). We at NIWA reaffirm the institute's commitment to continuing its long-term atmospheric measurements, which are crucial for climate research.

NIWA sites at Lauder and Arrival Heights, Antarctica, will maintain their multi-instrument measurement programmes to monitor radiation, ozone and a range of atmospheric trace gases. National and international scientific expertise will continue to oversee and support these activities. NIWA is increasing its investment in instrumental capacity at both sites, and is continuing the long-term observations made on behalf of overseas institutions.

The review of NIWA's Lauder staff retains three of six atmospheric-scientist positions and the five technicians, and creates one new measurement-scientist post; two of the departing scientists will still take part in an emeritus capacity. A measurement scientist and a technician based at other NIWA sites will lend further support.

The review outcome is a reprioritization in favour of long-term monitoring, rather than a major change in Lauder's research direction (see go.nature.com/vnnb37).

Olaf Morgenstern, Richard McKenzie, Vanessa Sherlock *National Institute of Water and Atmospheric Research, New Zealand.*
olaf.morgenstern@niwa.co.nz

Embrace complexity but not jargon

In advocating that science writers should not shun scientific jargon, Trevor Quirk proposes the wrong means to the right end for improving science communication (*Nature* **487,** 407; 2012).

The science writer's tough task is to help readers to grasp the nuances and complex concepts behind obscure specialist terminology by translating it into simple, accurate language. There is nothing patronizing about this. Esoteric terms can develop in any discipline — even some Olympic sports use mysterious labels such as 'keirin' and 'repechage'. And the general reader should not be expected to decode words used by scientists who have years of training under their belts.

The public benefits from skilful clarification, as do scientists.

Anita Makri *The Science and Development Network (SciDev.Net), London, UK.*
anita.makri@scidev.net

Bernard Lovell

(1913–2012)

Physicist and radar pioneer who created the famous Jodrell Bank radio telescope.

With vision, inspiration and determination, Bernard Lovell created the Jodrell Bank Observatory in Cheshire, UK, and the great radio telescope there that bears his name. Through his leadership, generations of astronomers developed the tools of radio astronomy that have revolutionized astrophysics and cosmology.

Lovell was born in Oldland Common near Bristol, UK, and educated at the local Kingswood Grammar School. He was attracted to science through a lecture by physicist Arthur Tyndall. He joined Tyndall as a research student at the University of Bristol, where his meticulous work on the resistance of thin metallic films earned him a PhD in 1936. Taking up a lectureship at the University of Manchester, UK, Lovell became interested in research on cosmic rays, joining the group of Patrick Blackett, who was developing cloud-chamber techniques for the rays' detection.

Lovell followed Blackett into government science, becoming a key figure in the development of radar in the Second World War. Lovell's group designed an airborne radar system that operated at a wavelength of 10 centimetres, with high positional accuracy. Lovell also designed large, airborne scanning antennas that required big holes to be cut in aircraft fuselages. The radars, vital for navigation and able to detect surfaced German U-boats, were installed in the aircraft of Coastal Command and Bomber Command and had a dramatic effect in the battle of the Atlantic.

While working at one of the Chain Home radar stations that guarded the UK coast, Lovell saw the background of sporadic echoes against which the aircraft signals had to be distinguished, and wondered where they came from. He discussed with Blackett whether they might be reflections from ionized clouds in the atmosphere, perhaps formed by cosmic-ray showers, an idea he pursued after the war.

Returning to Manchester in 1945, Lovell used his wartime experiences to develop the Jodrell Bank Observatory. An army radar brought to Manchester was moved some 30 kilometres out of the city to a university botany site to escape electrical interference from trams. The source of the echoes Lovell had seen while at the Chain Home radar was soon revealed: meteor trails. (No echoes have ever been detected from cosmic rays.)

Lovell wanted a larger antenna. He and a small team built a parabolic reflector 66 metres in diameter, made of wires stretched between scaffold poles. Robert Hanbury Brown used this dish to discover radio waves from the Andromeda galaxy. It was the start of astronomy at Jodrell Bank. That dish was fixed and could only stare upwards. Lovell's ambition was to build a steerable reflector at least as big. He started on plans for the iconic 76-metre-diameter dish that we know as the Lovell Telescope.

PHOTOSHOT

Building such an innovative, huge instrument required extraordinary determination and powers of persuasion. The telescope's original cost more than doubled from £259,000 in 1952 to £640,000 in 1956 (today equivalent to £13 million, or US$21 million). This was mainly because the discovery by Dutch astronomers in 1951 of a spectral line at the relatively short wavelength of 21 centimetres from hydrogen in interstellar space demanded a more precise reflecting surface. The UK government and the university met most of the increase, leaving a debt of £50,000 for which Lovell became personally responsible. He was faced with an enquiry by the government's Public Accounts Committee, and with the possibility of imprisonment.

The launch of the Sputnik satellite by the Soviet Union in 1957 transformed this embarrassing situation. Typically, Lovell rose to the occasion. The new telescope obtained echoes from the Sputnik launch vehicle, a feat unmatched anywhere else in the Western world. Suddenly everyone realized the telescope's importance. Lord Nuffield paid off the debt, half personally and half from the Nuffield Foundation, and serious astronomy could begin.

The telescope fulfilled Lovell's vision of opening a new window on the Universe. It had illustrious roles in early space exploration, in the discovery of quasars and the first gravitational lens, and in the discovery and study of pulsars, or neutron stars, which have provided the most precise tests of Einstein's general theory of relativity. After 55 years and two major refurbishments, the telescope is still working at the frontiers of knowledge.

Jodrell Bank Observatory, the academic institution set up in the telescope's shadow, is established as an international centre for astrophysics and technology. It was chosen last year to host the headquarters of the Square Kilometre Array, which will be the world's largest radio telescope when it is built in South Africa and Australia in coming decades.

Warm and generous, Lovell took particular pleasure in the millions of people who came to the observatory's visitor centre, which he initiated 41 years ago. His instinct for good science and his empathy with staff and students led him to allocate telescope time without controversy. He retained a close interest in the observatory after he retired, but as the directors who succeeded him, we can attest that he never told us what to do with his telescope.

Lovell received many honours, and served as president of the Royal Astronomical Society (1969–71) and of the British Association for the Advancement of Science (1975–76). But he was most pleased when the telescope was given his name to mark its 30th anniversary.

Outside science, Lovell was an accomplished musician, playing the organ regularly in the church of his home village of Swettenham in Cheshire. He was a keen cricketer and horticulturalist, establishing arboretums adjacent to the telescope and to his home. He was much loved and respected by his colleagues. ■

Francis Graham Smith, Rodney Davies and Andrew Lyne *are emeritus professors of physics at the University of Manchester, UK, and former directors of the Jodrell Bank Observatory (in 1981–88, 1988–97 and 1997–2006, respectively).*
e-mail: fgs@jb.man.ac.uk

Is irisin a human exercise gene?

ARISING FROM P. Boström *et al.* *Nature* **831,** 463–468 (2012)

Boström *et al.* report that exercise training induces the expression of the *FNDC5* gene in human muscle, producing irisin, which can convert white fat into brown fat, so enhancing metabolic uncoupling and hence caloric expenditure, and propose that this is a new health promoting hormone[1]. This assertion is based on experimental evidence that exogenous FNDC5 induces uncoupling protein 1 (UCP1) expression in white subcutaneous adipocytes; overexpression of FNDC5 in liver (elevating systemic irisin) prevents diet-induced weight gain and metabolic dysfunction and stimulates oxygen consumption in mice; and *FNDC5* mRNA expression levels double after exercise training in eight human skeletal muscle samples. However, the *UCP1* induction was lower than observed during *Brite*[2] formation or the level associated with an improved diabetes profile in humans[3]. Here we demonstrate that muscle *FNDC5* induction occurs only in a minority of subjects—whereas all types of exercise training programmes[4–7], in the vast majority of people, yield some gain in cardiovascular or metabolic health, in our analysis of ~200 subjects muscle *FNDC5* was increased only in highly active elderly subjects, whereas *FNDC5* expression was unrelated to metabolic status, which casts doubt over the general relevance of skeletal muscle FNDC5 to human health.

There are >1,000 genes 'activated' in human skeletal muscle by exercise training[8–10], all of which may contribute to the resulting improved health. Our aim was to consider whether *FNDC5* expression was one such factor in a much larger set of exercise cohorts. We failed to confirm *FNDC5* gene activation by aerobic exercise in younger subjects (Fig. 1a) despite increases in mitochondrial gene expression[8,9,11] and fitness (Fig. 1b), or in a resistance training study ($n = 43$, Fig. 1c) in 20–80 year olds[12]. Furthermore, improvements in insulin sensitivity, fasting glucose and blood pressure were not related to *FNDC5* expression (for example, Fig. 1d, e). As exercise provides health gains to all age groups, this raises doubts that FNDC5 can be responsible for these benefits.

Despite the lack of induction of *FNDC5* mRNA in most people, it is possible that a subset of subjects will induce *FNDC5* expression and that this may convey some health benefits. Indeed, analysis of a group of individuals ($n = 20$) from a 2007 gene-chip data (GEO accession GSE9103) set found that older, but not younger, active subjects had 30% greater *FNDC5* expression than sedentary controls (Fig. 1f). However, in a larger human cohort ($n = 118$)[13] with similar demographics, *FNDC5* expression was not positively linked to diabetes status (Fig. 2a–d). Thus, confidence in the premise that muscle *FNDC5* expression is linked to human metabolic health should remain tentative.

An enormous variety of exercise training regimens[4–7] improve glycaemic control, and our analysis is derived from subjects within the same age range and sampling periods as the eight subjects presented by Boström *et al.*[1]. Our detection of *FNDC5* is based on a gene-chip probe set (detecting the major common exon). The inter-person dynamic range of *FNDC5* expression was ~6-fold, whereas the intra-person dynamic range was up to 3-fold; demonstrating that the

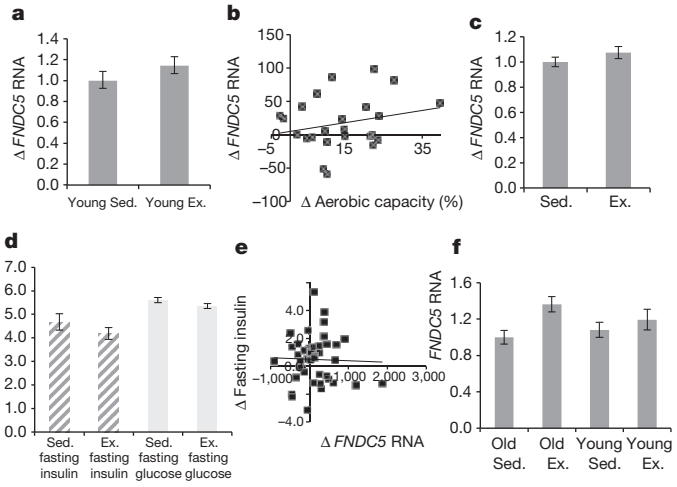


Figure 1 | Irisin is not routinely activated by exercise in humans. **a**, mRNA expression in young sedentary males before and 24 h after 6 weeks intense endurance cycling ($n = 24$); first published in 2005 (ref. 8) (FDR using SAMR >10%). **b**, Correlation between *FNDC5* and the change in aerobic capacity after 6 weeks supervised training ($n = 24$, $R^2 = 0.05$, $P = 0.37$)[8]. **c**, *FNDC5* mRNA expression in subjects ($n = 43$) before and after supervised resistance training[12] ($n = 43$, $P = 0.6$). **d**, Change in fasting insulin ($P = 0.03$) and glucose ($P = 0.04$) levels after 20 weeks of training. **e**, Correlation of the change in fasting insulin versus the change in *FNDC5* mRNA levels after 20 weeks of training ($R^2 = 0.01$, $P = 0.7$). **f**, *FNDC5* mRNA expression in young and older sedentary and age-matched endurance trained subjects ($n = 10$ per group, $P = 0.52$ and $P = 0.02$, respectively, derived from gene-chip study (GEO accession GSE9103)). Data are MAS5 normalized, quality controlled for outliers. Significance analysis of microarrays paired or unpaired *t*-tests were used as appropriate. Sed., sedentary; Ex., exercise trained. *y* axis units are arbitrary units. Error bars indicate standard error.

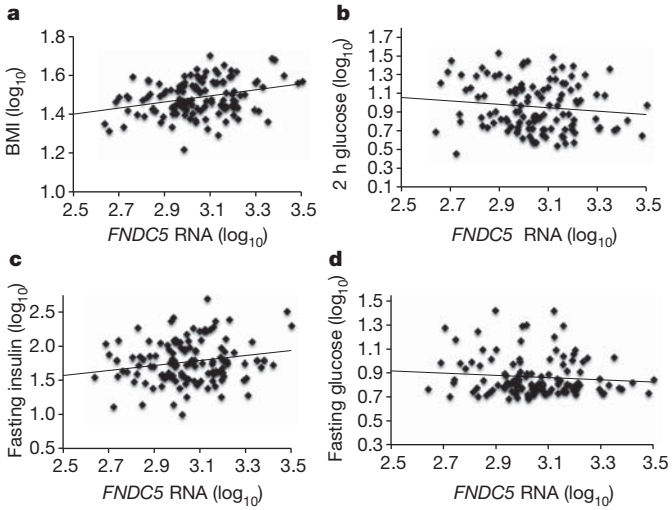


Figure 2 | Irisin expression is not related to diabetes status in humans. Plots correlating body mass index (BMI; **a**), 2-h glucose (**b**), fasting insulin (**c**) and fasting glucose (**d**) with skeletal muscle *FNDC5* mRNA expression in 118 diabetes medication-free subjects aged ~55 years demonstrates that *FNDC5* mRNA expression does not relate to peripheral insulin sensitivity (for example, fasting glucose ($R^2 = 0.01$, not significant), fasting insulin ($R^2 = 0.042$, not significant)) or glucose response during a glucose tolerance test ($R^2 = 0.016$, not significant)[13]. Furthermore, rather than being expressed at higher levels in the leaner subjects, there was a statistically significant positive association between BMI and *FNDC5* mRNA abundance ($R^2 = 0.1$, $P = 0.004$), albeit with a very modest degree of shared variance (~10%). Data are MAS5 normalized, quality controlled for outliers and log-transformed. Linear regression analysis was carried out and adjusted for four dependent tests.

detection method was entirely consistent with real-time qPCR. Our large sample size indicates that the response of *FNDC5* mRNA to exercise in human muscle is limited. Although we cannot rule out that FNDC5 protein is induced by exercise, given that the author's explanation was based on transcriptional events, we feel that the other 1,000 exercise-modulated genes merit more attention.

Boström *et al.*[1] did find that FNDC5-expressing adenoviral particles improved glycaemic control, increased oxygen consumption and modestly prevented weight gain in mice. The authors propose that these benefits were due to irisin stimulating UCP1 expression in subcutaneous adipocytes and thereby promoting substantial increases in oxygen consumption. This enhanced basal metabolic rate should increase food intake; yet this was not found, indicating that the modest reduction in mouse body mass had to be accompanied by reduced energy consumption elsewhere to maintain energy balance. Moreover, if FNDC5 does increase uncoupling, one would expect endurance-trained individuals to show the greatest uncoupling and lowest metabolic efficiencies; however, in reality the reverse is true.

James A. Timmons[1], Keith Baar[2], Peter K. Davidsen[1] & Philip J. Atherton[3]

[1]University of London, Royal College Street, London NW1 0TU, UK.
email: Jamie.Timmons@gmail.com
[2]University of California Davis, Davis, California 95616, USA.
[3]University of Nottingham, Derby Royal Infirmary, Derby DE22 3DT, UK.

Received 13 February; accepted 15 June 2012.

1. Boström, P. *et al.* A PGC1-α-dependent myokine that drives brown-fat-like development of white fat and thermogenesis. *Nature* **831,** 463–468 (2012).
2. Petrovic, N. *et al.* Chronic peroxisome proliferator-activated receptor γ (PPARγ) activation of epididymally derived white adipocyte cultures reveals a population of thermogenically competent, UCP1-containing adipocytes molecularly distinct from classic brown adipocytes. *J. Biol. Chem.* **285,** 7153–7164 (2010).
3. Timmons, J. A. & Pedersen, B. K. The importance of brown adipose tissue. *N. Engl. J. Med.* **361,** 415–416 (2009).
4. Boule, N. G. *et al.* Effects of exercise training on glucose homeostasis: the HERITAGE family study. *Diabetes Care* **28,** 108–114 (2005).
5. Holten, M. K. *et al.* Strength training increases insulin-mediated glucose uptake, GLUT4 content, and insulin signaling in skeletal muscle in patients with type 2 diabetes. *Diabetes* **53,** 294–305 (2004).
6. Keller, P. *et al.* Using systems biology to define the essential biological networks responsible for adaptation to endurance exercise training. *Biochem. Soc. Trans.* **35,** 1306–1309 (2007).
7. Babraj, J. A. *et al.* Extremely short duration high intensity interval training substantially improves insulin action in young healthy males. *BMC Endocr. Disord.* **9,** 3 (2009).
8. Timmons, J. A. *et al.* Human muscle gene expression responses to endurance training provide a novel perspective on Duchenne muscular dystrophy. *FASEB J.* **19,** 750–760 (2005).
9. Keller, P. *et al.* A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. *J. Appl. Physiol.* **110,** 46–59 (2011).
10. Melov, S., Tarnopolsky, M. A., Beckman, K., Felkey, K. & Hubbard, A. Resistance exercise reverses aging in human skeletal muscle. *PLoS ONE* **2,** e465 (2007).
11. Vollaard, N. B. J. *et al.* Systematic analysis of adaptations in aerobic capacity and submaximal energy metabolism provides a unique insight into determinants of human aerobic performance. *J. Appl. Physiol.* **106,** 1479–1486 (2009).
12. Phillips, B. *et al.* Resistance exercise training improves age-related declines in leg vascular conductance and rejuvenates acute leg blood flow responses to feeding and exercise. *J. Appl. Physiol.* **112,** 347–353 (2011).
13. Gallagher, I. J. *et al.* Integration of microRNA changes *in vivo* identifies novel molecular features of muscle insulin resistance in type 2 diabetes. *Genome Med.* **2,** 9 (2010).

Author Contributions J.A.T. and P.K.D. carried out the data analysis while J.A.T., K.B. and P.J.A. drafted the article. J.A.T., P.K.D., K.B. and P.J.A. edited the final article.

Competing Financial Interests Declared none.

doi:10.1038/nature11364

Boström *et al.* reply

REPLYING TO J. A. Timmons, K. Baar, P. K. Davidsen & P. J. Atherton *Nature* **488,** http://dx.doi.org/10.1038/nature11364 (2012)

Timmons *et al.*[1] report that, using human muscle biopsies from exercised subjects, they fail to detect a robust increase in *FNDC5* mRNA using gene expression arrays. Although our studies were mostly focused on mouse systems[2], we observed a twofold increase in *FNDC5* mRNA levels in a cohort of older, obese subjects after a 10-week protocol of endurance exercise. There are several explanations for the differences between the results from Timmons *et al.*[1] and our[2] data.

First, gene expression arrays are not considered fully quantitative and failure to observe increases on Affymetrix chips normally doesn't exclude the possibility of an actual increase in mRNA. This also complicates the use of gene expression probe signals for correlation analyses as performed by Timmons *et al.*[1]. Second, the data presented by Timmons *et al.*[1] actually do display a significant increase in *FNDC5* probe signal with exercise in the cohort most similar to ours (that is, subjects older than 60 years). The actual fold increase might be substantially larger if quantitative polymerase chain reaction (qPCR) analysis had been performed. Finally, in the cohort described by Timmons *et al.*[1] (their ref. 4; we cannot access the data set of their ref. 5 online), PGC1-α—a well established marker for exercise-induced changes in muscle[3–6]—is not increased (not a part of their 'training-responsive transcriptome'). The lack of induction of PGC1-α in the analyses of ref. 4 of Timmons *et al.* is a potential cause for concern and may be indicative of the susceptibility of type II errors using gene expression arrays, and thus limit the conclusions of their quantitative analysis of *FNDC5*. If one does not see an induction of a crucial positive control, why would one expect to find an induction of *FNDC5*, especially as *FNDC5* is regulated by PGC1-α?

In summary, we appreciate that small cohorts of exercised individuals may display different results, given the complexity of the intervention. However, both our[2] data and that of Timmons *et al.*[1] support increased *FNDC5* mRNA levels with exercise, at least in older subjects.

Pontus Boström[1], Jun Wu[1], Mark P. Jedrychowski[2], Anisha Korde[1], Li Ye[1], James C. Lo[1], Kyle A. Rasbach[1], Elisabeth Almer Boström[3], Jang Hyun Choi[1], Jonathan Z. Long[1], Shingo Kajimura[4], Maria Cristina Zingaretti[5], Birgitte F. Vind[6], Hua Tu[7], Saverio Cinti[5], Kurt Højlund[6], Steven P. Gygi[2] & Bruce M. Spiegelman[1]

[1]Dana-Farber Cancer Institute and Harvard Medical School, 3 Blackfan Circle, CLS Building, Floor 11, Boston, Massachusetts 02115, USA.
email: bruce_spiegelman@dfci.harvard.edu
[2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
[3]Renal Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.
[4]UCSF Diabetes Center and Department of Cell and Tissue, Biology, University of California, San Francisco, California 94143, USA.

[5]Department of Experimental and Clinical Medicine, Università Politecnica delle Marche, Electron Microscopy Unit-Azienda Ospedali Riuniti, Ancona 60020, Italy.
[6]Diabetes Research Center, Department of Endocrinology, Odense University Hospital, DK-5000, Odense, Denmark.
[7]LakePharma, Inc., 530 Harbor Blvd, Belmont, California 94002, USA.

1. Timmons, J. A., Baar, K., Davidsen, P. K. & Atherton, P. J. Is Irisin a human exercise gene? *Nature* **488,** http://dx.doi.org/10.1038/nature11364 (2012).
2. Boström, P. *et al.* A PGC1-α-dependent myokine that drives brown-fat-like development of white fat and thermogenesis. *Nature* **831,** 463–468 (2012).
3. Perry, C. G. R. *et al.* Repeated transient mRNA bursts precede increases in transcriptional and mitochondrial proteins during training in human skeletal muscle. *J. Physiol.* **588,** 4795–4810 (2010).
4. Ristow, M. *et al.* Antioxidants prevent health-promoting effects of physical exercise in humans. *Proc. Natl Acad. Sci. USA* **106,** 8665–8670 (2009).
5. Vissing, K. *et al.* Effect of sex differences on human MEF2 regulation during endurance exercise. *Am. J. Physiol. Endocrinol. Metab.* **294,** E405–E415 (2008).
6. Little, J. P., Safdar, A., Bishop, D., Tarnopolsky, M. A. & Gibala, M. J. An acute bout of high-intensity interval training increases the nuclear abundance of PGC-1α and activates mitochondrial biogenesis in human skeletal muscle. *Am. J. Physiol. Int. Comp. Physiol.* **300,** R1303–R1310 (2011).

doi:10.1038/nature11365

ENVIRONMENTAL SCIENCE

Scorecard for the seas

An index assessing the health of the oceans gives a global score of 60 out of 100. But the idea that a single number can encompass both environmental status and the benefits that the oceans provide for humans may prove controversial. SEE ARTICLE P.615

DEREK P. TITTENSOR

Just like the business pages of many newspapers at present, reports about the state of our oceans all too often read like a scandal sheet. However, according to research published by Halpern *et al.*[1] on page 615 of this issue, residents of Germany or Seychelles may have something to cheer about — with a score of 73 out of 100, they are among the top of the class for inhabited countries in an index that provides a score of the overall condition of marine ecosystems. But 32% of coastal nations receive a score of less than 50.

Although broad-scale ecosystem indicators are not novel[2], the one that Halpern and colleagues describe is noteworthy for several reasons. The authors fuse markedly different goals into a single composite index that not only consists of measures of ocean health, but also takes into account the goods, services and benefits that the oceans provide for humans. Furthermore, their index is spatially explicit, being calculated for each country (see Fig. 2 of the paper[1]) that has a marine exclusive economic zone (waters up to 322 kilometres offshore). This provides a yardstick against which management of the oceans can be compared, and thereby creates a 'league table' of national stewardship of marine resources.

The authors began by defining ten goals (and eight sub-goals) that describe both a sustainable marine realm and what the ocean can provide for people. These range from extractive uses, such as food provision, to ecological attributes, such as biodiversity, and also include less tangible benefits, such as carbon storage and 'sense of place' (Fig. 1). The researchers quantified each goal in terms of its current status (against a defined reference point); its recent trend; the pressures likely to affect it in the future; and its resilience. The goals were then synthesized into a single index of ocean health and benefits to give a value between 0 and 100. Thus, the index provides a numerical representation of the fine line between maintaining ocean ecosystems and extracting from them resources, economic benefits and livelihoods for humans.

There is no single objective way to amalgamate the ten disparate goals used in this index. The default scheme that Halpern and colleagues present is to weight each goal equally. Canada, for example, scores highly on artisanal (small-scale) fishing opportunities, coastal protection and biodiversity, low on mariculture, tourism and recreation and 'lasting special places', and moderately on the remaining goals and sub-goals. With equal weighting, these individual grades give an overall index score of 70. However, the authors also present schemes in which the goals are weighted according to different sets of values — from conservationist or utilitarian perspectives, for example — thereby allowing different users to choose the scheme most appropriate for their needs.

There also remains unavoidable subjectivity in the choice of goals and reference points for the index. For example, the authors' decision to set the reference point for fisheries to 75% of multi-species maximum sustainable yield, or that for mariculture to China's yield (the maximum observed), are judgement calls with which some may disagree. In addition, some of the data they use seem to be less reliable proxies of their goal than others (international arrivals as a measure of ocean-related tourism, for example). The authors acknowledge these data disparities, but point out that their study provides an opportunity to identify areas in which additional data collection may prove fruitful.

Some of Halpern and colleagues' results may raise eyebrows in the marine community. The authors' global index value of 60 seems high, given the extensive evidence for our detrimental and accelerating impact on ocean ecosystems. The value of 83 for the global-biodiversity goal also seems remarkably optimistic in light of the voluminous literature that catalogues the changes humans have wrought on the oceans[3,4] — research to which many of the authors of the current paper have contributed.

However, such dissonance can be reconciled through a recalibration of expectations — away from an index that measures the pristineness or ecological integrity of the marine realm, and towards one that quantifies an integrated system in which humans and the non-human ecosystem are considered equally

P. NARRA/NATURE PICTURE LIBRARY

Figure 1 | The ocean's bounty. To develop a single index to define the health of the oceans, Halpern and colleagues[1] assessed ten goals that encompass the ecological integrity of the marine realm and the goods, services and benefits that the oceans provide for humans.

important parts. In such a system, appropriate use of extractive resources or substantial mariculture production will result in a higher index score. Whether or not this is the right approach depends on your opinion about the role the oceans should fulfil for humans. But there is a genuine concern that improvements in human-related indices (such as tourism and recreation, or coastal economies) may mask deterioration in the fundamental ecological health underpinning the marine environment — and may provide excuses for inaction on this front.

An additional consideration is that the principle of distilling such complex and nuanced information down to a single value may sit uneasily with some people. An alternative strategy that might avoid conflating opposing ideas about what constitutes a healthy ocean would be to separate this single index into two — one that measures the provision of goods and benefits for humans and another that evaluates the health of the ocean as the distance to a 'more pristine' state. But this would lack the simplicity of a single, easily interpretable number by which performance can be benchmarked, and which can act as both a carrot and a stick. A single index allows the oceans to be assessed and compared in a similar way to, for example, the use of gross domestic product as an indicator of a country's standard of living, and brings with it similar benefits and limitations.

The debate about whether a single index is a reasonable goal to aim for is certainly worth having, and one could argue incessantly about the best way to construct such a metric. But as Voltaire's aphorism says, the perfect is the enemy of the good, and to have something on the table is certainly better than nothing. Halpern *et al.* have synthesized an extraordinary diversity of data in their work towards this laudable goal. A single index that can be communicated, plotted, monitored over time and transparently compared between countries, regions and oceans may help to bring ocean management into greater prominence in the media, and in a more readily interpretable format. Although scepticism remains as to whether efforts such as that of Halpern and colleagues will spur governments and regulatory bodies to further action, they will at least enable us to monitor the future progress or, perhaps more likely, the deterioration of our oceans. ■

Derek P. Tittensor *is at the United Nations Environment Programme World Conservation Monitoring Centre, Cambridge CB3 0DL, UK; the Microsoft Research Computational Science Laboratory, Cambridge; and Dalhousie University, Halifax, Nova Scotia, Canada.*
e-mail: derekt@mathstat.dal.ca

1. Halpern, B. S. *et al. Nature* **488,** 615–620 (2012).
2. Butchart, S. H. M. *et al. Science* **328,** 1164–1168 (2010).
3. Worm, B. *et al. Science* **314,** 787–790 (2006).
4. Halpern, B. S. *et al. Science* **319,** 948–952 (2008).

IMMUNOLOGY

Licensed in the lungs

In multiple sclerosis, the body's own immune cells attack the brain and spinal cord. But how they get there from peripheral tissues has been a mystery. Surprisingly, the lungs might be a key transit point. SEE LETTER P.675

RICHARD M. RANSOHOFF

The immune system has one cardinal function: to protect the body against pathogens. To accomplish this task, immune cells develop in such a way that they are activated only by foreign structures. This discrimination is imperfect, however, and when it goes wrong, our bodies can suffer autoimmune inflammation and disease. In many cases, this inflammation is largely or entirely limited to a single organ — multiple sclerosis, for example, affects the brain, optic nerves and spinal cord of the central nervous system (CNS). Understanding how autoimmune cells migrate to and accumulate within an affected organ is vital to developing effective treatment strategies. On page 675 of this issue, Odoardi and colleagues[1] use a rat model of multiple sclerosis to show that, before their entry into the CNS, inflammatory autoimmune cells transiently settle in the lungs — an organ not previously associated with immune-cell trafficking to the CNS.

Multiple sclerosis affects around 2.5 million people worldwide. The disease targets myelin, a membranous sheath that wraps around the axon fibres of nerve cells. Damage to myelin leads to myriad symptoms, including sensory disturbance, impaired balance and difficulty thinking. To study multiple sclerosis, researchers have developed several animal models, one of which is called adoptive-transfer experimental autoimmune encephalomyelitis (EAE). In this model, rats are first immunized with myelin basic protein (MBP), a major component of myelin. The immunization activates certain immune cells (T cells) that are specific for MBP, and induces them to proliferate. These T cells can then be collected from the immunized rats and injected into other rats, where they act as autoimmune cells and cause inflammation that centres on the spinal cord. Because these 'encephalitogenic' T cells specifically attack myelin, and not other CNS structures[2], the rat disease closely mimics certain aspects of the human condition.

However, researchers have been puzzled by the fact that it takes four to five days following transfer of the encephalitogenic T cells for rats to develop disease, despite the cells being activated *in vitro* before transfer. The research group presenting the current paper also previously reported another strange phenomenon of the model: the T cells must be infused into the recipient rat's circulation to induce EAE; direct administration into the cerebrospinal fluid, which bathes the CNS, not only fails to hasten disease onset but is in fact ineffective in inducing the disease[3].

In this earlier study, the authors showed that the injected T cells migrate from the bloodstream to immune-system organs, including lymph nodes and the spleen, and that the gene-expression pattern of the cells changes during this time: a sharp reduction in the expression of activation- and proliferation-related genes is accompanied by a striking increase in the expression of genes involved in cell migration[3] (Fig. 1). Then, at the onset of disease, millions of T cells accumulate abruptly and simultaneously in the CNS[3]; the same research group has also characterized the events taking place at this time of CNS entry[4]. So the first steps following T-cell infusion, and the events immediately preceding disease onset, had both been established — but what happens to the T cells between those time points remained unclear.

Odoardi and colleagues set out to clarify this 'black box'. They used encephalitogenic T cells that express green fluorescent protein, which allows the cells to be tracked, and they developed the ability to image these fluorescent cells in the spinal cord of living animals using a technique called two-photon microscopy[5], an advance that has also been made by other research groups. Imaging of the intact brain is likewise possible[6,7], and together these methods have revolutionized our understanding of the processes by which immune cells gain entry to the CNS during disease and immune responses[8].

The authors first asked whether the injection of activated encephalitogenic T cells could itself result in systemic inflammation that might affect the CNS or its associated vasculature, so as to induce chemical signals that attract the T cells to that region. To assess this possibility, the researchers conducted an experiment in which they joined the circulatory systems of two rats, one of which had received

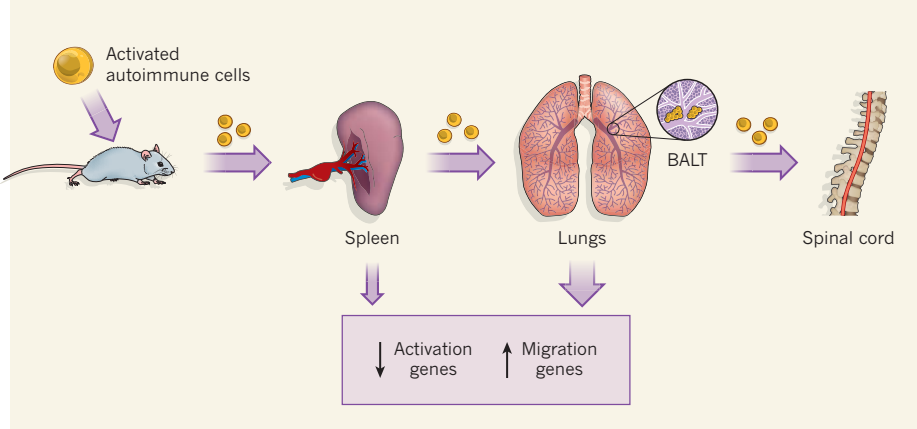


Figure 1 | Autoimmune cells transit the airways. In a commonly used rat model of multiple sclerosis, it takes four to five days following injection of disease-causing autoimmune cells into the bloodstream for signs of disease to appear. Equally puzzling is the fact that the autoimmune cells do not cause disease if they are injected directly into the cerebrospinal fluid. It was previously shown[3] that cells injected into the bloodstream move to the spleen, where they begin to downregulate the expression of proliferation- and activation-related genes and to upregulate the expression of genes involved in cell migration. Odoardi and colleagues now demonstrate[1] that the cells then move to the lungs, where they accumulate in bronchus-associated lymphoid tissue (BALT; clusters of immune cells near the walls of the lung's air passages). Here, the changes in gene expression are consolidated, and the authors propose that this 'licensing' step allows the cells to migrate to the spinal cord, where they begin to cause disease.

encephalitogenic T cells 48 hours earlier, and then monitored the arrival of the cells in the CNS of each animal. They found this to occur at approximately the same time in both rats, suggesting that the CNS of the rat that had received the cells earlier was not preconditioned to harbour an inflammatory reaction.

The researchers next sought to identify where the transferred cells reside before arriving in the CNS. Surprisingly, the two-photon microscopy experiments revealed the vast majority of the T cells to be in the rats' lungs. The cells were initially located in the lung bronchi (air passages) and alveoli (terminal air sacs), before accumulating in dense immune-cell clusters known as bronchus-associated lymphoid tissue (BALT; Fig. 1). The authors also show that, unlike encephalitogenic T cells injected into the bloodstream, T cells injected into the lungs' bronchi by means of the trachea can rapidly leave the lungs and cause disease. This finding suggests that the time the T cells spend in the lung environment 'licenses' them in a way that allows them to move to the CNS. They found, for example, that the changes in gene-expression profile previously observed in these cells were accentuated during the time spent in the lungs.

Importantly, the authors also found encephalitogenic T cells in BALT of rats that had been injected with the cells 2–3 months earlier, as newborns, and that had not developed signs of disease. When the researchers stimulated these dormant T cells by introducing MBP as an aerosol into the rats' tracheas, the cells became activated and accumulated in the CNS, causing disease.

Interpretation of this latter finding is assisted by considering the concept of tissue-resident immune memory cells, an idea stemming from the immune system's remarkable ability to mount an accelerated response to a pathogen to which it was previously exposed even decades before. Although it has long been known that memory T cells circulate through blood and lymphoid tissues, it is now clear that vast numbers of memory T cells also persist in other tissues, particularly those exposed to the outside world, such as the skin, gut and lung[9]. Not surprisingly, the distribution of tissue-resident memory cells mirrors sites of pathogen predilection; for example, influenza-specific memory cells are found mainly in the lungs.

Odoardi and colleagues have now shown that tissue-resident immune memory cells are not limited to pathogen-responsive populations but also include autoimmune cells, which can be activated to emigrate and mediate disease in a distant organ. Although the relevance of these findings to human disease remains speculative, distinct possibilities are evident. For example, healthy humans lack substantial BALT, but cigarette smoking, which is a potent risk factor for developing multiple sclerosis, is also known to induce BALT formation[10–12]. Furthermore, disease activity in patients with multiple sclerosis can be triggered by respiratory infections[13]. Both of these observations might be explained by a process in which myelin-specific autoimmune cells transit the lungs before establishing disease in the CNS, with some of these cells forming a resident lung population that can be subsequently reactivated. ■

Richard M. Ransohoff *is at the Neuroinflammation Research Center, Cleveland Clinic, Cleveland, Ohio 44195, USA.*
e-mail: ransohr@ccf.org

1. Odoardi, F. *et al. Nature* **488,** 675–679 (2012).
2. Kojima, K. *et al. J. Exp. Med.* **180,** 817–829 (1994).
3. Flügel, A. *et al. Immunity* **14,** 547–560 (2001).
4. Bartholomäus, I. *et al. Nature* **462,** 94–98 (2009).
5. Denk, W., Strickler, J. H. & Webb, W. W. *Science* **248,** 73–76 (1990).
6. Laudanna, C. & Constantin, G. *J. Immunol. Methods* **273,** 115–123 (2003).
7. Johnston, B. *et al. J. Immunol.* **164,** 3337–3344 (2000).
8. Ransohoff, R. M. *Nature* **462,** 41–42 (2009).
9. Purwar, R. *et al. PLoS ONE* **6,** e16245 (2011).
10. Moyron-Quiroz, J. E. *et al. Nature Med.* **10,** 927–934 (2004).
11. Richmond, I. *et al. Thorax* **48,** 1130–1134 (1993).
12. van der Mei, I. A. *et al. Neurol. Clin.* **29,** 233–255 (2011).
13. Sibley, W. A., Bamford, C. R. & Clark, K. *Lancet* **1,** 1313–1315 (1985).

OCEAN SCIENCE

Ancient burial at sea

A study reveals cyclic changes in the rate of burial of biogenic calcium carbonate at the Pacific ocean floor 43 million to 33 million years ago, as Earth exited a warm 'greenhouse' state to become an ice-capped planet. SEE ARTICLE P.609

HEATHER STOLL

From the highest mountains to the coastal plains, rocks are continually dissolved by rain and washed into the sea. The resulting dissolved ions are used by calcifying organisms in the ocean to make shells; when these organisms die, their shells 'rain' down on the sea floor, forming great piles of calcium carbonate. On page 609 of this issue, Pälike *et al.*[1] provide a record of the accumulation of calcium carbonate in the Pacific Ocean over the past 53 million years, based on sedimentary cores drilled out of the sea floor (Fig. 1). This reveals the dramatic shifts in carbonate burial over this period in a new level of detail, and also their relationship with climate.

During periods when more mountains dissolve, calcium carbonate accumulates at progressively deeper regions of the ocean floor, ensuring a balance between the rate at which ions are added to and removed from

IODP/USIO

Figure 1 | Cores for thought. Pälike *et al.*[1] have drilled cores of sediment from the floor of the Pacific Ocean. The image shows part of one such core, which was taken from a water depth of 4,000 metres. The muds change from dark (older, deeper sediments) to pale (younger, upper sediments), reflecting the surge of white calcium carbonate that has been deposited from 33 million years ago. The section shown is 450 centimetres in length.

the ocean. The burial of this carbonate at the bottom of the ocean removes carbon dioxide from the atmosphere, cleaning up the CO_2 that is continuously leaked into the atmosphere by volcanoes. But if the burial rate of carbonate outpaces the rate of CO_2 addition, atmospheric levels of CO_2 drop, which causes the climate to cool.

The authors show that about 33 million years ago, calcium carbonate began to accumulate in much deeper parts of the ocean than had been the case previously, suggesting that rocks on land were dissolving much faster. This could have been because of the major collision that occurred between India and Asia at around this time. The collision shoved up high-elevation mountains[2] in the reaches of the tropical rain belt, and this has long been surmised to have fuelled a greater rate of rock dissolution[3]. At the same time, the climate plunged into an 'icehouse' phase — a cool period characterized by the presence of continental ice sheets. The abrupt onset of glaciation on Antarctica lowered sea levels[4] and diverted some of the calcium carbonate that would have accumulated in shallow coral reefs into the deep ocean.

Pälike and colleagues' second key finding is that surprisingly large and geologically rapid oscillations (with a period of 1 million to 2 million years) occurred in the depth of calcium carbonate accumulation on the Pacific sea floor during the 10 million years before the icehouse plunge. However, it is difficult to explain these oscillations by invoking changes in the supply of solutes (dissolved ions) from continents to the ocean — mountains are not built and unbuilt in a day, or even in a million years. Furthermore, ice caps had not reached sufficient size to significantly lower sea level at this time.

Another possible explanation is that cyclic increases in atmospheric CO_2 and temperature periodically accelerated the dissolution rate of rocks. But if this had occurred, the additional solutes supplied at peak dissolution times would have remained dissolved in the ocean, because the higher concentration of atmospheric CO_2 would have made the ocean more acidic. In other words, there would have been no change in the depth at which carbonate accumulated on the sea floor. So what could explain the seemingly indecisive entry into the icehouse implied by the oscillations?

Pälike *et al*. propose an answer that involves the scores of hungry microbes that feast on the remains of algae raining down into deep water from the sunny ocean surface. These microbes respire CO_2, which, in turn, dissolves the calcium carbonate in the shells descending from the surface before it can accumulate on the sea floor. The authors suggest that, if this microbial CO_2 production had been curbed periodically (either by a shift in the composition of algal populations to types that are more resistant to microbial degradation, or by a cooling of the ocean that slowed this degradation[5]), then algal remains would have been buried rather than converted into CO_2. The result would have been a periodic accumulation of carbonate in the deep ocean, as observed.

To substantiate this proposed mechanism, future studies must explore the causes of the proposed synchronous global changes in marine ecology and/or identify evidence that global temperature oscillations affected respiration rates. Because the suggested mechanism could act as positive feedback for climate cooling by removing CO_2 from the atmosphere at faster rates during cold periods, mechanisms that could explain the reversal of the cycle on million-year timescales must also be found.

Alternatively, there may be ways in which the delivery of solutes to the ocean can be rapidly increased without relying on increases in CO_2 and temperature. For example, rates of rock dissolution are also strongly enhanced by intense rainfall. In tropical regions, the amount of rainfall for a given hemisphere increases when Earth's orbit is highly elliptical and the summer season coincides with the perihelion for that hemisphere (the point of the orbit at which the hemisphere is closest to the Sun). In the Irrawaddy River system[6] in southeast Asia, weathering intensity has been shown to increase with these precipitation cycles, rather than with global temperature. During the interval from 55 million to 35 million years ago, when highly soluble rocks in India were entering the tropical rainfall belt[7], modulation of the monsoon system due to changes in the ellipticity of Earth's orbit (which have a long periodicity of 1.2 million to 2.5 million years[8]) could have contributed to intense variations in solute flux.

Still, large changes in ocean ecology or solute fluxes are required to explain the oscillations in the Pacific record only if those oscillations are representative of changes also occurring throughout the Atlantic and other oceans. The global current in the deep ocean accumulates CO_2 from microbial respiration along its flow path, and therefore curtails carbonate accumulation in the ocean basin at the 'end of the line'. So, if the direction of the current reverses, carbonate accumulation becomes shallower in ocean basins close to the end of the current's path, and deeper at basins close to the start. It is therefore important to determine in upcoming drilling expeditions whether the rapid cycles in Pacific carbonate burial also occurred in the Atlantic at precisely the same time. If the cycles are opposite to each other, that would suggest that the same quantities of solutes were precipitating from the global ocean throughout the period concerned, but the precipitation was alternating between the basins, as has occurred during the glacial cycles of the past million years.

Verifying that ancient carbonate-accumulation events are globally synchronous on timescales of a million years or so requires precise correlations of sediments from one ocean to the next, necessitating careful integration of geochemical, sedimentological and palaeontological data. This is exactly the kind of detailed, interdisciplinary collaboration on which the international Integrated Ocean Drilling Program — of which Pälike and colleagues' study was a part — is based. So I, for one, am optimistic that the answer to this question is close at hand. ■

Heather Stoll *is in the Department of Geology, University of Oviedo, 33005 Oviedo, Spain.*
e-mail: hstoll@geol.uniovi.es

1. Pälike, H. *et al. Nature* **488,** 609–614 (2012).
2. Dupont-Nivet, G., Hoorn, C. & Konert, M. *Geology* **36,** 987–990 (2008).
3. Raymo, M. E. & Ruddiman, W. F. *Nature* **359,** 117–122 (1992).
4. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. *Nature* **433,** 53–57 (2005).
5. Olivarez Lyle, A. & Lyle, M. W. *Paleoceanography* **21,** PA2007 (2006).
6. Colin, C., Turpin, L., Bertaux, J., Desprairies, A. & Kissel, C. *Earth Planet. Sci. Lett.* **171,** 647–660 (1999).
7. Kent, D. V. & Muttoni, G. *Proc. Natl Acad. Sci. USA* **105,** 16065–16070 (2008).
8. Laskar, J. *et al. Astron. Astrophys.* **428,** 261–285 (2004).

OPTICS

Mixing waves in a diamond

Use of an ultra-high-intensity X-ray laser has allowed X-ray and optical waves to be mixed in a diamond sample. The effect paves the way to studying the microscopic optical response of materials on an atomic scale. SEE ARTICLE P.603

NINA ROHRINGER

The invention of the optical laser introduced the field of nonlinear optics — the study of the nonlinear interaction of light with matter. The field kicked off with the observation of second-harmonic generation[1], a nonlinear process that converts two photons of identical frequency into a single photon that has twice the frequency of the initial photons. Recently developed X-ray free-electron lasers (XFELs)[2,3], which produce ultrashort X-ray pulses of extremely high intensity, hold similar promise in the X-ray domain of the electromagnetic spectrum. On page 603 of this issue, Glover *et al.*[4] report how they have used an XFEL device to observe X-ray and optical sum-frequency generation, a nonlinear optical effect that may allow the microscopic optical properties of materials to be measured with atomic resolution.

If a medium is irradiated with low-intensity light of a specific frequency, the medium's optical response is typically linear. In this regime, the induced electric polarization of the medium has frequency components that match those of the impinging light field — the medium's optically active (polarizable) electric charge density oscillates in phase with the light's electric field. Linear light–matter interactions, which include absorption and scattering, do not alter the medium's optical properties.

However, for light of sufficiently high intensity, the optical properties change and the medium displays a nonlinear response: the induced polarization has frequency components that differ from those of the electric field. As a result, a medium that is irradiated with a laser beam containing, say, two frequency components (w_1 and w_2) can emit light with a frequency (w) that is either the sum of the components ($w = w_1 + w_2$) or the difference between them ($w = w_1 - w_2$). The same effects, known respectively as sum-frequency generation and difference-frequency generation, can occur if two temporally and spatially overlapping laser beams of distinct frequencies (w_1 and w_2) are shone on the sample. With optical light, the nonlinear response to high orders — emitted light frequencies that are multiples of those of the incident light — has been observed and exploited to create coherent (laser-like), short-wavelength radiation in the ultraviolet and extreme-ultraviolet regimes.

At X-ray wavelengths, the induced polarization is usually too small for sum-frequency generation to occur. Glover *et al.*[4] circumvent this problem by combining optical and X-ray light. They overlapped X-ray pulses from the Linac Coherent Light Source XFEL[2] with pulses from an optical laser, shone them on a diamond sample and demonstrated sum-frequency generation of the X-ray and optical pulses. This effect was predicted[5] in 1971, but had not previously been observed because there were no X-ray sources of sufficiently high intensity. By measuring the efficiency with which the sum-frequency signal was generated, the authors determined the optically induced microscopic polarization and the associated electric field. They then compared the results with theory, and confirmed the expectation that the optically induced polarization in diamond is associated with charges in the crystal's covalent bonds.

The effect of optical and X-ray wave mixing in crystals can be interpreted as optically modulated X-ray diffraction[6]. Optical light of frequency w_0 induces a polarization in the medium, resulting in a temporally oscillating redistribution of the polarizable charge density. Typically, only valence electrons are polarizable when irradiated with optical light. X-rays of frequency w_x interacting with the optically active, oscillatory part of the charge density will scatter off it inelastically, with their frequency being Doppler shifted to $w = w_x + w_0$. Most of the X-rays, however, will scatter elastically (their frequency w_x is preserved) off the optically unaffected, static component of the charge density. This elastic scattering gives rise to a standard Bragg intensity peak on a detector. The inelastic process generates an intensity peak that is slightly shifted from the elastic Bragg peak (Fig. 1).

In standard X-ray crystallography, X-ray illumination and scattering at different crystal orientations yield information about a crystal's three-dimensional charge density. In much the same way, the optically modulated, scattered X-rays can allow the three-dimensional, optically induced charge-density changes to be reconstructed[5,6]. Glover and colleagues' experiment is a first proof of principle of this reconstruction: sum-frequency signals were recorded for only one crystal orientation. But extending the experiment to several crystal orientations, which together will allow determination of the three-dimensional, induced charge-density variation, should be straightforward.

X-ray and optical wave mixing in crystals therefore enables optical polarization to be measured on a microscopic scale. Light-activated microscopic polarization and the associated electric field vary widely on an atomic scale, and determine the overall macroscopic optical properties of materials, including the refractive index. Although conceptually familiar, these microscopic details had never been measured before, because standard optical methods give only a coarse-grained view of them. Glover and colleagues demonstrate that X-rays make

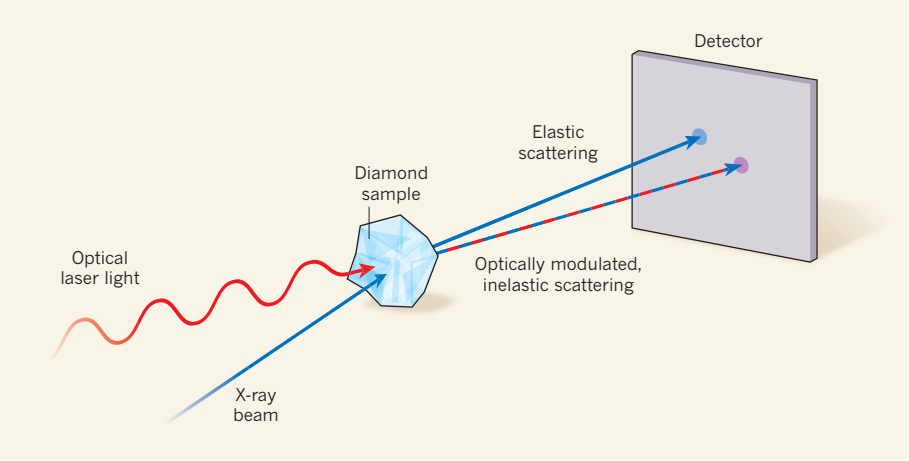


Figure 1 | Optically modulated X-ray diffraction. Glover *et al.*[4] directed optical laser light and a beam of X-ray light from the Linac Coherent Light Source into a diamond sample. Most of the beam's X-rays are scattered elastically (their wavelength is preserved) and produce the standard Bragg intensity peak on the detector. The rest of the X-rays undergo inelastic scattering, resulting in optically modulated X-rays which have a frequency that is the sum of the X-ray frequency and the laser frequency. These rays generate an intensity peak that is slightly shifted from the standard peak.

it possible to resolve and quantify such details.

The ability to use optical and X-ray sum-frequency generation to probe light–matter interactions at a microscopic scale will potentially advance the fields of optics, energy research and materials science. If optical lasers that emit pulses at high repetition rates (on the megahertz regime) are available at X-ray synchrotron facilities, it may be possible to explore the effect using these more accessible X-ray sources instead of an XFEL. This would allow extensive studies of stationary, microscopic optical properties. But to resolve dynamic light-induced microscopic processes in real time — a challenge that still lies ahead of us — the power of XFELs is indispensable. ■

Nina Rohringer *is at the Max Planck Institute for the Physics of Complex Systems, Dresden, and in the Center for Free-Electron Laser Science, 22607 Hamburg, Germany.*
e-mail: nina.rohringer@asg.mpg.de

1. Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. *Phys. Rev. Lett.* **7,** 118–119 (1961).
2. Emma, P. *et al. Nature Photon.* **4,** 641–647 (2010).
3. Ishikawa, T. *et al. Nature Photon.* **6,** 540–544 (2012).
4. Glover, T. E. *et al. Nature* **488,** 603–608 (2012).
5. Eisenberger, P. M. & McCall, S. L. *Phys. Rev. A* **3,** 1145–1151 (1971).
6. Freund, I. & Levine, B. F. *Phys. Rev. Lett.* **25,** 1241–1245 (1970).

EPIGENETICS

Actors in the cell reprogramming drama

The transformation of skin cells into stem cells is a fascinating but poorly understood process. At last, the molecular characters underlying the initial steps have been revealed. SEE LETTER P.652

KYLE M. LOH & BING LIM

Coercing cells into adopting new fates — for example, turning skin into muscle[1], brain[2] or heart[3] — is one of the most impressive achievements of modern biology. Prominent among these feats is the transformation of skin cells into induced pluripotent stem cells (iPSCs)[4]. In contrast to skin cells (fibroblasts) and other differentiated cells, which cannot generate any other cell type except their own, iPSCs have the remarkable ability to give rise to any kind of cell in the body[4]. Yet the molecular changes that drive this cellular metamorphosis have remained a mystery. For the first time, Doege *et al.*[5] (page 652) provide a mechanistic explanation for how the earliest phases of such a dramatic change in cell fate might be executed.

Cellular identities were once thought immutable[6]: when a cell became differentiated, it could not return to a stem-cell state. In differentiated cells, genes typically expressed in other cell types are repressed, thereby consigning the cells to their present fate and preventing excursions into other lineages[7,8]. Likewise, stem-cell pluripotency genes are often inactive in differentiated cells[4]. However, the artificial introduction of several transcription factors (such as Oct4, Sox2 and Klf4)[4] can reverse such lineage restrictions, inducing many differentiated cell types to become iPSCs.

Gene expression is jointly controlled by two classes of protein: transcription factors and epigenetic regulators. Transcription factors act by binding directly to DNA, whereas epigenetic regulators can influence gene expression in various ways, for example by altering histones[9] (proteins that package DNA). Doege *et al.* identify two epigenetic regulators — Parp1 and Tet2 — that stimulate the expression of dormant pluripotency genes in fibroblasts, thereby initiating the reprogramming of these cells into iPSCs. The authors show that, during cell reprogramming, Parp1 and Tet2 induce the removal of a repressing mark (known as H3K27me3) from histones associated with pluripotency genes, as well as the addition of a different mark (H3K4me2) that is thought to have an activating effect[10]. In particular, they find that this occurs in the pluripotency genes *Nanog* and *Esrrb*, days before the genes become active (Fig. 1).

Doege and colleagues show that these epigenetic changes presage future increased expression of *Nanog* and *Esrrb* and correlate with enhanced binding of Oct4 to these genes, which presumably further stimulates their expression. Strikingly, both endogenous Parp1 and Tet2 are essential for reprogramming — if either is abrogated, iPSC formation fails totally[5,11]. This finding suggests that the reprogramming transcription factors cannot efficiently bind to and reawaken their target genes unless they collaborate with endogenous epigenetic regulators, which preconfigure the repressed genes for activation days before their expression is increased.

Transcription factors such as Oct4 and Sox2 have long occupied the pantheon of cardinal reprogramming factors[6], but now Doege *et al.* show that epigenetic regulators are equally important. However, these regulators are not widely introduced into cells alongside transcription factors to induce reprogramming[6]. So, where in the reprogramming ensemble are the obligatory epigenetic regulators? The present report provides an elegant explanation: that the reprogramming transcription factors liaise with endogenous epigenetic regulators to

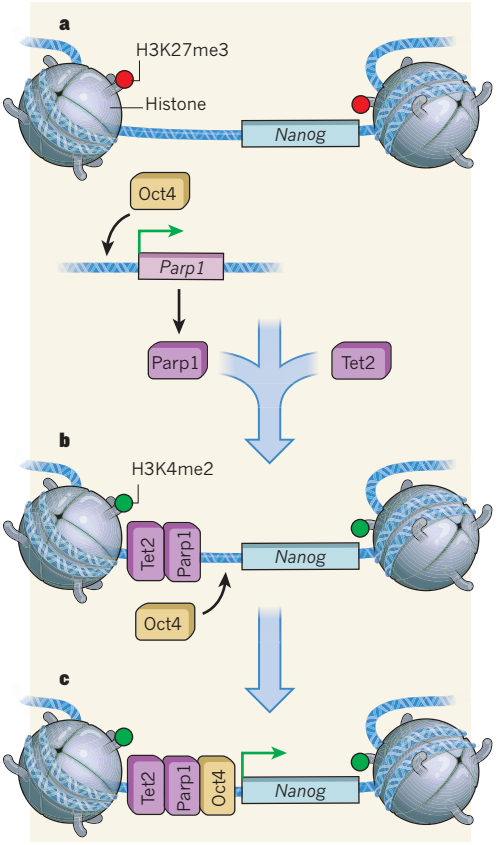


Figure 1 | Cell metamorphosis. Skin fibroblasts and other differentiated cells can be reprogrammed to become stem cells by the introduction of transcription-factor proteins such as Oct4. These reprogramming factors induce the expression of *Nanog* and other genes that are inactive in differentiated cells and active in stem cells, although the mechanisms underlying such gene reactivation have been obscure. **a**, In fibroblasts, the *Nanog* gene is surrounded by histone proteins that carry a chemical modification (H3K27me3) that is commonly associated with repressed genes. Doege *et al.*[5] report that the exogenously added Oct4 induces the expression of the protein Parp1. **b**, Early during reprogramming, this protein and another protein, Tet2, bind to the regulatory region of *Nanog*, facilitating depletion of the repressive H3K27me3 modification and the acquisition of another histone mark (H3K4me2) that is often found on genes poised for reactivation[10]. **c**, Such changes probably help Oct4 to bind strongly to, and so enhance the expression of, *Nanog*. Several days later, *Nanog* expression becomes markedly high. Subsequent cellular events, which remain unclear, lead to the production of full-fledged stem cells.

execute reprogramming. These regulators may already be present in the starting cell[12], or their expression can be stimulated by the added transcription factors — indeed, Doege *et al.* show that Oct4 induces Parp1 expression in fibroblasts (Fig. 1). However, some cell-lineage switches do explicitly require co-introduction of a particular epigenetic regulator[13]. Therefore, efforts to reprogram cells to currently unattainable lineages might benefit from the addition of relevant epigenetic modifiers.

The authors' findings provoke further questions. Although Parp1 and Tet2 reactivate pluripotency genes such as *Nanog* and *Esrrb*, what extinguishes the expression of fibroblast-specific genes? Moreover, is there some hierarchical order in which the awakened pluripotency proteins activate more of their dormant kin?

Another unresolved issue is how epigenetic regulators 'know' which genes need to be reactivated. Do they derepress gene expression broadly across the genome, or are they guided to specific genes by the reprogramming transcription factors? It is known that Oct4 binds to Parp1 in stem cells[14], which suggests that it might target Parp1 to reactivate specific pluripotency genes. Are certain epigenetic regulators (for example, Parp1) universally required to convert any cell type into any other one? Some fundamental epigenetic machinery, such as the TrxG H3K4 methyltransferase protein complex[12], might be broadly required to activate the expression of alternative-lineage genes for any reprogramming event. By contrast, cell-type-specific epigenetic regulators might direct transformation into selected lineages — for example, the regulator Baf60c specifically induces conversion into heart cells[13].

The main import of Doege and colleagues' article is that it begins to clarify, at surprisingly early stages of the process and in unprecedented detail, the mechanisms by which cell types can be switched from one to another. This supersedes previous phenomenological descriptions of the reprogramming process and provides insight into its underlying mechanisms. If we can ascertain the fundamental principles underlying such changes in cell identity, the terra incognita of cellular reprogramming may at last be mapped, expanding the ways in which diverse cell types can be generated for cell-based therapies. ■

Kyle M. Loh *is in the Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA.* **Bing Lim** *is at the Stem Cell & Developmental Biology Group, Genome Institute of Singapore, 138672 Singapore, and at the Beth Israel Deaconess Medical Center, Boston, Massachusetts.*
e-mails: kyleloh@stanford.edu; limb1@gis.a-star.edu.sg

1. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51,** 987–1000 (1987).
2. Vierbuchen, T. *et al. Nature* **463,** 1035–1041 (2010).
3. Ieda, M. *et al. Cell* **142,** 375–386 (2010).
4. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
5. Doege, C. A. *et al. Nature* **488,** 652–655 (2012).
6. Graf, T. *Cell Stem Cell* **9,** 504–516 (2011).
7. Enver, T. & Greaves, M. *Cell* **94,** 9–12 (1998).
8. Loh, K. M. & Lim, B. *Cell Stem Cell* **8,** 363–369 (2011).
9. Rosenfeld, M. G., Lunyak, V. V. & Glass, C. K. *Genes Dev.* **20,** 1405–1428 (2006).
10. Koche, R. P. *et al. Cell Stem Cell* **8,** 96–105 (2011).
11. Lai, Y.-S. *et al. Proc. Natl Acad. Sci. USA* **109,** 3772–3777 (2012).
12. Ang, Y.-S. *et al. Cell* **145,** 183–197 (2011).
13. Takeuchi, J. K. & Bruneau, B. G. *Nature* **459,** 708–711 (2009).
14. Pardo, M. *et al. Cell Stem Cell* **6,** 382–395 (2010).

ASTRONOMY

Collision course

Four billion years from now, the Andromeda galaxy will have a close encounter with the Milky Way. The two galaxies will commence a dance of disruption that will, over the course of another two billion years, lead to their complete union.

R. BRENT TULLY

In a series of three papers published in *The Astrophysical Journal*, van der Marel and collaborators[1–3] discuss the timing and dynamics of the 'imminent' — or at least inevitable — collision between the Milky Way and the neighbouring Andromeda galaxy*. The inevitable aspect of this situation has been known for a long time. In 1959, Kahn and Woltjer estimated[4] the mass of the Milky Way–Andromeda system through what has come to be called the 'timing argument', which invoked the assumption that the two galaxies would ultimately collide.

The argument posits that the two galaxies — or rather, their current contents — started in the Big Bang in close proximity, flew apart with cosmic expansion but were arrested by their mutual attraction, and are now falling back towards each other. On the basis of the simplistic assumption that the two are moving directly towards each other, Kahn and Woltjer estimated a mass for the system that is in good agreement with current projections.

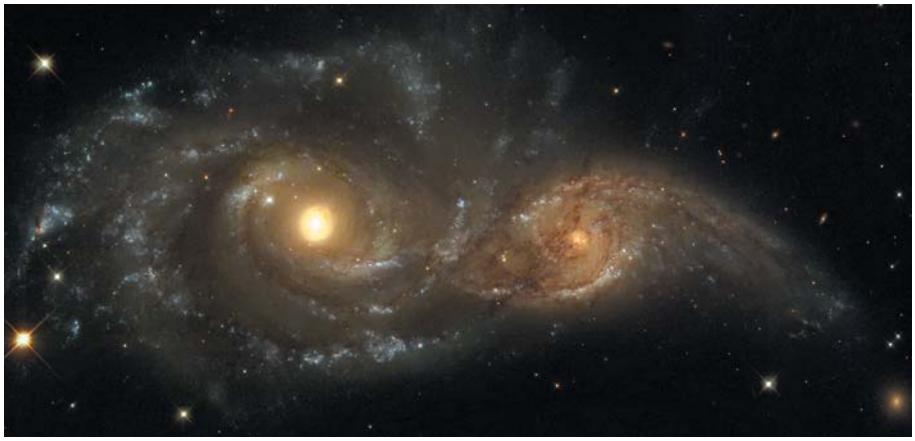
Getting down to details, there is the possibility that the galaxies orbit one another with angular momenta caused by tidal forces from such prominent players as the Maffei, Messier 81 and Centaurus A galaxies. The question since 1959 has been whether the two galaxies would collide on first return or fly past each other. The answer lies in the relative translational velocities of the two systems. How fast is Andromeda's position displacing on the sky? We have accurate knowledge of how fast it is moving in the line of sight, and we know well enough its distance from Earth. If we know its displacement rate on the sky (proper motion), then we can calculate its orbit.

We are only just beginning to measure the proper motions of galaxies. They are relatively well determined for the Milky Way's immediate neighbours, such as the Large and Small Magellanic Clouds. It turns out that the proper motions have been measured for two neighbours of the Andromeda galaxy, Messier 33 and Index Catalogue 10, because they emit water masers[5,6] — radio signals associated with the excitation of water molecules in star-forming regions. The observations of these masers are made using radio interferometers with baselines (extending from Hawaii to Maine) that afford angular resolution at the level of 10 microarcseconds. But masers in Andromeda have been found only recently, so a measurement of its proper motion had to come from the displacement of the galaxy's stars in the manner obtained by van der Marel and colleagues.

Here is the challenge. At a good observation site on the ground, a star is blurred to a disk with a diameter of about 1 arcsecond. The expected annual proper motion of Andromeda is only a few parts in 100,000 of this blurry dimension. Van der Marel and his team[1–3] helped themselves by making observations with the Hubble Space Telescope, which allowed them to obtain a resolution of 0.1 arcseconds. Given typical galaxy orbits, the authors predict displacements of Andromeda's star centres at the level of one part in 1,000–10,000 of the measured star diameters per year. It does not help that the stars they examined were in crowded regions (the brightest stars among the billions superimposed on each other in the Andromeda galaxy) and that the background references for measuring the motions are galaxies of irregular shapes that therefore have ill-defined centres; these background reference galaxies are so far away that they are effectively motionless.

To improve the statistics, the authors were able to make measurements of several thousand stars against a static background of several hundred distant galaxies. To look for changes in star positions, they made observations of three fields in Andromeda at separations of 5 to 7 years. The final accuracy of the measured proper motion of the Andromeda galaxy on the sky was 12 microarcseconds

*This News & Views article was published online on 22 August 2012.



NASA, ESA & THE HUBBLE HERITAGE TEAM (STSCI)

Figure 1 | A cosmic encounter. Van der Marel and colleagues[3] predict that the Milky Way and the Andromeda galaxy will pass each other in 4 billion years. This encounter will be similar to that currently occurring between the spiral galaxies NGC 2207 and IC 2163, which are seen here in an image obtained by the Hubble Space Telescope.

per year, and the authors claim a detection of motion at greater than 99% certainty.

So what is Andromeda's orbit? Not surprisingly, the orbit inferred by van der Marel and co-workers has large error bars[1,2]. However, an essential point seems to be well founded. The first passage of Andromeda about the Milky Way is going to be close enough to make a big mess. A perfect bullseye is within the range of orbital uncertainty. The central probability in a wide distribution of orbits is that the two systems will pass each other 4 billion years from now with their nuclei separated by about 30 kiloparsecs (about 100,000 light years). For a scale comparison, the Sun lies about 8.3 kpc from the centre of our Galaxy. It is not known where the Sun will be in its orbit at the time of this closest passage, because by that time the Solar System will have made 14 more laps around the centre of the Milky Way — too many laps to be precise about details.

Four billion years from now, our progeny will see, if they still have dark skies and keen eyes, quite a spectacle as Andromeda fills the horizon — just imagine being a resident in one of the members of the pair of colliding spiral galaxies NGC 2207 and IC 2163 (Fig. 1). The most probable condition of the Andromeda–Milky Way encounter involves alignment of the spin orientation of the galaxies with the direction of the orbital motions. The orbit of Andromeda will lie close to the plane of the Milky Way. These are conditions that give rise to large tidal effects.

The two galaxies will rip past each other at almost 600 kilometres per second at the nearest approach; after that the galaxies will rapidly slosh together. Van der Marel and colleagues[3] declare that the systems will merge roughly 6 billion years from now (2 billion years after first passage), when the centres come within less than 25 kpc of each other. The outcome of this wreck will almost certainly be a single elliptical galaxy comprising the stars and the total mass of the two currently separate systems — this is what computer simulations predict for the merging of large, equal-mass spirals. On the basis of simulations, the authors demonstrate that the Sun, which will by then be nearing the exhaustion of hydrogen fuel in its core, will probably lie at a considerable distance from the centre of the merged system. The chance of the Sun actually colliding with another star is small.

Van der Marel and colleagues describe a curious coda. Messier 33, the third-largest galaxy in the Local Group of galaxies, is currently taking a wide swing around Andromeda. It will surf Andromeda's gravity in the ride towards the Milky Way. However, Messier 33 has sufficient orbital angular momentum that it is unlikely to be absorbed by the merging pair soon. Well after the two big galaxies coalesce, the witness Messier 33 will probably be circling the star pile in a slowly decaying orbit towards an eventual demise. ■

R. Brent Tully *is at the Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA.*
e-mail: tully@ifa.hawaii.edu

1. Sohn, S. T., Anderson, J. & van der Marel, R. P. *Astrophys. J.* **753,** 7 (2012).
2. van der Marel, R. P. *et al. Astrophys. J.* **753,** 8 (2012).
3. van der Marel, R. P., Besla, G., Cox, T. J., Sohn, S. T. & Anderson, J. *Astrophys. J.* **753,** 9 (2012).
4. Kahn, F. D. & Woltjer, L. *Astrophys. J.* **130,** 705–717 (1959).
5. Brunthaler, A., Reid, M. J., Falcke, H., Greenhill, L. J. & Henkel, C. *Science* **307,** 1440–1443 (2005).
6. Brunthaler, A., Reid, M. J., Falcke, H., Henkel, C. & Menten, K. M. *Astron. Astrophys.* **462,** 101–106 (2007).

MICROBIOLOGY

Antibiotics and adiposity

Mice receiving low doses of certain antibiotics gain weight and accumulate fat. This could be because some gut bacteria survive the treatment better than others, shifting digestion towards greater energy provision. SEE ARTICLE P.621

HARRY J. FLINT

It has been known for more than 50 years that prolonged treatment with low doses of certain antibiotics can promote growth in farm animals, such as chickens and pigs. This practice is still extensively employed, although concerns over the spread of bacterial resistance to clinically important antibiotics have led to restrictions on its use in some regions, especially Europe. However, the mechanisms behind this growth enhancement have never been entirely clear. On page 621 of this issue, Cho *et al.*[1] report a remarkable effect of low-dose antibiotic treatment in promoting fat deposition in growing mice, and propose that this increased adiposity results from perturbations in the species composition of the microorganisms that inhabit the mammalian intestinal tract.

The gut microbiota influences many aspects of nutrition and health, and previous animal studies have implicated the composition of this diverse microbial community as a factor in adiposity and weight gain[2]. To discover how antibiotics might affect this picture, Cho and colleagues treated mice with low doses of antibiotics (subtherapeutic antibiotic therapy, or STAT). They show that the treatment did not lead to an increase in food intake or to changes in hormones that influence satiety, which rules out these mechanisms as possible explanations for the additional fat accumulation that they observed in the treated mice. Instead, the authors report that STAT causes substantial changes in the balance of species within the gut microbiota and in the expression of microbial genes that convert carbohydrates to short-chain fatty acids (SCFAs). Although the researchers tend to consider the antibiotics used in their STAT experiments as a group, their results indicate that different antibiotics have distinct effects on the gut microbiota, and that penicillin most

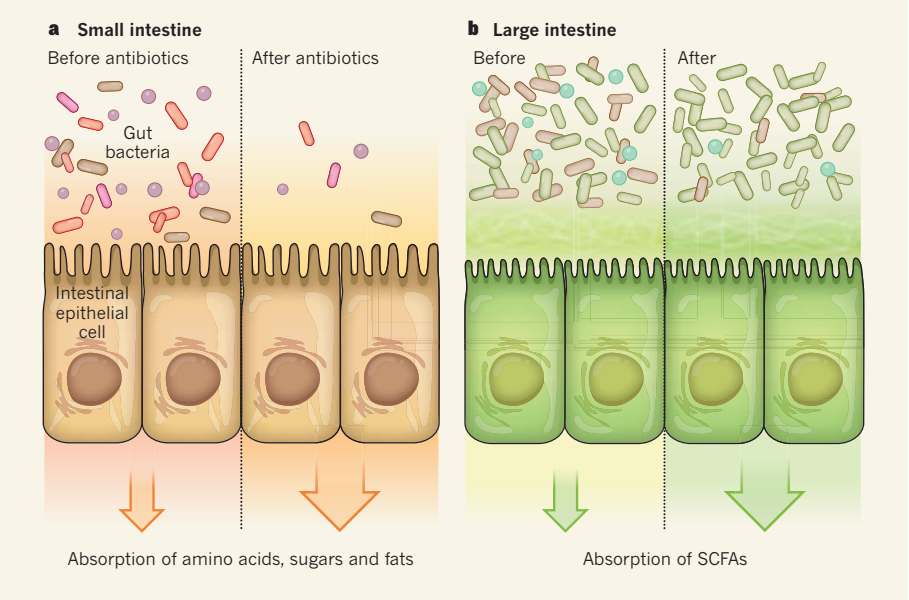


Figure 1 | Gut microbiota and energy gain. **a**, In the mammalian small intestine, food is digested into sugars, amino acids and fats that are absorbed across the intestinal epithelium for use by the host. In the small intestine, subtherapeutic doses of antibiotics might reduce the total number of bacteria or suppress bacterial activity, which could reduce production of bacterial toxins or competition from bacteria for nutrients, resulting in more food energy being left for the host. **b**, In the large intestine, dietary substrates that are hard to digest are broken down by microbial fermentation to form short-chain fatty acids (SCFAs). Cho and colleagues[1] suggest that, here, antibiotic treatment does not reduce overall bacterial numbers but rather shifts the microbial composition towards species that perform this fermentation, resulting in more efficient energy retrieval from ingested food, and more energy and SCFA absorption by the host.

consistently promotes adiposity.

Most diets contain a substantial amount of fibre, which is not digested in the small intestine but passes into the large intestine, where it provides a source of energy for resident microorganisms. Because oxygen is in short supply in this region, anaerobic fermentation by these microbes results in the formation of SCFAs that are absorbed across the mucosal lining of the gut (Fig. 1). Because SCFAs are also an additional source of energy for the host animal, changes in the efficiency of microbial fermentation and SCFA absorption can influence the recovery of calories from the diet. Naively, one might expect that antibiotic administration would tend to suppress gut bacterial numbers and activity, thereby reducing SCFA formation and the supply of energy to the host. On the contrary, the low levels of antibiotics used by Cho *et al.* had little effect on total bacterial numbers and actually resulted in increased SCFA concentrations in the mouse faeces. However, the authors did not demonstrate that this was the result of increased microbial fermentation rather than decreased SCFA absorption.

The two phyla that comprise the majority of bacteria in the mammalian intestine are the Bacteroidetes and the Firmicutes. Cho and colleagues report an increase in the proportion of Firmicutes in some groups of mice receiving STAT, and they propose that this taxonomic alteration makes the gut microbiota more efficient at extracting energy from the diet. In support of this, they detected less residual energy in the faeces of the treated animals. This finding echoes a previous study[2] demonstrating a higher proportion of Bacteroidetes in faecal samples from lean mice than in samples from mice that are genetically predisposed to obesity. Although it is now thought that interpretations based on phylum-level shifts are too simplistic[3,4], it is interesting to note that the Firmicutes include some species that have a key role in the degradation of recalcitrant dietary substrates[5,6].

Antibiotic treatment may also have other effects on energy provision for the host. For example, changes in the transit of food through the gut are known to affect SCFA absorption and fibre degradation[7], and it is possible that microbial signals differently influence the muscles that control transit through the gut. And quite apart from any changes in energy supply, it is also possible that certain gut microbes (or their metabolic products) regulate adipocyte (fat cell) development and lipid metabolism during the early stages of mammalian growth. Intriguingly, germ-free mice, which lack any microorganisms in their guts, can show either greater or lesser adiposity than mice with normal gut bacteria, depending on the diet supplied[8]. This illustrates the subtlety and complexity of the interactions between diet, microbes and host that influence adiposity.

Despite Cho and colleagues' demonstration that changes in gut microbiota composition accompany increased fat accumulation in response to STAT, the existence of particular groups of 'fat bugs' within the guts of mammals that might promote adiposity remains debatable. Several other mechanisms have been considered previously to explain the growth-enhancing effects of antibiotics in non-ruminant farm animals[9]. One is the suppression by antibiotics of toxin-producing microorganisms and subclinical infections, particularly in the post-weaning period, which might reduce the energy expended by the animal on gut-tissue replacement and immune responses. However, this seems less likely to be a factor in 'clean' laboratory animals, such as the mice in Cho and colleagues' studies, than in farm animals.

Another intriguing possibility, not considered by Cho *et al.*, is that a major suppression of bacterial activity by antibiotics in the small intestine might remove competition from bacteria for digestible food components, and thereby increase nutrient absorption in this region and energy flow to the host (Fig. 1). This might also increase the amount of fibre reaching the large intestine. It is quite possible, therefore, that the key effects of STAT upon adiposity and weight gain occur in the small intestine rather than in the large intestine.

Future investigations to decipher the mechanisms linking antibiotic-induced perturbations of the gut microbiota and fat deposition will need to consider nutritional physiology in more detail. This could be achieved, for example, by examining events in different regions of the intestinal tract together with the flow of food being digested and SCFA fluxes, in addition to determining gut microbiota composition and metabolite concentrations. Despite some important differences in digestive physiology and gut microbiology between mice and humans, these studies promise to enhance our understanding of the development of adiposity in people, and such insights could even have implications for our clinical use of antibiotics, especially during early childhood. ■

Harry J. Flint *is at the Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen AB21 9SB, UK.*
e-mail: h.flint@abdn.ac.uk

1. Cho, I. *et al. Nature* **488,** 621–626 (2012).
2. Turnbaugh, P. J. *et al. Nature* **444,** 1027–1031 (2006).
3. Murphy, E. F. *et al. Gut* **59,** 1635–1642 (2010).
4. Jumpertz, R. *et al. Am. J. Clin. Nutr.* **94,** 58–65 (2011).
5. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. *ISME J.* **6,** 1535–1543 (2012).
6. Flint, H. J., Scott, K., Duncan, S., Louis, P. & Forano, E. *Gut Microbes* **3,** http://dx.doi.org/10.4161/gmic.19897 (2012).
7. Stephen, A. M., Wiggins, H. S. & Cummings, J. H. *Gut* **28,** 601–609 (1987).
8. Fleissner, C. K. *et al. Br. J. Nutr.* **104,** 919–929 (2010).
9. Butaye, P., Devriese, L. A. & Haesebrouck, F. *Clin. Microbiol. Rev.* **16,** 175–188 (2003).

ARTICLE

doi:10.1038/nature11340

X-ray and optical wave mixing

T. E. Glover[1], D. M. Fritz[2], M. Cammarata[3], T. K. Allison[4], Sinisa Coh[5,6], J. M. Feldkamp[2], H. Lemke[2], D. Zhu[2], Y. Feng[2], R. N. Coffee[2], M. Fuchs[7], S. Ghimire[7], J. Chen[7,8], S. Shwartz[8], D. A. Reis[7,8,9], S. E. Harris[8,10] & J. B. Hastings[2]

Light–matter interactions are ubiquitous, and underpin a wide range of basic research fields and applied technologies. Although optical interactions have been intensively studied, their microscopic details are often poorly understood and have so far not been directly measurable. X-ray and optical wave mixing was proposed nearly half a century ago as an atomic-scale probe of optical interactions but has not yet been observed owing to a lack of sufficiently intense X-ray sources. Here we use an X-ray laser to demonstrate X-ray and optical sum-frequency generation. The underlying nonlinearity is a reciprocal-space probe of the optically induced charges and associated microscopic fields that arise in an illuminated material. To within the experimental errors, the measured efficiency is consistent with first-principles calculations of microscopic optical polarization in diamond. The ability to probe optical interactions on the atomic scale offers new opportunities in both basic and applied areas of science.

Light–matter interactions have advanced our understanding of atoms, molecules and materials, and are also central to a number of areas of applied science. Although optical interactions have received a great deal of study, the microscopic details of how light manipulates matter are poorly understood in many circumstances. A material's optical response is complex, being determined by coupled many-body interactions that vary on the scale of atoms rather than on the scale of a long-wavelength applied field. Data are needed to combat this complexity, and so far it has not been possible to probe the microscopic details of light–matter interactions.

X-ray and optical wave mixing, specifically sum-frequency generation (SFG), was proposed nearly half a century ago as an atomic-scale probe of light–matter interactions[1,2]. The process is, in essence, optically modulated X-ray diffraction: X-rays inelastically scatter from optically induced charge oscillations and probe optically polarized charge in direct analogy to how standard X-ray diffraction probes ground-state charge. Furthermore, the optically induced microscopic field is determined because it is closely related to the induced charge[3–6]. So far it has not been possible to measure these two quantities directly. X-ray and optical wave mixing has frequently been discussed[1,2,4,7–12], but it has not yet been demonstrated owing to a lack of sufficiently intense X-ray sources. More generally, although there have been theoretical studies of nonlinear X-ray scattering[13–18], experimental observations have largely been confined to the spontaneous processes of X-ray parametric down-conversion[19–23] and resonant inelastic X-ray scattering[24,25].

X-ray free-electron lasers offer unprecedented brightness and new scientific opportunities[26]. Here we use an X-ray laser to demonstrate X-ray/optical SFG through the nonlinear interaction of the two fields in single-crystal diamond. Optically modulated X-ray diffraction from the (111) planes generates a sum-frequency (X-ray plus optical) pulse. The measured conversion efficiency (3×10^{-7}) determines the (111) Fourier component of the optically induced charge and associated microscopic field that arise in the illuminated sample. To within experimental errors of ~40%, the measured charge density is consistent with first-principles calculations of microscopic optical polarization in diamond. The ability to measure atomic-scale charges and fields induced by light should contribute to a better understanding of materials and create new ways to study photo-triggered dynamics.

X-ray and optical wave mixing

X-ray and optical wave mixing is an atomic-scale probe of optical interactions: X-rays provide atomic spatial resolution and light makes it possible selectively to probe optically polarized valence charge[1,2]. Sum-frequency radiation is produced when two applied fields simultaneously drive a coherent electronic response; the second field must polarize charge before the polarization due to the first field decays. For X-ray/optical SFG, the simultaneously polarized charge is equal to optically polarized valence charge because only this charge has significant polarizability for both the optical and X-ray fields. Tightly bound core electrons can only be polarized by the X-ray radiation. More generally, as the 'optical' wavelength is varied through the visible to the extreme-ultraviolet (EUV) and soft-X-ray regimes, the polarized charge corresponds to different charge components in a material. Therefore, various charge components can be selectively probed by adjusting the 'optical' wavelength. The SFG technique can, for instance, be extended to probe the full valence charge distribution by mixing X-rays with EUV radiation of a frequency high enough that all valence electrons respond uniformly as free electrons[8,20,23] (but low enough that the polarizability of tightly bound core charge is negligible).

X-ray/optical SFG is a parametric scattering process analogous to standard X-ray diffraction: kinematics are determined by energy and momentum conservation, and the generation of an SFG field is described by the wave equation. In direct analogy to standard X-ray crystallography, inelastically scattered sum-frequency X-rays probe a specific Fourier (that is, reciprocal-space) component of the charge density. Specifically, given an optical wavevector k_o and reciprocal lattice vector G, X-ray/optical SFG measures the $(Q = k_o + G)$th Fourier component of the optically induced change, $\delta\rho_o(Q)$, to the valence charge density.

X-ray/optical SFG, and the closely related process of X-ray/optical difference-frequency generation, probes the linear optical susceptibility on a microscopic length scale. Higher-order X-ray/optical mixing

[1]Advanced Light Source Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [2]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [3]Institut de Physique de Rennes, UMR 6251 UR1-CNRS, University Rennes 1, 35042 Rennes Cedex, France. [4]JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA. [5]Department of Physics, University of California, Berkeley, California 94720, USA. [6]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [7]PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [8]Department of Applied Physics, Stanford University, Stanford, California 94305, USA. [9]Department of Photon Science, Stanford University, Stanford, California 94305, USA. [10]Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA.

processes can probe nonlinear optical susceptibilities. X-rays, for instance, could scatter from optically induced charge oscillating at twice the optical frequency and thereby be modified by two units of the optical photon energy; this would make it possible to study the atomic-scale details of optical second-harmonic generation.

Observation of X-ray/optical SFG

We choose diamond as the material in which to demonstrate X-ray/optical SFG: its macroscopic optical[27] and microscopic structural[28] properties are well known and estimates exist for the X-ray/optical SFG efficiency in diamond, although these estimates differ by a factor of ~100 (refs 1, 9, 11). An X-ray pulse (8,000 eV, ~80 fs) and an optical pulse (1.55 eV, ~2 ps, ~10^{10} W cm^{-2}) simultaneously illuminate a diamond sample slightly detuned from the optimum orientation for elastic Laue geometry diffraction from the (111) planes (Fig. 1a). Scattered X-rays are wavelength resolved using a Si(220) crystal. The combination of sample detuning and wavelength discrimination reduced the intensity of elastically scattered (background) light reaching our detector by seven to eight orders of magnitude relative to optimum elastic diffraction conditions.

An SFG signal was detected at the expected sample and analyser angles. The signal was observed to depend on the simultaneous presence of X-ray and optical beams, and on optimization we obtained a signal rate of ~4,000 detected photons per second. An energy analyser scan (Fig. 1b) confirms a detected photon energy equal to the sum of the input X-ray and optical photon energies and indicates a rocking-curve width (17 ± 2 μrad) in excellent agreement with the intrinsic analyser rocking-curve width (~17 μrad). This indicates that the SFG beam is spectrally narrow and well collimated compared with the energy spread (~310 meV) and angle spread (~17 μrad) transmitted by the analyser (called the analyser acceptance). Simulations discussed below are in agreement with this observation.

The analyser was fixed in position and the variation of SFG power with sample angle was measured (Fig. 1c). The signal is sharply peaked at the angle satisfying energy and momentum conservation, and the measured rocking-curve width (8.1 ± 0.7 μrad) is equal, to within the errors, to half the measured analyser width (17 ± 2 μrad). A sample rotation of $\delta\theta$ causes a $2\delta\theta$ rotation of the output beam, which indicates that the sample rocking-curve width is set by the limited energy and angle acceptance of the analyser. This observation supports the above conclusion obtained from consideration of the analyser rocking curve (Fig. 1b), namely that the SFG beam is narrow in energy and angle compared with the analyser acceptance. These observations place a lower limit of ~8 μrad on the SFG rocking-curve width because the SFG intensity is constant over an angular range (~8 μrad) limited by the finite analyser acceptance. Simulations discussed below predict an SFG rocking-curve width of 20 μrad.

Figure 1d shows how the SFG signal varies with rotation of the optical polarization vector. We expect the efficiency to vary as $(\varepsilon_o \bullet G_{111})^2$, where ε_o is the optical polarization vector and G_{111} is the (111) lattice vector. This scaling reflects the quadratic dependence of SFG efficiency on the linearly induced optical charge, which in turn varies as $\varepsilon_o \bullet G_{111}$. The SFG signal exhibits the expected dependence: it is maximized when ε_o is in the scattering plane and it is zero when ε_o is normal to the scattering plane.

The SFG efficiency was determined by measuring the incident X-ray pulse energy (photons per pulse) and, under both elastic Laue and SFG conditions, the outgoing X-ray pulse energy. The measured

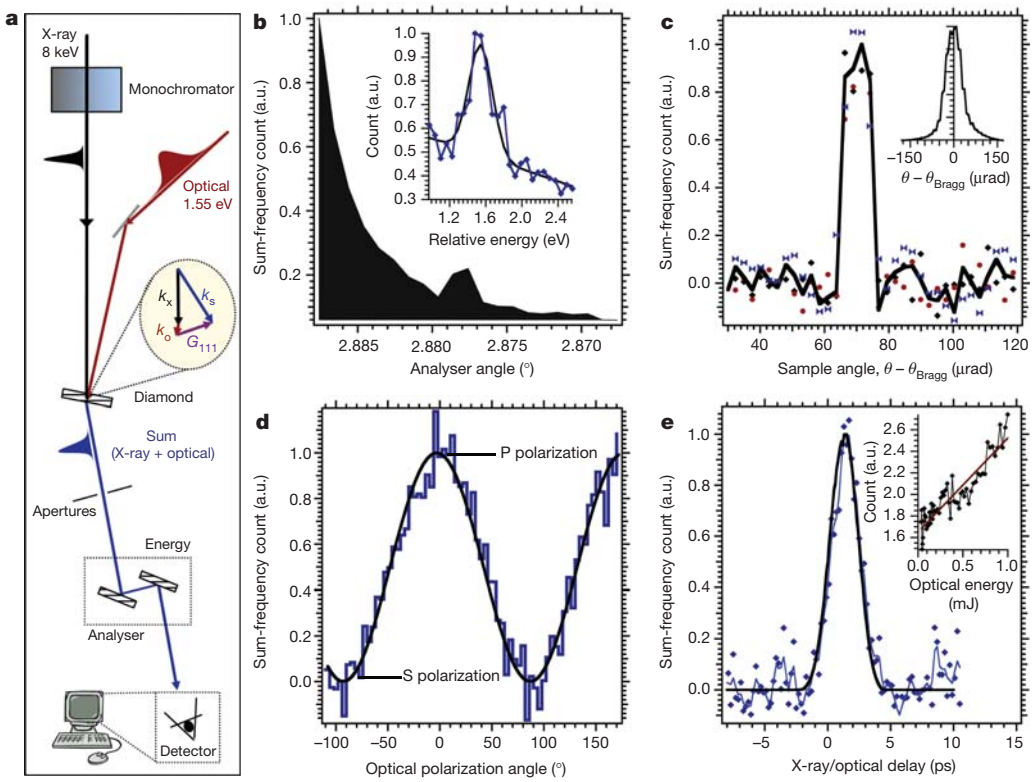


Figure 1 | X-ray/optical SFG experiment. **a**, Experimental layout for generation of X-ray (8-keV) and optical (1.55-eV) SFG in diamond. a.u., arbitrary units. **b**, SFG signal versus energy analyser angle. Inset, energy relative to 8 keV. The black line is a Gaussian fit (310 ± 35 meV, 17 ± 2-μrad full-width at half-maximum). The SFG energy exceeds the X-ray energy by one optical photon. **c**, SFG signal versus diamond-sample angle: average of three data scans (width, ~8 μrad; black line) and individual scans (markers). Energy and momentum conservation is satisfied at $\theta - \theta_{Bragg} \approx 70$ μrad. Inset, sample rocking curve for elastic diffraction (width, ~65 μrad). **d**, SFG signal (measured, blue; expected variation, black) versus angle of optical polarization vector (ε_o). The signal is maximum when ε_o is in the diffraction plane (0°) and is zero when ε_o is normal to the diffraction plane (±90°). **e**, SFG signal versus X-ray/optical time delay (raw data, blue markers; one-point-smoothed data, blue curve). Black cross-correlation curve (2.5-ps full-width at half-maximum): solution of wave equation for an 80-fs X-ray pulse and a 1.7-ps optical pulse. Inset, SFG signal versus optical intensity. The red line is a fit to a linear dependence on optical intensity.

efficiency is 3×10^{-7} and the estimated uncertainty is a multiplicative factor of two. The SFG signal strength is a bilinear function of the input beam intensities, varying as the product of the optical and input X-ray intensities. Because the efficiency is defined as the number of SFG photons per input X-ray photon, it is independent of input X-ray intensity but is linearly dependent on optical intensity.

Finally, an X-ray/optical cross-correlation is demonstrated. To within errors of ~50%, the SFG signal is observed to vary linearly with optical intensity (Fig. 1e, inset). The measured variation in the SFG signal with X-ray/optical relative time delay indicates a cross-correlation width of ~2.5 ps (Fig. 1e). By solving the wave equation, we compute the SFG intensity as a function of relative time delay to determine that the measured cross-correlation is reproduced by a simulation with an 80-fs X-ray pulse and a 1.7-ps optical pulse, the second of which is consistent with the experimental optical pulse duration (2 ± 0.5 ps). The simulated cross-correlation is also shown in Fig. 1e. The cross-correlation width is set by two factors: the duration of the (long) optical pulse and the fact that the X-ray and optical pulses move through the material at different speeds and in different directions.

Wave equation model

The generation of a sum-frequency electric field is described by the time- and space-dependent wave equation. The slowly-varying-envelope approximation leads to the following equation relating the sum-frequency field (E_s) and bilinear current (J_{NL}) envelopes (Supplementary Information, section 3):

$$\cos(\theta_s)\frac{dE_s}{dz} + \sin(\theta_s)\frac{dE_s}{dx} + \frac{1}{c}\frac{dE_s}{dt} + \beta E_s = -\frac{2\pi}{c}J_{NL} \qquad (1)$$

The z and x axes are respectively the inward-facing surface normal and an arbitrary line parallel to the sample surface, β is the X-ray absorption coefficient and θ_s is the angle between the z axis and the sum-frequency wavevector.

X-rays see diamond as a periodic collection of free electrons ($h\nu_x \gg h\nu_{binding}$, where $h\nu_x$ is the X-ray photon energy and $h\nu_{binding}$ is the electron binding energy; h is Planck's constant), and the bilinear current, the source of E_s, is written as a charge times a velocity[29]. We find that the dominant current is the 'Doppler' current (Supplementary Information, section 4), which represents inelastic scatter as X-rays (E_x) impart a first-order velocity to an optically induced charge density ($\delta\rho_o$):

$$J_{NL}(k_{sum} = Q + k_x, \omega_{sum} = \omega_o + \omega_x) = \frac{e}{im\omega_x}\delta\rho_o(Q,\omega_o)\varepsilon_s \bullet E_x(k_x,\omega_x) \qquad (2)$$

In equation (2), ε_s is the sum-frequency polarization vector, e and m are respectively the electron charge and mass, k_x and ω_x are respectively the X-ray wavevector and frequency, and $\delta\rho_o(Q,\omega_o)$ is the Fourier component of the optically induced charge oscillating with spatial wavevector $Q = k_o + G_{111}$ and optical frequency ω_o; it is induced by, and is therefore a function of, the external applied optical field, E_{light}. The input pulses E_x and E_{light} are modelled as Gaussians in space and time with linear chirp parameters determined by their angle–energy spread. With J_{NL} as given in equation (2) as the source term, we solve the equation obtained by taking a two-dimensional Fourier transform of equation (1).

The power distributions in angle–energy space for the calculated SFG and the driving bilinear current power distributions are shown in Fig. 2a–c for crystal lengths of 1, 10 and 500 μm (the last of which corresponds to our sample). The induced current, which is proportional to the space-time product of the input beams, has an angle–energy power distribution determined by convolution of the Fourier-space input-beam profiles. The angle–energy width of the bilinear current (~4 μrad, ~1 eV) is determined primarily by the angle–energy width of the input X-ray beam; its energy spread (1 eV) is large compared with that of the optical beam (<0.1 eV), and its transverse wavevector spread is approximately twice that of the optical beam. The angle–energy width of the weakly scattered SFG beam is a fraction of the bilinear current width. In direct analogy with conventional weak X-ray scattering, this fraction is determined by the effective crystal length and is less than one for frequency conversion over a long enough crystal. Figure 2a shows that for crystal lengths less than ~1 μm, the current SFG process accepts the full angle–energy width of the bilinear current. Figure 2b, c shows that the SFG angular width is not reduced for crystal lengths up to 500 μm but that the energy width is reduced, to ~140 meV at 500 μm.

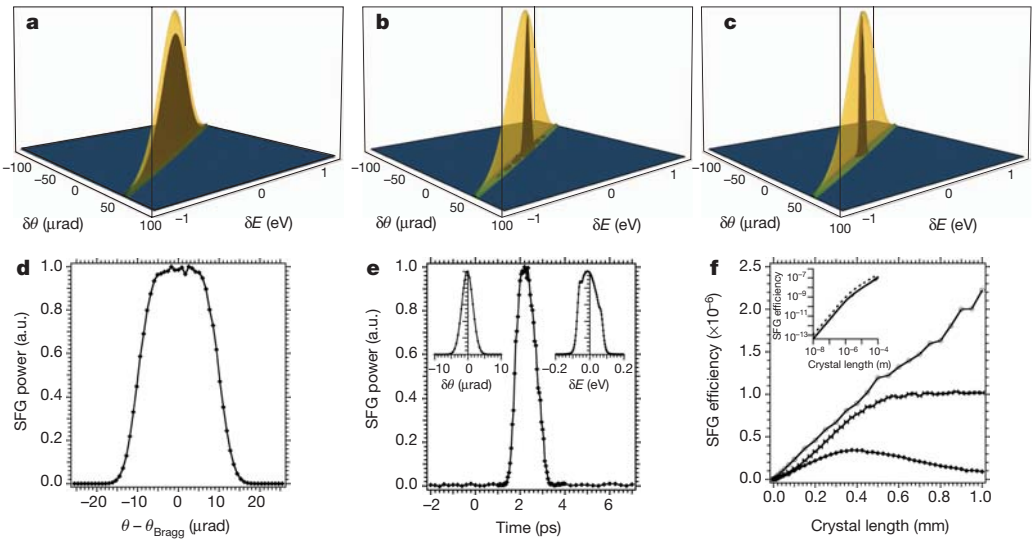


Figure 2 | Wave equation simulations. All quoted widths are full-widths at half-maximum. **a**–**c**, SFG power distribution (brown) in angle–energy space for sample thicknesses of 1 μm (**a**), 10 μm (**b**) and 500 μm (our sample; **c**). Deviations $\delta\theta$ and δE from the central SFG angle and energy are plotted. The respective widths are 3.7 μrad and 960 meV (**a**), 3.7 μrad and 210 meV (**b**), and 3.6 μrad and 140 meV (**c**). The driving bilinear current is also shown (3.7 μrad and 1 eV; gold). **d**, SFG power versus deviation from the optimum sample angle, for a 500-μm crystal. The rocking-curve width is 20 μrad. **e**, SFG pulse duration (~1 ps), angle spread (~4 μrad, left inset) and energy spread (~140 meV, right inset), for a 500-μm crystal. **f**, SFG power versus crystal length with X-ray absorption (diamonds), without absorption for a 1-ps optical pulse (bowties) and without absorption for a long (200-ps) optical pulse (circles). Inset, log-scale plot. The dashed line represents the data for the long optical pulse. The other two cases (solid lines) are indistinguishable.

The calculated SFG rocking-curve width, temporal width and efficiency variation with crystal length are shown in Figs 2d–f. The rocking-curve width (20 μrad; Fig. 2d) is consistent with the experimentally determined lower limit (~8 μrad) and, as with conventional weak X-ray diffraction, is determined by both the effective crystal length and the angle–energy spread of the input beams. The SFG temporal width (~1 ps; Fig. 2e) is larger than the product (~80 fs) of perfectly overlapped X-ray (~80 fs) and optical (~2 ps) pulses because these two pulses move through the crystal at different speeds and in different directions. With regard to dependence on crystal length, the SFG intensity decreases for crystals longer than ~400 μm owing to X-ray absorption (Fig. 2f). In the absence of absorption, the X-ray/optical velocity mismatch combined with our non-collinear geometry (different beam directions) leads to saturation of SFG efficiency as a function of crystal length (Fig. 2f) because the pulses eventually separate from one another. Without the velocity mismatch, the increase of efficiency with crystal length no longer saturates (Fig. 2f). The efficiency also improves with increased input-beam collimation, monochromaticity or both, to $\sim 10^{-3}$ for X-ray energy and angle widths of 0.1 meV and 1 μrad and optical energy and angle widths of 10 meV and 1 mrad, respectively. This assumes a 500-μm-long crystal and neglects loss due to absorption.

Equations (1) and (2) show that the fundamental unknown specifying E_s/E_x and, therefore, the SFG efficiency is the optically induced charge $\delta\rho_o(Q)$; it is determined by requiring that the wave equation model reproduce the measured SFG efficiency (3×10^{-7}), a requirement satisfied by $\delta\rho_o(Q) \approx 0.84 \times 10^{-4}$ electrons per cubic ångström ($e^- \text{Å}^{-3}$). Furthermore, we define a nonlinear susceptibility (χ_{NL}) such that J_{NL} is written as a product of E_x, E_{light} and χ_{NL} rather than in terms of $\delta\rho_o(Q)$. The measured efficiency corresponds to a nonlinear susceptibility of $\chi_{NL}(\omega_o + \omega_x, G_{111}) \approx 1.6 \times 10^{-14}$ e.s.u. Finally, the measured quantity $\delta\rho_o(Q)$ determines the microscopic field $\delta E_o(Q)$ associated with this charge because the two are related by Gauss's law[3–6]. Here $\delta E_o(Q)$ is the Qth Fourier component of the microscopic field arising in diamond as a result of the application of the external field E_{light} (refs 3–6). We find that $|\delta E_o(Q = k_o + G_{111})| \approx 0.47$ in units of the macroscopic internal optical field and ~0.18 in units of the unscreened external optical field. The uncertainty in $\delta\rho_o(Q)$, $\chi_{NL}(G_{111})$ and $\delta E_o(Q)$ is ~40%, determined by the estimated uncertainty of a factor of two in the measured efficiency. The connections between the efficiency, induced charge, nonlinear susceptibility and induced microscopic field are discussed in Supplementary Information, section 2.

Optically polarized valence charge

Associating optically polarized charge with the charge components of interest in a material is important for X-ray/optical SFG pump–probe applications. Here we consider atomic-scale optical polarization in diamond. Models that assign all optical polarizability to covalent bond charge, here termed bond charge models, have had success describing the macroscopic optical response of covalent semiconductors[30,31]. To test whether this idea can be extended to describe atomic-scale polarizability, we compare measured charge densities with those calculated from first principles. We find that the Fourier-space induced charge computed using density functional theory[32,33] ($\delta\rho_o(G_{111}) = 1.07 \times 10^{-4}\, e^- \text{Å}^{-3}$; Methods; see also http://www.quantum-espresso.org) is, to within the ~40% experimental error, consistent with the measured charge density ($\sim 0.84 \times 10^{-4}\, e^- \text{Å}^{-3}$).

A real-space picture of the induced charge is required to assess whether light primarily polarizes bond charge. Such a real-space reconstruction is beyond the scope of the experiments reported here. Instead, we compute the real-space induced charge distribution using density functional theory. The computed ground-state valence charge density (Fig. 3a, d) shows that the covalent bond is characterized by two charge accumulation peaks, one on each side of the bond midpoint, a feature confirmed by X-ray scattering data[28]. The corresponding induced charge calculation (Fig. 3b, e) shows that optical polarization is predominantly expressed at the charge accumulation peaks of the covalent bond; this can be seen by comparing the positions of peak bond charge density and peak optically induced charge density (Fig. 3c, f). We find that the induced charge varies linearly with optical field strength and, for the optical field strength of the current experiment, corresponds to a 0.2% modification to the charge density at the charge accumulation peak. The density functional theory calculations therefore support a view that light predominantly perturbs chemical bonds in the lattice. We note that EUV radiation (90 ± 30 eV) predominantly polarizes charge at the bond midpoint rather than at the charge accumulation peaks[23].

Finally, we have constructed a simple semi-empirical bond charge model in the spirit of ref. 1 (Supplementary Information, section 1) and find that to within our estimated measurement uncertainty, a multiplicative factor of ~1.4, the bond charge model predicts a charge density ($\sim 1.3 \times 10^{-4}\, e^- \text{Å}^{-3}$) in reasonably good agreement with the measured charge density ($\sim 0.84 \times 10^{-4}\, e^- \text{Å}^{-3}$). The bond charge model uses measured data: the macroscopic optical response and the microscopic bond charge distribution. An assessment of this model is therefore primarily an assessment of the assumption that optical polarizability resides in the bond charge. Although confirmation in the form of an experimental real-space reconstruction of the induced charge is required, reasonable agreement between measured densities and those predicted by the bond charge model lends further support to the identification of bond charge as the optically polarizable charge.

Probing dynamics with X-ray wave mixing

Ultrafast X-ray diffraction (UXD) is a well-established technique that has made important contributions to advancing our understanding of molecular and material dynamics[34–38]. The method uses time-delayed X-ray pulses to follow ionic cores after sudden excitation, often by an optical pulse. Ionic cores are followed because compact charge about the nucleus, or 'core' charge, typically dominates the electron density in a material. Importantly, information on the chemically relevant valence charge is typically not obtained in experiments using ultrafast X-ray diffraction. X-ray wave mixing provides a way to probe the valence charge density in a material. It can be used to probe both the initial perturbation of a light-driven process as well as the subsequent valence charge dynamics. We first consider the initial perturbation.

The perturbation initiating a light-driven process is often poorly understood and, so far, has not been directly measurable. An X-ray pulse can accompany a perturbing optical pulse to probe the initial valence charge distortion directly by means of X-ray/optical SFG. A real-space reconstruction of the induced charge can, in principle, be obtained by measuring multiple Fourier components and using crystallographic phasing methods[39,40]. As with conventional crystallography, the achievable spatial resolution is, through Bragg's law, set by the scattering angle[40]. A less challenging method is to measure a small set of SFG Fourier amplitudes to distinguish between competing model descriptions of the initial perturbation. For instance, magneto-optical Kerr measurements (S. Roy *et al.*, manuscript in preparation) suggest that optical radiation perturbs the helical spin structure in antiferromagnetic dysprosium thin films. However, the underlying mechanism is unclear: light might directly couple to localized spin-aligned *f* electrons or the direct coupling may instead be to delocalized metallic electrons. By measuring a small number of X-ray/optical SFG Fourier amplitudes, it should be possible to determine the degree to which the initial optical excitation is localized and therefore to distinguish between these possibilities.

Beyond probing initial optical perturbations, the evolution of valence charge in a material can be probed with a time-delayed pulse pair: an X-ray/optical pair to probe optically responsive charge or an X-ray/EUV pair to probe a significantly larger fraction of the valence charge[8,20,23]. Unfavourable frequency scaling of the electronic polarizability suggests that the efficiency of X-ray/EUV SFG is significantly

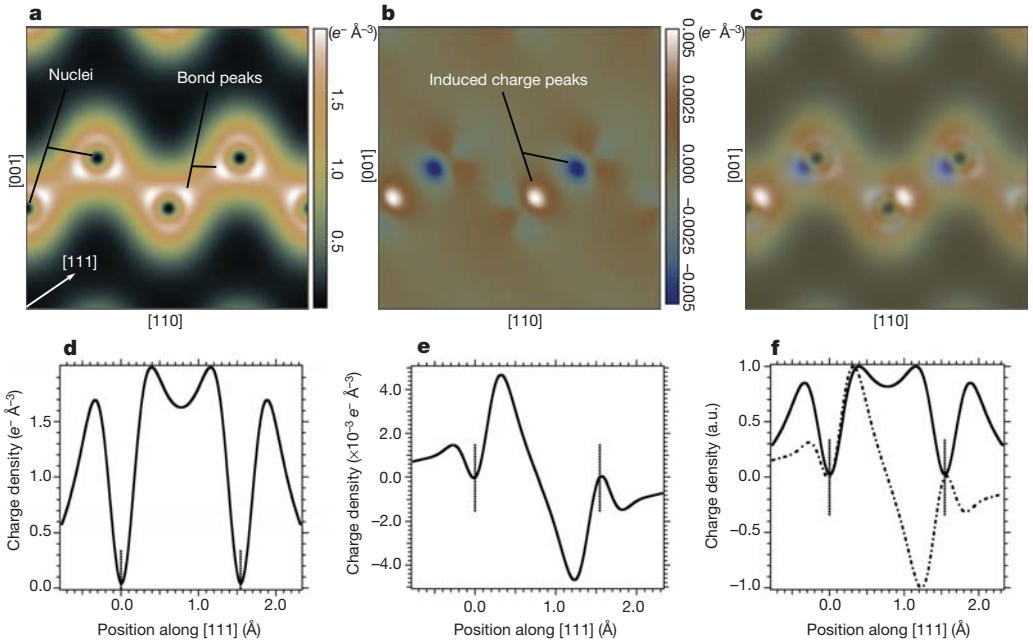


Figure 3 | Density functional theory calculations of real-space valence charge density in diamond. **a**, Ground-state charge density (colour scale) showing the covalent bond charge accumulation peaks. **b**, Induced charge density produced by a constant field (2.7×10^8 V m^{-1}) in the [111] direction. **c**, Induced charge image (**b**) overlaid on the unperturbed charge image (**a**), demonstrating that optical polarization is primarily found at the charge accumulation peaks. **d**–**f**, Charge density along the [111] direction and through the bond peaks for **a**–**c**, respectively. The bond peaks (0.39 and 1.16 Å; **d**) and induced charge peaks (0.32 and 1.23 Å; **e**) are evident; **f** shows normalized densities for both the ground-state (solid) and the induced (dashed) charge distributions.

lower than that of X-ray/optical SFG. We calculate that the bilinear current (the square root of the efficiency) for X-ray (8,000 eV)/EUV (100 eV) SFG is ~5,000 times smaller than that for X-ray (8,000 eV)/optical (1.55 eV) SFG (Supplementary Information, section 4). Although X-ray/EUV SFG could in principle be used to construct time-dependent valence charge density maps, the expected low efficiency may make this an application to which future light sources are better suited. For experiments in which the significantly lower signal rate is not acceptable, X-ray/optical SFG can be used to probe a selected portion of the valence charge.

A material's optical polarizability is not uniformly distributed across the valence charge and, rather than probing the full valence charge distribution, X-ray/optical SFG selectively probes the optically polarizable component. To the extent that the optically polarizable charge can, through measurement and theory, be correctly associated with charge components of interest in a material, it should be possible to obtain novel dynamical information. For instance, assuming, as suggested above, that the optically polarizable charge in diamond is bond charge, chemical bond dynamics can be studied by measuring the time evolution of select X-ray/optical SFG Bragg peaks. Ultrafast X-ray diffraction experiments using the temporal behaviour of conventional Bragg peaks have offered insight into, among other things, charge ordering dynamics in complex materials[35], coherent atomic motion associated with both phonon[36] and phonon–polariton[37] propagation, and bond softening in optically excited materials[38]. Analogous experiments that probe valence charge should provide additional insight. For instance, the coupling between ionic cores and chemical bonds could be studied: induced coherent ionic-core oscillations will induce chemical bond oscillations that may differ from the core oscillations in any combination of phase, amplitude and frequency. Similarly, bond delocalization or breaking would have a clear signature in the evolution of selected X-ray/optical SFG Bragg peaks.

Finally, to comment on the viability of X-ray/optical SFG at existing synchrotrons, we estimate that if the same experimental apparatus (including a 1-kHz optical laser) were used in conjunction with a third-generation synchrotron such as the Advanced Photon Source, it would take ~4,000 times longer to obtain data with the same statistics (Supplementary Information, section 6). By using high-repetition-rate (1-MHz) optical lasers, it may be possible to approach the count rate of the current experiment at a third-generation synchrotron (Supplementary Information, section 6).

Conclusion

We have demonstrated X-ray/optical SFG, a new tool for probing light–matter interactions on the atomic scale. To within experimental errors of ~40%, the measured conversion efficiency (3×10^{-7}) is consistent with first-principles calculations of atomic-scale optical polarization in diamond. This measurement, in combination with *ab initio* simulations, supports the extension to atomic length scales of the long-held supposition that light selectively polarizes chemical bonds in covalent semiconductors. This conclusion is tentative and must be confirmed by an experimental real-space reconstruction of the optically polarized charge.

More generally, X-ray wave mixing creates new opportunities for probing matter. X-ray/optical SFG provides access to two closely related and previously inaccessible quantities: the induced charges and the associated microscopic fields that arise when light illuminates a material. The ability to measure optically induced microscopic fields provides novel data for benchmarking condensed-matter theories. Similarly, X-ray/optical SFG and the closely related X-ray/EUV SFG process create new opportunities to study dynamics by directly probing valence charge on atomic time and length scales.

METHODS SUMMARY

Experiment. Experiments were performed at the Linac Coherent Light Source[26] (LCLS). The 8-keV LCLS beam (~1.7 μrad, ~20-eV full-width at half-maximum, ~80 fs) passed through a Si(111) monochromator and was directed to a (100)-surface-cut diamond sample oriented for Laue geometry diffraction from the (111) planes. X-ray/optical SFG was observed using 800-nm optical pulses ($\sim 1.5 \times 10^{10}$ W cm^{-2}) stretched to a duration (~2 ps) long compared with the X-ray/optical timing jitter (~200 fs). Switching from elastic diffraction to SFG was accomplished by rotating the diamond sample by ~70 μrad, moving the

analyser crystal to the expected SFG beam position and then rotating the analyser to increase the central pass energy by 1.55 eV ($h\nu_{optical}$, where $\nu_{optical}$ is the optical frequency). Scattered X-rays were wavelength-resolved with a Si(220) channel-cut crystal.

Variability in wave equation predictions. Input beams are modelled as linearly chirped Gaussians with spatial and temporal chirp parameters adjusted to match angle and energy widths: $\sim 4.1 \pm 1$ mrad and 68 ± 10 meV for the optical beam and $\sim 1.7 \pm 1$ μrad and 1 ± 0.1 eV for the X-ray beam. The SFG efficiency varies by less than 2% for the stated variations in all parameters except for X-ray energy. We estimate an uncertainty of 10% in the X-ray bandwidth (1 ± 0.1 eV) and determine that this corresponds to an uncertainty of 10% in the predicted SFG efficiency, which is small compared with the uncertainty (of a factor of two) in the measured SFG efficiency. We considered two X-ray pulse models: an ~80-fs (actual duration) pulse with a 1-eV bandwidth accounted for by a linear frequency chirp, and a transform-limited (~2-fs) pulse with a 1-eV bandwidth. We find that the SFG efficiency and pulse duration are insensitive to these choices, varying by less than 2%.

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

Received 14 February; accepted 18 June 2012.

1. Freund. I & Levine, B. F. Optically modulated X-ray diffraction. *Phys. Rev. Lett.* **25,** 1241–1245 (1970).
2. Eisenberger, P. M. & McCall, S. L. Mixing of X-ray and optical photons. *Phys. Rev. A* **3,** 1145–1151 (1971).
3. Jackson, J. D. *Classical Electrodynamics* (Wiley, 1975).
4. Arya, K. & Jha, S. S. Microscopic optical fields and mixing coefficients of x-ray and optical frequencies in solids. *Pramana* **2,** 116–125 (1974).
5. Adler, S. L. Quantum theory of the dielectric constant in real solids. *Phys. Rev.* **126,** 413–420 (1962).
6. Pine, A. S. Self-consistent-field theory of linear and nonlinear crystalline dielectrics including local-field effects. *Phys. Rev.* **139,** A901–A911 (1965).
7. Woo, J. W. F. & Jha, S. S. Inelastic scattering of x rays from optically induced charge-density oscillations. *Phys. Rev. B* **6,** 4081–4082 (1972).
8. Freund, I. Nonlinear X-ray diffraction: determination of valence electron charge distributions. *Chem. Phys. Lett.* **12,** 583–588 (1972).
9. Van Vechten, J. A., &. Martin, R. M. Calculation of local effective fields: optical spectrum of diamond. *Phys. Rev. Lett.* **28,** 446–449 (1972).
10. Freund, I&. Levine, B. F. Surface effects in the nonlinear interaction of X-ray and optical fields. *Phys. Rev. B* **6,** 3059–3060 (1973).
11. Arya, K. & Jha, S. S. Microscopic optical fields in diamond and germanium: molecular-orbital approach. *Phys. Rev. B* **10,** 4485–4487 (1974).
12. Johnson, D. L. Local field effects and the dielectric response matrix of insulators: A model. *Phys. Rev. B* **9,** 4475–4484 (1974).
13. Schweigert, I. V. & Mukamel, S. Coherent ultrafast core-hole correlation spectroscopy: X-ray analogues of multidimensional NMR. *Phys. Rev. Lett.* **99,** 163001 (2007).
14. Tanaka, S. & Mukamel, S. Probing exciton dynamics using Raman resonances in femtosecond X-ray four-wave mixing. *Phys. Rev. A* **67,** 033818 (2003).
15. Nazarkin, A., Podorov, S., Uschmann, I., Forster, E. & Sauerbrey, R. Nonlinear optics in the angstrom regime: hard-X-ray frequency doubling in perfect crystals. *Phys. Rev. A* **67,** 041804(R) (2003).
16. Tanaka, S. & Mukamel, S. Coherent X-ray Raman spectroscopy: a nonlinear local probe for electronic excitations. *Phys. Rev. Lett.* **89,** 043001 (2002).
17. Hudis, E., Shkolnikov, P. L. & Kaplan, A. E. X-ray stimulated Raman scattering in Li and He. *Appl. Phys. Lett.* **64,** 818–820 (1994).
18. Freund, I. & Levine, B. F. Parametric conversion of X rays. *Phys. Rev. Lett.* **23,** 854–857 (1969).
19. Eisenberger, P. & McCall, S. L. X-ray parametric conversion. *Phys. Rev. Lett.* **26,** 684–688 (1971).
20. Danino, H. & Freund, I. Parametric down conversion of X rays into the extreme ultraviolet. *Phys. Rev. Lett.* **46,** 1127–1130 (1981).
21. Yoda, Y., Suzuki, T., Zhang, X. W., Hirano, K. & Kikuta, S. X-ray parametric scattering by a diamond crystal. *J. Synchrotron Radiat.* **5,** 980–982 (1998).
22. Adams, B. *et al.* Parametric down conversion of X-ray photons. *J. Synchrotron Radiat.* **7,** 81–88 (2000).
23. Tamasaku, K. Sawada, K. Nishibori, E. & Ishikawa, T. Visualizing the local optical response to extreme-ultraviolet radiation with a resolution of λ/380. *Nature Phys.* **7,** 705–708 (2011).
24. Platzman, P. M. & Isaacs, E. D. Resonant inelastic X-ray scattering. *Phys. Rev. B* **57,** 11107–11114 (1998).
25. Kotani, A. & Shin, S. Resonant inelastic X-ray scattering spectra for electrons in solids. *Rev. Mod. Phys.* **73,** 203–246 (2001).
26. Emma, P. *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photon.* **4,** 641–647 (2010).
27. Palik, E. D. (ed.) *Handbook of Optical Constants of Solids* 313–334 (Academic, 1998).
28. Spackman, M. A. The electron distribution in diamond: a comparison between experiment and theory. *Acta Crystallogr. A* **47,** 420–427 (1991).
29. Shen, Y. R. *The Principles of Nonlinear Optics* 5–40 (Wiley, 1984).
30. Phillips, J. C. Covalent bond in crystals. I. Elements of a structural theory. *Phys. Rev.* **166,** 832–838 (1968).
31. Levine, B. F. Electrodynamic bond-charge calculation of nonlinear optical susceptibilities. *Phys. Rev. Lett.* **22,** 787–790 (1969).
32. Giannozzi, P. *et al.* Quantum Espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21,** 395502 (2009).
33. Souza, I. Iniguez, J. & Vanderbilt, D. First principles approach to insulators in finite electric fields. *Phys. Rev. Lett.* **89,** 117602 (2002).
34. Ihee, H. *et al.* Ultrafast X-ray diffraction of transient molecular structures in solution. *Science* **309,** 1223–1227 (2005).
35. Möhr-Vorobeva, E. *et al.* Nonthermal melting of a charge density wave in $TiSe_2$. *Phys. Rev. Lett.* **107,** 036403 (2011).
36. Lindenberg, A. M. *et al.* Time-resolved X-ray diffraction from coherent phonons during a laser-induced phase transition. *Phys. Rev. Lett.* **84,** 111–114 (2000).
37. Cavalleri, A. *et al.* Tracking the motion of charges in a terahertz light field by femtosecond X-ray diffraction. *Nature* **442,** 664–666 (2006).
38. Fritz, D. M. *et al.* Ultrafast bond softening in bismuth: mapping a solid's interatomic potential with X-rays. *Science* **315,** 633–636 (2007).
39. Wu, J. S., Spence, J. C. H., O'Keeffe, M. & Groy, T. L. Application of a modified Oszlanyi and Suto ab initio charge-flipping algorithm to experimental data. *Acta Crystallogr. A* **60,** 326–330 (2004).
40. Coppens, P. *X-Ray Charge Densities and Chemical Bonding* (Oxford Univ. Press, 1997).

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

Acknowledgements We thank J. D. Jackson, D. Vanderbilt, and J. C. H. Spence for commenting on various aspects of this work. T.E.G. was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. The work of S.E.H. and S.S. was supported by the US Air Force Office of Scientific Research and the US Army Research Office. D.A.R. and S.G. were supported as part of the AMOS programme within the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, US Department of Energy. Portions of this research were carried out at the LCLS at SLAC National Accelerator Laboratory. LCLS is an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. Preliminary experiments were performed at the Advanced Light Source at Lawrence Berkeley National Laboratory.

Author Contributions T.E.G. and J.B.H. had the idea for the x-ray/optical SFG project. D.M.F., T.E.G., J.B.H., M.C., T.K.A., J.M.F. and D.A.R. contributed to the experiment design. D.M.F., M.C., H.L., D.Z. and R.N.C. were responsible for the X-ray pump–probe instrument and the optical laser. D.M.F., M.C., T.E.G., J.B.H., H.L., D.Z., T.K.A., J.M.F., Y.F., R.N.C., D.A.R., S.S., M.F., S.G. and J.C. collected data. S.E.H., S.S., T.E.G. and T.K.A. contributed to the wave equation calculations. Data analysis was done by T.E.G., H.L. and J.M.F. Construction of the bond charge model was done by T.E.G. Density functional theory calculations were done by S.C. T.E.G. wrote the manuscript. All authors contributed to the work presented here and to the final paper.

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

METHODS

Experiment. Experiments were performed at the LCLS[26], an X-ray free-electron laser that provided an 8-keV, 60-Hz beam with angle and energy spreads of respectively ~1.7 μrad and ~20 eV (full-widths at half-maximum). The ~80-fs pulse (full-width at half-maximum) contained $\sim 5 \times 10^{10}$ photons. Complications from spectral jitter were mitigated by converting spectral fluctuations to intensity fluctuations: X-rays were passed through a two-bounce Si(111) monochromator (~1-eV bandpass) and a non-destructive Compton scattering diagnostic recorded the shot-by-shot beam intensity to normalize intensity fluctuations. Monochromatic X-rays (~1.7 μrad, 1 eV) were directed to a (100)-surface-cut diamond sample oriented for 8-keV, (111) Laue geometry diffraction. A Si(220) channel-cut crystal was used to energy-resolve detected X-rays. X-ray/optical SFG was observed using 800-nm optical pulses (1.5×10^{10} W cm^{-2}) stretched to a duration (~2 ps) long compared with the X-ray/optical timing jitter (~200 fs). X-ray/optical overlap was achieved by maximizing the reduction in X-ray diffraction from an optically perturbed Bi sample.

Switching from elastic diffraction to SFG was accomplished by first rotating the diamond sample by $\delta\phi_{111} \approx 70$ μrad (less grazing). A two-theta arm then rotated the analyser to the expected SFG beam position, ~370 μrad relative to the elastically scattered beam position before the diamond rotation. Finally, the central pass energy of the analyser was increased (such that the optical photon energy was $h\nu_{\mathrm{optical}} = 1.55$ eV) by rotating the crystal itself ($\delta\phi_{220} \approx 85$ μrad). The sample and analyser detunings, each ~10 rocking-curve half-widths, reduced the intensity of elastically scattered (background) light reaching the detector by seven to eight orders of magnitude relative to optimized elastic scattering conditions ($\delta\phi_{111} = \delta\phi_{220} = 0$). We did not make use of the ~500-μrad angular separation between the elastic and SFG beams.

The X-ray flux reaching our detector was $\sim 3 \times 10^7$ photons per pulse under optimized elastic diffraction conditions, which is less than the initial $\sim 5 \times 10^{10}$ photons per pulse due to loss associated with both absorption and five crystal reflections (two Si(111), the diamond sample and two Si(220)). The SFG flux reaching our detector was ~65 photons per pulse.

Variability in wave equation predictions. Input beams were modelled as linearly chirped Gaussians with spatial and temporal chirp parameters adjusted to match the corresponding angle and energy widths: The optical beam had an energy spread of 68 meV and an in-vacuum angle spread of ~4.1 mrad. A linear frequency chirp stretched the optical pulse to a duration (~2 ps) long compared with the X-ray/optical timing jitter (~200 fs). The X-ray angle spread (~1.7 μrad) was set by the LCLS divergence whereas the energy spread (~1 eV) was set by the Si(111) monochromator. Two X-ray pulse models were considered: one with a pulse duration equal to the actual LCLS duration (~80 fs) and a 1-eV bandwidth accounted for by a linear frequency chirp, and one with a transform-limited pulse with a duration (~2 fs) determined by the 1-eV bandwidth. We find that the SFG efficiency and pulse duration are insensitive to these choices; each varies by less than 2%. Insensitivity of efficiency to the specific duration of the (short) X-ray pulse might be expected because for either pulse duration the X-ray pulse is short compared with the optical pulse and therefore sees the same optical field strength. Similarly, pulse duration insensitivity seems reasonable because the X-ray and optical pulses move through the crystal at different speeds; this causes the relative X-ray/optical time delay to vary linearly across the crystal. The magnitude of this 'time shift' (>1 ps) is large compared with either X-ray pulse duration considered (80 or 2 fs). The actual LCLS pulse is probably a random distribution of transform-limited spikes within a ~80-fs temporal envelope.

Finally, the predicted efficiency depends on the degree to which the input beams are collimated and/or monochromatic; here we consider how much the predicted efficiency varies owing to uncertainty in the angle–energy spread of the input beams. We estimate the uncertainties in the X-ray energy, laser wavelength, X-ray angle and laser angle 'widths' to be ±0.1 eV, ±10 meV, ±1 μrad and ±1 mrad, respectively. The variation in the predicted SFG efficiency is less than 2% for the stated variations in all parameters except for X-ray energy. The variation in efficiency with X-ray energy spread is approximately linear. The bandwidth of the X-ray beam should be known to good precision because it is set by the Si(111) monochromator acceptance. We estimate an uncertainty of ~10% in the X-ray bandwidth (1 ± 0.1 eV), which corresponds to a variation of ~10% in the predicted efficiency. This uncertainty is small compared with the uncertainty in the measured SFG efficiency, which is a factor of two as determined by variability in the measured X-ray/SFG pulse energy.

ARTICLE

doi:10.1038/nature11360

A Cenozoic record of the equatorial Pacific carbonate compensation depth

A list of authors and their affiliations appears at the end of the paper.

Atmospheric carbon dioxide concentrations and climate are regulated on geological timescales by the balance between carbon input from volcanic and metamorphic outgassing and its removal by weathering feedbacks; these feedbacks involve the erosion of silicate rocks and organic-carbon-bearing rocks. The integrated effect of these processes is reflected in the calcium carbonate compensation depth, which is the oceanic depth at which calcium carbonate is dissolved. Here we present a carbonate accumulation record that covers the past 53 million years from a depth transect in the equatorial Pacific Ocean. The carbonate compensation depth tracks long-term ocean cooling, deepening from 3.0–3.5 kilometres during the early Cenozoic (approximately 55 million years ago) to 4.6 kilometres at present, consistent with an overall Cenozoic increase in weathering. We find large superimposed fluctuations in carbonate compensation depth during the middle and late Eocene. Using Earth system models, we identify changes in weathering and the mode of organic-carbon delivery as two key processes to explain these large-scale Eocene fluctuations of the carbonate compensation depth.

The Pacific, as the world's largest ocean, is intricately involved in the prominent changes in the global carbon cycle and climate system that took place during the Cenozoic[1]. The equatorial Pacific makes a disproportionally large global contribution to biogenic sediment burial in the pelagic realm due to equatorial upwelling, and thus plays an important part in climate regulation[1]. Expeditions 320 and 321 of the Integrated Ocean Drilling Program (IODP), the 'Pacific Equatorial Age Transect' (PEAT), exploited the northward Pacific plate trajectory during the Cenozoic to recover a continuous sediment sequence from the equatorial Pacific. Cores from eight sites were obtained, from the sea floor to basaltic basement (that is, aged between 53 and 18 million years (Myr) before present), near the past position of the Equator at successive crustal ages on the Pacific plate[2]. Together with previous Deep Sea Drilling Project and Ocean Drilling Program drill cores, these sediments allow reconstruction of changes in the state, nature and variability of the global carbon cycle and climate system in unprecedented detail, from directly after the period of maximum Cenozoic warmth, through the onset of major glaciations, to the present.

Atmospheric CO_2 concentrations and climate are regulated on geological timescales by volcanic and metamorphic outgassing, weathering feedbacks involving the weathering of silicate and organic carbon (C_{org})-bearing rocks, and carbonate and C_{org} burial[3,4]. The integrated effect of these processes dictates the carbonate saturation state of the oceans that is reflected in the carbonate compensation depth (CCD), which has been highly influential in understanding past changes in the marine carbon cycle[5,6]. The CCD is a sediment property and occurs in the oceans where the downward flux of carbonate rain delivered from calcifying organisms is balanced by dissolution (Fig. 1), so that little or no carbonate is preserved in sediments below this depth. The CCD can be reconstructed using sediment cores[5–8].

Here we determine the evolution of the CCD by establishing the variation of net carbonate accumulation rates with respect to palaeodepth at multiple sites, using a common chronology and stratigraphic correlation. We then use carbon cycle models to explore viable mechanisms that may be responsible for the observed CCD variations; we highlight the importance of the supply of weathering-derived solutes to the deep ocean, and changes in the partitioning—into labile and refractory components—of organic matter delivered to deep-sea sediments.

A Cenozoic CCD record

Our new reconstruction of how the CCD in the equatorial Pacific region evolved through the Cenozoic is shown in Fig. 2a, and is compared with reconstructions of foraminiferal benthic $\delta^{18}O$ and $\delta^{13}C$ (Fig. 2b and c) and atmospheric CO_2 reconstructions (Fig. 2d) (Supplementary Fig. 1 shows this as a plot of mass accumulation rate against age and palaeo-latitude, Supplementary Fig. 2 shows an enlarged version for the Eocene). The evolving palaeo-position of

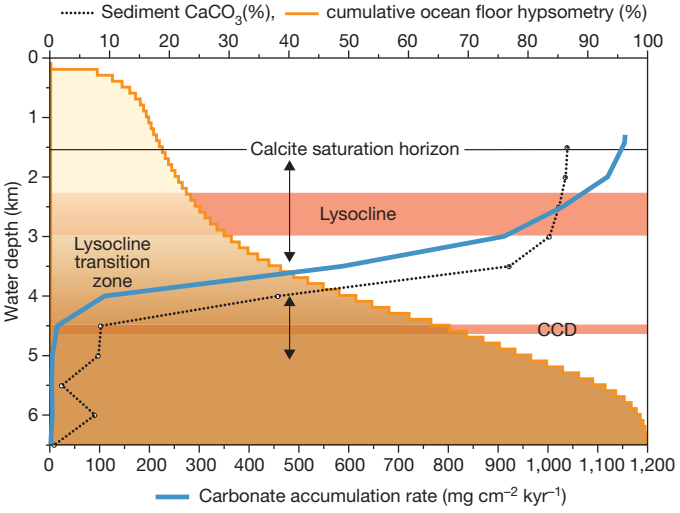


Figure 1 | Illustration of the position of the CCD and lysocline, and their relationship to ocean bathymetry, carbonate accumulation rate and $CaCO_3$ content. This figure shows the relationship between CCD, sediment $CaCO_3$ content (dotted black line), carbonate accumulation rate (blue line) and lysocline, in comparison with cumulative ocean floor hypsometry (orange line). The CCD, a sediment property, is defined as where carbonate rain is balanced by carbonate dissolution. Previously, it has been operationally defined to coincide with a fixed content of $CaCO_3$ (for example, 10%) in sediments[5], or where the carbonate accumulation rate interpolates to zero[6] (this second definition is advantageous as it is independent of non-carbonate supply or dilution effects). The lysocline is the horizon where dissolution becomes first noticeable (a sediment property), and is typically below the calcite saturation horizon.

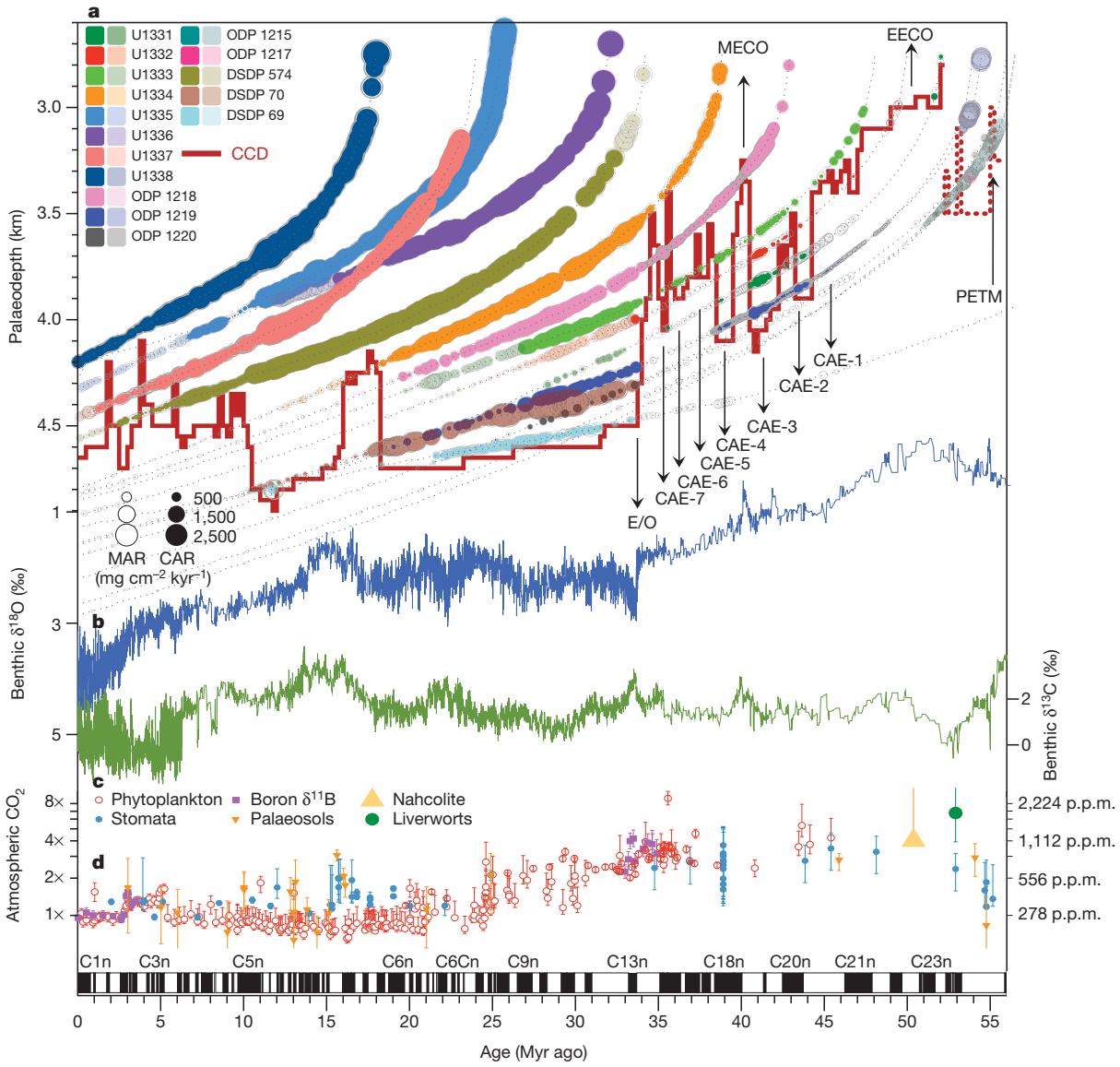


Figure 2 | CCD and carbonate accumulation rate reconstruction compared with published benthic foraminiferal $\delta^{18}O$ and $\delta^{13}C$ values and atmospheric CO_2. **a**, Equatorial Pacific accumulation rate history as a function of geological age at the backtracked and unloaded palaeo-water depth, and using a palaeomagnetic polarity age scheme[2] plotted as black and white bars below the figure. Circle area is scaled by accumulation rate: carbonate accumulation rate (CAR; filled circles), total mass accumulation rate (MAR; open circles). Data are plotted with a lighter colour outside a ±3.5° band around the palaeo-equator. The position of the equatorial Pacific CCD is indicated by a solid red line (dashed red line marks reconstruction from off-equatorial sites). See text for abbreviations. **b**, **c**, Benthic oxygen (**b**; blue curve, left-hand vertical axis) and carbon (**c**; green curve, right-hand vertical axis) isotope values from a global compilation[9], reported relative to the VPDB (Vienna Pee Dee Belemnite) standard. **d**, Atmospheric CO_2 compilation and error bars from refs 16, 40; left-hand vertical axis, log CO_2 scale relative to pre-industrial CO_2 (1× = 278 p.p.m.v.); right-hand vertical axis, log CO_2 scale in absolute values. Error bars are as in ref. 16; for example, for boron $\delta^{11}B$ error bars reflect long-term analytical reproducibility or internal precision, whichever is larger (at 95% confidence). PETM, Palaeocene-Eocene Thermal Maximum.

the CCD is determined as where the carbonate accumulation rates interpolated across the palaeo-depth transect reach zero as a function of depth. Overall, there is an intriguing correspondence of CCD deepening and deep ocean cooling[9], with a deepening from 3–3.5 km during the early Cenozoic to 4.6 km at present, consistent with findings of an overall Cenozoic increase in silicate weathering[10–13]. Superimposed on this overall deepening are repeated large CCD fluctuations during the middle and late Eocene.

We find that during the earliest Eocene (~56–53 Myr ago), the equatorial CCD generally occurred at a palaeo-depth of around 3.3–3.6 km, with superimposed 'hyperthermal'-related CCD shoaling events[9,14]. Between ~52 and 47 Myr ago, the CCD reached depths as shallow as 3 km, coincident with the Early Eocene Climatic Optimum (EECO)[15], when atmospheric CO_2 concentrations, including their uncertainty bounds, reached an estimated ~1,100–3,000 parts per million by volume (p.p.m.v.)[16], and the lowest benthic oxygen isotope values (indicating peak deep-ocean temperature) throughout the Cenozoic were attained[15].

From ~46 to 34 Myr ago, our record reveals a fluctuating and highly variable CCD[8], resolving 5–7 CCD deepenings and carbonate accumulation events (CAEs) with durations of several hundred thousand years to 1 Myr, interrupted by rapid CCD shoalings with an amplitude of ~0.5–1.0 km (CAE-1 to CAE-7; Fig. 2). The largest-magnitude fluctuation of the CCD during the middle and late Eocene coincided with carbonate accumulation event 3 (CAE-3)[8], followed by a near 1-km shoaling that is coeval with the Middle Eocene Climatic Optimum (MECO)[17,18]. A sustained large deepening (>1 km) then terminated a late Eocene interval of variability and

coincided with the Eocene/Oligocene (E/O) transition[19]. This terminal early Oligocene CCD deepening was contemporaneous with ice sheet growth, sea-level fall and a shift in carbonate deposition from shallow to deep waters[19,20]. The deeper (~4.6 km) and apparently more stable CCD of the Oligocene was interrupted again in the Miocene at ~18.5 Myr ago by a ~600-m shoaling, lasting for around 2.5 Myr, which was previously described as the "carbonate famine"[6]. The CCD deepened again to around 4.7 km at ~16 Myr ago as a consequence of increased carbonate productivity[6]. For this deepening event at ~16 Myr ago, the analysis of the PEAT data alone is complicated by the latitudinal movement of some sites (U1332–U1334) outside the Pacific equatorial zone (Supplementary Fig. 1). However, the interpretation is supported by previous work[6] and corroborated by large increases in carbonate mass accumulation rates at shallower depths of Sites U1335, U1337 and U1338 (Fig. 2). An interval with the deepest equatorial Pacific Cenozoic CCD of around 4.8 km was terminated by shoaling at ~10.5 Myr ago in a rapid "carbonate crash" event[21]. For the remainder of the Neogene, the equatorial Pacific CCD resided around 4.5 km depth with superimposed fluctuations in carbonate accumulation related to shorter-term climatic oscillations on Milankovitch timescales (20–400 kyr), as well as to the periodic deposition of diatom mats[2].

Eocene CCD fluctuations

Two features of the equatorial Pacific CCD behaviour during the Cenozoic stand out and demand further investigation. First, the Eocene CCD resided at an average depth of around 3.5 km, or ~1 km shallower than during post-Eocene time, with a geologically rapid and permanent deepening at the E/O transition. The E/O CCD deepening has previously been ascribed to ice expansion on Antarctica and sea-level fall, driving a shelf-to-basin shift in carbonate partitioning[19,20]. In contrast, for Eocene CAEs we lack evidence[22] for large bi-polar fluctuations in ice volume, so that alternative mechanisms must be explored.

Second, the middle to late Eocene was characterized by five major CCD fluctuations that lasted between 250 kyr and 1 Myr and had amplitudes between 200 and 900 m. The ~900-m CCD shoaling at ~40.5 Myr ago, a globally recognized feature[17] associated with the transition from CAE-3 to MECO, approaches the amplitude of (but is opposite in sign to) the CCD shift during the E/O transition from greenhouse to icehouse. We currently lack a definite confirmation that other CAEs are represented in records outside the equatorial Pacific, primarily owing to the scarcity of continuous, well-dated pelagic sedimentary records at an array of palaeo-depth positions, but we note that data from ODP Site 929E in the equatorial Atlantic tentatively support our hypothesis of CAEs as global features (Supplementary Fig. 3). In the equatorial Pacific, CAEs coincided with increased biogenic silica accumulation rates and shifts between calcareous and siliceous microfossils[8,18]. CAEs also coincided with enhanced burial of C_{org} in an Eocene Pacific Ocean otherwise characterized by productivity similar to today but with much faster water-column recycling; these increases in C_{org} burial start from a much lower Eocene baseline of C_{org} preserved in sediments, averaging only one-tenth of the present day value[8,23–25].

Carbon system modelling

To quantitatively explore the potential of different processes to generate the CCD signal reconstructed for the middle to late Eocene, we made use of steady-state results from an Earth system model of intermediate complexity (GENIE)[26,27], and investigated non-steady-state behaviour of shorter-lived processes with the palaeo-configuration of the LOSCAR box model[28] (Supplementary Information). Overall, the results from both models significantly reduce the number of possible mechanisms that are consistent with the reconstructed CCD history, either by demonstrating that a number of possible processes are not able to sustain large CCD changes over long enough time periods (>250 kyr), or by inconsistency with other proxy observations.

We start by assuming that middle–late Eocene CCD fluctuations represent alternating steady states of marine carbon cycling and that all carbonate weathering and climate feedbacks had time to operate and equilibrate. The justification for this is that the typical silicate weathering compensation time is of the order of ~0.1 Myr (refs 4, 29), much shorter than the duration of the reconstructed CCD fluctuations (0.25–1 Myr; Fig. 2). We then test the sensitivity of a range of established hypotheses for changing the CCD.

Processes we have investigated that either do not appear to be consistent with reconstructed amplitudes and durations of CCD shifts or are inconsistent with other proxy evidence (Supplementary Information) include: (1) C_{org} surface export rain ratio changes, which we exclude because the resultant CCD variations are too small in comparison to what we observe; (2) a shift in carbonate deposition between the shelf and deep ocean, for which a repeated large-scale oscillation in ice mass would be required that has so far not been observed; (3) a shift of deep ocean ventilation between a dominant Southern Ocean and dominant North Pacific source, which would lead to opposite CCD behaviour in different ocean basins; and (4) changes in deep-sea temperatures and the Mg and Ca concentration of sea water[4,9,30–32], which both modify the stability of calcium carbonate. For these, the CCD can be affected only by relatively subtle changes in the offset between CCD and lysocline because of the need to ultimately re-balance sources and sinks. We note that we cannot completely rule out the potential for ventilation changes to be compatible with our CCD data, and this will need to be resolved by future drilling in the North Atlantic.

Next, we focus on two model scenarios that do have the potential to sustain CCD changes of the required duration and amplitude: (1) perturbations to continental weathering and variations in solute input to the deep ocean, driving synchronous changes in CCD and lysocline depth globally, and/or (2) changes in the partitioning of C_{org} flux between labile and refractory components, affecting both deep-sea carbonate dissolution and the thickness of the lysocline transition zone (and hence partially decoupling the CCD from the lysocline).

To explore the first mechanism, we computed the steady-state CCD position in the equatorial Pacific through a range of atmospheric CO_2 values relative to pre-industrial modern ($1 \times CO_2 = 278$ p.p.m.v.) and against a range of solute weathering fluxes of Ca and HCO_3^- to the deep ocean in GENIE. Because in GENIE the solute flux to the deep ocean is the total weathering flux minus what is deposited on the shelves, changes in solute flux implicitly model either a change of the total flux, or shelf–basin partitioning. Our results (Fig. 3a) indicate that, for a given value of atmospheric CO_2, changes in solute flux to the deep ocean are in principle able to achieve changes in the equatorial Pacific CCD of the amplitude suggested by observations (several hundred metres to >1 km). For the Cenozoic, this supports the initial correlation between increased silicate weathering rates and CCD deepening.

However, other carbon cycle impacts resulting from the assumed driver of changes in solute supply must also be considered. On <1-Myr timescales, tectonic uplift is too slow, and there is no convincing evidence for repeated large-scale sea level fluctuations during the middle-to-late Eocene that could alternately shift the locus of carbonate deposition between shelf and deep ocean. Therefore, increasing solute supply should be coupled to increased weathering, a warmer climate and higher CO_2, unless changes in orbital configuration significantly enhance or reduce monsoon circulation (at constant CO_2), which could affect weathering fluxes via effects on precipitation intensity and distribution (which is not tested here). Our modelling reveals that increasing atmospheric CO_2 with fixed weathering (that is, with no weathering–temperature feedback enabled), results in a shallower CCD. This is a consequence of non-linearities in the carbonate system and reflects a deepening of the lysocline at the expense of the CCD and contraction of the lysocline

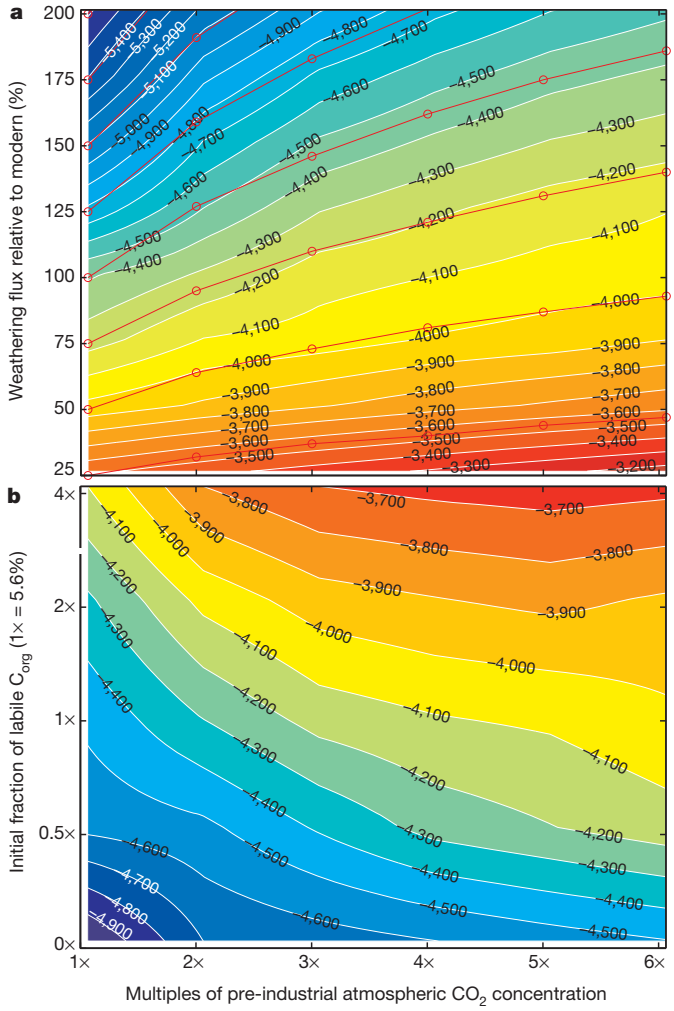


Figure 3 | CCD predicted by the GENIE steady state model. **a**, Eocene CCD plotted as contours (numbers show depth in m) as a function of atmospheric CO_2 concentrations (horizontal axis) and total net deep-sea weathering flux (compared to modern value of 10 Tmol yr^{-1}; vertical axis). All models were run without direct feedback on silicate weathering. The red lines indicate the increase in weathering flux when moving from 1× to 6× CO_2 silicate weathering feedback (using a partitioning of total initial weathering into silicate and carbonate weathering in a 1:1 ratio). **b**, As **a**, but plotting CCD as function of CO_2 (horizontal axis) and initial fraction of sediment labile C_{org} (vertical axis). This model ensemble was run with 100% net weathering compared to modern, and a surface $CaCO_3$:C_{org} rain ratio of 0.2.

transition zone. When this is combined with the response of increased weathering by activating the full silicate weathering feedback, we find an unexpected result: when progressively increasing the rate of prescribed CO_2 outgassing in a series of GENIE experiments, atmospheric CO_2 and weathering flux happen to co-vary in such a way that they result in a largely invariant CCD depth (Fig. 3a).

This result is subject to a number of model uncertainties and assumptions, including the degree of nonlinearity of weathering with climate, for which we have assumed a simple global-average response[29], and a 1:1 partitioning between carbonate and silicate weathering. The negative feedback in regulating atmospheric CO_2 on geological timescales due to weathering of continental materials had been included in the original BLAG[33] modelling study and is also used in GENIE and LOSCAR, but there are significant uncertainties attached to this parameterization. For example, the strength of this feedback has been shown to lead to significant variations in the carbon cycle response to weathering, but remains poorly constrained[29]. Although different weathering formulations were not tested with GENIE, we predict that the weak relationship between CCD and atmospheric CO_2 (Fig. 3a) probably indicates that additional silicate weathering changes in response to temperature will result in only small deviations from an invariant CCD after reaching steady state. Our results illustrate the non-intuitive nature of CCD behaviour and that higher marine carbonate deposition under a warmer, higher-CO_2 climate need not require a deeper observed CCD.

Independent observations of sediment composition and changes of dominant microfossil groups lead us to also assess whether changes in the behaviour of organic matter driving dissolution during the middle–late Eocene played a role. The pertinent observations are shifts between siliceous and calcareous microfossils and an increase of C_{org} burial flux during CAE-3, with simultaneous increased burial fluxes of calcareous and biosiliceous sediment[1,2,23–25]. For example, sediment smear slides from the equatorial Pacific reveal a major increase in diatom content from 0% to near 50% near the end of this event[24]. The relative changes in biogenic silica are much larger than those in calcareous plankton, and indicate major temporal reorganization of biotic composition rather than simple changes in productivity levels. In addition, the amount of C_{org} preserved in Eocene equatorial Pacific sediments is an order of magnitude smaller than today, despite a productivity that was not very different to the present[8,24], suggesting a more nutrient enriched deep ocean.

We investigate the viability of a 'sediment labile organic matter' hypothesis to help explain the CCD fluctuations. We repeat the GENIE net-weathering supply analysis but now change the partitioning between C_{org} that is redissolved in the upper water column and C_{org} that reaches the sea floor but is available to microbes to drive carbonate dissolution. This modifies the ocean's dissolved inorganic carbon as well as the net carbonate preservation flux[34]. This hypothesis builds on, but differs from, the classic glacial CO_2 rain ratio hypothesis[35], which postulated that changes in the $CaCO_3$ flux to ocean sediments, at a fixed particulate organic carbon (POC) supply, could be an effective way of changing atmospheric CO_2. However, the rebalancing of sedimentation versus weathering—carbonate compensation—while helping to drive a potential 60 p.p.m. fall in atmospheric CO_2 for a 40% decrease in $CaCO_3$:POC export ratio, also leads to a CCD that changed relatively little (Supplementary Fig. 5). This mechanism may also not be compatible with the consequence of any POC 'ballasting' by $CaCO_3$ (refs 36, 37).

Results for modelling the sediment labile organic matter hypothesis (Fig. 3b) indicate that a smaller initial fraction of labile organic matter results in a net increase in carbonate accumulation in regions of high productivity such as the eastern equatorial Pacific, Southern Ocean and equatorial Atlantic, owing to reduced water column dissolution. Carbonate compensation acts to mitigate the CCD changes but with a spatially heterogeneous pattern, leaving a deeper CCD in the eastern equatorial Pacific (Supplementary Fig. 9) and Atlantic, but resulting in smaller changes in the western Pacific. The modelled changes in carbonate accumulation in the eastern equatorial Pacific in this scenario indicate that CCD changes of ~300–600 m are possible for a 2–4-fold change in the initial fraction of sediment labile organic matter (Fig. 3b), roughly equivalent to the amplitude of all but the largest MECO-associated fluctuations in the record[17]. Although GENIE does not explicitly distinguish labile from refractory carbon delivered to the sediments, we are effectively assuming that surface biological productivity and the total flux of POC to deep-sea sediments need not change, but that it is either predominantly available for oxidation and driving carbonate dissolution (labile), or it is largely preserved and buried (refractory) and does not drive substantial additional carbonate dissolution.

CCD buffering

Overall, our modelling also reveals that the CCD is remarkably well buffered against short- and long-term perturbations of the global carbon cycle. Only a few of the mechanisms commonly envisaged

as important in controlling the CCD in practice have the capacity to change the amplitude of the steady-state CCD sufficiently and for long enough to be compatible with our reconstruction. In contrast, our labile organic matter hypothesis is consistent with shifts between siliceous and calcareous microfossil groups, and changes in C_{org} preservation and burial[24], and results in a large enough CCD shift to explain variability during the middle–late Eocene, perhaps in concert with simultaneous weathering flux changes. We note that the initial depth of the CCD may amplify these changes as its response to forcing is a function of the total amount of calcite available to be dissolved over a given depth range of sea floor (Supplementary Information and Supplementary Fig. 4). This suggests a smaller compensation capacity during Eocene time (due to a shallower CCD), which results in a higher sensitivity of carbonate preservation for a given carbon perturbation.

The Pacific CCD record presented here offers a new view of the evolution of Cenozoic ocean carbonate chemistry and provides the basis for future quantitative tests of multiple possible controlling mechanisms. The close correspondence of deep-ocean temperatures derived from benthic foraminiferal $\delta^{18}O$ records and the equatorial Pacific CCD is intriguing and suggests a close coupling of climate and carbon cycle feedbacks during the Cenozoic, tied to an overall increase in weathering during that time.

METHODS SUMMARY

Carbonate measurements were performed by coulometry[2], and supplemented with data from ODP and DSDP Legs (Supplementary Table 1). Accumulation rates were determined by using high-resolution age models and bulk dry density measurements. Stratigraphic correlation of sites was achieved through bio- and magnetostratigraphy, X-ray fluorescence data and physical property measurements[38], adjusted to the age model of the PEAT expeditions[2] revised to new site correlations[38]. Present-day site positions were backtracked using published stage poles[2]. Palaeo-depths were computed including backstripping and using standard methodology[6]. The CCD was semiquantitatively determined by plotting available carbonate accumulation rate data in 250-kyr windows, and fitting a regression line through carbonate accumulation rates decreasing with depth. GENIE Earth system modelling was based on Eocene boundary conditions from previous studies[26,27] using a Palaeocene palaeobathymetry (model SVN revision 7491). Scenarios were investigated as open system runs and with enabled climate feedback (temperature responsive to greenhouse gas forcing) until steady state conditions were achieved (~150 kyr). Ensembles were run on the University of Southampton high-performance computing system IRIDIS3. All scenarios were run for atmospheric CO_2 ranging from 1× to 6× pre-anthropogenic (1× = 278 p.p.m.v.). The scenario in Fig. 3a ('weathering') varied total weathering fluxes from 25% to 200% of modern DIC values in 25% steps (100% = 10 Tmol yr^{-1})[39]. For all runs bioturbation was switched off to speed up the achievement of steady state. All models were run adding a background wetland CH_4 flux at 5× pre-industrial levels, and with a constant detrital flux of 0.18 g cm^{-2} kyr^{-1}. The net-weathering scenario in Fig. 3a was then re-run with varying values for the initial fraction of labile organic carbon 'POM2' (standard GENIE value ~5.6%, additional runs with zero, half, double and quadruple standard value), detailed in a previous publication[34]. Additional GENIE scenarios are described in Supplementary Information.

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

Received 22 February; accepted 26 June 2012.

1. Lyle, M. *et al.* Pacific Ocean and Cenozoic evolution of climate. *Rev. Geophys.* **46,** RG2002, http://dx.doi.org/10.1029/2005RG000190 (2008).
2. Pälike, H. *et al.* Expedition 320/321 summary. *Proc. IODP* **320/321**, doi:10.2204/iodp.proc.320321.2010 (2010).
3. Broecker, W. S. & Peng, T.-H. The role of $CaCO_3$ compensation in the glacial to interglacial atmospheric CO_2 change. *Glob. Biogeochem. Cycles* **1,** 15–29 (1987).
4. Ridgwell, A. & Zeebe, R. The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth Planet. Sci. Lett.* **234,** 299–315 (2005).
5. Van Andel, T. H., Heath, G. R. & Moore, T. C. Jr. Cenozoic history and paleoceanography of the central equatorial Pacific Ocean: a regional synthesis of Deep Sea Drilling Project data. *Geol. Soc. Am.* **143,** 1–134 (1975).
6. Lyle, M. Neogene carbonate burial in the Pacific Ocean. *Paleoceanography* **18,** 1059, http://dx.doi.org/10.1029/2002PA000777 (2003).
7. Peterson, L. C. & Backman, J. Late Cenozoic carbonate accumulation and the history of the carbonate compensation depth in the western equatorial Indian ocean. *Proc. ODP Sci. Res.* **115,** 467–508 (1990).
8. Lyle, M. W., Olivarez Lyle, A., Backman, J. & Tripati, A. Biogenic sedimentation in the Eocene equatorial Pacific—the stuttering greenhouse and Eocene carbonate compensation depth. *Proc. ODP Sci. Res.* **199,** 1–35 (2005).
9. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279–283 (2008).
10. Edmond, J. M. Himalayan tectonics, weathering processes, and the strontium isotope record in marine limestones. *Science* **258,** 1594–1597 (1992).
11. Lear, C. H., Elderfield, H. & Wilson, P. A. A Cenozoic seawater Sr/Ca record from benthic foraminiferal calcite and its application in determining global weathering fluxes. *Earth Planet. Sci. Lett.* **208,** 69–84 (2003).
12. Misra, S. & Froelich, P. N. Lithium isotope history of Cenozoic seawater: changes in silicate weathering and reverse weathering. *Science* **335,** 818–823 (2012).
13. Peucker-Ehrenbrink, B. & Ravizza, G. The marine osmium isotope record. *Terra Nova* **12,** 205–219 (2000).
14. Leon-Rodriguez, L. & Dickens, G. R. Constraints on ocean acidification associated with rapid and massive carbon injections: the early Paleogene record at Ocean Drilling Program Site 1215, equatorial Pacific Ocean. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **298,** 409–420 (2010).
15. Zachos, J. C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686–693 (2001).
16. Beerling, D. J. & Royer, D. L. Convergent Cenozoic CO_2 history. *Nature Geosci.* **4,** 418–420 (2011).
17. Bohaty, S. M., Zachos, J. C., Florindo, F. & Delaney, M. L. Coupled greenhouse warming and deep-sea acidification in the middle Eocene. *Paleoceanography* **24,** PA2207, http://dx.doi.org/10.1029/2008PA001676 (2009).
18. Spofforth, D. J. A. *et al.* Organic carbon burial following the middle Eocene climatic optimum in the central western Tethys. *Paleoceanography* **25,** PA3210, http://dx.doi.org/10.1029/2009PA001738 (2010).
19. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433,** 53–57 (2005).
20. Merico, A., Tyrrell, T. & Wilson, P. A. Eocene/Oligocene ocean de-acidification linked to Antarctic glaciation by sea-level fall. *Nature* **452,** 979–982 (2008).
21. Lyle, M. W., Dadey, K. & Farrell, J. The late Miocene (11–8 Ma) eastern Pacific carbonate crash: evidence for reorganization of deep-water circulation by the closure of the Panama Gateway *Proc. ODP Sci. Res.* **138,** 821–837 (1995).
22. Edgar, K. M., Wilson, P. A., Sexton, P. F. & Suganuma, Y. No extreme bipolar glaciation during the main Eocene calcite compensation shift. *Nature* **448,** 908–911 (2007).
23. Moore, T. C. Jr & Jarrard, R. D. Olivarez Lyle, A. & Lyle, M. W. Eocene biogenic silica accumulation rates at the Pacific equatorial divergence zone. *Paleoceanography* **23,** PA2202, http://dx.doi.org/10.1029/2007PA001514 (2008).
24. Olivarez Lyle, A. & Lyle, M. W. carbon and barium in Eocene sediments: possible controls on nutrient recycling in the Eocene equatorial Pacific Ocean *Proc. ODP Sci. Res.* **199,** 1–33 (2005).
25. Olivarez Lyle, A. & Lyle, M. W. Missing organic carbon in Eocene marine sediments: is metabolism the biological feedback that maintains end-member climates? *Paleoceanography* **21,** PA2007, http://dx.doi.org/10.1029/2005PA001230 (2006).
26. Panchuk, K., Ridgwell, A. & Kump, L. R. Sedimentary response to Paleocene-Eocene Thermal Maximum carbon release: a model-data comparison. *Geology* **36,** 315–318 (2008).
27. Ridgwell, A. & Schmidt, D. N. Past constraints on the vulnerability of marine calcifiers to massive carbon dioxide release. *Nature Geosci.* **3,** 196–200 (2010).
28. Zeebe, R. E. LOSCAR: Long-term Ocean-atmosphere-Sediment CArbon cycle Reservoir Model v2.0.4. *Geoscientific Model Dev.* **5,** 149–166 (2012).
29. Uchikawa, J. & Zeebe, R. E. Influence of terrestrial weathering on ocean acidification and the next glacial inception. *Geophys. Res. Lett.* **35,** L23608, http://dx.doi.org/10.1029/2008GL035963 (2008).
30. Coggon, R. M., Teagle, D. A. H., Smith-Duque, C. E., Alt, J. C. & Cooper, M. J. Reconstructing past seawater Mg/Ca and Sr/Ca from mid-ocean ridge flank calcium carbonate veins. *Science* **327,** 1114–1117 (2010).
31. Stuecker, M. F. & Zeebe, R. E. Ocean chemistry and atmospheric CO_2 sensitivity to carbon perturbations throughout the Cenozoic. *Geophys. Res. Lett.* **37,** L03609, http://dx.doi.org/10.1029/2009GL041436 (2010).
32. Tyrrell, T. & Zeebe, R. History of carbonate ion concentration over the last 100 million years. *Geochim. Cosmochim. Acta* **68,** 3521–3530 (2004).
33. Berner, R. A., Lasaga, A. C. & Garrels, R. M. The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. *Am. J. Sci.* **283,** 641–683 (1983).
34. Ridgwell, A. *et al.* Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling. *Biogeosciences* **4,** 87–104 (2007).
35. Archer, D. & Maier-Reimer, E. Effect of deep-sea sedimentary calcite preservation on atmospheric CO_2 concentration. *Nature* **367,** 260–263 (1994).
36. Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S. & Wakeham, S. G. A new, mechanistic model for organic carbon fluxes in the ocean: based on the quantitative association of POC with ballast minerals. *Deep Sea Res. II* **49,** 219–236 (2001).
37. Ridgwell, A. An end to the "rain ratio" reign? *Geochem. Geophys. Geosyst.* **4,** 1051, http://dx.doi.org/10.1029/2003GC000512 (2003).
38. Westerhold, T. *et al.* Revised composite depth scales and integration of IODP Sites U1331–U1334 and ODP Sites 1218–1220. *Proc. IODP* **320/ 321,** 1–137 (2012).

39. Archer, D. Modeling the calcite lysocline. *J. Geophys. Res. C* **96,** 17037–17050 (1991).
40. Pagani, M. *et al.* The role of carbon dioxide during the onset of Antarctic glaciation. *Science* **334,** 1261–1264 (2011).

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

Acknowledgements This research used samples and data provided by IODP. We thank the masters and crew of IODP Expeditions 320 and 321. H.P. acknowledges support from the Philip Leverhulme Prize, the BIK-F, and NERC grants NE/H000089/1, NE/H020136/1, NE/G003270/1, NE/F003641/1, NE/H022554/1 and NE/I006168/1. We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. We thank M. Palmer and D. Teagle for discussions. E.J.R. is a Visiting Fellow at the Research School of Earth Sciences, The Australian National University.

Author Contributions H.P. and A.R. wrote the manuscript. H.P., A.R., C.O.J.C. and R.E.Z. contributed to the modelling work. All authors contributed to data analysis, interpretation, manuscript editing or discussions.

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

Heiko Pälike[1], Mitchell W. Lyle[2], Hiroshi Nishi[3], Isabella Raffi[4], Andy Ridgwell[5], Kusali Gamage[6], Adam Klaus[6], Gary Acton[7], Louise Anderson[8], Jan Backman[9], Jack Baldauf[2], Catherine Beltran[10], Steven M. Bohaty[1], Paul Bown[11], William Busch[12], Jim E. T. Channell[13], Cecily O. J. Chun[1,14,15], Margaret Delaney[16], Pawan Dewangan[17], Tom Dunkley Jones[18,19], Kirsty M. Edgar[1,20], Helen Evans[21], Peter Fitch[18], Gavin L. Foster[1], Nikolaus Gussone[22], Hitoshi Hasegawa[23], Ed C. Hathorne[24], Hiroki Hayashi[25], Jens O. Herrle[14,15], Ann Holbourn[26], Steve Hovan[27], Kiseong Hyeong[28], Koichi Iijima[29], Takashi Ito[30], Shin-ichi Kamikuri[31,32], Katsunori Kimoto[33], Junichiro Kuroda[34], Lizette Leon-Rodriguez[35], Alberto Malinverno[21], Ted C. Moore Jr[36], Brandon H. Murphy[37], Daniel P. Murphy[1,2], Hideto Nakamura[31], Kaoru Ogane[38], Christian Ohneiser[39], Carl Richter[40], Rebecca Robinson[41], Eelco J. Rohling[1], Oscar Romero[42], Ken Sawada[31], Howie Scher[43], Leah Schneider[44], Appy Sluijs[45], Hiroyuki Takata[46], Jun Tian[47], Akira Tsujimoto[48], Bridget S. Wade[49,50], Thomas Westerhold[51], Roy Wilkens[52], Trevor Williams[21], Paul A. Wilson[1], Yuhji Yamamoto[32], Shinya Yamamoto[53], Toshitsugu Yamazaki[54] & Richard E. Zeebe[55]

[1]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK. [2]Department of Oceanography, Texas A&M University, College Station, Texas 77840-3146, USA. [3]The Center for Academic Resources and Archives, Tohoku University Museum, Tohoku University, Aramaki Aza Aoba 6-3, Aoba-ku, Sendai 980-8578, Japan. [4]Dipartimento di Geotecnologie per l'Ambiente e il Territorio, DiGAT–CeRS Geo, Universitá ''G. DAnnunzio'', Campus Universitario, via dei Vestini 31, Chieti Scalo, Italy. [5]School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK. [6]Integrated Ocean Drilling Program, Texas A&M University, 1000 Discovery Drive, College Station, Texas 77845-9547, USA. [7]Department of Geology, University of California, Davis, One Shields Avenue, Davis, California 95616, USA. [8]Department of Geology, University of Leicester, Leicester LE1 7RH, UK. [9]Department of Geological Sciences, Stockholm University, SE-10691 Stockholm, Sweden. [10]UPMC-Université Paris 06, ISTeP, UMR 7193-4 Place Jussieu, 75252 Paris Cedex 05, France. [11]Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. [12]Earth and Environmental Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, Louisiana 70148, USA. [13]Department of Geological Sciences, University of Florida, 241 Williamson Hall, Gainesville, Florida 32611-2120, USA. [14]Institute of Geosciences, Goethe University Frankfurt, Altenhöferallee 1, D-60438 Frankfurt, Germany. [15]Biodiversity and Climate Research Centre (BIK-F), and Senckenberg Gesellschaft für Naturforschung, D-60325 Frankfurt, Germany. [16]Ocean Sciences, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [17]National Institute of Oceanography, Dona Paula, Goa 403 004, India. [18]Imperial College London, Department of Earth Science and Engineering, South Kensington Campus, London SW7 2AZ, UK. [19]School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [20]School of Earth and Ocean Sciences, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK. [21]Lamont-Doherty Earth Observatory of Columbia University, PO Box 1000, 61 Route 9W, Palisades, New York 10964, USA. [22]Institut für Mineralogie, Westfälische Wilhelms-Universität Münster, Corrensstrasse 24, 48149 Münster, Germany. [23]Department of Natural History Science, Graduate School of Science, Hokkaido University, N10W8, Kita-ku, Sapporo 060-0810, Japan. [24]GEOMAR, Helmholtz Centre for Ocean Research Kiel, Wischhofstrasse 1-3, 24148 Kiel, Germany. [25]Interdisciplinary Faculty of Science and Engineering, Shimane University, 1060 Nishikawatsucho, Matsue City, Shimane 690-8504, Japan. [26]Institut für Geowissenschaften, Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 40, 24098 Kiel, Germany. [27]Department of Geoscience, Indiana University of Pennsylvania, 114 Walsh Hall, Indiana, Pennsylvania 15705, USA. [28]Deep Sea Resources Research Center, Korea Ocean Research and Development Institute, ANSAN PO Box 29, Seoul 425-600, Korea. [29]Institute of Biogeosciences, Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka 237-0061, Japan. [30]Faculty of Education, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan. [31]Faculty of Science, Division of Natural History Sciences, Hokkaido University, Kita-10, Nishi-8, Kita-ku, Sapporo 060-0810, Japan. [32]Center for Advanced Marine Core Research, Kochi University, Kochi 783-8502, Japan. [33]Research Institute for Global Change (RIGC), JAMSTEC, 2-15 Natsushima-Cho, Yokosuka 237-0061, Japan. [34]Institute for Frontier Research on Earth Evolution (IFREE), JAMSTEC, 2-15 Natsushima-cho, Yokosuka 237-0061, Japan. [35]Department of Earth Science, Rice University, 6100 Main Street, MS-126, Houston, Texas 77005, USA. [36]Department of Geological Sciences, University of Michigan, 1100 North University, Ann Arbor, Michigan 48109-1005, USA. [37]Earth and Planetary Sciences, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [38]Institute of Geology and Paleontology, Tohoku University, Aoba 6-4, Aramaki, Aoba-ku, Sendai City 980-8578, Japan. [39]Department of Geology, University of Otago, PO Box 56, Dunedin, New Zealand. [40]School of Geosciences, University of Louisiana, PO Box 44530, Lafayette, Louisiana 70504-0002, USA. [41]Graduate School of Oceanography, University of Rhode Island, South Ferry Road, Narragansett, Rhode Island 02882, USA. [42]Instituto Andaluz de Ciencias de la Tierra, Universidad de Granada, Campus Fuentenueva, 18002 Granada, Spain. [43]Department of Earth and Ocean Sciences, 701 Sumter Street, EWS 617 University of South Carolina, Columbia, SC 29208, USA. [44]Department of Geosciences, Pennsylvania State University, 509 Deike Building, University Park, Pennsylvania 16802, USA. [45]Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Laboratory of Palaeobotany and Palynology, Budapestlaan 4, 3584CD Utrecht, The Netherlands. [46]BK21 Coastal Environmental System School, Division of Earth Environmental System, Pusan National University, San 30 Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea. [47]State Key Laboratory of Marine Geology, Tongji University, Siping Road 1239, Shanghai 200092, China. [48]Faculty of Education, Shimane University, 1060 Nishikawatsucho, Matsue City, Shimane 690-8504, Japan. [49]Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843-3115, USA. [50]School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK. [51]Center for Marine Environmental Sciences (MARUM), University of Bremen, 28359, Bremen, Germany. [52]Hawaii Institute of Geophysics and Planetology, University of Hawaii at Manoa, 1680 East West Road, Honolulu, Hawaii 96822, USA. [53]Yamanashi Institute of Environmental Sciences, 5597-1 Kenmarubi, Kamiyoshida, Fujiyoshida, Yamanashi 403-0005, Japan. [54]Geological Survey of Japan, AIST, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan. [55]School of Ocean and Earth Science and Technology, Department of Oceanography, University of Hawaii at Manoa, 1000 Pope Road, MSB 504, Honolulu, Hawaii 96822, USA.

METHODS

Data generation. Carbonate measurements were performed by coulometry during IODP Expeditions 320/321[2], and supplemented with previously published data from ODP and DSDP Legs (Supplementary Table 1). Mass and carbonate accumulation rates were determined by using high-resolution age models and bulk dry density measurements. Stratigraphic correlation of sites was achieved through bio- and magnetostratigraphy, XRF data and physical property measurements[38], and all data are adjusted to the age model of the PEAT expeditions[2]. Present-day site positions were backtracked using published stage poles[2]. Palaeo-depths were computed assuming subsidence proportional to the square root of age, following previous approaches[6], starting from an assumed ridge crest depth of 2.75 km, and taking into account sediment loading. Uncertainties in the palaeo-depth history are a function of (1) age control, (2) knowledge of the palaeo-depth of the ridge crest, (3) the subsidence history of drill sites, and (4) the sediment loading history for each site. For this study, we assumed a palaeo-depth of the ridge of 2.75 km, fitted subsidence parameters according to the determined basement age and present-day depth and backstripped the sediment loading following previous work[6]. Age models for individual sites are based on an integrated palaeomagnetic and biostratigraphic framework largely following the Expedition dates[2], but revised to new site correlations[38]. Biostratigraphic ages from previous ODP and DSDP sites were updated to this age framework. The new carbonate compensation depths were semiquantitatively determined by plotting available carbonate accumulation rate data in 250-kyr windows, and fitting a regression line through the carbonate accumulation rates that are decreasing with depth.

Earth system modelling. GENIE Earth system modelling was based on previous studies[26,27] using a Palaeocene palaeobathymetry, using model SVN revision 7491 (contact authors for detailed access information). Scenarios were investigated as open system runs and with enabled climate feedback (temperature responsive to greenhouse gas forcing) until steady state conditions were achieved (~150 kyr), using the following Eocene boundary conditions[27]: a solar constant reduced by 0.46% for Palaeogene time (1,361.7 W m^{-2}); a reduced salinity of 33.9 p.s.u.; a constant $CaCO_3$:C_{org} ratio of 0.2, and with seawater concentrations of Mg ≈ 30 mmol kg^{-1} and Ca ≈ 15 mmol kg^{-1}. Ensembles were run on the Southampton high performance computing system IRIDIS3. All scenarios were run for atmospheric CO_2 concentrations ranging from 1× to 6× pre-anthropogenic (1 × = 278 p.p.m.v.). Bottom water temperatures in the model runs corresponding to increasing CO_2 levels are (in °C) 6.6, 9.4, 11.1, 12.3, 13.3 and 14.2, respectively. For all runs bioturbation was switched off to speed up the achievement of steady state. All models were run adding a background wetland CH_4 flux at 5× pre-industrial levels, and with a constant detrital flux of 0.18 g cm^{-2} kyr^{-1}. The scenario in Fig. 3a ('weathering') varied total weathering fluxes from 25% to 200% of modern DIC values in 25% steps (100% = 10 Tmol yr^{-1})[39]. The net-weathering scenario in Fig. 3a was then also re-run with varying values for the initial fraction of labile organic carbon parameter 'POM2' (detailed in ref. 34, standard GENIE value ~5.6%, additional runs with zero, half, double and quadruple standard value). Results from further GENIE scenario runs not resulting in large CCD changes (rain ratio and Mg/Ca changes) are detailed in Supplementary Information, and we also include a description of transient model runs using the LOSCAR box model[28]. For the supplementary 'rain ratio' scenario we varied $CaCO_3$:C_{org} ratios from 0.1 to 0.225 in 0.025 steps, using a fixed 50% modern weathering supply to the deep ocean. A supplementary 'Mg/Ca' scenario varied seawater Mg and Ca concentrations using previously published values[32] for a Mg/Ca range from 1.3 to 5.1, also using a 50% weathering flux compared to modern.

Palaeo-depth uncertainties. The vertical uncertainty of palaeo-depth trajectories is likely to be largest for the earliest part of our records, attributable to the initially more rapid thermal cooling and subsidence at the palaeo-ridge (basement age error ±0.5 Myr), and due to absolute uncertainties in the palaeo-depth of the ridge crest (depth error ±350 m)[41]. Reconstructions become more robust for each site moving forward in time, as the thermal subsidence rate attenuates. We therefore estimate the uncertainty of absolute site palaeo-depths to be of the order of several hundred metres in the early part of the reconstruction for each site, and ~250 m for the remaining record. The palaeo-depth-transect approach however, means that most of the included sites originate from the same Pacific plate ridge segment, thereby reducing the relative error of depth reconstructions.

41. Calcagno, P. & Cazenave, A. Subsidence of the sea-floor in the Atlantic and Pacific Oceans — regional and large-scale variations. *Earth Planet. Sci. Lett.* **126,** 473–492 (1994).

ARTICLE

doi:10.1038/nature11397

An index to assess the health and benefits of the global ocean

Benjamin S. Halpern[1,2], Catherine Longo[1], Darren Hardy[1], Karen L. McLeod[3], Jameal F. Samhouri[4], Steven K. Katona[5], Kristin Kleisner[6], Sarah E. Lester[7,8], Jennifer O'Leary[1], Marla Ranelletti[1], Andrew A. Rosenberg[5], Courtney Scarborough[1], Elizabeth R. Selig[5], Benjamin D. Best[9], Daniel R. Brumbaugh[10], F. Stuart Chapin[11], Larry B. Crowder[12], Kendra L. Daly[13], Scott C. Doney[14], Cristiane Elfes[15,16], Michael J. Fogarty[17], Steven D. Gaines[8], Kelsey I. Jacobsen[8], Leah Bunce Karrer[5], Heather M. Leslie[18], Elizabeth Neeley[19], Daniel Pauly[6], Stephen Polasky[20], Bud Ris[21], Kevin St Martin[22], Gregory S. Stone[5], U. Rashid Sumaila[6] & Dirk Zeller[6]

The ocean plays a critical role in supporting human well-being, from providing food, livelihoods and recreational opportunities to regulating the global climate. Sustainable management aimed at maintaining the flow of a broad range of benefits from the ocean requires a comprehensive and quantitative method to measure and monitor the health of coupled human–ocean systems. We created an index comprising ten diverse public goals for a healthy coupled human–ocean system and calculated the index for every coastal country. Globally, the overall index score was 60 out of 100 (range 36–86), with developed countries generally performing better than developing countries, but with notable exceptions. Only 5% of countries scored higher than 70, whereas 32% scored lower than 50. The index provides a powerful tool to raise public awareness, direct resource management, improve policy and prioritize scientific research.

Human activities such as overfishing, coastal development and pollution have altered marine ecosystems and eroded their capacity to provide benefits now and in the future[1–3]. Yet people benefit directly or indirectly from these activities by extracting food, visiting coastal areas, making a living, or continuing centuries-old traditions. In a world with over seven billion people, nearly half of whom live near the coast[4], we urgently need new analytical approaches to guide how to balance multiple competing and potentially conflicting public goals[5,6] and connect human development with the ocean's capacity to sustain progress[7,8]. Assessments that neglect overall condition in favour of scrutiny of individual goals or stressors cannot adequately serve this role.

Recent initiatives, such as the US National Ocean Policy and EU (European Union) Maritime Strategy, emphasize using comprehensive ecosystem-based management to address the needs of both humans and nature[9,10]. Although such frameworks rely heavily on the concept of ocean health, few guidelines exist for how to measure it[10,11]. Indeed, even though ecosystem health is generally described as the well-being of coupled human–natural systems[12–18], most ecosystem assessments focus solely on the negative impacts of humans on nature[19]. Few synthetic measures exist to assess clearly and quantitatively the health of coupled ocean systems[20]. Without a framework to define and guide the measurement of ocean health, policy and management will resort to assessments that are less transparent, more subjective and that lack standardization across locations and through time[21,22].

Hundreds of specific indicators exist to measure various aspects of ocean condition[23]. A comprehensive index must simultaneously evaluate widely disparate metrics, allowing for an integrated assessment of changes in, for example, fish stocks, extinction risks, coastal jobs, water quality and habitat restoration. Building on and incorporating a wide range of existing indicators (see Supplementary Information for further details), we developed and implemented a systematic approach for measuring overall condition of marine ecosystems that treats nature and people as integrated parts of a healthy system. We thus provide a standardized, quantitative, transparent and scalable measure that can be used by scientists, managers, policy makers and the public to better understand, track and communicate ecosystem status and design strategic actions to improve overall ocean health. Each of ten goals (and their component parts) comprising the index (Fig. 1) can be considered separately or aggregated into an overall score for a region, country, or the entire ocean, and compared across these scales, provided that data sources are consistent. Although tracking individual components of health and benefits is useful[24,25], combining them into a synthetic measure using a concise set of indicators facilitates communication and allows direct comparison among management objectives. Here we provide a robust

[1]National Center for Ecological Analysis and Synthesis, 735 State St Suite 300, Santa Barbara, California 93101, USA. [2]Center for Marine Assessment and Planning, University of California, Santa Barbara, California 93106, USA. [3]COMPASS, Oregon State University, Department of Zoology, Corvallis, Oregon 97331-2914, USA. [4]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, Washington 98112, USA. [5]Conservation International, 2011 Crystal Drive, Arlington, Virginia 22202, USA. [6]Sea Around Us Project, Fisheries Centre, University of British Columbia, Vancouver, British Colombia V6T 1Z4, Canada. [7]Marine Science Institute, University of California, Santa Barbara, California 93106, USA. [8]Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA. [9]Nicholas School of the Environment, Duke University, Durham, North Carolina 27708, USA. [10]Center for Biodiversity and Conservation, American Museum of Natural History, New York, New York 10024, USA. [11]Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775, USA. [12]Center for Ocean Solutions and Hopkins Marine Station, Stanford University, Monterey, California 93940, USA. [13]College of Marine Science, University of South Florida, St Petersburg, Florida 33705, USA. [14]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [15]Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93106, USA. [16]IUCN Global Species Programme/Conservation International, Biodiversity Assessment Unit, 2011 Crystal Drive, Arlington, Virginia 22202, USA. [17]Northeast Fisheries Science Center, Woods Hole, Massachusetts 02453, USA. [18]Department of Ecology and Evolutionary Biology & Center for Environmental Studies, Box 1943, Brown University, Providence, Rhode Island 02912, USA. [19]COMPASS, University of Washington, School of Aquatic and Fisheries Sciences, Seattle, Washington 98195, USA. [20]Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, St Paul, Minnesota 55108, USA. [21]New England Aquarium, Central Wharf, Boston, Massachusetts 02110, USA. [22]Department of Geography, Rutgers University, 54 Joyce Kilmer Drive, Piscataway, New Jersey 08854, USA.

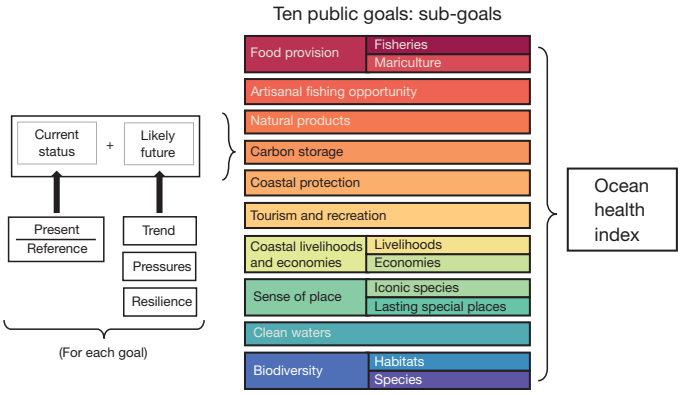


Figure 1 | Conceptual framework for calculating the index. Each dimension (status, trend, pressures and resilience) is derived from a wide range of data. Dimensions combine to indicate the current status and likely future condition for each of ten goals (see equations in Methods Summary and equations (1) and (4) in Methods). Colour scheme is also used in Figs 3–6.

framework to assess ocean health and motivate better data collection to strengthen future iterations of the index.

In developing the index we addressed six major challenges (see Supplementary Information for further details): (1) identify a modest number of widely accepted goals to assess ocean health and benefits at any scale; (2) develop models that measure, with reasonable accuracy, how well each goal is achieved; (3) define robust reference points for each model; (4) incorporate sustainability into the index; (5) ensure that the index is responsive to real differences and changes in ocean health and benefits; and (6) allow flexibility to adapt to constraints (or future improvements) of data availability, quality and quantity. Although the index can be implemented at any scale, here we focus on global and exclusive economic zone (EEZ) scales.

State of the global ocean

The index score for the ocean within EEZ boundaries is 60 out of 100, providing an important benchmark and indicating substantial room for improvement throughout the index's portfolio of ten public goals (Fig. 1). Because EEZs include nearly all continental shelf area and produce the vast majority of food, natural resources, recreation, livelihoods and other benefits to humans[18], this assessment captures most of the concerns of the public, policy makers and resource managers. High seas areas may be added as data become available.

Index scores varied greatly by country, ranging from 36 to 86, with many West African, Middle Eastern and Central American countries scoring poorly, and parts of Northern Europe, Canada, Australia, Japan and various tropical island countries and uninhabited regions scoring highly (Fig. 2 and Supplementary Table 27). Of all EEZs, 32% had an index score <50 whereas only 5% had a score >70. Developed countries tended to score higher than developing countries; index scores are significantly correlated with the Human Development Index[26], an independent measure of development status ($r = 0.57$, $P < 0.0001$, $n = 141$; Supplementary Fig. 8). This is because developed countries tend to have stronger economies, better regulations and infrastructure to manage pressures, and greater capacity to pursue sustainable resource use. Yet some developed countries such as Poland and Singapore scored poorly (index score of 42 and 48, respectively), with low scores for several goals (Fig. 3), whereas developing countries like Suriname (index = 69) and Seychelles (index = 73) scored relatively well because many goals had very high scores (Supplementary Table 27). Individual goal scores varied markedly within and among countries, in turn driving index scores, as illustrated by the United States, China and Poland (Fig. 3).

Three key points affect the interpretation of index scores. First, results for individual goals may seem counterintuitive because we assessed ocean health through the lens of coupled human–natural systems. For example, extractive goals such as 'natural products' score best when harvest levels are high but sustainable, with inherent impacts on nature captured as pressures on other goals. Furthermore, a composite picture of ocean health across multiple goals may not match expectations based on the status of an individual goal. Many goals scored low globally, in particular 'food provision', 'natural products' and 'tourism and recreation' (all <50; Fig. 2 and Supplementary Table 27), whereas other goals scored higher (>75), including 'carbon storage', 'clean waters' and 'biodiversity'. Conclusions based on a single goal will deviate from those derived from the index's portfolio assessment. For example, Russia scored very low for 'food provision' and 'natural products' and very high for 'clean waters' and 'biodiversity', and had an overall index score of 67 (Fig. 3). Because detailed production function models currently do not exist for most goals, it is difficult to know if differences in goal scores are due to direct trade-offs among goals, poor management of low-scoring goals, poorer quality data for some goals, or reference points that are not directly comparable among goals. However, trade-offs probably occur among many goals, such that simultaneously achieving perfect scores on all goals would be difficult.

Second, the index represents the health of coupled human–natural systems. This portfolio includes goals that tend to be more highly valued by preservationists and non-extractive users—protecting places where biodiversity can flourish (existence value) and preserving a sense of place (cultural or aesthetic values)—and also those valued more highly by extractive users—providing food and natural resources. Jarvis Island, an uninhabited, relatively pristine location, received the highest score (index = 86) because many extraction-based goals are not applicable and the island is afforded a high level of protection, whereas

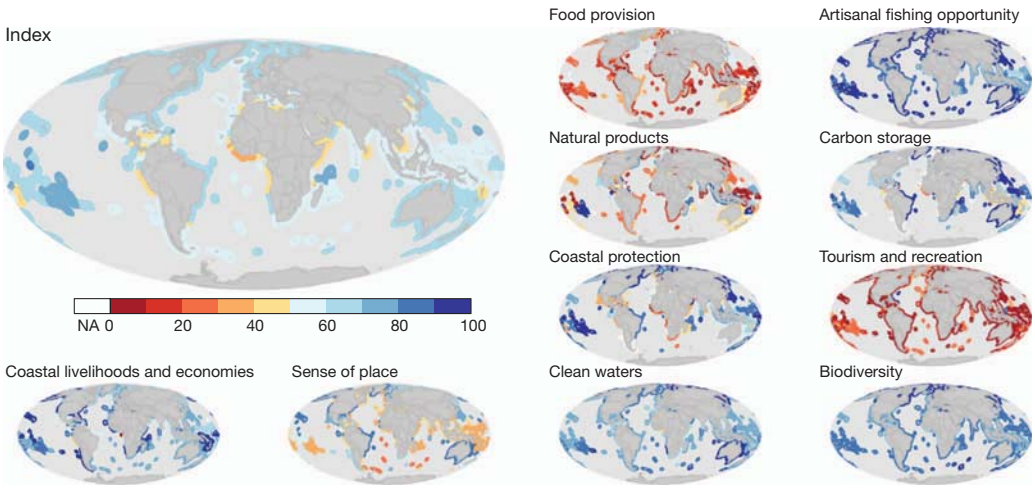


Figure 2 | Map of index and individual goal scores per country. All waters within 171 exclusive economic zones (EEZs), that is, up to 200 nautical miles, were assessed and are represented on the map. See Supplementary Table 24 for details and Supplementary Fig. 2 for sub-goal maps. NA, not available.

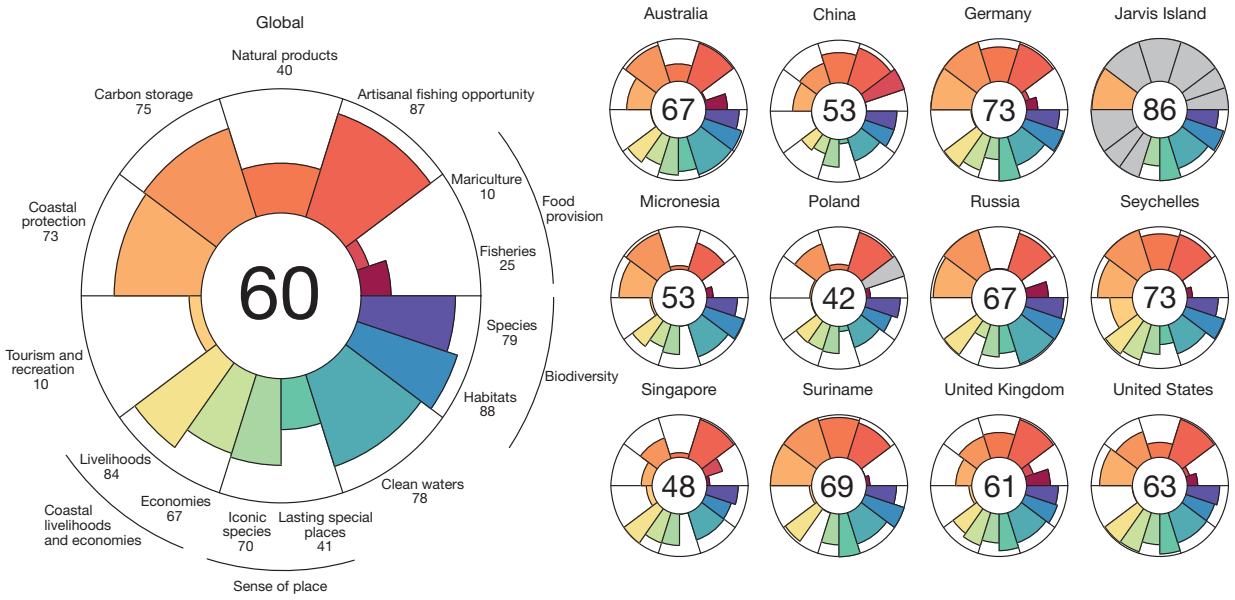


Figure 3 | Index scores (inside circle) and individual goal scores (coloured petals) for global area-weighted average of all countries and for several representative countries. The outer ring is the maximum possible score for each goal, and a goal's score and weight (relative contribution) are represented by the petal's length and width, respectively, except for 'food provision' sub-goals which are weighted by relative actual yield despite equal width of petals (see Supplementary Table 24 for per-country weights). All plots use equal weighting for all goals. Figures 3–6 use consistent goal-specific colour schemes. Grey indicates that a goal is not relevant to that reporting region.

Germany scored highly (index = 73) because eight goals performed well (excepting 'food provision' and 'tourism and recreation'). Our approach to scoring ocean health departs from a purely protectionist one that would aim to maintain natural systems with minimal human impact. The index credits sustainable non-extractive and extractive use, except in places where such uses are prohibited (for example, no-take reserves), as well as preservationist goals.

Third, the index allows transparent assessment of how societal values influence perspectives on ocean health. Although we weighted goals equally to avoid presuming societal values, we recognize that people value ocean benefits differently. To evaluate potential consequences of unequal weighting, we calculated index scores for four potential weighting schemes that approximate preservationist, non-extractive use, extractive use, and strongly extractive use value sets (see Supplementary Information for further details). Resulting global index scores ranged from 56 to 67 across value sets (Fig. 4 and Supplementary Table 30; country-level average maximum difference ± standard error (s.e.) = 7.1 ± 0.2). For a few countries, most notably Romania, Russia, French Guiana, Micronesia and Denmark, changing weights created important differences, altering index scores by up to 27 (Supplementary Table 30). Monte Carlo simulations of thousands of possible value sets produced similar results (index = 60.1 ± 0.003 (s.e.); min = 50; max = 70; Supplementary Fig. 5 and Supplementary Table 30). The preservationist perspective produced the highest index score, primarily because extraction-based goals generally scored low whereas non-extractive goals scored higher. Because goal weights can influence index scores, it is critical to determine societal values (weights) before index calculation. Choosing a single unequal weighting scheme for this global analysis would not have been appropriate as these weights will vary by country, region and community.

Exploring the index

Variation among country-level index and individual goal scores offers novel insights into causes and consequences of different levels of ocean health (Fig. 5). Index scores had a largely unimodal distribution, which is expected in composite indices[27]. No country scored above 86 and most scored below 70. 'Natural products', 'carbon storage' and 'coastal protection' drove variation among countries because of flatter distributions and greater range in values, whereas 'food provision' and 'tourism and recreation' most influenced overall index scores because of their consistently low values (Supplementary Table 31 and Supplementary Information). 'Tourism and recreation' in particular proved difficult to model given limited data, such that scores for this goal are probably artificially low for many countries. 'Biodiversity' scores may seem surprisingly high, but this result accurately reflects that relatively few known marine species risk extinction (see http://www.iucnredlist.org) and that the reference point for this goal is not pristine abundance but instead stable populations of all species (see Supplementary Information). Diving deeper into the index, current status is the main driver of individual goal scores, but with scores notably reduced by negative trends for 'biodiversity' and 'carbon storage' due to decreasing species status and increasing habitat loss, and by pressures greater than resilience for 'tourism and recreation' and 'coastal livelihoods and economies' (Fig. 6 and Supplementary Information).

Countries with identical or similar scores provide examples of how multiple paths exist for achieving any given index score. For example, the United States and United Kingdom scored 63 and 62, respectively, but the scores arose from very different individual goal scores. The United Kingdom scored substantially higher for 'food provision' and 'natural products' whereas the United States scored higher for 'coastal protection' and 'coastal livelihoods and economies' (Fig. 3 and Supplementary Table 27).

Comparing individual goal scores provides guidance for improving overall ocean health, both globally and nationally. Despite the successes of several developed countries in managing their fisheries[28], sustainable global food provision from wild-caught fisheries and mariculture is far below what could be delivered if wild stocks were more sustainably harvested and sustainable mariculture production was increased (country-level 'food provision' scores = 15 ± 1.2 (s.e.); range = 0–72). Coastal habitat loss, which affects multiple goals ('carbon storage', 'coastal protection', and 'biodiversity'), also reduces index scores in many countries, particularly in West Africa, Central America and the Caribbean. Enhanced protection and restoration of mangroves, salt marshes, coral reefs and seagrass beds, for example,

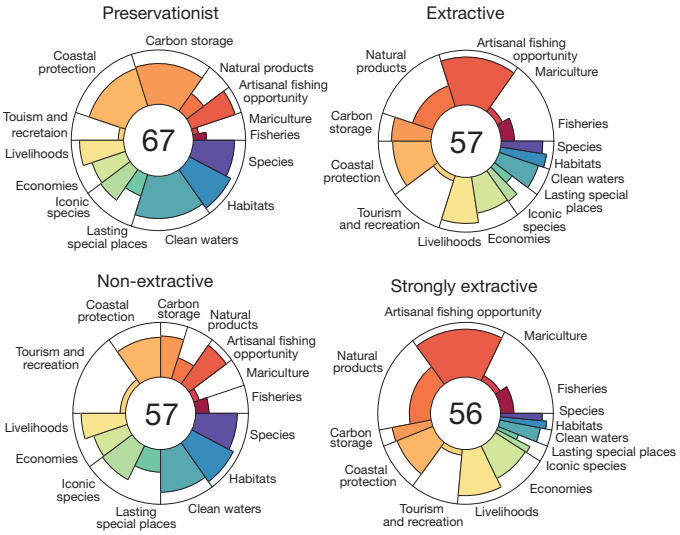


Figure 4 | Global index scores with goals weighted unequally based on four different potential value sets. Value sets are illustrative rather than prescriptive; labels for the value sets are approximations and should not be interpreted literally. See Supplementary Table 4 for weights used in each value set.

could significantly improve ocean health by addressing multiple goals. More effective and comprehensive protection of coastal areas and species, as is being pursued under the Convention on Biological Diversity Aichi Biodiversity Targets for 2020 (ref. 29), would directly benefit 'sense of place' and 'biodiversity' goals, and indirectly benefit most other goals by increasing ecological resilience and thus the likelihood of future goal delivery[30]. Efforts to promote coastal livelihoods, environmentally sensitive urbanization of the coastal zone and improved sanitation infrastructure would improve 'coastal livelihoods and economies', 'tourism and recreation', and 'clean water' goals. Simulating specific management scenarios could provide guidance on which actions would have the greatest impact.

Sustainability into the future

Sustainable delivery of each goal is foundational to our definition of a healthy ocean and approach to modelling the index. The status of many goals incorporates a penalty for pursuing a goal in a way that hampers its future delivery, whereas the 'likely future state' augments scores for goals expected to improve in the near-term future (see Supplementary Information). About half of the goals are getting worse, on average, with negative trends, and pressures roughly equal to resilience (Fig. 6). This assessment could be overly optimistic if existing regulations are not being implemented effectively and existing pressures increase with time. Longer-term trends (>10 years) in goals for which sufficient data exist corroborate recent trends, showing that some sub-goals are broadly improving (for example, 'mariculture' and 'lasting special places') whereas the 'fisheries' sub-goal is declining (Supplementary Fig. 3). Neither method for capturing sustainability actually models future status of goals; such models would be complex and currently do not exist. Given sufficient data, however, both approaches provide meaningful indications of sustainability.

To measure resilience we relied on best available global measures, such as Worldwide Governance Indicators[31], that rarely incorporated information on the implementation or effectiveness of regulations, both key elements of good governance. Therefore, the index incorporates existence rather than outcome of resilience measures, and the projected future improvements are probably optimistic. Many sub-national regulations were not included in this analysis because of the need for globally consistent data. Future regional-scale assessments will incorporate more refined data on governance effectiveness.

A new frame of reference

The index produces results that may be surprising, as the approach deviates from the conventional view of humans as largely exogenous, negative drivers of change in oceans[32]. Humans undoubtedly have substantial negative impacts on the ocean, and index scores are negatively (albeit weakly) correlated with coastal human population ($r = -0.20$; $P = 0.01$; Supplementary Fig. 6) and cumulative impact scores[1] within each region ($r = -0.20$, $P = 0.009$; Supplementary Fig. 9). Yet the regional variation around these relationships shows that all possible combinations of population, impact and provision of benefits exist. Although focus on benefits to people is not new to management or science[18,33], it has yet to become the common currency of assessment.

The index draws from and builds on the ecosystem services perspective in two key ways. Most goals have direct analogues to ecosystem services, but a few do not, in particular the 'coastal livelihoods and economies' and 'artisanal fishing opportunity' goals, which are a function of multiple services and other socio-economic dynamics. Providing a consistent framework for including these highly relevant societal goals not captured by the services framework creates greater opportunity and utility for using the index in management settings. Second, the index represents an absolute measure of ocean health in that all goals are judged against reference points that describe what is possible or desirable in a particular place, whereas ecosystem service assessments typically quantify delivery of services without setting targets. We developed methods for setting reference points for each goal[34], with important implications for resulting goal scores (see Supplementary Information for further details), allowing for measurement of absolute state of ocean health and benefits.

Discussion

The index provides a robust, widely applicable tool for ongoing assessment of ocean health with respect to well-accepted societal goals and a key benchmark against which to compare future progress and inform comprehensive ocean policy. As with any indicator, the index assesses rather than models current and future conditions, and so it cannot predict the future. However, it can be used to simulate the consequences of a range of potential actions, providing a powerful tool to inform decisions about how to use or protect ocean ecosystems.

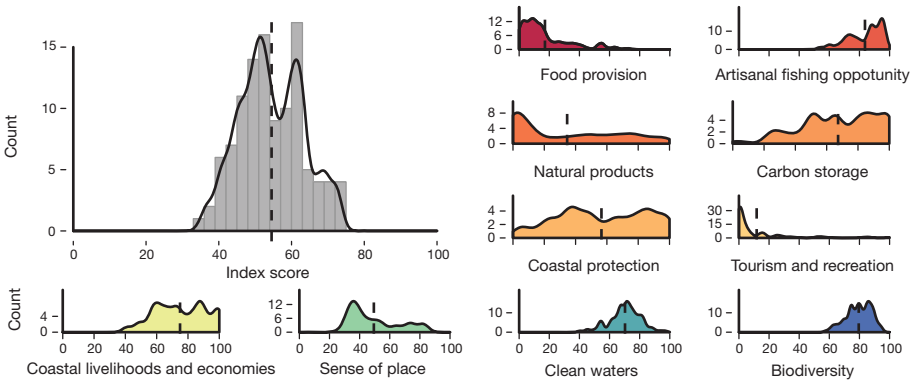


Figure 5 | Distribution of scores across countries for the index and each goal. Histogram plots are smoothed across five point bins; dashed vertical line is the arithmetic mean and so differs from the area-weighted mean in other figures. Note different scales on *y* axes.

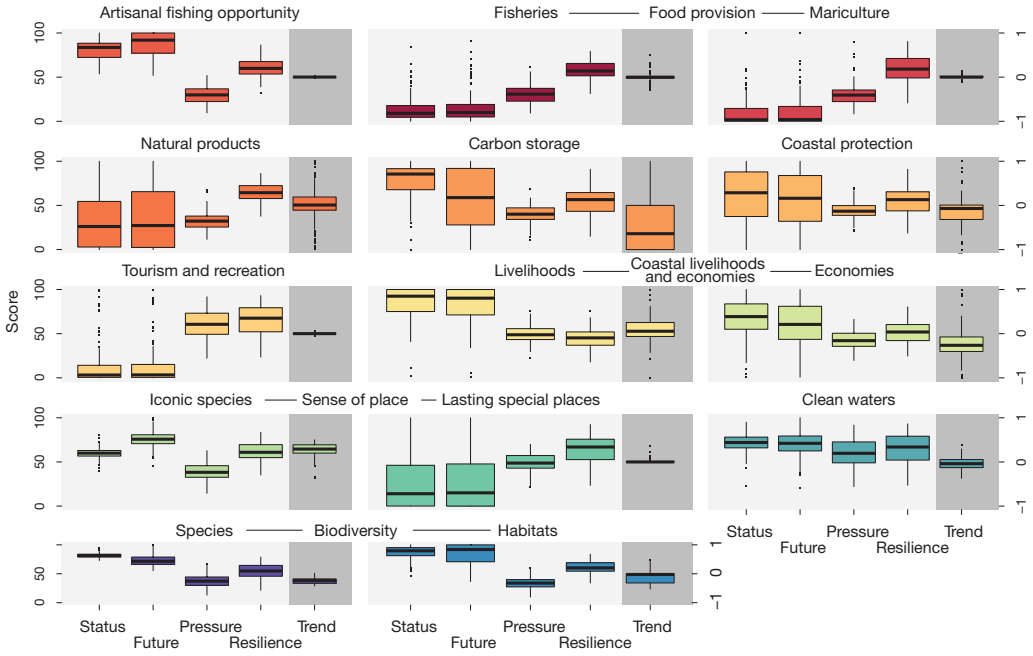


Figure 6 | Box and whisker plots for status, likely future state, pressures, resilience (light grey area), and trend (dark grey area) for country-level scores for each goal and sub-goal. Histograms are shown separately for each sub-goal. Note the different scale for trend (right *y* axis).

Global-scale analyses are useful for global comparisons but tend to be locally imprecise because of inherent challenges in using available global data sets. Future finer-scale applications will allow full exploration of how to best use and refine the index. By calling attention to specific data layers (and gaps), the index can stimulate better measurements, more focused management and, hopefully, accelerate progress towards a healthier ocean.

Developing the index required many assumptions and compromises (see Supplementary Information); here we elaborate on three. First, we limited the index to ten constituent goals primarily for parsimony and ease of communication while maintaining a structure complementary to other ecosystem benefit typologies such as in the Millennium Ecosystem Assessment[18]. We recognize that this structure significantly influences our results. Second, gaps existed in many data sets that we used, requiring proxies or models to fill those gaps (see Supplementary Information). For example, international arrivals data provide a modest proxy for coastal tourism ('tourism and recreation' goal) and undervalue the goal in nations with significant domestic tourism. Likewise, no global data exist for important stressors such as illegal fishing, habitat loss rates and point-source pollution. By identifying these data gaps, the index can help motivate future data collection. In other cases, we had to forgo better quality, region-specific data to maintain global consistency. Future iterations of the index, including those at finer geographic scales, can incorporate new data as available. Better data will in turn allow for construction of improved models that show greater fidelity to each goal's intent, but may also cause scores to change simply because of improved data rather than a change in ocean health.

Finally, key knowledge gaps remain, particularly regarding reference points. The 'mariculture' sub-goal provides an example, where production data are available with appropriate global coverage but sustainability indicators are incomplete. More importantly, underlying production models do not exist to provide appropriate reference levels for any given location. We therefore had to assume that the best-scoring country (China) was the best possible case (that is, reference) and compare all other countries to it, depressing the score for many countries and lowering their 'food provision' scores, particularly if they had high mariculture production.

The composite nature of the index provides guidance on many potential avenues for improving ocean health that cut across multiple goals. In being both quantitative and comprehensive across sectors and goals, the index provides a mechanism for decision makers to evaluate and prioritize actions relative to particular goals given an awareness of potential trade-offs within the full portfolio of goals. More specifically, results indicate that better enforcement of marine protected areas or water quality laws would result in higher resilience scores, lower pressure scores, and ultimately improved status for multiple goals. More efficient use of existing natural and human resources would improve system resilience and therefore status scores (for example, see ref. 35), as would arise through more sustainable harvest of fisheries and mariculture production. Finally, investing in better data collection and reporting to allow more accurate calculation of the index would provide a mechanism for adaptive management, where individuals and institutions learn from their experiences to make more informed decisions. Although such recommendations are not novel from a sectoral perspective, the index offers new potential to leverage benefits from such actions across multiple sectoral goals.

The index allows clear and rapid communication of vast quantities of information. Resource managers, policy makers and the public typically gravitate towards specific issues. By demonstrating how and where these issues fit into a broader context, the index creates an important opportunity to transform the dialogue on how we manage our interactions with the ocean and meets a need unfulfilled by tracking single-sector outcomes. Indeed, pursuing options where several goals improve slightly may provide better outcomes than aiming to improve significantly a single goal, which could lead to inefficient or even unwanted outcomes[36]. We are currently implementing the index at regional scales in the United States, Brazil and Fiji and will regularly update the global assessment, allowing us to assess how the index responds to specific management actions and better understand and evaluate progress and trade-offs that emerge from individual management decisions. The transparency, comparability and target-driven quantitative assessments provided by the index are valuable to management at all scales, making the index an important tool for decision making from local to international levels.

METHODS SUMMARY

We define the index as the condition of ten widely accepted public goals for ocean ecosystems (Fig. 1 and Supplementary Information), which include but are not limited to established ecosystem services (for example, 'coastal livelihoods and economies' is not an ecosystem service)[18]. The index (I) score is the weighted sum of ten goal-specific index scores (I_i):

$$I = \sum_{i=1}^{N} \alpha_i I_i$$

where α_i is the goal-specific weight ($\sum \alpha_i = 1$; default is $\alpha_i = 1/N$) (see Supplementary Information) and I_i is the average value of present and likely future status, $I_i = (x_i + \hat{x}_{i,F})/2$, for each goal i. The present status of goal i (x_i) is its present status value (X_i) relative to a reference point ($X_{i,R}$) uniquely chosen for each goal following guiding principles (see Supplementary Information and ref. 34), and rescaled 0–100. The likely future status ($\hat{x}_{i,F}$)is a function of present status (x_i), recent (~5 year) trend (T_i), pressures (p_i), and factors that promote resilience (r_i), such that

$$\hat{x}_{i,F} = (1+\delta)^{-1}[1 + \beta T_i + (1-\beta)(r_i - p_i)]x_i$$

where the discount rate $\delta = 0$ and the weighting term $\beta = 0.67$, giving trend twice the importance of the difference between resilience and pressures in determining likely future state (see Supplementary Information). We tested the sensitivity of results to assumptions about δ and β and found minimal differences for near-term timeframes (see Supplementary Information). Assessment of the likely future status captures whether the present status is likely to persist, improve or decline in the near-term future, based on current status (x_i) and trends, and is therefore an indication rather than prediction of the near-term future. Ecological pressures fall into five broad categories—pollution, habitat destruction, species introductions, fishing and climate change—and are weighted equally to social pressures (such as poverty, political instability and corruption), with resilience measures such as international treaties and ecological resilience included when they address pressures relevant to a particular goal (see Supplementary Information). The inclusion of these factors ensures that the index is responsive to changes that are reflected more slowly in the current state.

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

Received 13 January; accepted 5 July 2012.
Published online 15 August 2012.

1. Halpern, B. S. *et al.* A global map of human impact on marine ecosystems. *Science* **319,** 948–952 (2008).
2. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293,** 629–637 (2001).
3. Pandolfi, J. M. *et al.* Global trajectories of the long-term decline of coral reef ecosystems. *Science* **301,** 955–958 (2003).
4. Center for International Earth Science Information Network. *Socioeconomic Data and Applications Center (SEDAC)* http://sedac.ciesin.columbia.edu/ (2005).
5. Kareiva, P., Lalasz, R. & Marvier, M. Conservation in the Anthropocene: Beyond solitude and fragility. *Breakthrough J.* No. 2 29–37 (2011).
6. Kareiva, P., Watts, S., McDonald, R. & Boucher, T. Domesticated nature: shaping landscapes and ecosystems for human welfare. *Science* **316,** 1866–1869 (2007).
7. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461,** 472–475 (2009).
8. Rockström, J. *et al.* Planetary boundaries: exploring the safe operating space for humanity. *Ecol. Soc.* **14,** 32 (2009).
9. National Ocean Council. Draft National Ocean Policy Implementation Plan, http://www.whitehouse.gov/sites/default/files/microsites/ceq/national_ocean_policy_draft_implementation_plan_01-12-12.pdf (2012).
10. Administration of Barack H. Obama. Executive Order 13547: Stewardship of the ocean, our coasts, and the great lakes (2010).
11. McLeod, K. L. & Leslie, H. M. (eds). *Ecosystem-Based Management for the Oceans* (Island, 2009).
12. Munoz-Erickson, T. A., Aguilar-Gonzalez, B. & Sisk, T. D. Linking ecosystem health indicators and collaborative management: a systematic framework to evaluate ecological and social outcomes. *Ecol. Soc.* **12,** 6 (2007).
13. Rapport, D. J. *et al.* Ecosystem health: the concept, the ISEH, and the important tasks ahead. *Ecosyst. Health* **5,** 82–90 (1999).
14. Cairns, J., McCormick, P. V. & Niederlehner, B. R. A proposed framework for developing indicators of ecosystem health, *Hydrobiologia* **263,** 1–44 (1993).
15. Kelly, J. R. & Harwell, M. A. Indicators of ecosystem recovery. *Environ. Manage.* **14,** 527–545 (1990).
16. Rapport, D. J., Costanza, R. & McMichael, A. J. Assessing ecosystem health. *Trends Ecol. Evol.* **13,** 397–402 (1998).
17. Costanza, R. in *Ecosystem health: New Goals for Environmental Management* 239 (Island, 1992).
18. Millennium Ecosystem Assesment. *Ecosystems and Human Well-Being: Synthesis Report* (Millennium Ecosystem Assessment, 2005).
19. Emerson, J. *et al. The State of the Nation's Ecosystems: Measuring the Land, Waters, and Living Resources of the United States* (Island, 2008).
20. Kershner, J., Samhouri, J. F., Andrew, J. & Levin, P. S. Selecting indicator portfolios for marine species and food webs: a Puget Sound case study. *PLoS ONE* **6,** e25248 (2011).
21. Rice, J. C. & Rochet, M. J. A framework for selecting a suite of indicators for fisheries management. *ICES J. Mar. Sci.* **62,** 516–527 (2005).
22. Niemeijer, D. & de Groot, R. S. A conceptual framework for selecting environmental indicator sets. *Ecol. Indic.* **8,** 14–25 (2008).
23. The Heinz Center. *The State of the Nation's Ecosystems 2008: Measuring the Lands, Waters and Living Resources of the United States* (Island, 2008).
24. Butchart, S. H. M. *et al.* Global biodiversity: indicators of recent declines. *Science* **328,** 1164–1168 (2011).
25. Alder, J. *et al.* Aggregate performance in managing marine ecosystems of 53 maritime countries. *Mar. Policy* **34,** 468–476 (2010).
26. United Nations Development Programme. Human Development Report, http://hdr.undp.org/en/media/HDR_2010_EN_Complete_reprint.pdf (Palgrave Macmillan, 2010).
27. Glover, D. M., Jenkins, W. J. & Doney, S. C. *Modeling Methods for Marine Science* (Cambridge Univ. Press, 2011).
28. Worm, B. *et al.* Rebuilding global fisheries. *Science* **325,** 578–585 (2009).
29. Convention on Biological Diversity. Strategic plan for biodiversity 2011–2020, including Aichi biodiversity targets, http://www.cbd.int/doc/strategic-plan/2011-2020/Aichi-Targets-EN.pdf (2010).
30. Worm, B. *et al.* Impacts of biodiversity loss on ocean ecosystem services. *Science* **314,** 787–790 (2006).
31. World Bank Group. *Worldwide Governance Indicators* http://info.worldbank.org/governance/wgi/index.asp (2010).
32. Shackeroff, J. M., Hazen, E. L. & Crowder, L. B. in *Ecosystem-Based Management for the Oceans* (eds McLeod, K. L. & Leslie, H. M.) 33 (Island, 2009).
33. Kareiva, P. *et al. Natural Capital: Theory and Practice of Mapping Ecosystem Services* (Oxford Univ. Press, 2011).
34. Samhouri, J. F. *et al.* Sea sick? Setting targets to assess ocean health and ecosystem services. *Ecosphere* **3,** 1–18 (2012).
35. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478,** 337–342 (2011).
36. Newton, A. C. Implications of Goodhart's law for monitoring global biodiversity loss. *Conserv. Lett.* **4,** 264–268 (2011).

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

Acknowledgements B. Wrigley and H. Wrigley provided the founding grant. Additional financial and in-kind support was provided by the Pacific Life Foundation, Thomas W. Haas Fund of the New Hampshire Charitable Foundation, the Oak Foundation, Akiko Shiraki Dynner Fund for Ocean Exploration and Conservation, Darden Restaurants Inc. Foundation, Conservation International, New England Aquarium, National Geographic, and the National Center for Ecological Analysis and Synthesis, which supported the Ecosystem Health Working Group as part of the Science of Ecosystem-Based Management project funded by the David and Lucile Packard Foundation. We would like to thank L. Bergen, J. Bort, B. Bronson, C. Costello, N. Crone, A. Dickson, J. Francis, A. Ghermandi, R. Haskell, L. Kaufman, K. Law, L. Madin, P. Nuñes, D. Obura, L. Onofri, J. Packard, R. Portela, N. Rao, J. Regetz, S. Running, K. Selkoe, L. Speer, B. Spitzer, P. Stevick, H. Tallis, H. Tausig, S. Troeng and D. Zeyen for helpful discussions and logistical support during development of the index. Individual authors also acknowledge additional support from NSF, NASA, NOAA, Stanford's Center for Ocean Solutions and the Jaffe Family Foundation. K.K., D.P., U.R.S. and D.Z. thank the Sea Around Us Project, with support from the Pew Charitable Trusts. Data reported in this paper are tabulated in the Supplementary Information and archived at http://ohi.nceas.ucsb.edu/data. Results can be explored and visualized at http://oceanhealthindex.org.

Author Contributions B.S.H., C.L., K.L.M., J.F.S., S.K.K., S.E.L., A.A.R., D.R.B., F.S.C., L.B.C., K.L.D., S.C.D., M.J.F., S.D.G., L.B.K., H.M.L., S.P., K.S.M., B.R. and G.S.S. conceived and planned the study. B.S.H., C.L., D.H., S.E.L., J.O., M.R., C.S., E.R.S., B.D.B., C.E., K.I.J., D.P., U.R.S. and D.Z. aided with data collection and database development. B.S.H., C.L., D.H., K.L.M., J.F.S., S.K.K., K.K., S.E.L., A.A.R., C.S., E.R.S., F.S.C., K.L.D., S.C.D., C.E., M.J.F., S.D.G., L.B.K., H.M.L., U.R.S. and D.Z. helped with data interpretation. B.S.H., C.L., D.H., K.L.M., J.F.S., S.K.K., K.K., S.E.L., J.O., M.R., A.A.R., C.S., E.R.S., B.D.B., D.R.B., S.C.D., M.J.F., S.D.G. and S.P. developed index and goal-specific models. B.S.H., C.L., D.H., A.A.R., S.C.D. and M.J.F. conducted analyses. B.S.H., C.L., K.L.M., J.F.S., B.D.B. and E.N. developed the figures. B.S.H. drafted the initial manuscript and provided overall project management. All authors edited 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 the paper. Correspondence and requests for materials should be addressed to B.S.H. (halpern@nceas.ucsb.edu).

METHODS

We measured ocean health as a function of ten widely held public goals (see Supplementary Information for further details) for what the ocean can provide to people (Fig. 1). The index therefore recognizes linkages between human societies and ocean ecosystems, and that people are part of coastal and ocean systems. Full details for how scores were calculated for the overall index and each goal, the data used in each case, and how each data layer was processed are provided in the Supplementary Information. Here we provide a summary of the general approach and models used, with details for model equations and parameters provided in Supplementary Table 33.

The objective (utility function) of the index is to maximize its value (I), where I is determined as a linear weighted sum of the scores for each of the public goal indices (I_1, I_2, . . ., I_{10}) and the appropriate weights for each goal (α_1, α_2,..., α_{10}), such that:

$$I = \alpha_1 I_1 + \alpha_2 I_2 + \ldots \alpha_{10} I_{10} = \sum_{i=1}^{N} \alpha_i I_i \quad (1)$$

where $\sum \alpha_i = 1$. Each goal score, I_i, is a function of its present status x_i and an indication of its likely near-term future status $\hat{x}_{i,\mathrm{F}}$:

$$I_i = \frac{x_i + \hat{x}_{i,\mathrm{F}}}{2} \quad (2)$$

The present status of goal i, x_i, is its present value, X_i, relative to a reference point, $X_{i,\mathrm{R}}$, uniquely chosen for each goal and rescaled 0–100 such that:

$$x_i = \frac{X_i}{X_{i,\mathrm{R}}} \quad (3)$$

For goals, where exceeding the reference point is possible but not desirable because it is unsustainable, the calculation of the present state accounts for this.

The reference point, $X_{i,\mathrm{R}}$, can be determined in four potential ways, depending on the conceptual and data constraints for each goal[34]. Reference points can be estimated mechanistically using a production function (for example, maximum sustainable yield for fisheries), spatially by means of comparison with another region (for example, highest-ranked country represents the best possible known case), temporally using a past benchmark (for example, historical habitat extent), or in some cases via known (for example, zero pollution) or established (for example, 20% of waters set aside in marine protected areas) targets. Past benchmarks can either be a fixed point in time or a moving target (for example, 5 years before most current data).

For six of the ten goals, production (or delivery) of the goal involves activities by people that can negatively feed back on the potential of the goal to be realized (for example, overfishing ultimately reduces the total catch that is available). The six goals include 'food provision', 'artisanal fishing opportunity', 'natural products', 'tourism and recreation', 'coastal livelihoods and economies', and 'sense of place' (for example, visiting cultural sites can have a negative impact on them). This type of sustainability is built into the status assessment for the goals for which it can be assessed and assumed to be neutral in other goals (for example, 'sense of place') for which we currently do not have research or data to inform how this feedback works.

The estimate of a goal's likely near-term future status, $\hat{x}_{i,\mathrm{F}}$, is a function of four dimensions: present status; recent trend (over the past ~5 years) normalized to a reference value (T_i); current cumulative pressures to the goal (p_i); and social and ecological resilience to negative pressures (measured as a function of governance and social institutions in place to protect or regulate the system, and the ecological condition of locations; r_i). Trend is calculated as the slope of the change in status based on recent data. Whenever possible, trends were calculated as the slope of annual data over the previous 5 years; we included values from previous years and slopes calculated on as few as two data points (in very few cases) when faced with incomplete data sets. Because status values ranged from 0 to 100, trend primarily ranged from −1.0 to 1.0; we constrained values outside this range to these range end points because such steep slopes are usually a result of extremely unusual events or insufficient data.

The role of resilience and pressure dimensions is to improve our understanding of the likely near-term future condition by incorporating additional information beyond that provided by the recent trend, T_i. Our approach identifies those factors that negatively affect a goal as pressures and those that positively affect a goal as resilience (see 'Calculating pressures' and 'Calculating resilience' sections below). The recent trend captures the direction and rate of change based on conditions in the recent past. However, some pressure or resilience measures that were in existence in the past may have a cumulative effect that has not yet manifested itself (for example, for p: fishing pressure may have increasingly negative impacts as successive year classes of fish become increasingly less abundant; for r: establishment of a marine protected area may require a number of years before its benefits become apparent). In addition, the recent trend does not capture the effect of current levels of resilience (r) and pressures (p). The expectation of a likely future condition suggested by the trend will become more or less optimistic depending on the effects of r and p. If the effects are equal they cancel each other out.

Both r_i and p_i are scaled such that $0 \leq r_i, p_i \leq 100$, with 100 being the maximum value in both cases. The trend (T_i) is constrained to $-1.0 \leq T_i \leq 1.0$ (as noted above). The likely future status is then defined as:

$$\hat{x}_{i,\mathrm{F}} = (1+\delta)^{-1}[1 + \beta T_i + (1-\beta)(r_i - p_i)]x_i \quad (4)$$

where the discount rate (δ) is set to 0 initially. β represents the relative importance of trend versus resilience and pressure terms in determining the likely trajectory of the goal status into the future. We assume $\beta = 0.67$ based on the idea that the direct measure of trend is a better indicator of future (5 years) condition than indirect measures of pressure and resilience. This assumption makes trend twice as important; it is not possible to derive this weight empirically, and so we tested the sensitivity of the results to this assumption. Because we presume a roughly 5-year horizon for the likely future status, we assume $\delta = 0$; we tested the sensitivity of results to this assumption.

It is important to note that with high-resolution spatial and temporal data that perfectly measure all four dimensions within a goal, the likely future condition would approach the current status as current status approaches its maximum value. In other words, the likely future status cannot exceed the maximum possible value for status for each goal, which is 100 in nearly all cases. In reality, data are rarely perfect, creating potential situations where likely future condition exceeds 100. To address these cases, we implemented two rules that follow logical constraints. First, if current status = 100, then trend is set = 0.0, as any trend >0.0 in those cases must be due to incomplete or imperfect data. Second, given that $x_i^{\max}$ is equal to the maximum attainable status given realistic constraints, then if poor data quality or other practical constraints lead to $\hat{x}_{i,\mathrm{F}} > x_i^{\max}$ we set $\hat{x}_{i,\mathrm{F}} = x_i^{\max}$.

The maximum possible index value (U) is the sum of the maximum possible values for each goal indicator. Because this maximum value is the best possible value today and in the future, $r > p$ and $T = 0$, such that $x_i^{\max} = X_{i,\mathrm{R}}$ where the reference state has been normalized to 100. U is then:

$$U = \sum_{i=1}^{N} \alpha_i x_i^{\max} \quad (5)$$

We can therefore calculate the index (I):

$$\mathrm{I} = \frac{\sum_{i=1}^{N} \alpha_i I_i}{U} \quad (6)$$

This formulation also allows for assessments to be conducted using goals that are of interest/use for a particular location. For example, few extractive use public goals are relevant to or valued in uninhabited regions, so calculation of the index for these areas is based on a subset of goals. In this way, uninhabited areas that are pristine could score very highly, even though they are not delivering more direct market-based benefits to people.

Calculating pressures. To calculate pressures for each goal (p_i) we evaluate both ecological (p_{E}) and social pressures (p_{S}), such that:

$$p_i = \gamma(p_{\mathrm{E}}) + (1-\gamma)(p_{\mathrm{S}}) \quad (7)$$

where γ is the relative weight for ecological versus social pressures and is set equal to 0.5. At global scales, little evidence exists to support unequal weighting of ecological and social pressures for most goals; furthermore, unequal weighting would require unique values for each goal and there is currently no empirical work to guide such decisions. At local or regional scales there may be clear evidence for unequal weights per goal and γ should be adjusted accordingly.

We assessed five broad, globally relevant categories of ecological stressors: fishing pressure (p_{f}), habitat destruction (p_{hd}), climate change (including ocean acidification) (p_{c}), water pollution (p_{p}), and species introductions (invasive species and genetic escapes) (p_{sp}) (Supplementary Table 2). The five categories are intended to capture known pressures to the social–ecological system associated with each goal, that is, impacts that are assumed to significantly affect the ecological and social state of a system, and are derived from other systems of categorizing classes of stressors[1,37,38]. Because many ecological stressors within these categories

have specific consequences for goals, we assessed and ranked separately each ecological stressor within these categories.

To account for the cumulative effect of stressors, we summed the weighted intensities of each stressor within a pressure category (p_k) and divided by the maximum weighted intensity that could be achieved by the worst stressor (max = 3.0) such that:

$$p_k = \frac{\sum_i^M w_i s_i}{3} \quad (8)$$

where w_i is the stressor-specific sensitivity weights (from Supplementary Table 25) and s_i is the data-derived intensity of the associated stressor (which is scaled 0–1). If $p_k > 100$, we set the value equal to 100. This formulation assumes that any cumulative pressure load greater than the maximum intensity of the worst stressor is equivalent to maximum stressor intensity. The intensity data layers for stressors come from a wide range of sources (see Supplementary Table 23).

Overall ecological pressures (p_E) are then calculated as the weighted average of the pressure categories relevant to each goal, with weights set as the maximum rank in each pressure category (w_{i_max}), such that:

$$p_E = \frac{\sum_i^N (w_{k\_max} p_k)}{\sum w_{i\_max}} \quad (9)$$

Stressors that have no impact drop out rather than being assigned a rank of zero, which would affect the average score.

For social pressures, we primarily used data from the Worldwide Governance Indicators (WGI), a composite of hundreds of different measures that assesses in very broad but comprehensive terms the social structure and functioning of countries, scoring them along six component composite indicators: control of corruption, government effectiveness, political stability, regulatory quality, rule of law, voice and accountability. We averaged scores for all six components of the WGI and then rescaled them 0–100, with pressures then assessed as (1 − WGI). For the 'coastal livelihoods and economies' goal, we used one additional data layer to approximate social pressure: the global competitiveness index (GCI). If additional social pressure layers are identified for other goals in the future, they would be averaged with the WGI score in this same manner. Social pressures are therefore:

$$p_S = \frac{\sum_i^N z_i}{N} \quad (10)$$

where z_i are the social pressure measures specific to the goal (in most cases, only the WGI score). Unequal weighting may be appropriate in some cases but is difficult to assess currently. Finally, to combine the social and ecological pressures, we assumed that each should have the potential to contribute equally to the overall pressure score (as described in equation (7)).

Calculating resilience. To calculate resilience for each goal (r_i) we assess three types of measures: ecological integrity (Y_E), goal-specific regulations aimed at addressing ecological pressures (G), and social integrity (Y_S). The first two measures address ecological resilience whereas the third addresses social resilience. When all three aspects are relevant to a goal, resilience is calculated as:

$$r_i = \gamma\left(\frac{Y_E + G}{2}\right) + (1-\gamma)Y_S \quad (11)$$

where the three types of measures are all scaled 0–100, and γ is assumed to be 0.5. We chose $\gamma = 0.5$ so that the weight of resilience components that address ecological systems versus social systems were equivalent, based on the same rationale as for ecological pressures versus social pressures, with the intent, as best as possible, to have resilience measures directly matched with pressures.

Goal-specific regulations (G) are intended to describe the factors that set rules and regulations to address ecological pressures, and are measured as laws and other institutional measures related to a specific goal. Governance is a function of (1) institutional structures that address the intended objective; (2) a clear process for implementing the institution being in place; and (3) whether the institution has been effective at meeting stated objectives[39]. At global scales it is very difficult to assess these three elements; we usually only had information on whether institutions exist. However, in some cases we had detailed information on institutions that enabled us to assess whether they would contribute to effective management, and thus, increased ocean health. In those latter cases, we gave more weight to those measures. Specifically, we calculated G as a weighted average:

$$G = \frac{\sum w_i G_i}{\sum w_i} \quad (12)$$

where G_i is the specific regulatory measure (data set), and w_i is the weight for each i data set used to assess G based on the quality of information contained in the data sets with regard to estimates of regulation effectiveness (see Supplementary Table 3). For habitat resilience, fishing resilience, and CITES signatories (Convention on the International Trade of Endangered Species), any country without a score is given $G_i = 0$; otherwise, any country without data for G_i is excluded from equation (12) for that country.

37. Halpern, B. S. *et al.* Mapping cumulative human impacts to California Current marine ecosystems. *Conserv. Lett.* **2,** 138–148 (2009).
38. Nellemann, C., Hain, S. & Alder, J. in *Dead Water: Merging of Climate Change with Pollution, Over-Harvest, and Infestations in the World's Fishing Grounds* (United Nations Environment Programme, 2008).
39. Donoghue, J. *et al.* Theme plenary session: implementation, compliance and effectiveness. Annual Meeting of the American Society of International Law: Implementation, Compliance and Effectiveness (ASIL, 1997).

ARTICLE

doi:10.1038/nature11400

Antibiotics in early life alter the murine colonic microbiome and adiposity

Ilseung Cho[1,2], Shingo Yamanishi[1], Laura Cox[3], Barbara A. Methé[4], Jiri Zavadil[5,6], Kelvin Li[4], Zhan Gao[3], Douglas Mahana[3], Kartik Raju[3], Isabel Teitler[3], Huilin Li[7], Alexander V. Alekseyenko[1,6] & Martin J. Blaser[1,2,3]

Antibiotics administered in low doses have been widely used as growth promoters in the agricultural industry since the 1950s, yet the mechanisms for this effect are unclear. Because antimicrobial agents of different classes and varying activity are effective across several vertebrate species, we proposed that such subtherapeutic administration alters the population structure of the gut microbiome as well as its metabolic capabilities. We generated a model of adiposity by giving subtherapeutic antibiotic therapy to young mice and evaluated changes in the composition and capabilities of the gut microbiome. Administration of subtherapeutic antibiotic therapy increased adiposity in young mice and increased hormone levels related to metabolism. We observed substantial taxonomic changes in the microbiome, changes in copies of key genes involved in the metabolism of carbohydrates to short-chain fatty acids, increases in colonic short-chain fatty acid levels, and alterations in the regulation of hepatic metabolism of lipids and cholesterol. In this model, we demonstrate the alteration of early-life murine metabolic homeostasis through antibiotic manipulation.

Antibiotics, discovered in the early twentieth century, came into widespread use after the Second World War, with substantial public health benefits. Antibiotic use has increased markedly, now approximating one antibiotic course per year in the average child in the United States[1,2]. However, there is increasing concern that antibiotic exposure may have long-term consequences[3–5].

For more than 50 years we have known that the administration of low doses of antibacterial agents promotes the growth of farm animals, consequently, in the United States, the largest use of antibiotics and related antimicrobial substances is within farms, with low doses fed to large numbers of animals used for food production to increase weight gain by as much as 15%[6,7]. These effects are broad across vertebrate species, including mammals (cattle, swine, sheep) and birds (chickens, turkeys), and follow oral administration of the agents, either in feed or water, indicating that the microbiota of the gastrointestinal (GI) tract is a major target. That the effects are observed with many different classes of antibacterial agents (including macrolides, tetracyclines, penicillins and ionophores) indicates that the activity is not an agent-specific side effect, nor have the effects been observed with antifungals or antivirals.

The vertebrate GI tract contains an exceptionally complex and dense microbial environment, with bacterial constituents that affect the immune responses of populations of reactive host cells[8] and stimulate a rich matrix of effecter mechanisms involved in innate and adaptive immune responses[9]. The GI tract also is a locus of hormone production, including those involved in energy homeostasis (such as insulin, glucagon, leptin and ghrelin) and growth (for example, glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide 1 (GLP-1))[10]. Alterations in the populations of the GI microbiota may change the intra-community metabolic interactions[11], modify caloric intake by using carbohydrates such as cellulose that are otherwise indigestible by the host[12], and globally affect host metabolic, hormonal and immune homeostasis[13]. Full (therapeutic) dose antibiotic treatments alter both the composition of the gastrointestinal microbiota[14] and host responses to specific microbial signals[15]. In combination with dietary changes, antibiotic administration has been associated with changes in the population structure of the microbiome. However, the effects of exposure to subtherapeutic antibiotic dosages have not been described.

Early studies of the effects of gut microbiota on metabolism were limited by the use of culture-based technologies that interrogated <5% of the extant GI tract microbes[16]. Culture-independent investigation of small-subunit ribosomal RNA sequences allows the microbial population structure[17] of the gut microbiota to be characterized with greater resolution. Despite inter-individual differences, substantial similarities exist[18] among mammalian species in the GI microbiota at higher taxonomic levels and functional pathways, indicating a basis for the conserved responses to early-life subtherapeutic antibiotic treatment (STAT) within farms. Previous work has shown that obesity leads to variation in the GI microbiome[12,19]; we use the insights provided from modern agricultural practices to suggest an alternative approach, using a murine model of STAT to explore how antibiotic exposure modulates host metabolic phenotypes.

Early-life STAT increases adiposity

We exposed C57BL/6J mice at weaning to penicillin, vancomycin, penicillin plus vancomycin, chlortetracycline, or no antibiotic in their drinking water at levels in the mid-range of US Food and Drug Administration (FDA)-approved levels for subtherapeutic antibiotic use in agriculture[6,7]. After a 7 week exposure, the observed weights were within the expected range of growth for female C57BL/6J mice, and there was no significant difference in overall growth between the STAT and control mice (Fig. 1a). However, by dual energy X-ray absorptiometry (DEXA) scanning, (Fig. 1b) total fat mass was significantly higher in all four groups of STAT mice than in the control group (Fig. 1c). Per cent body fat also was increased in most STAT groups compared to controls (Fig. 1d). Lean weight was not significantly ($P = 0.24$) different in the STAT mice (15.0 ± 0.1 g (mean $\pm$ standard error)) compared to controls (15.4 ± 0.3 g)

[1]Department of Medicine, New York University School of Medicine, New York, New York 10016, USA. [2]Medical Service, VA New York Harbor Healthcare System, New York, New York 10010, USA. [3]Department of Microbiology, New York University School of Medicine, New York, New York 10016, USA. [4]J. Craig Venter Institute, Rockville, Maryland 20850, USA. [5]Department of Pathology, New York University School of Medicine, New York, New York 10016, USA. [6]Center for Health Informatics and Bioinformatics, New York University School of Medicine, New York, New York 10016, USA. [7]Department of Population Health, New York University School of Medicine, New York, New York 10016, USA.

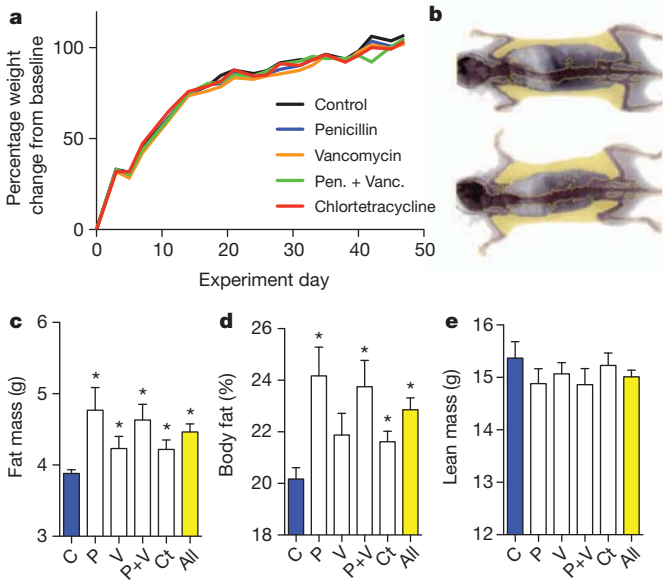


Figure 1 | Weight and body composition of control and STAT mice. **a**, Weight gain did not differ between control and STAT mice ($n = 10$ mice per group). **b**, Representative DEXA show per cent body fat in control (22.9%; top) and STAT (32.0%; bottom) mice. **c**, Total fat mass was significantly increased ($^*P < 0.05$) in all STAT groups compared to controls. **d**, Per cent body fat was significantly increased in all STAT groups (all $P < 0.05$) except vancomycin. **e**, Lean mass was lower in STAT mice, but not significantly different from controls. Data are presented as mean ± s.e.m. For all figures: all, all antibiotics; C, controls; Ct, chlortetracycline; P, penicillin; P+V, penicillin plus vancomycin; V, vancomycin.

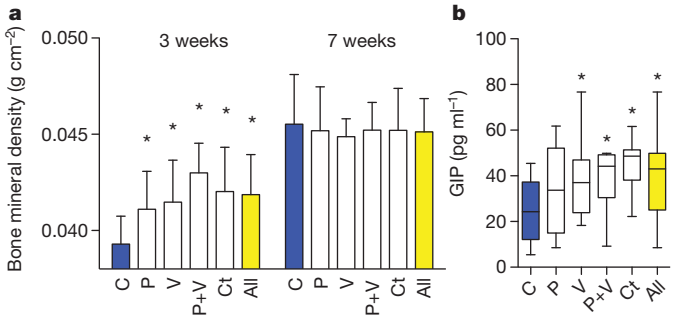


Figure 2 | Bone development and serum GIP measurements. **a**, After 3 weeks of STAT, bone mineral density was significantly increased in each group ($n = 10$ mice per group) compared to controls ($^*P < 0.05$) but did not persist at 7 weeks. **b**, Serum GIP levels measured at death were significantly increased in the vancomycin, penicillin plus vancomycin, and chlortetracycline groups and in the aggregate antibiotic group compared to controls ($P < 0.05$). Data are presented as mean ± s.e.m. Box plots show median ± interquartile range (IQR) and 95% ranges (whiskers).

(Fig. 1e). Thus, 7 weeks after starting the intervention, each of several STAT exposures changed body composition but not overall weight. Repeat STAT experiments showed similar results (Supplementary Fig. 2). There were no significant differences in calculated feed efficiency, expressed as weight gained per unit of feed consumed, between the STAT and control mice. In a larger, confirmatory experiment to assess when the morphometric changes appear, we began examining the mice immediately at weaning. Both the STAT males and females showed significantly increased early life growth rates (Supplementary Fig. 3A), and fat mass in the STAT animals began to diverge from controls by 16–20 weeks, in both males and females. Between 8 and 26 weeks, there were significantly increased rates of fat accumulation in both female and male STAT mice (0.042 and 0.045 g week^{-1}, respectively) compared to controls; female mice also showed significantly increased total mass (Supplementary Fig. 3B). These studies confirm the increased adiposity associated with STAT in females, show parallel effects in males, and indicate that the morphometric changes begin at the earliest time (day 22) of measurement.

Bone mineral density is increased in early-life growth

Bone mineral density (BMD) was evaluated by DEXA scanning in the control and STAT mice after 3 and 7 weeks of exposure. At 3 weeks, the BMDs in each of the five STAT groups (and overall) was significantly increased compared to controls (Fig. 2a). By 7 weeks, the BMDs in all mice increased, without significant differences between the STAT (0.045 ± 0.002 mg cm^{-2}) and control mice (0.046 ± 0.003 mg cm^{-2}) (Fig. 2a). Thus, an early bone developmental phenotype observed in each of the STAT groups normalized by 7 weeks. Parallel observations have been made in other STAT experiments (data not shown).

Increased GIP in STAT mice

To examine metabolic correlates to the changes in body composition, we assessed GIP, an incretin synthesized by small intestinal K cells[20] with receptors located on adipocytes that stimulates lipoprotein lipase activity[21]. GIP was significantly elevated in the STAT mice (39.1 ± 2.5 pg ml^{-1}) compared to controls (24.4 ± 4.2 pg ml^{-1}), with the levels ranging from 34.3 ± 6.1 pg ml^{-1} in the penicillin group to 44.9 ± 3.7 pg ml^{-1} in the chlortetracycline group (Fig. 2b). Increases in GIP levels are consistent with existing mouse models of obesity, such as in IRS-1/GIPR and Kir6.2/GIPR double knockout mice that show increased levels of GIP, adiposity and serum glucose[22,23]. GIP elevation in STAT mice provides a mechanism for the observed adiposity increase[24,25], but also could be secondary to the metabolic changes. There were no significant differences for fasting insulin-like growth factor (IGF)-I, insulin, peptide YY, leptin, or ghrelin levels between control and STAT mice (Supplementary Fig. 4). Glucose tolerance tests performed during week 6 of the experiment showed a trend towards hyperglycaemia in STAT mice (Supplementary Fig. 5).

STAT does not alter overall gut microbial census

To determine whether the STAT exposure leading to these metabolic changes affected the GI tract microbiome, microbial DNA extracted from faecal and caecal samples collected from the mice during the week of euthanasia or at necropsy, respectively, were studied. DNA concentrations measured from both caecal (77.62 ± 33.51 ng µl^{-1}) and faecal (23.79 ± 14.41 ng µl^{-1}) samples were not significantly different between control ($n = 10$) and STAT mice ($n = 10$ per group). The census in the STAT and control mice, determined through quantitative PCR using 338F/518R universal primers (Supplementary Table 1), showed no significant differences in bacterial counts or fungal census among the STAT and control groups (Fig. 3a). These data indicated that STAT exposure did not lead to substantial changes in the overall microbial census, and next led to us to conduct an assessment of the composition of the populations.

STAT alters the composition of intestinal microbiota

To assess microbial populations in the STAT and control microbiomes, we analysed the relative distribution of taxonomic groups based on 16S rRNA v3 region sequence data. The extracted DNA was subjected to 454 pyrosequencing, yielding 555,233 readable sequences (5,784 ± 676 sequences per sample with mean length 188 ± 3 bp). The sequences were analysed at multiple (phylum to genus) taxonomic levels (Supplementary Fig. 6 and Supplementary Table 2). In both faecal and caecal samples, the ratio of Firmicutes to Bacteria was significantly elevated in the STAT mice compared to controls (Fig. 3b and Supplementary Fig. 7). Weighted Unifrac analysis of the dominant taxa (present in >1% of the total population) showed a nonrandom clustering of STAT and control mice ($P < 0.05$) (Fig. 3c). Importantly, deep branching was identified, with the mean weight of mice on the two major branch points on the heat map being

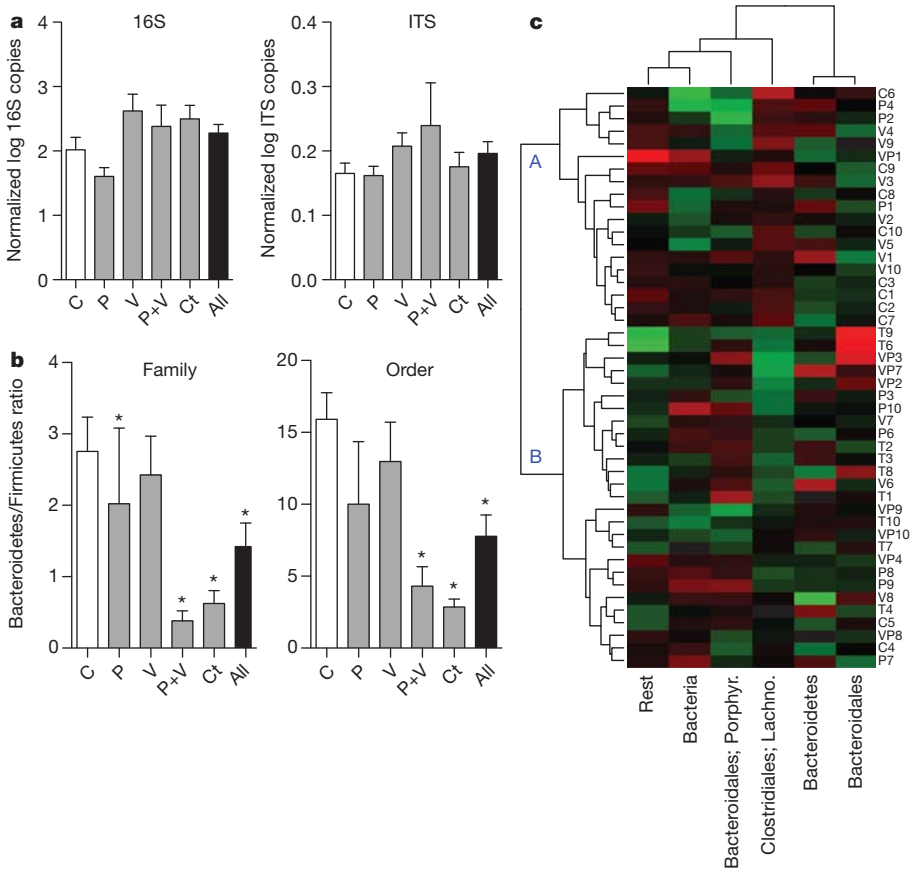


Figure 3 | Changes in the faecal gut microbiome after 50 days of STAT. **a**, There were no significant differences in microbial census between the STAT and control groups ($n = 10$ mice per group) evaluated by qPCR with universal primers for 16S rRNA and internal transcribed spacer (ITS). **b**, By 454-pyrosequencing, Firmicutes were shown to be increased in the STAT mice at multiple taxonomic levels. (Controls $n = 10$, penicillin $n = 9$, vancomycin $n = 10$, penicillin plus vancomycin $n = 8$, chlortetracycline $n = 10$; $^*P < 0.05$.) Data are presented as mean ± s.e.m. **c**, Heat map of specimens showing relative abundance of bacteria present at >1% at the family taxonomic level. Hierarchical clustering based on Euclidean distance identified nonrandom branch distributions of control and STAT mice ($P < 0.05$). Lachno., Lachnospiraceae; Porphyr., Porphyromonadaceae.

significantly ($P < 0.05$) different (21.4 ± 3.1% (point A) versus 23.0 ± 2.8% (point B); Fig. 3c). A major contributor to the observed differences is increased Lachnospiraceae representation in the STAT mice. Minor taxonomic groups may have roles in the development of metabolic phenotypes, and STAT-associated increases in several minor taxa are consistent with this possibility (Supplementary Table 3). Additional rarefaction curves and Unifrac analyses, quantile plots, PCoA representations, and heat maps generated using non-Euclidean distance metrics (Supplementary Figs 8–11 and http://www.med.nyu.edu/medicine/labs/blaserlab/PDFs/Cho-et-al-Online-Figures.pdf) demonstrated consistent shifts within the microbial populations in the STAT-exposed groups. Although STAT did not change the overall bacterial census, even the minimal antibiotic doses caused shifts in taxonomic composition, such as the Lachnospiraceae bloom. We noted an increase in the relative concentrations of Firmicutes compared to Bacteroidetes in the STAT mice compared to controls, which accompanied the observed increases in adiposity. This observation extends previous findings[26,27] of relative increases in the Firmicute population in *ob/ob* mice that are genetically prone to obesity. However, observations at such high taxonomic strata may not sufficiently describe the changes associated with obesity[28,29]; variations in DNA extraction efficiency and PCR-based sequencing of complex, heterogeneous microbial communities may bias census estimates of specific bacterial taxa. Furthermore, although the overall phenotypes (increased adiposity and hepatic lipogenesis) are consistent in the STAT groups, the intermediate steps may be more host- and treatment-specific. Our findings indicate that specific STAT exposures can be used as probes of microbiome structure and function.

STAT exposure alters gut microbiome SCFA metabolism

Because of the central role of short-chain fatty acid (SCFA) synthesis in colonic metabolism[30,31], we examined the effect of STAT exposure on the gene counts of prokaryotic genes butyryl coA transferase (BCoAT) and formyltetrahydrofolate synthetase (FTHFS) that are involved in butyrate and acetate synthesis, respectively (Supplementary Fig. 12). Quantitative PCRs (qPCRs) for total bacteria, and degenerate qPCRs for BCoAT and FTHFS, were performed on caecal specimens in control and STAT mice. At 3 weeks, there were significant decreases in BCoAT gene copy numbers in the penicillin plus vancomycin, chlortetracycline, and aggregate groups. By 6 weeks, BCoAT copy numbers had increased in all the groups compared to the 3-week values, but there was no longer an aggregate difference. For FTHFS overall, there were no significant differences between control and STAT mice overall at 3 or 6 weeks, although there was variation within the antibiotic groups (Fig. 4a). Exploring the inter-antibiotic differences further, we noted that there were several patterns in the FTHFS qPCR with different melting curve peaks (Supplementary Fig. 13), indicating differences in the microbial population. In total, these results provide evidence that STAT treatment is dynamically affecting composition of genes related to SCFAs, probably in antibiotic-specific ways, but with overall conserved effects. Both BCoAT and FTHFS have a role in metabolism of carbohydrates into SCFAs[32] and have been used to assess the functional characteristics of complex communities[33]; the observed changes in gene copy numbers in young STAT mice relative to controls as well as changes at different time points provide evidence that the STAT colonic microbiome alters SCFA metabolism. Our finding that the copy numbers of these

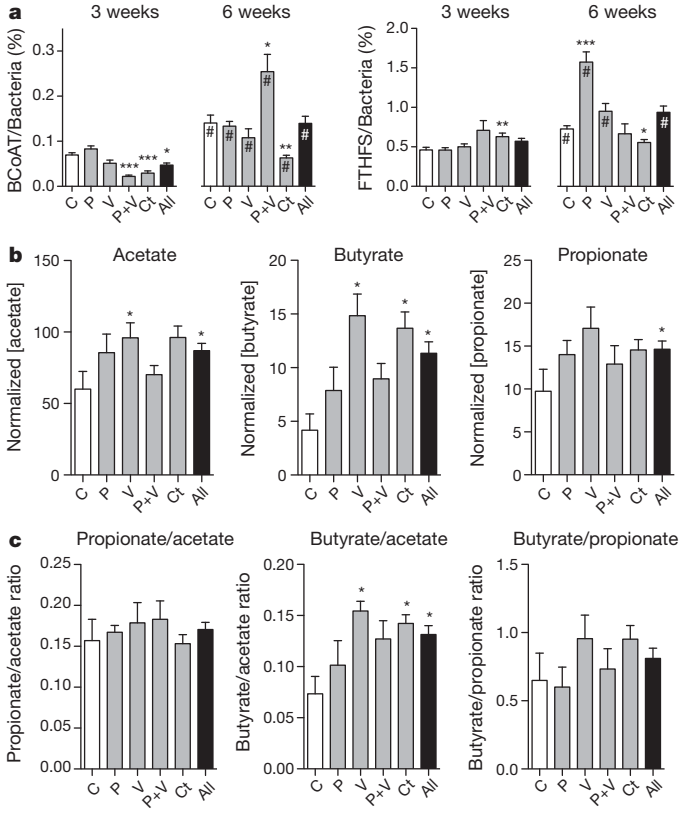


Figure 4 | Caecal SCFA production after STAT exposure. **a**, Quantitative PCR was performed for butyryl CoA transferase (BCoAT) and formyltetrahydrofolate synthetase (FTHFS) at experiment weeks 3 and 6 on STAT and control groups ($n = 10$ mice per group). At 3 weeks, BCoAT was diminished in two STAT groups and the aggregate group, a difference that persisted only in the chlortetracycline group. FTHFS copies do not show a consistent pattern. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ comparing STAT to controls; hash symbol indicates significant difference between 3 and 6 weeks. **b**, SCFA concentration analysed by gas chromatography (GC) shows increases in SCFAs in each of the STAT groups compared to controls. **c**, The ratio of butyrate relative to acetate is significantly higher in the STAT mice than controls. Data are presented as mean ± s.e.m.

two genes increase during growth (between 3 and 6 weeks) in both control and STAT mice (Fig. 4) suggests a greater dependence on SCFA synthesis pathways with maturation. Relative differences in the extent of these changes and in the melting curve peak populations may reflect changes in microbial populations. Furthermore, gene copy number (Fig. 4) in conjunction with melting curve peak patterns (Supplementary Fig. 13) may be useful for studying community genotypes of metabolic potential.

Direct measurements of SCFAs in the caecal contents of control and STAT mice demonstrate substantial increases in acetate, propionate and butyrate in all STAT groups (Fig. 4b); ratios of butyrate to acetate are also significantly altered by STAT exposure (Fig. 4c). These findings provide evidence that STAT exposure perturbs not only the composition of the GI microbiome but also the metabolic capabilities of the microbiome, specifically with respect to SCFAs. Increased SCFA concentrations and butyrate/acetate levels provide mechanisms for the STAT-induced adiposity phenotypes. SCFAs directly provide energy to colonocytes, and absorption into the portal circulation stimulates adipogenesis[30,34]. Metabolic cage experiments examining metabolic balance show no significant difference in caloric intake between control and STAT mice but lower caloric output in faecal pellets from STAT mice (Supplementary Fig. 14), providing evidence for selection for microbiota that can extract calories from otherwise indigestible constituents.

STAT alters hepatic metabolism of fatty acids and lipids

In confirmatory experiments using the identical STAT penicillin protocol, liver tissue was collected from both control and STAT mice. Standard histological analysis and triglyceride measurements showed no significant differences between control and STAT mice (Supplementary Fig. 15a, b). There were no significant differences in 16S rRNA gene counts, as detected by qPCR, from the liver specimens, providing evidence that bacterial translocation is not altered in the STAT mice (Supplementary Fig. 15c). Microarray analyses surveyed for differences in >45,000 genes, identifying 466 that by *t*-test or post translational modification statistics (PTM) were significantly up- or downregulated (397 significantly differed from controls by both tests) in the STAT compared to control mice. Focusing on pathways related to fatty acid metabolism and lipid metabolic processes, 22 and 47 genes, respectively, were differentially expressed between STAT and control mice (Fig. 5a). When specific genes were mapped to murine hepatic metabolic pathways, there were consistent changes in the same direction for several pathways, including upregulation in pathways related to lipogenesis and triglyceride synthesis (Fig. 5b). The changes in gene expression observed by microarray analyses were extended by qPCR assays of the same genes (Supplementary Fig. 16). These microarray and qPCR findings demonstrate substantial changes in the regulation of hepatic lipid, cholesterol and triglyceride metabolism that result from STAT-induced intraluminal intestinal changes.

Increased STAT adiposity is not metabolically altered

In the same confirmatory experiments (Supplementary Fig. 3), visceral adipose tissue dissected from control and STAT (penicillin) mice had no significant differences in adipocyte counts (Supplementary Fig. 17a, c) or in immunohistochemical staining for $CD68^{+}$ macrophages (Supplementary Fig. 17b, d). These findings were extended by *Gapdh*-normalized (Supplementary Fig. 17e) metabolic gene qPCR analyses. There were no significant differences in leptin, adiponectin, resistin, sterol regulatory element binding transcription factor 1 (*Srebf1*, also called *SREBP1c*), peroxisome proliferator activated receptor γ (*Pparg*, also called *PPARγ2*), and fatty acid synthase (*Fasn*, also called *FAS*) levels between the control and STAT groups (Supplementary Fig. 17f). These findings provide evidence that STAT adipose tissue shows no substantial physiological difference compared to controls in cell density, local inflammation, or metabolic potential as demonstrated by quantitative PCR. Increased adiposity seems to be a downstream phenomenon primarily mediated by changes in the gut and liver.

Discussion

By developing a model to assess adiposity, we show that each of the several STAT approaches tested affects the adiposity of post-weaning C57BL/6J mice. Similarly, there was a consistent early change in bone development. Particularly in the dynamic phases of growth in young animals, STAT alterations of the microbiome may affect pluripotent cells that can become osteoblasts, adipocytes, or myocytes. Although the STAT model does not precisely replicate the weight gain observed in farm animals, possibly due to differences in the husbandry of laboratory mice in a clean environment versus animals raised on a farm, the broad effects of exposure demonstrated in this model provide evidence that altering the microbiome may have substantial consequences. Such changes in early-life body composition may be due to altered host responses[35,36] and/or shifts in the metabolic characteristics of the gut microbiome. We postulate that the STAT exposures selected for microbiota with increased metabolic activity that were able to extract a higher proportion of calories from dietary complex carbohydrates that were relatively indigestible in the control mice. The increased SCFA concentrations are the metabolic products of this activity, which then may be delivered in increased quantities through the portal circulation to the liver, enabling enhanced lipogenesis (Supplementary Fig. 18). Enhanced caloric absorption has been implicated as a mechanism for increased weight gain in other murine obesity models[12]. The observed increase in

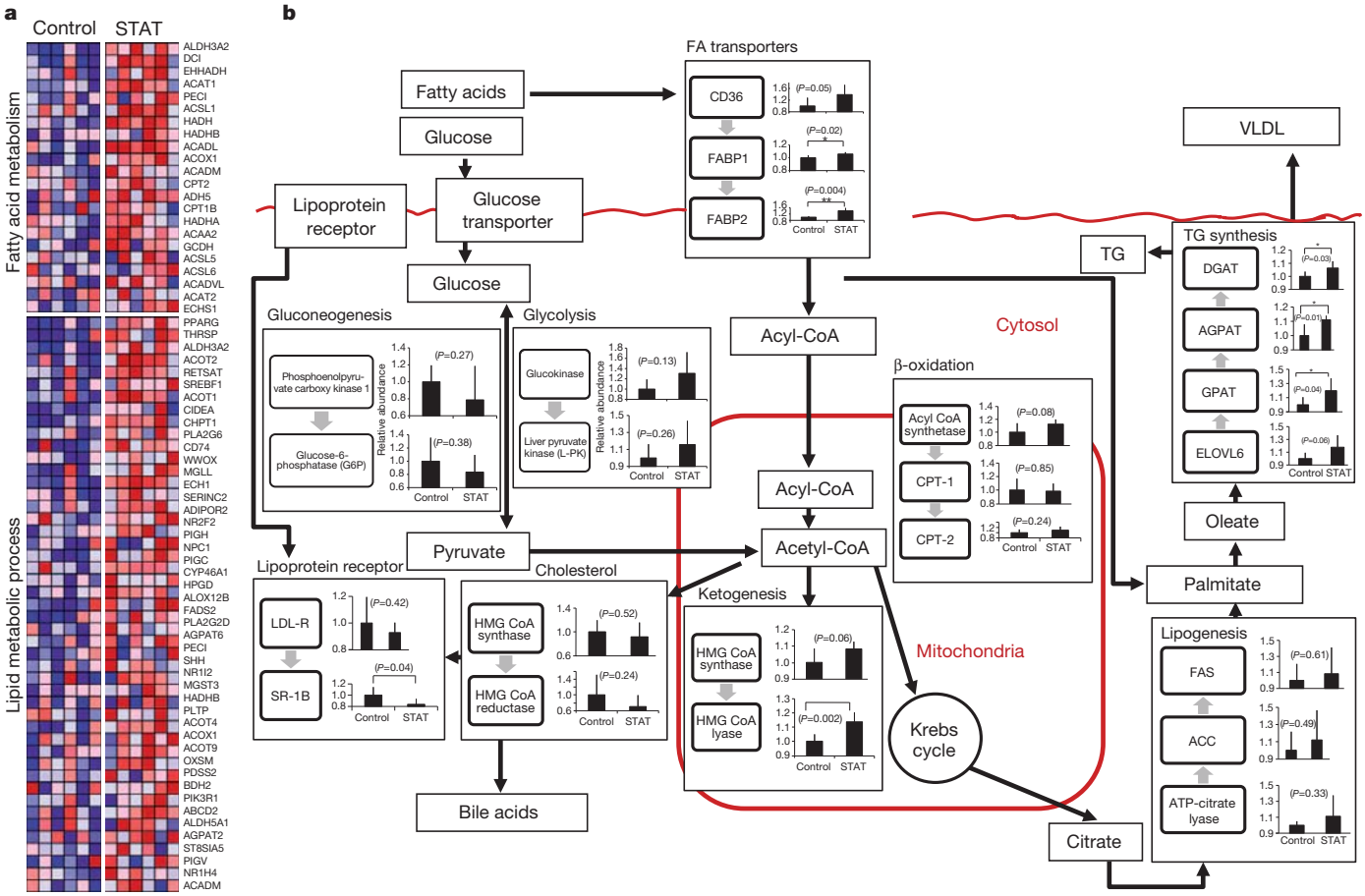


Figure 5 | Differentially regulated genes related to hepatic lipogenesis, identified through microarray and quantitative PCR analyses. a, Microarray analysis of liver specimens surveyed for differences in >45,000 genes; 397 genes were significantly up- or downregulated. Heat maps generated by gene set enrichment analysis (GSEA)[43] identify differences between the STAT and control mice ($n = 6$ mice per group), including pathways related to fatty acid metabolism and lipid metabolic processes. **b**, Mapping of metabolic genes detected by microarray onto specific pathways, including those related to lipogenesis and triglyceride synthesis, show consistent changes with STAT. Data are presented as mean ± s.e.m. White bars, controls; black bars, STAT; $*P < 0.05$, $**P < 0.01$. TG, triglycerides; VLDL, very low-density lipoproteins.

adiposity is similar across several different STAT exposures, despite the different antibiotics used. This is consistent with the observation that a wide range of FDA-approved antibacterial agents are used for effective growth promotion in the agricultural industry[6,7,37]. Our studies of the intestine, liver and adipose tissue demonstrate that the active effect of the antibiotics is on the microbiota affecting the downstream liver; mechanistically, the adipose tissue is passively accepting the increased lipid load produced from more proximal activities.

Why might STAT exposures have such substantial effects on recipient animals? We propose that STAT represents a compounded perturbation, as described previously[38], that has more serious consequences for long-term alterations of community state, possibly generating a different assemblage of species. The STAT-induced increase in SCFA production is a change in substrate availability that is a characteristic[39] of ecological disturbance. Such overharvesting, at a developmentally sensitive time, may push the ecosystem beyond the normative recovery that usually follows infrequent disturbances[38]. In complex, co-evolved ecosystems, the intricate interactions among participants conform to equilibria[40] that promote the robustness of a community[41]. However, when such equilibria are substantially perturbed, changes in host health and disease may result[42]. Other causative mechanisms for the changes in host phenotype could reflect that maturation of the gut microbiome is temporally altered by STAT exposure or that host responses are altered by microbiome changes. Ultimately, the interaction between the microbiome's metabolic capabilities and numerous host reactive cells could account for the observed changes, including the increase in GIP. In our study, we confirm in a tractable experimental model the decades-long observations in farm animals that STAT exposure changes host development, and we demonstrate specific techniques useful for studying the metabolic effects of microbiome manipulation. Our study also indicates the possibility that modulation of the infant human gut microbiome by antibiotics could have long-term metabolic consequences affecting adiposity and bone development.

METHODS SUMMARY

Female C57BL/6J mice were given penicillin, vancomycin, penicillin plus vancomycin, or chlortetracycline (1 μg antibiotic per g body weight) via drinking water, or no antibiotics (control). Body weight was serially measured and body composition determined using dual energy X-ray absorptiometry (DEXA). At death, blood, caecal contents, liver and visceral adipose tissue were collected, and serum hormones measured. DNA was extracted from caecal contents and faecal pellets, and 16S rRNA gene v3 regions were barcoded and sequenced, using 454-FLX Titanium chemistry. Quality-filtered sequences were processed through the QIIME pipeline and analysed in the R statistical environment. Quantitative PCR assessed taxa and metabolic genes of interest, and expression profiling of hepatic RNA was performed by microarray.

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

Received 1 April 2011; accepted 6 July 2012.
Published online 22 August 2012.

1. McCaig, L. F. & Hughes, J. M. Trends in antimicrobial drug prescribing among office-based physicians in the United States. *J. Am. Med. Assoc.* **273,** 214–219 (1995).

2. Kozyrskyj, A. L., Ernst, P. & Becker, A. B. Increased risk of childhood asthma from antibiotic use in early life. *Chest* **131,** 1753–1759 (2007).
3. Blaser, M. J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nature Rev. Microbiol.* **7,** 887–894 (2009).
4. Dethlefsen, L. & Relman, D. A. Microbes and Health Sackler Colloquium: Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108,** 4554–4561 (2010).
5. Manichanh, C. *et al.* Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res.* **20,** 1411–1419 (2010).
6. Butaye, P., Devriese, L. A. & Haesebrouck, F. Antimicrobial growth promoters used in animal feed: effects of less well known antibiotics on gram-positive bacteria. *Clin. Microbiol. Rev.* **16,** 175–188 (2003).
7. Ozawa, E. Studies on growth promotion by antibiotics. *J. Antibiot.* **8,** 205–214 (1955).
8. Abreu, M. T., Fukata, M. & Arditi, M. TLR signaling in the gut in health and disease. *J. Immunol.* **174,** 4453–4460 (2005).
9. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
10. Hansotia, T. & Drucker, D. J. GIP and GLP-1 as incretin hormones: lessons from single and double incretin receptor knockout mice. *Regul. Pept.* **128,** 125–134 (2005).
11. Gesta, S., Tseng, Y. H. & Kahn, C. R. Developmental origin of fat: tracking obesity to its source. *Cell* **131,** 242–256 (2007).
12. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
13. Reikvam, D. H. *et al.* Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS ONE* **6,** e17996 (2011).
14. Robinson, C. J. & Young, V. B. Antibiotic administration alters the community structure of the gastrointestinal micobiota. *Gut Microbes* **1,** 279–284 (2010).
15. Wlodarska, M. *et al.* Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated *Citrobacter rodentium*-induced colitis. *Infect. Immun.* **79,** 1536–1545 (2011).
16. McCracken, V. J., Simpson, J. M., Mackie, R. I. & Gaskins, H. R. Molecular ecological analysis of dietary and antibiotic-induced alterations of the mouse intestinal microbiota. *J. Nutr.* **131,** 1862–1870 (2001).
17. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
18. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Rev. Microbiol.* **9,** 279–290 (2011).
19. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
20. Buffa, R. *et al.* Identification of the intestinal cell storing gastric inhibitory peptide. *Histochemistry* **43,** 249–255 (1975).
21. Miyawaki, K. *et al.* Inhibition of gastric inhibitory polypeptide signaling prevents obesity. *Nature Med.* **8,** 738–742 (2002).
22. Tsukiyama, K. *et al.* Gastric inhibitory polypeptide is the major insulinotropic factor in K_{ATP} null mice. *Eur. J. Endocrinol.* **151,** 407–412 (2004).
23. Zhou, H. *et al.* Gastric inhibitory polypeptide modulates adiposity and fat oxidation under diminished insulin action. *Biochem. Biophys. Res. Commun.* **335,** 937–942 (2005).
24. Yip, R. G., Boylan, M. O., Kieffer, T. J. & Wolfe, M. M. Functional GIP receptors are present on adipocytes. *Endocrinology* **139,** 4004–4007 (1998).
25. Yamada, Y. & Seino, Y. Physiology of GIP—a lesson from GIP receptor knockout mice. *Horm. Metab. Res.* **36,** 771–774 (2004).
26. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
27. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3,** 213–223 (2008).
28. Murphy, E. F. *et al.* Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut* **59,** 1635–1642 (2010).
29. Fleissner, C. K. *et al.* Absence of intestinal microbiota does not protect mice from diet-induced obesity. *Br. J. Nutr.* **104,** 919–929 (2010).
30. Wong, J. M., de Souza, R., Kendall, C. W., Emam, A. & Jenkins, D. J. Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* **40,** 235–243 (2006).
31. Hong, Y. H. *et al.* Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology* **146,** 5092–5099 (2005).
32. Lovell, C. R. & Leaphart, A. B. Community-level analysis: key genes of CO_2-reductive acetogenesis. *Methods Enzymol.* **397,** 454–469 (2005).
33. Henderson, G., Naylor, G. E., Leahy, S. C. & Janssen, P. H. Presence of novel, potentially homoacetogenic bacteria in the rumen as determined by analysis of formyltetrahydrofolate synthetase sequences from ruminants. *Appl. Environ. Microbiol.* **76,** 2058–2066 (2010).
34. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70,** 567–590 (1990).
35. Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101,** 15718–15723 (2004).
36. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307,** 1915–1920 (2005).
37. Jukes, T. H. Antibiotics in animal feeds. *N. Engl. J. Med.* **282,** 49–50 (1970).
38. Paine, R. T., Tegner, M. J. & Johnson, E. A. Compounded perturbations yield ecological surprises. *Ecosystems* **1,** 535–545 (1998).
39. Pickett, S. T. & White, P. S. *The Ecology of Natural Disturbance and Patch Dynamics* (Academic, 1985).
40. Blaser, M. J. & Kirschner, D. The equilibria that allow bacterial persistence in human hosts. *Nature* **449,** 843–849 (2007).
41. Sole, R. V. & Montoya, J. M. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268,** 2039–2045 (2001).
42. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nature Rev. Genet.* **13,** 260–270 (2012).
43. 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).

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

Acknowledgements This work was supported in part with grants from the NIH (T-RO1-DK090989, 1UL1-RR029893, UL1-TR000038), the Diane Belfer Program in Human Microbial Ecology, the Philip and Janice Levin Foundation, the Michael Saperstein Fellowship, and institutional funds provided by the J. Craig Venter Institute, and the NYU Genome Technology Center. We thank N. Javitt for advice and J. Chung for technical assistance.

Author Contributions I.C. and M.J.B. designed the study; I.C., L.C., S.Y., Z.G., D.M., I.T. and K.R. performed experiments; B.A.M. and K.L. performed sequencing and sequencing analysis; J.Z. performed microarray analyses; I.C. and H.L. performed statistical interpretation and analyses; A.V.A. performed bioinformatics analyses and interpretation; I.C. and M.J.B. took primary responsibility for writing 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 the paper. Correspondence and requests for materials should be addressed to M.J.B. (martin.blaser@med.nyu.edu).

METHODS

Mouse husbandry. Female C57BL/6J mice were obtained at weaning (21 days of life) from Jackson Laboratories and allowed to adjust to the NYU animal facility for 1 week. Mice were weighed at the beginning of each experiment and distributed five per cage so that mean weights in each cage were equal. In each experiment, each study group (control or antibiotic(s)) was composed of ten mice. The mice were allowed ad libitum access to food and water and were maintained on a 12-h light/dark cycle and fed standard laboratory chow (Purina Mills International Diet no. 5001). The protocol was approved by the New York University School of Medicine Institutional Animal Care and Use Committee (IACUC).

Antibiotic dosing. Beginning on day 28 of life, mice were given standard water (pH 6.8) or water containing one of the following antibiotic regimens: penicillin VK, vancomycin, penicillin VK plus vancomycin, and chlortetracycline. Doses were adapted from the FDA Green Book and generally were in the mid-range of those approved for use in agriculture[6,7,37]. To simplify administration, the antibiotics were added to drinking water at a dose of 1 μg antibiotic per g body weight of the mice in the cage, based on the calculation that daily water intake is proportional to body weight. In previous studies, the daily water intake in C57BL/6J mice averaged 15 ml water per 100 g body weight[44]. Water containers were changed twice weekly to supply fresh antibiotics.

Mouse measurements and phenotypes. Three times a week, each mouse was weighed twice on an electronic scale tared between all measurements. The quantity of feed consumed per cage was measured and average daily feed intake calculated. Feed efficiency was calculated by dividing mean weight gain by mean daily feed consumed per mouse. Faecal pellets were collected every week and stored at −20 °C until processing. Body composition was determined using dual energy X-ray absorptiometry (DEXA) with a Lunar PIXImus II mouse densitometer (GE Medical Systems) at weeks 3 and 7. Each mouse was scanned individually and data collected on fat composition, lean body mass, per cent body fat, and bone mineral density. MRI experiments were performed on a 7 tesla MRI system gated on respiratory rate. During the scan, anaesthesia was maintained with 1.5% isoflurane in oxygen. The animal body temperature was maintained at 37 °C (±1 °C) during the scan by directing a thermostatically controlled warm air source. Images captured were separated into fat and lean tissue using the IDEAL Dixon method based on chemical shift properties[45] and total body fat percentage (volume/volume) was calculated by merging fat and lean images in MRIcron software (http://www.mccauslandcenter.sc.edu/mricro/mricron/). Weight values were validated by comparisons with scale weight and fat% values with MRI-determined fat% (Supplementary Fig. 1). Mice were killed by CO_2 inhalation and cervical dislocation and blood was collected via cardiac puncture. The blood was separated into serum and whole cells. Necropsy was performed on each mouse and caecal contents, liver, serum and visceral adipose tissue were collected. All samples were snap-frozen and stored at −80 °C until processing.

Statistical comparison of growth rates. From the experiment shown in Supplementary Fig. 4, the plots of group means of total, fat and lean mass over time were fitted with the following linear spline model with a common knot at week 26:

$$E(Y) = \beta_0 + \beta_1 \text{group} + \beta_2 \text{week} + \beta_3 (\text{week} - 26)_+ + \beta_4 \text{group} \times \text{week} + \beta_5 \text{group} \times (\text{week} - 26)_+$$

where group = 0 indicates the control group and group = 1 indicates the STAT group, and $(x)_+$ is defined as a function that equals x when x is positive and is equal to zero otherwise. With this model we performed the group comparisons at baseline, before week 26 and to examine patterns of change over time through tests H_0: $\beta_1 = 0$, H_0: $\beta_4 = 0$ and H_0: $\beta_4 = \beta_5 = 0$, respectively. The MIXED procedures of SAS software (version 9.2; SAS Institute, Inc.) were used to perform the tests and calculate the estimates.

Hormone measurements. Blood cells and serum were separated through centrifugation (3,000*g* for 10 min at 4 °C) and the serum frozen at −70 °C. Serum specimens (100 μl) were examined using the Millipore Mouse Gut Hormone Panel (Millipore Corp.) using a Luminex 200 (Millipore) analyser. The panel was custom-designed to assess for glucose-dependent insulinotropic peptide (GIP), insulin, IGF-1, peptide YY, ghrelin and leptin.

Glucose tolerance test. Control and STAT mice were fasted for 12 h during study week 6 (week 10 of life) and weights determined by electronic scale. Blood glucose at baseline was measured using an Accu-Chek Compact Plus glucometer (Roche). The mice then were challenged with 2 mg glucose per g body weight, given intraperitoneally, as previously described[46], and blood glucose levels determined 30, 60 and 120 min later. Area under curve (AUC) calculations were used to compare the STAT and control groups between 30 and 60 min.

DNA extraction. DNA extractions from caecal contents and faecal pellets were performed using the Qiagen QIAmp DNA stool extraction kit (Qiagen) per the manufacturer's protocol. Total extracted DNA was quantified using a Nanodrop 1000 (Thermo Scientific). PCR to confirm bacterial DNA extractions was performed using the 8F/1510R bacterial primers for 16S rRNA[47,48].

Quantitative PCR amplification. After genomic DNA extraction and quantification, samples were prepared for amplification and sequencing at the JCVI Joint Technology Center (JTC). Genomic DNA sample concentrations were normalized to ~5–10 ng μl^{-1}. The v3 region of the 16S rRNA gene was amplified using forward primer 341F (5′-CCTACGGGAGGCAGCAG-3′) and reverse primer 534R (5′-ATTACCGCGGCTGCTGG-3′) to which a 'CG' on the 5′ end of the primer followed by symmetrical 10-nucleotide barcodes were included as part of the primer design (5′- CGN(10)+16S primer-3′). Barcoded primer designs were completed using a set of algorithms developed at the JCVI. This design allowed for the inclusion of a unique barcode to each sample at the time of PCR so that the tagged samples could be multiplexed for sequencing. A total of 96 barcodes were used. Every effort was made to prevent contamination of PCR reactions with exogenous DNA including a set of reactions in a laminar flow hood. PCR reactions were completed as follows (per reaction): 1 μl of gDNA, 1× final concentration of Accuprime PCR Buffer II (Invitrogen), 250 mM betaine, 200 nM forward and reverse primers, 0.5 units of Accuprime Taq DNA Polymerase High Fidelity (Invitrogen), and nuclease-free water to bring the final volume to 10 μl. PCR cycling conditions were: initial denaturation of 2 min at 95 °C followed by 25 cycles of 10 s at 95 °C, 20 s at 63 °C and 30 s at 72 °C. A negative control (water blank) reaction also was included and examined after 30 cycles. PCR reactions were visualized on 1% agarose gels and quantified using a Tecan SpectraFluor Plus (Tecan Group Ltd) before normalization and pooling of samples for sequencing.

454-FLX read quality filtering. After the completion of sequencing, a read processing pipeline involving a set of modular scripts designed at the JCVI was used for deconvolution, trimming and quality filtering[49]. First reads were deconvoluted, or assigned to samples, based on their unique 10-nucleotide barcode, allowing ≤1-nucleotide mismatch to the reference barcode sequence and ≤6 primer sequence mismatches, using the EMBOSS program fuzznuc to locate barcodes and primer sequences on each read. To eliminate spurious hits, the distance between detected barcode and primer were required to be ≤3 nucleotides. Barcodes then were trimmed from the sequence and remaining sequences which failed to have barcodes detected then were filtered. After deconvolution, primer sequences detected at >75% identity were trimmed and those without valid primers detected were eliminated. Reads with an average length of <100 nucleotides or with any Ns were removed. A BLASTN quality check[50] was performed against a small internal data set of 200 16S reads to remove sample reads substantially inconsistent with 16S gene sequences. The criteria were that >30% of the query must be covered by the alignment with a minimum of 60 nucleotides of identity. Passing reads were subsequently examined using a modified version of the RDP Chimera Check in which candidate chimaeras were identified based only on reads that were sequenced within the sample. Final sequence data have been deposited in the NIH Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). Accession numbers for all primary sequencing data are available from the NCBI under BioProject 168618.

Sequence analyses. The pooled sample was cleaned using the Agencourt AMPure system (Beckman Coulter Genomics). The A and B adapters necessary for sequencing with 454-FLX Titanium chemistry (Roche) were subsequently ligated to the sample pool after PCR following standard manufacturer protocols. Library construction, emPCR, enrichment, and 454 sequencing were performed by following the vendor's standard operating procedures with some modifications. Specifically, quantitative PCR was used to accurately estimate the number of molecules needed for emPCR. We also used automation (BioMek FX) to 'break' the emulsions after emPCR and we used butanol to enable easier sample handling during the breaking process.

Statistical and computational sequence analysis. Quality-filtered sequences (see Supplementary Methods) were pre-processed through the QIIME pipeline[51] involving: (1) clustering the sequences into operational taxonomical units (OTUs) using the UCLUST program[52] at a 97% similarity threshold; (2) taxonomically assigning each OTU by selecting a representative sequence from each OTU-cluster and running RDP Classifier[53] with 80% bootstrap confidence; (3) aligning representative sequences for each OTU with the Greengenes core-set alignment template using PyNAST[54]; (4) building a phylogenetic tree for the OTUs using the FASTTREE program[55]; and (5) calculating weighted Unifrac beta-diversity indices[56]. We extracted the OTU absolute abundance table and weighted Unifrac beta-diversity matrix[56,57] from the pipeline for further analysis in the R statistical programming environment[58,59]. The rarefactions for diversity indices and species richness were calculated in the R statistical programming environment[59,60] using Community Ecology Package Vegan. The OTU absolute abundances were converted to relative abundances by normalizing to total

sequence count per sample analysed. The resulting relative abundance matrix was used to produce heat maps and major taxa bar plots. (Code available on the NYU Center for Health Informatics and Bioinformatics website http://www.nyuinformatics.org/research/labs/microbiomics.) All data are presented as mean ± s.e.m., unless otherwise indicated for groups without normal data distribution (median ± IQR). Comparisons of medians between non-normally distributed groups were performed using the Mann–Whitney *U*-test or Kruskal–Wallis analysis of variance test for simultaneous comparisons of more than two groups. Multivariate analysis of taxonomic distributions among groups was performed using double-principal coordinate analysis (DPCoA), as described[61,62]. To compare the inter- and intragroup statistical differences in beta diversity, we used a two-sided *t*-test. We ensured that possible violations of *t*-test assumptions when comparing beta-diversity did not affect statistical decisions (Supplementary Tables 4 and 5). *P* values <0.05 were considered to be significant.

Comparison of intra- and intergroup weighted Unifrac distances. Statistical comparison of intra- and intergroup weighted Unifrac distances requires application of a statistical test appropriate for the data. In our analysis, we apply a series univariate two-sample *t*-tests corrected for multiple comparisons for this purpose. However, application of the *t*-test requires that certain assumptions are satisfied, of which normality is the most notable. We inspected normal quantile plots to determine the extent of the linear trend, which provides evidence for the similarity of the empirical distribution to normal. We formally supplement this visual diagnostic with normality testing. Normality testing for the purpose of satisfying test assumptions is difficult in statistical analyses because normality can only be rejected and never confirmed; thus, it usually is not done in typical analyses of this sort. However, we performed the Shapiro–Wilk's test for normality on inter- and intragroup weighted UniFrac distances. Although in many cases normality could be rejected depending on the desired alpha level (Supplementary Table 4), we were able to identify long regions of linearity in each case to establish that the peak of the distributions is approximately normally distributed. For final assurance that decisions based on *t*-test are not affected by possible violations of the normality assumption, we also performed Mann–Whitney tests which (after adjustment for multiple comparison adjustment) provided the same rejections of the null hypothesis as for the original *t*-tests (Supplementary Table 5).

Quantitative PCR. Quantitative PCR assays to assess for taxa of interest were performed on a Rotor Gene 3000 quantitative PCR cycler using the LightCycler FastStart DNA Master PLUS SYBR Green I kit (Roche) according to the manufacturer's instructions. Quantitative assays to assess total microbial census were performed for Bacteria using 16S rRNA and Fungi using ITS sequences. Quantitative assays were also performed for butyryl CoA transferase (BCoAT) and formyltetrahydrofolate synthetase (FTHFS), using degenerate primers. All primer sequences are provided in Supplementary Table 1. Values were normalized based on total Bacteria 16S copies.

Hepatic triglyceride measurements. For triglyceride measurements, hepatic tissue was homogenized at 4 °C in RIPA lysis buffer (Sigma-Aldrich) and lipids extracted using a chloroform/methanol (2:1) method, evaporated, and dissolved in 2-propanol[63]. Triglyceride concentration was assayed using the enzymatic hydrolysis of triglycerides by lipase to glycerol and free fatty acids and quantification by colorimetric assay (Sigma-Aldrich).

Hepatic gene expression profiling and data analysis. Total DNA and total RNA was extracted from the mouse liver by the Wizard Genomic DNA purification Kit (Promega) and RNeasy Mini Kit (Qiagen), respectively, according to the manufacturer's instructions. Total RNA was reverse transcribed to cDNA using the Verso cDNA Synthesis Kit (Thermo Scientific). To generate standards for each target gene expression analysis, the DNA or cDNA region of interest was PCR-amplified and the PCR product was cloned using the pGEM-Teasy vector system (Promega). qPCR was performed with a LightCycler 480 SYBR Green I Master (Roche) and run in a LightCycler 480 system (Roche). Target mRNA was normalized to GAPDH mRNA as an internal control in each sample.

Expression profiling of the STAT and control animal groups ($n = 6$ each) was performed using the Affymetrix Genechip system (Affymetrix). Total RNA quality and quantity were determined using the Agilent 2100 Bioanalyser and Nanodrop ND-1000. Total RNA (100 ng) was used to prepare cDNA following the Affymetrix 3′IVT Express Kit labelling protocol (Affymetrix). Standardized array processing procedures recommended by Affymetrix were performed, including hybridization, fluidics processing and scanning of the Affymetrix MG-430 2.0 arrays. GeneSpring GX11 software (Agilent Technologies) was used to normalize the raw data (Affymetrix CEL files) by Robust Multichip Average algorithm (RMA). Gene set enrichment analysis (GSEA)[43] was used to identify significantly enriched gene expression patterns underlying fatty acid and lipid metabolism, by querying the C2 (curated pathways) and C5 (Gene Ontologies) categories of the GSEA MolSig v.3 database. The microarray data have been submitted in the MIAME-compliant format to the NCBI GEO public database (http://www.ncbi.nlm.nih.gov/geo/) under the identifier GSE38880.

Histology and immunohistochemistry. Fat tissues of mice were collected and frozen at −80 °C before being thawed and embedded in optimum cutting temperature (OCT) medium. Sections were cut in 5 μm thickness for staining with haematoxylin and eosin. Immunohistochemistry was performed on frozen formalin-fixed tissues using rat anti-mouse CD68 (AbDSerotec). In brief, sections were thawed at room temperature for 30 min, fixed in 10% NBF for 15 min, rinsed in distilled water, and soaked in reaction buffer for 15 min. Antibody incubation and detection were carried out at 37 °C on a NeXes instrument (Ventana Medical Systems) using Ventana's reagent buffer, endogenous biotin blocking kit and detection kits unless otherwise noted. Endogenous peroxidase activity was blocked with hydrogen peroxide. Antibodies against CD68 were diluted 1:75 and were incubated for 30 min. CD68 was detected using a biotinylated rabbit anti-rat antibody (Vector Laboratories) diluted 1:4,000 and incubated for 30 min. After secondary antibody application, streptavidin-horseradish-peroxidase conjugate was applied. The complex was visualized with 3,3-diaminobenzidene and enhanced with copper sulphate. Slides were washed in distilled water, counterstained with haematoxylin, dehydrated and mounted with permanent media. Appropriate positive and negative controls were included with the sections studied.

Histological and immunohistochemical analyses. Histology and immunohistochemistry slides were digitally scanned at ×40. From each section (STAT, $n = 10$; control, $n = 10$), 10 high-power fields (HPF) were selected by a single observer. Copies of each of the images were provided to two independent individuals blind to the experimental design who counted adipocytes or $CD68^+$ macrophages in each HPF using ImageJ (http://rsbweb.nih.gov/ij/). The counts were compared between the two investigators and the mean number of adipocytes or $CD68^+$ macrophages per HPF for control and STAT mice were compared using the Student's *t*-test.

Gas chromatographic analysis. Mouse faecal pellets were collected at week 6 of the experiment and frozen until analysed. Single pellets were weighed and homogenized in 100 μl of deionized water for 3 min. The pH of the suspension was adjusted to 2–3 by adding 5 M HCl at room temperature for 10 min with intermittent shaking. The suspension was transferred into a polypropylene tube and centrifuged for 20 min at 3,000g, yielding a clear supernatant. The internal standard, 2-ethylbutyric acid (TEBA), was added into the supernatant at a final concentration of 1 mM. Chromatographic analysis used the Shimadzu QP-500 GC/MS system (Shimadzu). A fused-silica capillary column (30m, 0.52mm, 0.50μm) with a free fatty acid phase (DB-FFAP 125-3237, J&W Scientific, Agilent Technologies Inc.) was used for analysis. Helium was the carrier at a flow rate of 14.4 ml min^{-1}. The initial oven temperature (100 °C) was maintained for 30 s, raised to 180 °C at 8 °C min^{-1} and held for 60 s, then increased to 200 °C at 20 °C min^{-1} and held for 5 min. The flame ionization detector and injection port were kept at 240 and 200 °C, respectively. The flow rates of hydrogen, air, and nitrogen were 30, 300 and 20 ml min^{-1}, respectively. The injected sample volume for GC analysis was 1 μl, and each analysis had a run time of 32 min[64].

Metabolic cage measurements. C57BL/6J mice were singly housed in Techniplast Type 304 metabolic cages (Techniplast). Mice had unrestricted access to 45% HFD powdered chow and STAT or control water, as appropriate. There was a 2-day acclimatization period, followed by 3 days of measurements. Measurements recorded were body weight, food and water consumed, and faeces and urine produced. Total faeces and urine were collected for analysis. Calories consumed were calculated by weight of food consumed and 4,057 cal g^{-1} as per the manufacturer's instructions (Research Diets). Calories excreted were calculated based on total faeces produced and calorimetric analysis of a homogenate pellet incorporating all faeces produced per mouse from each 24-h period.

Bomb calorimetry. Calorimetric measurements were obtained using a Parr 6725 Semi-micro calorimeter. Faecal pellets obtained during week 6 of the experiment were dehydrated overnight at 56 °C with a silica gel desiccant. All pellets studied had minimum dry weights >12 mg. The calorimeter was calibrated using multiple benzoic acid standards intermixed with the experimental measurements in each day's run. Corrections were made for sulphur content and fuse length; final results were expressed as calories per gram dry weight.

44. Harkness, J. E. & Wagner, J. E. *The Biology and Medicine of Rabbits and Rodents* 3rd edn (Lea and Febiger, 1989).
45. Reeder, S. B. *et al*. Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): application with fast spin-echo imaging. *Magn. Reson. Med.* **54,** 636–644 (2005).
46. Andrikopoulos, S., Blair, A. R., Deluca, N., Fam, B. C. & Proietto, J. Evaluating the glucose tolerance test in mice. *Am. J. Physiol. Endocrinol. Metab.* **295,** E1323–E1332 (2008).
47. Favier, C. F., Vaughan, E. E., De Vos, W. M. & Akkermans, A. D. L. Molecular monitoring of succession of bacterial communities in human neonates. *Appl. Environ. Microbiol.* **68,** 219–226 (2002).

48. Martínez-Murcia, A. J., Acinas, S. G. & Rodriguez-Valera, F. Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. *FEMS Microbiol. Ecol.* **17,** 247–255 (1995).
49. Li, K. *et al.* ANDES: Statistical tools for the ANalyses of DEep Sequencing. *BMC Res. Notes* **3,** 199 (2010).
50. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
51. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
52. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
53. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
54. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26,** 266–267 (2010).
55. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).
56. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* **5,** 169–172 (2010).
57. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
58. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
59. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comp. Graph.* **5,** 299–314 (1996).
60. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
61. Pavoine, S. & Bailly, X. New analysis for consistency among markers in the study of genetic diversity: development and application to the description of bacterial diversity. *BMC Evol. Biol.* **7,** 156 (2007).
62. Pavoine, S., Dufour, A. B. & Chessel, D. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *J. Theor. Biol.* **228,** 523–537 (2004).
63. Hong, F. *et al.* Interleukin 6 alleviates hepatic steatosis and ischemia/reperfusion injury in mice with fatty liver disease. *Hepatology* **40,** 933–941 (2004).
64. Zhao, G., Nyman, M. & Jonsson, J. A. Rapid determination of short-chain fatty acids in colonic contents and faeces of humans and rats by acidified water-extraction and direct-injection gas chromatography. *Biomed. Chromatogr.* **20,** 674–682 (2006).

LETTER

doi:10.1038/nature11408

Graphene and boron nitride lateral heterostructures for atomically thin circuitry

Mark P. Levendorf[1]*, Cheol-Joo Kim[1]*, Lola Brown[1], Pinshane Y. Huang[2], Robin W. Havener[2], David A. Muller[2,3] & Jiwoong Park[1,3]

Precise spatial control over the electrical properties of thin films is the key capability enabling the production of modern integrated circuitry. Although recent advances in chemical vapour deposition methods have enabled the large-scale production of both intrinsic and doped graphene[1–6], as well as hexagonal boron nitride (*h*-BN)[7–10], controlled fabrication of lateral heterostructures in these truly atomically thin systems has not been achieved. Graphene/*h*-BN interfaces are of particular interest, because it is known that areas of different atomic compositions may coexist within continuous atomically thin films[5,10] and that, with proper control, the bandgap and magnetic properties can be precisely engineered[11–13]. However, previously reported approaches for controlling these interfaces have fundamental limitations and cannot be easily integrated with conventional lithography[14–16]. Here we report a versatile and scalable process, which we call 'patterned regrowth', that allows for the spatially controlled synthesis of lateral junctions between electrically conductive graphene and insulating *h*-BN, as well as between intrinsic and substitutionally doped graphene. We demonstrate that the resulting films form mechanically continuous sheets across these heterojunctions. Conductance measurements confirm laterally insulating behaviour for *h*-BN regions, while the electrical behaviour of both doped and undoped graphene sheets maintain excellent properties, with low sheet resistances and high carrier mobilities. Our results represent an important step towards developing atomically thin integrated circuitry and enable the fabrication of electrically isolated active and passive elements embedded in continuous, one-atom-thick sheets, which could be manipulated and stacked to form complex devices at the ultimate thickness limit.

Figure 1a summarizes our approach for the patterned regrowth and characterization. After growing the first film of graphene (G^1, superscript representing the growth order), we then deposit a protective layer of photoresist and pattern away unwanted areas. Next we grow a second layer of graphene or *h*-BN (G^2/*h*-BN^2) and transfer the hybrid film onto the desired substrate for further characterization. All of our growths are performed on Cu foil using standard literature recipes for both graphene and *h*-BN growths[4,5,8]. Introducing dopant gases during any graphene growth stage allows us to produce both intrinsic graphene (*i*-G; H_2 + CH_4) and *n*-doped graphene (*n*-G; $NH_3 + H_2 + CH_4$) in a single process run (see Supplementary Information). Figure 1b shows optical images of the Cu growth substrate at different steps of the process. After patterning the first graphene layer, the substrate was heated to 135 °C in air to enhance the contrast, as is shown in Fig. 1b (left panel); the areas of Cu protected by *i*-G^1 remain unoxidized whereas the unprotected areas do not[17]. Figure 1b (right panel) shows the same foil after reduction of CuO_x and subsequent synthesis of the second graphene, and demonstrates the homogeneity of the Cu/graphene substrate. Before further characterization, the presence of graphene is confirmed using micro-Raman spectroscopy (see Supplementary Information).

The fidelity of the transferred pattern and the quality of the junctions formed by our method were studied using dark-field transmission electron microscopy (DF-TEM)[18,19], where use of a specific objective aperture filter allows imaging of areas with corresponding lattice orientations. Multiple such images can then be coloured and overlaid to form a complete map of the film, resolving the grain structure and number of layers with nanometre-scale resolution, both near to and away from the junction area. Figure 1c shows a composite image of a graphene sample that includes both the first and second grown areas as well as the junction between the two. For this, growths of *i*-G^1/*i*-G^2 were transferred onto 10-nm-thick Si_3N_4 TEM membranes. Both G^1 and G^2 regions (see inset to Fig. 1c) comprise single crystals of similar sizes, which indicates that the polycrystalline structure of the graphene is determined mainly by the synthesis conditions rather than the growth order, and also that the structure of the first grown graphene is unaffected by our patterned regrowth. Significantly, the location of the junction closely follows the designed pattern within the resolution of our instrument (partially outlined area in Fig. 1c; see Supplementary Information).

Our DF-TEM data and electrical characterization, in addition to the growth conditions, suggest that the junction between G^1 and G^2 areas are laterally connected. Our recent work[20] has shown that a more-reactive growth environment produces graphene–graphene grain boundaries with high-quality lateral connections and minimal inter-domain electrical resistance, whereas a less-reactive growth environment results in poorly connected or even overlapped junctions. In our patterned regrowth process, we therefore use highly reactive growth conditions. DF-TEM images of the junction areas indeed show high-quality, continuous growth of graphene between G^1 and G^2 areas (Fig. 1d), indicating that the crystallinity is maintained uniformly across this region without amorphous carbon or voids, within the spatial resolution of DF-TEM (around 10 nm). Furthermore, the electrical characterization across these junctions shows a low junction resistance, similar to that of the high-quality, laterally connected graphene grain boundaries reported in ref. 20.

We also find a highly reactive environment promotes uniform lateral connectivity through increased nucleation at the junction area. In Fig. 1e, we plot grain size as a function of position within the box indicated in Fig. 1c. While both G^1 and G^2 regions show average grain sizes of about 1.7 μm, near the junction they are much smaller (the mean is approximately 0.33 μm). The moving average (Fig. 1e, red line) shows that the width of this area with smaller grains is approximately 2 μm, comparable to the average G^1 and G^2 grain size. This, in addition to the increased density of small bilayer areas (bright spots in Fig. 1c and d), suggests that the junction between G^1 and G^2 is formed by graphene that has nucleated and grown in the junction area rather than by direct stitching between grains nucleated away from the junction area. This suggests edge-nucleated growth, which is similar to a "seeding" effect reported previously[21–23]. In contrast, regrowth under less-reactive conditions that provide slower growth do not produce well-connected junctions in either graphene/graphene and graphene/*h*-BN films (see Supplementary Information).

[1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. [2]School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA. [3]Kavli Institute at Cornell for Nanoscale Science, Ithaca, New York 14853, USA.
*These authors contributed equally to this work.

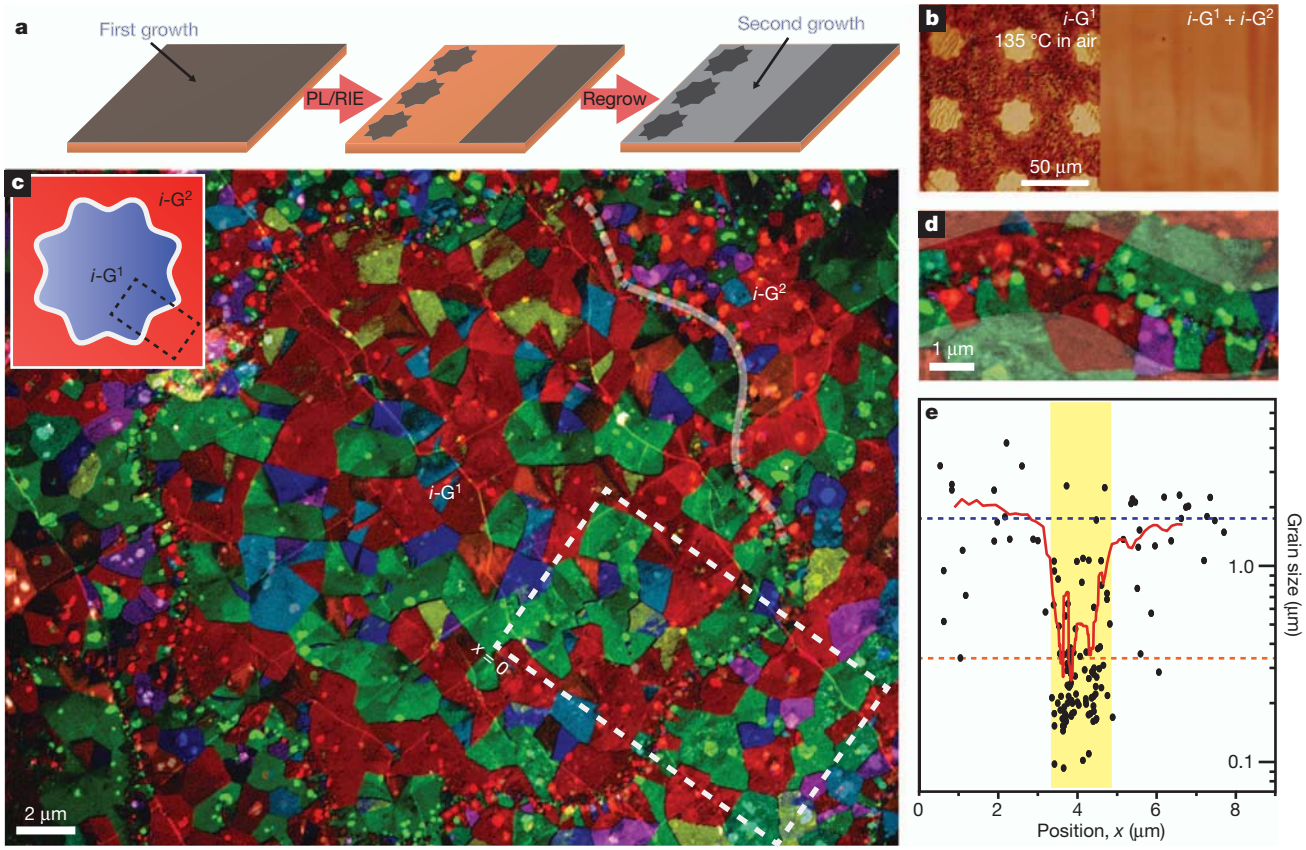


Figure 1 | Process schematic and DF-TEM characterization of graphene heterostructures. **a**, Schematic for formation of atomically thin lateral heterojunctions using photolithography (PL) and reactive ion etching (RIE). See main text for details. **b**, The left panel shows an optical image of a patterned Cu/G[1] foil oxidized to enhance contrast (the dark areas are Cu). The right panel shows an optical image after reduction of CuO_x and subsequent growth of intrinsic-G[2] (*i*-G[2]). **c**, False-colour DF-TEM image of an *i*-G[1]/*i*-G[2] patterned area (schematic in inset). **d**, Zoomed-in image of the junction region. **e**, Plot of grain size versus position in the box outlined in **c**. Dotted lines indicate average grain size away from (blue line) and near (highlighted region and orange line) the junction. Shading indicates the width of the region with smaller grains.

The technique described above is also applicable to the formation of insulator–metal lateral junctions using graphene and *h*-BN (Fig. 2), which we now discuss (see Methods). Figure 2a shows a false-colour DF-TEM image of a single layer *h*-BN sheet—as confirmed by Raman spectroscopy and X-ray photoelectron spectroscopy (see Supplementary Information)—with relatively large domains in excess of 1 μm. Electron diffraction also shows that the film consists of single crystals with hexagonal lattice structures (Fig. 2a, inset). This *h*-BN growth, however, yields films that are mechanically discontinuous, as indicated by the dark lines in Fig. 2a (arrows; also visible in bright-field TEM, see Supplementary Information). Instead, *h*-BN grown under more-reactive conditions (higher precursor flow rate, see Methods) results in a continuous film, while exhibiting all the known characteristics of *h*-BN sheets. This is confirmed by the X-ray photoelectron spectroscopy data in Fig. 2c (upper panels; B and N in a 1:1.08 atomic ratio) and Raman spectroscopy (lower panel; *h*-BN peak).

Careful control over the length of the *h*-BN[2] growth allows for the synthesis of *i*-G[1]/*h*-BN[2] sheets with high pattern fidelity, as shown by the optical image in Fig. 2b. The darker regions indicate areas of *i*-G[1] (stronger absorption), whereas the lighter regions are comprised of *h*-BN[2]. This, as well as the two-dimensional Raman image (inset in Fig. 2b) of the graphene 2D band intensity, confirms the successful pattern transfer to the *i*-G[1]/*h*-BN[2] hybrid sheet. Furthermore, DF-TEM of suspended films confirms a sharp junction, as is shown in Fig. 2d, where the grains of the *i*-G[1] growth (lower, coloured half) end abruptly at the interface with the *h*-BN[2] (upper half). Both the bright-field TEM (Supplementary Information) and scanning electron microscopy (SEM) images (Fig. 2e) show a mechanically continuous sheet that is cleanly suspended with no breaks or tears at the junction region, confirming the stability and integrity of these growths. We have observed similar mechanical continuity for both *i*-G[1]/*i*-G[2] and *n*-G[1]/*i*-G[2] growths (see Supplementary Information).

The composition of G/*h*-BN heterojunctions at the nanometre scale was investigated by cross-sectional imaging and chemical mapping of the *i*-G[1]/*h*-BN[2] interface using an aberration-corrected scanning transmission electron microscope (STEM) and electron energy loss spectroscopy (EELS). For this, using a focused ion beam, a thin ($\sim 30 \pm 5$ nm) slice containing a junction region is carved out from a film of *i*-G[1]/*h*-BN[2] transferred onto a Si/SiO_2 substrate (see Fig. 2f). Figure 2g (upper panel) shows the resultant EELS composition maps across a junction region, where a line heterojunction is clearly seen between two neighbouring regions with homogeneous atomic concentration. The left side indicates a high level of graphitic carbon (red; *g*-C), suggesting it belongs to graphene, whereas the other side has a high concentration of boron (green) originating from *h*-BN. Significantly, an increase in intensity of boron corresponds to a decrease in *g*-C, as shown by the line profiles of the concentration of boron and *g*-C in the junction region (Fig. 2g, lower), strongly suggesting that the G/*h*-BN junction is a lateral heterojunction with a compositional transition width of less than 10 nm (see Supplementary Information).

Successful synthesis of these hybrid films allows us to fabricate electrically isolated graphene devices in a single, atomically flat sheet, which we show in Fig. 3a. We observe conducting behaviour confined to the patterned graphene areas, with the *h*-BN[2] showing no conductivity within the limits of our equipment (sheet resistance $R_{sheet} > 400\ T\Omega\ \square^{-1}$, where $\square$ indicates a square of the material having no thickness). We also confirm that our *h*-BN is free of small

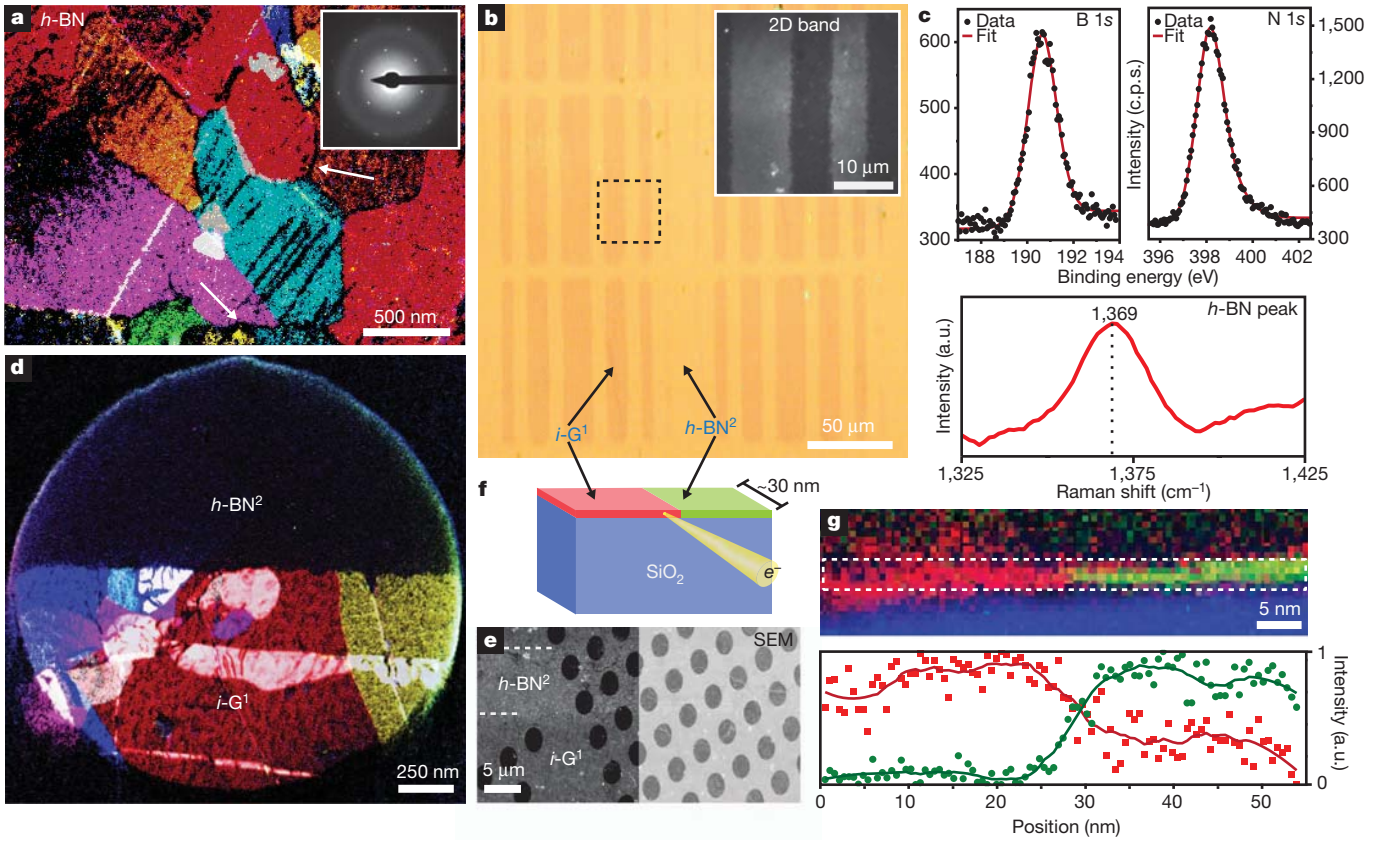


Figure 2 | *h*-BN/G heterostructure synthesis and structural characterization. **a**, False-colour DF-TEM image of an *h*-BN sheet grown in an environment of low reactivity with domains >1 μm. Arrows indicate regions where the *h*-BN film failed to connect. Inset, representative diffraction image taken from the *h*-BN sheet, indicating hexagonal crystal structure. **b**, Optical image of an *i*-G^1 (darker areas)/*h*-BN^2 (lighter areas) on a Si/SiO_2 substrate. Inset, Raman graphene 2D band (area indicated by the dotted box) showing a stark contrast between the regions. **c**, X-ray photoelectron spectroscopy data (upper panels, the transition from the 1s orbital) for an *i*-G^1/*h*-BN^2 sheet (*h*-BN^2 grown in a more reactive environment than **a**), showing a 1:1.08 B:N atomic ratio. Raman spectroscopy (lower panel) confirms the presence of the *h*-BN peak. **d**, False-colour DF-TEM image of a suspended *i*-G^1/*h*-BN^2 sheet with the junction region clearly starting at the sharp line that ends the *i*-G^1 area. **e**, SEM image of *i*-G^1/*h*-BN^2 film suspended over 2-μm-sized holes. Higher contrast (left panel) highlights *i*-G^1/*h*-BN^2 regions, whereas lower contrast (right panel) shows the suspended film. **f**, Schematic for cross-sectional STEM EELS. **g**, The upper panel shows the elemental map of an *i*-G^1/*h*-BN^2 junction region containing graphitic carbon (red; *g*-C), boron (green), and oxygen (blue). The lower panel shows the intensity profile of *g*-C and boron, indicating no voids or overlap in the junction region.

conducting pockets of *h*-BNC (boron-nitride-carbon) that might form during growth[10] using electrostatic force microscopy (EFM) (Supplementary Information and Fig. 3c, right). The EFM phase shift is highly uniform within both the *i*-G^1 and *h*-BN^2 regions; however, there is an abrupt change in the phase shift at the junction between *i*-G^1 and *h*-BN^2 owing to the different electrical conductivities of these materials, indicating very little carbon contamination bleeding into the *h*-BN^2 region.

These sheets are particularly useful for ultraflat three-dimensional electronics[24], where alternating *h*-BN and graphene regions can act as a wire array connected by lateral insulators. Because such an array maintains a uniform thickness throughout, the device will remain flat even after multiple transfers of such sheets without any post-processing, such as chemical mechanical polishing. We demonstrate this in Fig. 3b and c, where we fabricated a large sheet of *i*-G^1/*h*-BN^2 lines and performed multiple transfers onto a single substrate. After transferring the first sheet, a second was placed perpendicular to the first, with each layer contacted by electrodes. Optical images (Fig. 3b) show the structural uniformity of the final devices at different scales. The flatness and electrical properties of one such graphene–graphene crossed junction (inset in Fig. 3b,) is studied by atomic force microscopy (AFM) height (Fig. 3c, left panel) and EFM phase (Fig. 3c, right panel) measurements. Topographically, region A (*h*-BN on *h*-BN) is almost indistinguishable from region B (*h*-BN on graphene), as shown by the height histograms from each region. This contrasts with the EFM phase image, where both the bottom and top graphene strips are detected with nearly identical phase shifts. The small difference between the EFM phases of the two areas nevertheless indicates that the *h*-BN on top of the bottom graphene strip is acting as a dielectric film (see Supplementary Information). Electrical measurements (Fig. 3c, right panel) of these connections show a negligible addition of a contact resistance, confirming that such films behave as atomically flat three-dimensional interconnects. Although the vertical tunnelling current through one layer of *h*-BN is significant[25,26], few-layer *h*-BN formed by additional transfers could be used to electrically isolate devices vertically as well, and allow fabrication of other passive elements, such as capacitors[27].

Unlike G/*h*-BN structures, however, electrically conductive heterojunctions require characterization and optimization of the junction resistance. For this purpose, we fabricated arrays of devices that contained zero to four *i*-G^1/*i*-G^2 junctions. A false-colour SEM of an example cross-junction device is shown in Fig. 4a. The sheet resistance at the Dirac point (R_{Dirac}) for each device was measured using top-gated four-terminal measurements (Fig. 4b) and compiled for statistical comparison (Fig. 4c). Devices with (orange; 15 devices) and without (grey; 19 devices) junctions show both narrow distributions of R_{Dirac}, as well as similar medians (no junction: 6.6 kΩ □$^{-1}$; cross-junction: 6.9 kΩ □$^{-1}$). This shows that the electrical properties of heterojunction devices are similar to that of devices without junctions, despite the presence of many smaller grains in the junction area

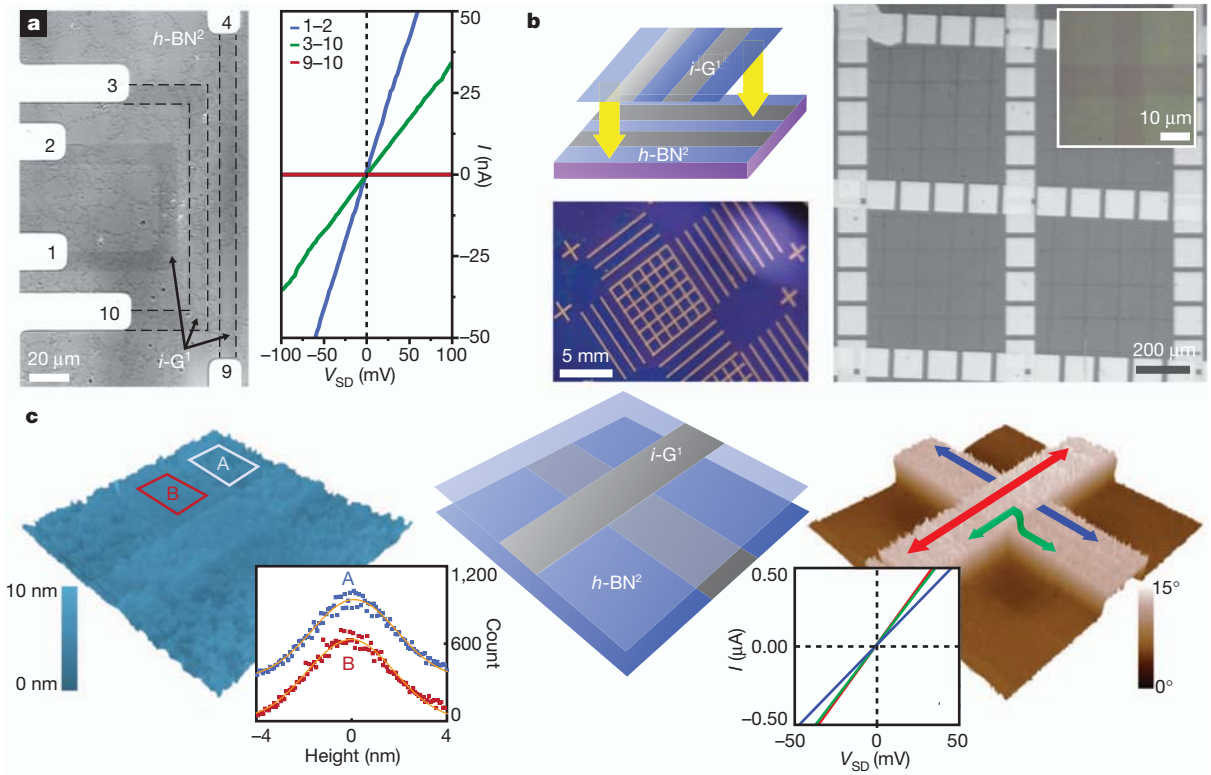


Figure 3 | *h*-BN/graphene electrical measurements. **a**, The left panel is an optical image of an *i*-G^1/*h*-BN^2 sheet with electrodes contacting graphene strips (outlined by dotted lines). The numbers label the electrodes (also shown in the right panel for the current–voltage (*I*–*V*) curves). The right panel shows two-terminal *I*–*V* characteristics of the indicated devices, with graphene showing conducting behaviour and *h*-BN exhibiting insulating characteristics ($R_{sheet} > 400\ T\Omega\ \square^{-1}$). **b**, The upper left panel is a schematic of a multiple transfer process for ultraflat three-dimensional interconnects. The lower left panel, right panel, and inset are optical images of increasing magnification of a final device substrate with each layer contacted by electrodes. **c**, The middle panel is a schematic of a G^1–G^1 cross-junction. The left panel is an AFM height image of the cross. Region A (*h*-BN on *h*-BN) is virtually indistinguishable from region B (*h*-BN on graphene), as indicated by the histogram (inset). The right panel is an EFM phase image of the same junction. Both graphene strips are visible. Two-terminal *I*–*V* characteristics (inset; $R_{sheet} < 3\ k\Omega\ \square^{-1}$) show no additional contact resistance owing to the graphene–graphene contact. Coloured arrows correspond to the *I*–*V* measurements in the inset.

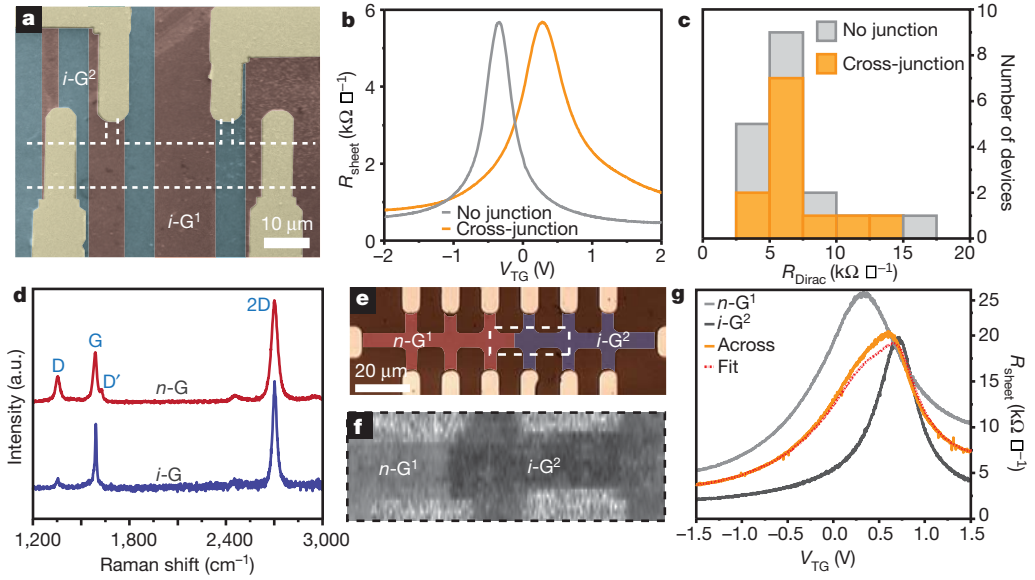


Figure 4 | Graphene junctions and heterostructures. **a**, SEM image with false-colour overlay of an *i*-G^1 (red)/*i*-G^2 (blue) cross-junction device before patterning. The final device area is indicated by the dashed lines. Cross-junction devices contained 1–4 junction regions. **b**, Four-terminal top gate (TG) dependence for devices without (grey) and with (orange) *i*-G^1/*i*-G^2 junctions, showing very similar peak resistances. **c**, Histograms of Dirac-point sheet resistance (R_{Dirac}) for devices with and without junction regions. **d**, Raman spectroscopy of *n*-doped (*n*-G) and *i*-G growths (both synthesized in second step in order to allow direct comparison). **e**, Optical image with false-colour overlay of a heterojunction device. **f**, Two-dimensional Raman mapping of the ratio of the integrated D′ and G peaks for the boxed area in **e**. **g**, Gate dependence of the junction region shows good agreement with an area-weighted average of the homogenous regions (red line, $\Delta R_{junction} \approx 0.15\ k\Omega\ \mu m$).

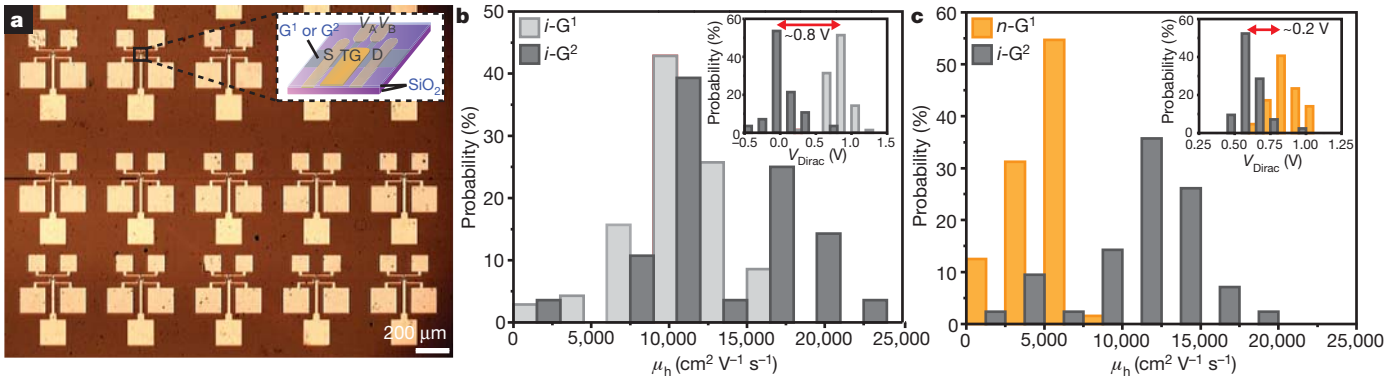


Figure 5 | Graphene device arrays and statistics. **a**, Optical image of an array of graphene transistors with device schematic (inset). **b**, Probability distribution of hole mobilities (μ_h) for i-G^1 (70 devices) and i-G^2 (28 devices) arrays. Inset, probability distribution of the Dirac point for each array showing a difference (0.8 V) in the mean position. **c**, Probability distribution of μ_h for n-G^1 (orange; 65 devices) and i-G^2 (grey; 43 devices) regions. Inset, the Dirac-point distributions are now much closer to the i-G^2 reference (0.2 V).

(Fig. 1). Additionally, field-effect carrier mobilities near the Dirac point remain high (>10,000 cm^2 V^{-1} s^{-1}) for both types of devices, which is consistent with the properties of electrically transparent, laterally connected grain boundaries grown under reactive growth conditions[20].

Our method thus allows for the growth of doped heterostructures, such as p–n junctions, within a single sheet of graphene, which could enable the production of active components with mechanically and electrically continuous junctions. Figure 4e shows a false-colour optical image of a device with two differently doped graphene areas (n-G^1/i-G^2). The n-doped graphene area exhibits an additional D′ peak (due to the presence of substitutional nitrogen dopants; Fig. 4d, red curve) that is not typically seen in i-G growths (blue curve). The location of the junction region was thus confirmed using two-dimensional Raman mapping of the ratio of the integrated D′ and G peaks (Fig. 4f), where the n-G^1 region is much brighter[5]. Electrical measurements within and across these regions again confirm electrical continuity across this region. Using top gates (5 nm Cr/45 nm Au, 100 nm of evaporated SiO_2) we measured the gate dependence in different regions, as shown in Fig. 4g (n-G^1/i-G^2). The device shows behaviour consistent with high-quality graphene, even across the junction region. The gate dependence of R_{sheet} in this area is in excellent agreement with an area-weighted average of the n-G^1 and i-G^2 areas with the addition of the small resistance $\Delta R_{junction} \approx 0.15\ k\Omega\ \mu m$ (consistent with the small junction resistance seen in Fig. 4c).

Multiple iterations of patterned regrowth would allow the formation of more complex circuits connected by lateral heterojunctions, whereas multiple transfers of these films would produce vertical heterojunctions and interconnects for increased functionality. To realize more complicated structures, it is essential that the electrical properties of sheets grown at different steps are affected by the processing in a reproducible manner. In Fig. 5, we show statistics for intrinsic (both i-G^1 and i-G^2) and n-doped graphene device arrays (optical image shown in Fig. 5a), which exhibit high-performance electrical properties. Namely, they show high field-effect mobilities (mean >10,000 cm^2 V^{-1} s^{-1}; Fig. 5b)[28] and low hysteresis (see Supplementary Information). Although we observe a slight difference between the i-G^1 and i-G^2 mobility distributions (Fig. 5b), these values are consistent with variations seen from growth to growth (see i-G^2 in Fig. 5c), suggesting that high electrical performances could be generally achieved even after multiple regrowth processes. In addition, the Dirac point shows a narrow distribution within each area for both intrinsic and n-doped graphene growths (see insets to Fig. 5b and c). There is a shift of around 0.8 V between the Dirac-point distribution of the i-G^1 and i-G^2 growths (inset to Fig. 5b); however, the magnitude of this shift is reduced when the first growth is n-doped (~0.2 V, inset to Fig. 5c). Although this is consistent with the expected effects of n-doping, the overall device appears to remain p-type, probably owing to the effects of the fabrication process. Thus, improvement in the transfer and fabrication steps should lead to increased control over not only the Dirac-point distributions, but also general homogeneity[29,30].

We propose that our patterned regrowth technique provides a versatile and scalable method for growing and integrating layered materials, beyond h-BN and graphene, for atomically thin circuitry. In particular, the addition of two-dimensional semiconducting materials, such as MoS_2 (ref. 31), would bring together the three key building blocks (insulator, metal and semiconductor) of modern integrated circuitry into a single, transferrable film. Furthermore, the devices made using this approach are likely to remain mechanically flexible and optically transparent, allowing transfer to arbitrary substrates for flexible, transparent electronics.

METHODS SUMMARY

Patterned regrowth. Synthesis of graphene and h-BN were carried out in a helium-leak-checked semiconductor grade tube furnace using 25-μm-thick Cu as the growth substrate. Graphene growths were performed at 1,000 °C, using the general methods reported in refs 4 and 5. For intrinsic graphene (i-G) flow rates were H_2: 100 standard cm^3 min^{-1}, CH_4: 6 standard cm^3 min^{-1}. For n-doped graphene (n-G) the flow rates were H_2: 100 standard cm^3 min^{-1}, CH_4: 2 standard cm^3 min^{-1} NH_3: 7 standard cm^3 min^{-1}. For the first growth (G^1), Cu substrates were annealed in H_2 at $T = 1{,}000$ °C for 1 h before introduction of CH_4. Photolithography was then performed directly on the Cu substrate in order to pattern a protective layer of photoresist. An O_2 reactive ion etch (100 W, 105 s) was used to remove the unwanted graphene from the surface. The patterned Cu was then immediately placed in Microposit Remover 1165 to thoroughly clean the surface of residual photoresist. Immediately after this, the substrates were placed directly back into the reaction chamber for the second graphene or h-BN growth. For the second growth, samples were exposed to growth reactants upon reaching the growth temperature. h-BN syntheses were carried out for 5–15 min by sublimation of the ammonia-borane, as in ref. 9. For the 'slow' growth h-BN (Fig. 2a), the effective flow rate from the precursor was about 1 standard cm^3 min^{-1}. For the 'fast' growth, the flow rate was increased to about 4 standard cm^3 min^{-1}. We did not use an additional carrier gas because this led to thicker films.

Received 16 March; accepted 10 July 2012.

1. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306,** 666–669 (2004).
2. Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438,** 197–200 (2005).
3. Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438,** 201–204 (2005).
4. Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324,** 1312–1314 (2009).
5. Zhao, L. *et al.* Visualizing individual nitrogen dopants in monolayer graphene. *Science* **333,** 999–1003 (2011).
6. Jin, Z., Yao, J., Kittrell, C. & Tour, J. M. Large-scale growth and characterizations of nitrogen-doped monolayer graphene sheets. *ACS Nano* **5,** 4112–4117 (2011).

7. Kubota, Y., Watanabe, K., Tsuda, O. & Taniguchi, T. Deep ultraviolet light-emitting hexagonal boron nitride synthesized at atmospheric pressure. *Science* **317,** 932–934 (2007).
8. Kim, K. K. *et al.* Synthesis of monolayer hexagonal boron nitride on Cu foil using chemical vapor deposition. *Nano Lett.* **12,** 161–166 (2012).
9. Song, L. *et al.* Large scale growth and characterization of atomic hexagonal boron nitride layers. *Nano Lett.* **10,** 3209–3215 (2010).
10. Ci, L. *et al.* Atomic layers of hybridized boron nitride and graphene domains. *Nature Mater.* **9,** 430–435 (2010).
11. Liu, Y., Bhowmick, S. & Yakobson, B. I. BN white graphene with ''colorful'' edges: the energies and morphology. *Nano Lett.* **11,** 3113–3116 (2011).
12. Pruneda, J. M. Origin of half-semimetallicity induced at interfaces of C-BN heterostructures. *Phys. Rev. B* **81,** 161409 (2010).
13. Miyamoto, Y., Rubio, A., Cohen, M. & Louie, S. Chiral tubules of hexagonal BC_2N. *Phys. Rev. B* **50,** 4976–4979 (1994).
14. Wehling, T. O. *et al.* Molecular doping of graphene. *Nano Lett.* **8,** 173–177 (2008).
15. Williams, J. R., Dicarlo, L. & Marcus, C. M. Quantum Hall effect in a gate-controlled p-n junction of graphene. *Science* **317,** 638–641 (2007).
16. Dresselhaus, M. S., Jorio, A., Hofmann, M., Dresselhaus, G. & Saito, R. Perspectives on carbon nanotubes and graphene Raman spectroscopy. *Nano Lett.* **10,** 751–758 (2010).
17. Chen, S. *et al.* Oxidation resistance of graphene-coated Cu and Cu/Ni alloy. *ACS Nano* **5,** 1321–1327 (2011).
18. Huang, P. Y. *et al.* Grains and grain boundaries in single-layer graphene atomic patchwork quilts. *Nature* **469,** 389–392 (2011).
19. Kim, K. *et al.* Grain boundary mapping in polycrystalline graphene. *ACS Nano* **5,** 2142–2146 (2011).
20. Tsen, A. W. *et al.* Tailoring electrical transport across grain boundaries in polycrystalline graphene. *Science* **336,** 1143–1146 (2012).
21. Yu, Q. *et al.* Control and characterization of individual grains and grain boundaries in graphene grown by chemical vapour deposition. *Nature Mater.* **10,** 443–449 (2011).
22. Sun, Z. *et al.* Growth of graphene from solid carbon sources. *Nature* **468,** 549–552 (2010).
23. Kim, Y.-J., Lee, J.-H. & Yi, G.-C. Vertically aligned ZnO nanostructures grown on graphene layers. *Appl. Phys. Lett.* **95,** 213101 (2009).
24. Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5,** 722–726 (2010).
25. Britnell, L. *et al.* Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12,** 1707–1710 (2012).
26. Lee, G.-H. *et al.* Electron tunneling through atomically flat and ultrathin hexagonal boron nitride. *Appl. Phys. Lett.* **99,** 243114 (2011).
27. Sanchez-Yamagishi, J. *et al.* Quantum Hall effect, screening, and layer-polarized insulating states in twisted bilayer graphene. *Phys. Rev. Lett.* **108,** 076601 (2012).
28. Meric, I. *et al.* Current saturation in zero-bandgap, top-gated graphene field-effect transistors. *Nature Nanotechnol.* **3,** 654–659 (2008).
29. Moser, J., Barreiro, A. & Bachtold, A. Current-induced cleaning of graphene. *Appl. Phys. Lett.* **91,** 163513 (2007).
30. Liang, X. *et al.* Toward clean and crackless transfer of graphene. *ACS Nano* **5,** 9144–9153 (2011).
31. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS_2 transistors. *Nature Nanotechnol.* **6,** 147–150 (2011).

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

Acknowledgements We thank P. McEuen and M. Spencer for discussions. This work was mainly supported by AFOSR grants (FA9550-09-1-0691 and FA9550-10-1-0410) and the NSF through the Cornell Centers for Materials Research (NSF DMR-1120296), which also provided the electron microscopy facilities. Additional funding was provided by the Alfred P. Sloan Foundation. L.B. was partially supported by a Fullbright scholarship; R.W.H. and P.Y.H. were supported by an NSF Graduate Research Fellowship. Device fabrication was performed at the Cornell NanoScale Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by the National Science Foundation (grant ECS-0335765).

Author Contributions M.P.L. and C.-J.K. contributed equally to this work. Synthesis, device fabrication, and electrical measurements and analysis were done by M.P.L. and C.-J.K. DF-TEM and data analysis were performed by L.B. and C.-J.K. EELS measurement and data analysis were conducted by P.Y.H. and D.A.M. Raman measurements and analysis were carried out by M.P.L. with assistance from R.W.H. J.P. designed the experiments and oversaw the research. M.P.L. and J.P. wrote the manuscript with assistance from R.W.H and input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P. (jpark@cornell.edu).

LETTER

doi:10.1038/nature11374

Potential methane reservoirs beneath Antarctica

J. L. Wadham[1], S. Arndt[1,2], S. Tulaczyk[3], M. Stibal[1], M. Tranter[1], J. Telling[1], G. P. Lis[1], E. Lawson[1], A. Ridgwell[1], A. Dubnick[4], M. J. Sharp[3], A. M. Anesio[1] & C. E. H. Butler[1]

Once thought to be devoid of life, the ice-covered parts of Antarctica are now known to be a reservoir of metabolically active microbial cells and organic carbon[1]. The potential for methanogenic archaea to support the degradation of organic carbon to methane beneath the ice, however, has not yet been evaluated. Large sedimentary basins containing marine sequences up to 14 kilometres thick[2] and an estimated 21,000 petagrams (1 Pg equals 10^{15} g) of organic carbon are buried beneath the Antarctic Ice Sheet. No data exist for rates of methanogenesis in sub-Antarctic marine sediments. Here we present experimental data from other subglacial environments that demonstrate the potential for overridden organic matter beneath glacial systems to produce methane. We also numerically simulate the accumulation of methane in Antarctic sedimentary basins using an established one-dimensional hydrate model[3] and show that pressure/temperature conditions favour methane hydrate formation down to sediment depths of about 300 metres in West Antarctica and 700 metres in East Antarctica. Our results demonstrate the potential for methane hydrate accumulation in Antarctic sedimentary basins, where the total inventory depends on rates of organic carbon degradation and conditions at the ice-sheet bed. We calculate that the sub-Antarctic hydrate inventory could be of the same order of magnitude as that of recent estimates made for Arctic permafrost. Our findings suggest that the Antarctic Ice Sheet may be a neglected but important component of the global methane budget, with the potential to act as a positive feedback on climate warming during ice-sheet wastage.

The production of methane (CH_4) by methanogenic archaea is common across many sub-surface environments, including the deep ocean[4–8], permafrost[9] and lake sediments[10–12] and is promoted by the presence of a suitable organic carbon (OC) substrate and an absence of higher-energy-yielding electron acceptors (for example, O_2 and SO_4^{2-}) with which to degrade organic matter. The release of this biogenic methane to the atmosphere is important in driving changes in global climate on geological, millennial and centenary timescales[13]. 'Geological' methane, produced largely via thermogenic processes in the deep subsurface, supplements the biogenic component[13]. This methane may be generated via the thermal breakdown of organic matter and by inorganic synthesis and outgassing from the mantle. The recent discovery that sub-ice-sheet environments are likely to be anoxic[14], are host to micro-organisms[15] and may contain significant reservoirs of OC identifies them as favourable sites for methanogenesis. Research has so far focused upon the potential biological conversion of overridden OC to methane beneath the Northern Hemisphere Pleistocene ice sheets[14,16], where high-pressure and low-temperature conditions permit methane to accumulate as hydrate. Methane accumulation beneath the Antarctic Ice Sheet has not yet been evaluated, despite the presence of extensive and deep sedimentary basins containing viable microbial populations[15] and OC available for conversion to methane.

Several factors suggest that there should be methane present beneath the Antarctic Ice Sheet. Geophysical data indicate extensive Antarctic sedimentary basins (ASBs) beneath the West and East Antarctic Ice Sheets (WAIS and EAIS), containing sedimentary drapes of up to 14 km in thickness (Supplementary Table 1). Many of these basins are located around the Antarctic periphery, but penetrate several 100–1,000 km into the Antarctic interior and are associated with the onset of accelerated motion in ice streams and their tributaries (Supplementary Information 1). The inferred origin of these sediments is marine, glaci-marine and crustal sedimentary sources (Supplementary Table 1). Melting conditions beneath about half of the ice sheet mean that sediments contain liquid water beneath the ice cover[17,18]. The widespread presence of either dissolved methane or geochemical evidence for methanogenesis in marine cores, rock cores and seeps around the Antarctic margin indicates that OC is commonly cycled to methane in ocean margin basins (Supplementary Information 1). It is reasonable to expect similar processes to prevail beneath the ice. We therefore evaluated the potential for methane generation and release from sedimentary basins buried beneath the Antarctic Ice Sheet.

We calculate that 50% of the WAIS (10^6 km^2) and 25% of the EAIS (2.5×10^6 km^2) overlie ASBs, where areas of the bed substantially below sea level have been inferred to contain thick marine sediments (Supplementary Information 1). Despite the difference in overriding substrate (for example, an ice cap rather than an ocean water column or atmosphere), comparable processes should operate in both ASBs, subseafloor sediments and the terrestrial deep biosphere. Typically, ASBs are capped by subglacial till (Fig. 1a). Oxygen supply from the release of air bubbles in melting basal ice is small and quenched by sulphide oxidation and microbial respiration in surface till, such that subglacial sediments are largely anoxic. Hence, most degradation of organic matter occurs anaerobically, first using NO_3^-, next Fe(III) and SO_4^{2-} as electron acceptors[14], and ultimately via disproportionation into CO_2 and methane[16]. Beneath the till layer, the sediment biogeochemical profile should resemble that of circum-Antarctic marine or terrestrial sediments at the onset of glaciation. In the former case, marine sulphate concentrations would be expected to decrease within the upper 150 m of the sediment column, beyond which methanogenesis is the dominant terminal process for OC decomposition[19] (Fig. 1). We calculate that about 21,000 petagrams of carbon (Pg C) was present in ASBs at the time of ice-sheet initiation (Supplementary Information 3e). This is more than ten times the magnitude of OC stocks in northern permafrost regions[20] and parallels the order of magnitude estimates made for reactive carbon in ocean sediments (assuming a 10% reactive OC content for ASBs)[21]. Subglacial ASBs thus constitute a substantial reservoir of OC that has been previously overlooked.

No data exist for rates of methanogenesis in sub-Antarctic marine and terrestrial sediments, but new experimental data from other subglacial environments demonstrate the potential for overridden organic matter beneath glacial systems to produce methane (see Supplementary Information). We obtained subglacial samples derived from the impurity-rich (silty) sections of basal ice exposed at the margins of the Greenland Ice Sheet and in Antarctic and Canadian Arctic glaciers. These samples display concentrations of OC an order of magnitude less (0.07–0.5%) than those recorded in the sub-seafloor, largely owing to inclusions of minerogenic glacial till with the overridden organic

[1]School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK. [2]Department of Earth Sciences – Geochemistry, Utrecht University, 3508 Utrecht, The Netherlands. [3]Earth and Planetary Sciences Department, University of California, Santa Cruz, California 95064, USA. [4]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton T6G 2E3, Canada.

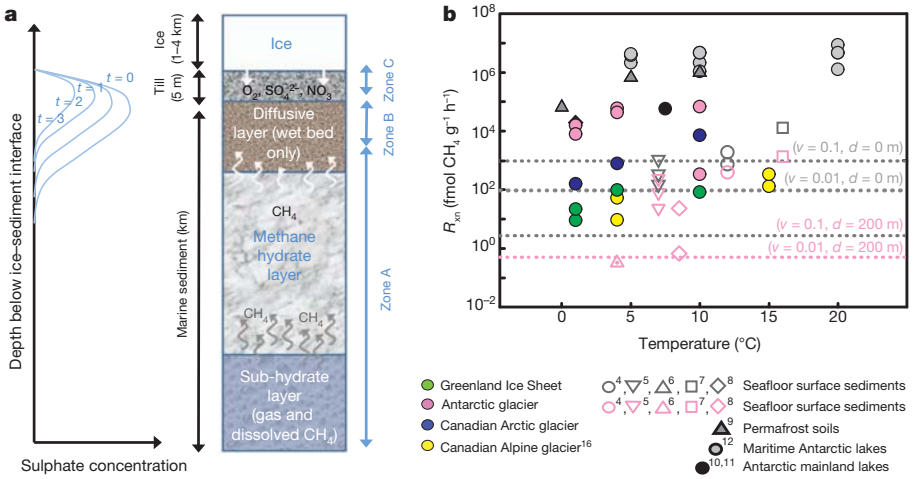


Figure 1 | Methane production in sub-Antarctic sediments. **a**, A conceptual model of biogeochemical processes in a sub-Antarctic marine sediment/till complex, indicating three main zones and assuming non-frozen basal conditions. Zone A is the methane production zone (after 16 kyr of glaciation), zone B is the sulphate reduction/anaerobic oxidation of methane zone (during the first 16 kyr of glaciation) and zone C is the sulphide oxidation (oxic/anoxic)/nitrate reduction zone. White arrows indicate methane diffusion. In parts of the ice sheet where frozen basal conditions prevail, zone A is likely to extend to the ice/sediment interface. The graph on the left also indicates the likely evolution of sulphate concentrations in the upper part of the marine sediment/till complex over time (t values indicate arbitrary, increasing periods of time), as the sulphate pool is depleted by methane oxidation and sulphate reduction (where $t = 3$ is of the order of 10 kyr). **b**, Comparison of rates of methane production (R_{xn}, in units of fmol CH_4 per gram dry weight of sediment per hour) measured experimentally for subglacial sediments (this paper) and other anoxic environments. For comparison, we also plot R_{xn} values used in the modelling for two parameter sets for surface sediments (depth $d = 0$; grey symbols) and sediments exceeding 200 m depth (pink symbols) and for $\nu = 0.1$ and $\nu = 0.01$ (where ν is the reactivity parameter in the OC degradation model, with high ν values indicating high organic matter reactivity). (The percentage total OC content of sediments, TOC(0,0) = 1%.)

matter[4]. Order-of-magnitude contrasts in rates of methanogenesis are observed, a reflection of large differences in OC bioavailability. The high rates of methanogenesis in Canadian Arctic (10^2–10^3 fmol CH_4 per gram dry weight of subglacial sediment per hour) and Antarctic sediments (10^3–10^4 fmol $CH_4\,g^{-1}\,h^{-1}$), which include overridden lake sediments, are consistent with the known availability of simple polymers and OC in lacustrine material[22], and approach values reported from perennially ice-covered Antarctic lakes[10,11]. Very low rates of methanogenesis (9–93 fmol $CH_4\,g^{-1}\,h^{-1}$ at 1–10 °C) in Greenland Ice Sheet sediments align with those reported for the Robertson Glacier in Canada (9–51 fmol $g^{-1}\,h^{-1}$ at 4 °C; ref. 16), both reflecting the more recalcitrant nature of carbon in overridden palaeosols (Fig. 1b). Remarkably, these minimum rates of activity align closely with rates of methane production reported in deep (>150 m) sub-seafloor sediments (Fig. 1b).

We used a well-established one-dimensional numerical hydrate model, developed to model methane hydrate accumulation beneath the sea floor[3]. We assume that the seaward extent of these ASBs is the present-day grounding line and that the WAIS and EAIS were present to an extent similar to that of the present day one million years (Myr) ago and 30 Myr ago, respectively. The 1-Myr value is the maximum estimate indicated by recent modelling (Supplementary Information 1). The potential to generate methane in ASBs depends critically upon the rate at which methane can be produced *in situ* relative to the rates at which it is lost during glaciation. Methane production rates in the model ($R_{xn}(z,t)$) are scaled to decreasing organic matter quality with sediment age, and hence depth, using a reactive continuum model[23]. Key parameters in the continuum model are the percentage surface total OC content of sediments (TOC(time $t = 0$ yr, depth $d = 0$ m), and the reactivity parameter ν. The exponent ν determines the shape of the vertical distribution of organic matter reactivity with depth (Supplementary Information 2f, Supplementary Figs 2 and 4).

Modelling in initial simulations assumes that advective gas transport (for example, via fluid flow or venting) is likely to be minimal at the ice-sheet bed (Supplementary Information 2f). Two end-member scenarios are modelled for the surface boundary condition in these cases: (1) the maximum flux scenario (methane loss occurs via diffusion at the sediment surface into the overlying subglacial water system), and (2) the zero flux scenario (closed-system conditions arising from either frozen basal conditions or saturation of the overlying water film with methane). We apply the zero flux scenario over 20% and 10% of the EAIS and WAIS basins respectively, based on the distribution of frozen basal conditions above ASBs[17] (Supplementary Information 1, and Fig. 2).

In sub-seafloor sediments, sulphate oxidizes up to 90% of the methane diffusing into the sulphate reduction zone by the anaerobic oxidation of methane (zone B, Fig. 1a)[24]. A key difference between sub-seafloor and sub-Antarctic sediments is that, whereas the former have a large overlying sulphate reservoir from sea water, subglacial sediments have a relatively small and finite sulphate reserve derived from the original marine material and ongoing sulphide oxidation (Supplementary Information 2e). We calculate that an initial sulphate pool in the subglacial sediment/till complex reduces to zero over about 16,000 years (16 kyr) by anaerobic oxidation of methane coupled to sulphate reduction (Supplementary Information 1, Fig. 1a). Hence, all model simulations assume that methanogenesis cannot occur in the upper 150 m of the sediment column during the first 16 kyr.

The potential to release subglacial methane rapidly following deglaciation is contingent upon whether methane generated in sediments during glaciation can be trapped as methane hydrate or as free gas beneath a hydrate layer. Hydrate forms when dissolved methane concentrations are in excess of saturation within the gas hydrate stability zone (GHSZ), where the latter spans the upper 270 m and 670 m of a ~1-km-deep sediment column beneath the WAIS and EAIS respectively (Fig. 3). The thinner GHSZ beneath the WAIS reflects the higher geothermal heat flux and thinner ice, both of which affect hydrate stability.

Model sensitivity analysis indicates that the amount of hydrate formed in ASBs is strongly sensitive to the initial surface OC content of sediment and the reactivity parameter ν in the reactive continuum model. The minimum and maximum values of ν indicate a predominance of low- and high-reactivity OC compounds respectively, while intermediate values indicate a more uniform distribution of different OC compounds with depth. For higher reactivity parameters (for example, $\nu > 0.3$), OC is exhausted rapidly in the upper sediment layers, where the anaerobic oxidation of methane dominates for the initial

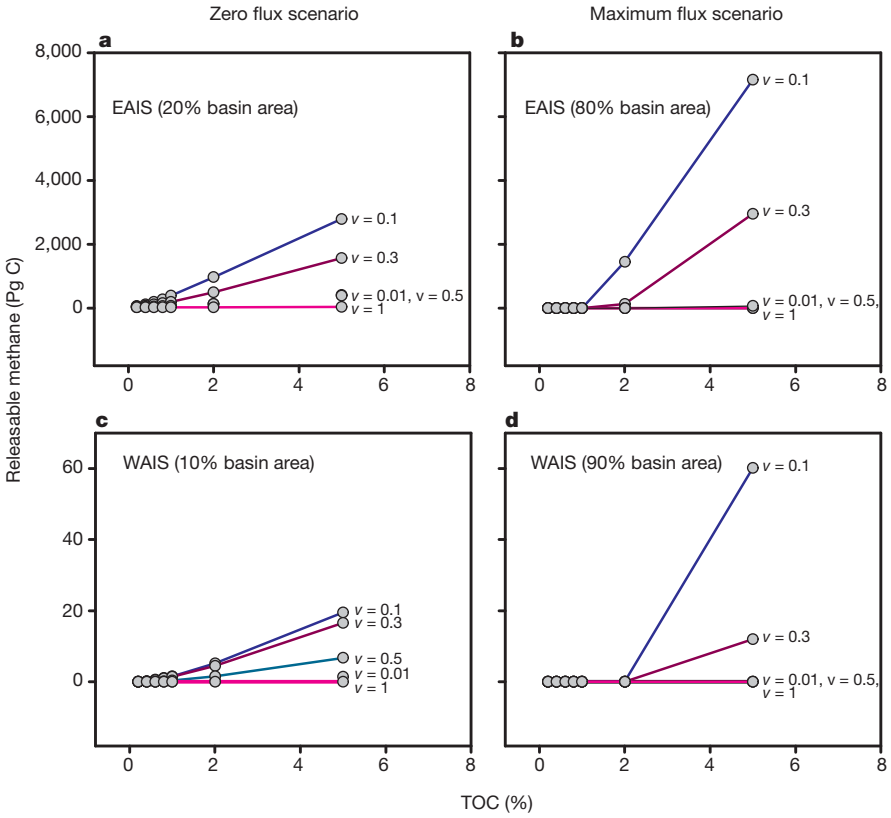


Figure 2 | Methane hydrate+gas accumulation potential beneath the ice sheet. This indicates the sensitivity of methane hydrate+gas accumulation ('releasable methane') beneath the Antarctic Ice Sheet to varying TOC(0,0) and ν values in the reactive continuum model under the zero flux scenario for the EAIS (**a**) and WAIS (**b**), and the maximum flux scenario for the EAIS (**c**) and WAIS (**d**). Methane hydrate+gas accumulation is calculated after 30 Myr (EAIS) and 1 Myr (WAIS) of simulation time. The high hydrate inventory for the maximum flux scenario reflects the greater area contribution of melting basal conditions for the WAIS and EAIS.

16 kyr, while for low reactivity values (for example, $\nu = 0.01$), slow rates of methanogenesis limit the amount of methane accumulation over million-year timescales. It is notable that intermediate values of ν give the highest methane hydrate+gas accumulation under both the maximum and zero flux scenarios (Supplementary Information 3, Fig. 2). In these cases, higher TOC concentrations penetrate deeper into the sediment column and rates of degradation are sufficient to generate substantial methane. We conclude that the slow burn of organic matter over million-year timescales produces the most methane hydrate in Antarctic sediments.

Model results highlight that methane hydrate formation beneath the ice sheet is strongly controlled by basal temperature conditions (for example, wet bed versus frozen bed). Significant hydrate is only likely to be formed in sectors of ASBs with frozen basal conditions (the zero flux scenario) unless the OC content is relatively high (>2%) (Fig. 2). The latter may occur locally, as indicated in Antarctic Ocean Drilling Program core records containing ancient organic-rich material (Supplementary Fig. 3). In East Antarctica, long duration glaciation, aerially extensive ASBs and frozen basal conditions in some sectors may promote substantial hydrate accumulation (Fig. 2, Supplementary Information 1). Assuming two plausible end-member values for ν (0.1 for typical marine sediments and 0.01 for highly recalcitrant OC typical of the deep rock biosphere; Supplementary Information 3a) and a surface TOC content of 1%, the model generates total hydrate+gas inventories of 70–390 Pg C for the EAIS (>90% as hydrate) (Fig. 2). Even if the duration of continuous ice cover in EAIS ASBs were shorter (for example, 5–10 Myr), it would still be possible to generate some hundreds of Pg C of methane hydrate (Supplementary Fig. 6). Both end-member parameter sets give modelled surface and deep methane production rates (R_{xn}) that compare well to those measured in marine sediments[4–6] and subglacial sediments (this paper and ref. 16) (Fig. 1b). The resulting hydrate is generated in ASBs with frozen conditions at the ice-sheet bed, occupying up to 30% pore space volume between 10 m and 100 m of the sediment surface (Fig. 3). The shallow depth of the latter contrasts with oceanic systems[13], and reflects (1) the lack of a sulphate pool in the upper sediment column after 16 kyr, designating methanogenesis as the terminal step for organic matter degradation, (2) no loss of methane at the sediment surface by diffusion and (3) lower methane solubility in shallow sediments. For the WAIS, shorter-duration glaciation and widespread melting conditions at the ice-sheet bed[17] result in zero methane hydrate formation in sediments for representative parameter sets (such as $\nu = 0.01$–0.1, TOC(0,0) = 1%), although there may be some local hydrate formation in association with organic-rich horizons (Fig. 2, Supplementary Tables 5 and 6).

Substantial parts of the WAIS contain sediment thicknesses of several kilometres (Supplementary Table 1) and exhibit enhanced geothermal heat flow and volcanism[18]. The thermal breakdown of organic matter or heavier hydrocarbons is common in marine basins with thick sediments and enhanced geothermal activity[25]. Hence, for the WAIS we consider the potential for thermogenic methanogenesis in deep sediments, with methane transported to the GHSZ by upward fluid flow. Because there is no sedimentation beneath the ice sheet (driving downward advection of sediments/porewaters), net upward fluid flow induced by geothermal heating is likely. We ran further model simulations for the WAIS (maximum flux scenario), applying a range of vertical fluid advection velocities representative of active geothermal sites (Supplementary Information 2f, and Fig. 4). Hydrate is produced throughout the entire GHSZ in simulations employing upward fluid flow velocities of just a few millimetres per year (Fig. 4). Typical upward fluid flow velocities at active vents and seeps exceed 10 mm yr^{-1} and can be 2,500 mm yr^{-1}, generating upward methane fluxes of over 3 mol CH_4 m^{-2} yr^{-1} (refs 26 and 27). However, this fluid flow has been shown to exhibit high spatial variability[26]. If geothermally active sites were present over 10% of the WAIS ASB area[18] and displayed low upward fluid flow rates (5 mm yr^{-1}) and more diffuse methane seepage (<1.0 mol m^{-2} yr^{-1}), about 90 Pg C of methane hydrate could be generated over 1 Myr (Fig. 4). If the geothermally active area reduced to 5% or 1% of the WAIS ASBs, then 47 Pg C and 9 Pg C of hydrate could be generated, respectively. Although they are speculative, these simulations indicate that it might be reasonable to expect local hydrate accumulations in WAIS ASBs.

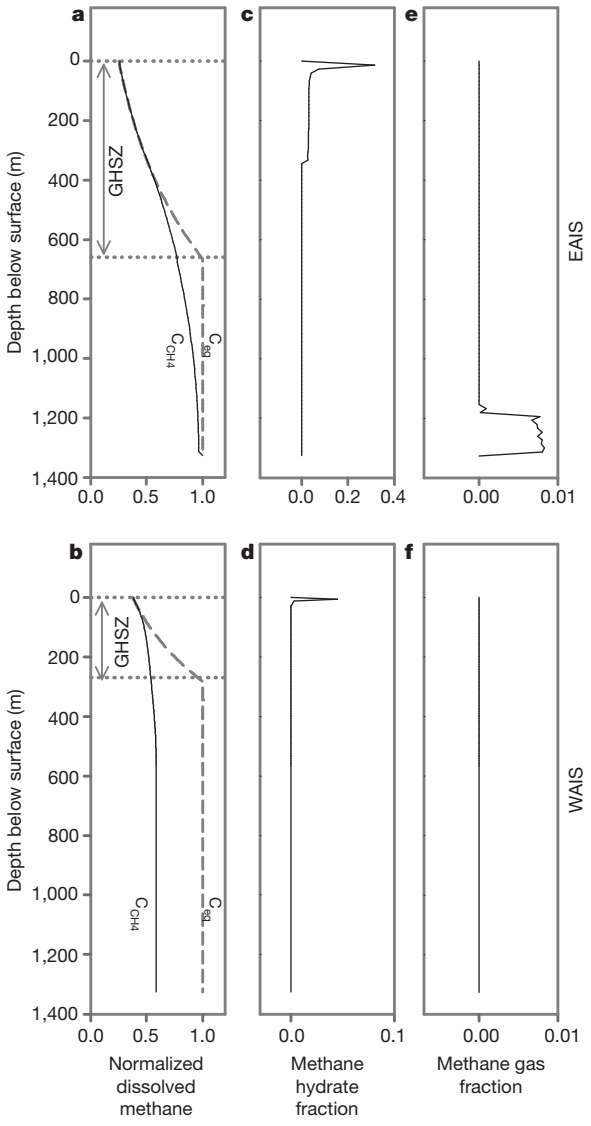


Figure 3 | Vertical profiles of methane solubility, dissolved methane, methane hydrate and methane gas in zero flux simulations. Concentrations in **a** and **b** are normalized to the maximum methane solubility defined at the three-phase equilibrium point for methane in pore waters C_{eq}, methane hydrate and methane gas are reported in units of percentage of pore volume (note the different scale for the *x* axis). Simulations assume 30 Myr (EAIS) and 1 Myr (WAIS) of glaciation. The upper and lower extent of the GHSZ is indicated. C_{eq} is the equilibrium concentration of methane in pore waters and C_{CH4} is the modelled dissolved concentration of methane in pore waters.

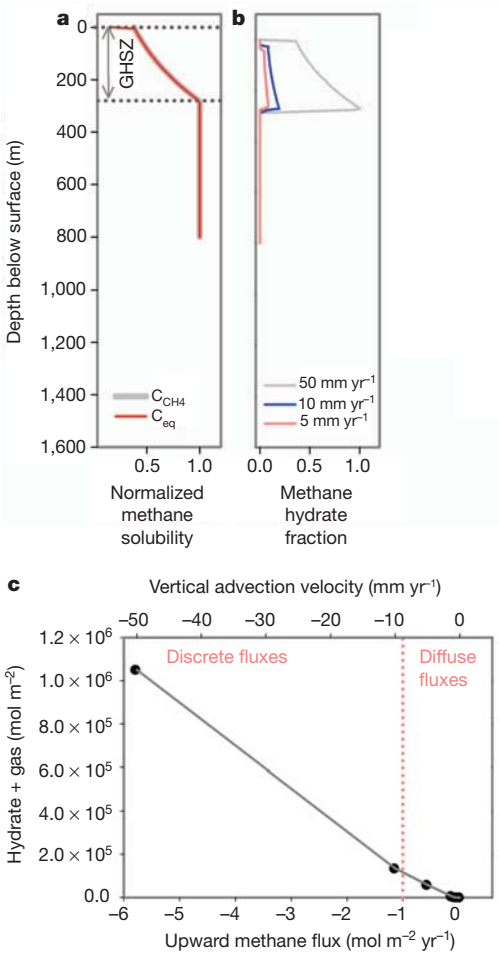


Figure 4 | Modelled thermogenic methane accumulation beneath WAIS over 1 Myr of glaciation under the maximum flux scenario ($v = 0.1$, TOC(0,0) = 1%). **a**, Vertical profiles of methane solubility (C_{eq}) and dissolved methane concentrations (C_{CH4}, for a vertical advection velocity of 50 mm yr^{-1} only) at $t = 1$ Myr (that is, after a million years of glaciation) where concentrations are normalized to the maximum C_{eq}. The upper and lower extent of the GHSZ is indicated. **b**, Methane hydrate formed after 1 Myr of glaciation and employing different vertical advection velocities (in the range 5–50 mm yr^{-1}). Zones of upward fluid flow velocities and methane fluxes for active vents ('discrete fluxes') versus more diffuse seepage ('diffuse fluxes') are indicated based upon refs 26 and 27.

Model results indicate that the potential hydrate reserve could be 70–390 Pg C (1.31–7.28 × 10^{14} m^3 methane gas) beneath the EAIS (biogenic production in frozen bed sectors) and some tens of Pg C (about 2 × 10^{13} m^3 of gas) beneath the WAIS (themogenic production in wet-based, geothermally active areas). This represents a sizeable reservoir of methane hydrate, of a similar order of magnitude to more recent estimates of Arctic permafrost and Arctic ocean hydrate reserves[13], but lower than those calculated globally for marine sediments[28]. The size of the subglacial hydrate inventory is highly sensitive to ice thickness and likely to be located at shallow depths (Fig. 3), highlighting the strong potential for deglaciation to trigger hydrate destabilization via its effect on the ice overburden pressure (the pressure added by an ice column on top of sediments). For example, rapid ice-sheet thinning and retreat during previous collapses of the WAIS[29] might result in changes in the extent of the GHSZ and the destabilization of any methane hydrate, followed by release via venting, similar to that occurring in the marine environment[26]. The impact of such methane hydrate destabilization upon atmospheric methane concentrations depends on the size, geographical distribution of the inventory and the pattern and rate of ice-sheet retreat. Ice-sheet retreat rates of the order of 1,000 km^2 yr^{-1} are possible for previous episodes of deglaciation (for example, the Laurentide retreat occurred at 1,200 km^2 yr^{-1}; ref. 30). These rates might be exceeded in WAIS ASBs owing to the grounding of the ice sheet below sea level and the presence of weak tills, which promote the fast flow of ice and dynamic instability[29]. If we assume that just 15 Pg C was present as methane hydrate over 10% of the WAIS ASB area (10^5 km^2), focused largely in marginal geothermally active zones[18], ice retreat of 1,000 km yr^{-1} could make 0.15 Pg C yr^{-1} of hydrate available for release to the atmosphere. This slightly exceeds the atmospheric turnover rate of 0.13 Pg C yr^{-1}, indicating its potential to affect atmospheric methane concentrations. For this to happen, hydrate destabilization would need to be a disequilibrium process, manifested by rapid methane transport to the sediment surface (for example, by venting[26]) at a rate that far exceeds that at which it can be oxidized. A hydrate inventory of the order of some tens of Pg C in WAIS marginal basins seems plausible, based on

our model results, and is comparable to inventories for other localized marine hydrate reserves, for example, 15 Pg C of methane hydrate over 26,000 km^2 at Blake Ridge and 23 Pg C as hydrate at Hydrate Ridge, Cascadia Margin[31].

We conclude that the Antarctic Ice Sheet could constitute a previously neglected component of the global methane hydrate inventory, although significant uncertainty exists. The predicted shallow depth of these reserves also makes them more susceptible to climate forcing than some other global hydrate reserves. If substantial methane hydrate and gas were present beneath the WAIS, hydrate destabilization during episodes of ice-sheet collapse could act as a positive feedback on global climate change during past and future ice-sheet wastage. Further validation of these assertions awaits deep drilling into ASBs via future international initiatives.

METHODS SUMMARY

We use a well-established one-dimensional numerical hydrate model[3] (see Supplementary Information 2f), modified via the inclusion of an OC degradation model, based upon the reactive continuum approach[23]. Within this, methane production rates $R_{xn}(z,t)$ are scaled to decreasing organic matter quality with sediment age, and hence depth. We assume physical properties for sediments similar to the properties previously used for ocean sediment modelling (listed in ref. 3). We prescribe an initial TOC depth profile of subglacial Antarctic marine sediments (TOC(z,0)) that is calculated from the steady-state solution of OC using the reactive continuum model and assuming a prescribed surface OC content, TOC(0,0) (Supplementary Information 3a), as well as an age/depth model, age(z). The age/depth model is constructed using a cubic spline interpolation of observed depth/age relationships in circum-Antarctic marine sediments[32]. See Supplementary Table 2 for site-specific parameters. We conduct a sensitivity analysis of hydrate accumulation to (1) OC content of sediments, (2) reactivity of organic matter in sediments, represented by ν in the reactive continuum model, (3) initial saturation of methane in sediment pore waters at $t = 0$ of the simulation, (4) ice thickness and (5) duration of glaciation (see Supplementary Information 2 and 3). Predictions of methane hydrate accumulation are made for representative parameter sets. Model results are scaled to estimates of total methane hydrate beneath the Antarctic Ice Sheet, assuming that ASBs underlie 50% of the WAIS area and 20% of the EAIS area. We assume that 10% of WAIS and 20% of EAIS ASBs exhibit frozen basal conditions at the sediment–ice interface, represented by our zero flux scenario in which no methane may diffuse out of the top of the sediment column. The remaining ASB areas are represented by our maximum flux scenario, in which methane is allowed to diffuse out of the top of the sediment column.

Received 10 October 2011; accepted 28 June 2012.

1. Priscu, J. *et al.* in *Polar Lakes and Rivers* (eds Vincent, W. F. & Laybourn-Parry, J.) 320 (Oxford University Press, 2009).
2. Ferraccioli, F., Armadillo, E., Jordan, T., Bozzo, E. & Corr, H. Aeromagnetic exploration over the East Antarctic Ice Sheet: a new view of the Wilkes Subglacial Basin. *Tectonophysics* **478,** 62–77 (2009).
3. Davie, M. K. & Buffett, B. A. A numerical model for the formation of gas hydrate below the seafloor. *J. Geophys. Res.* **106,** 497–514 (2001).
4. Wellsbury, P., Mather, I. & Parkes, R. J. Geomicrobiology of deep, low organic carbon sediments in the Woodlark Basin, Pacific Ocean. *FEMS Microbiol. Ecol.* **42,** 59–70 (2002).
5. Cragg, B. A. *et al.* Bacterial populations and processes in sediments containing gas hydrates (ODP Leg 146: Cascadia Margin). *Earth Planet. Sci. Lett.* **139,** 497–507 (1996).
6. Colwell, F. S. *et al.* Estimates of biogenic methane production rates in deep marine sediments at Hydrate Ridge, Cascadia margin. *Appl. Environ. Microb.* **74,** 3444–3452 (2008).
7. Parkes, R. J. *et al.* Bacterial biomass and activity in deep sediment layers from the Peru margin. *Phil. Trans. R. Soc. Lond. A* **331,** 139–153 (1990).
8. Yoshioka, H., Sakata, S., Cragg, B. A., Parkes, R. J. & Fujii, T. Microbial methane production rates in gas hydrate-bearing sediments from the eastern Nankai Trough, off central Japan. *Geochem. J.* **43,** 315–321 (2009).
9. Kotsyurbenko, O. R. *et al.* Acetoclastic and hydrogenotrophic methane production and methanogenic populations in an acidic West-Siberian peat bog. *Environ. Microbiol.* **6,** 1159–1173 (2004).
10. Franzmann, P. D., Roberts, N. J., Mancuso, C. A., Burton, H. R. & McMeekin, T. A. Methane production in meromictic Ace Lake, Antarctica. *Hydrobiologia* **210,** 191–201 (1991).
11. Smith, R., Miller, L. & Howes, B. The geochemistry of methane in Lake Fryxell, an amictic, permanently ice-covered, antarctic lake. *Biogeochemistry* **21,** 95–115 (1993).
12. Ellis-Evans, J. C. Methane in maritime Antarctic freshwater lakes. *Polar Biol.* **3,** 63–71 (1984).
13. Archer, D. Methane hydrate stability and anthropogenic climate change. *Biogeosciences* **4,** 521–544 (2007).
14. Wadham, J. L., Tranter, M., Tulaczyk, S. & Sharp, M. Subglacial methanogenesis: a potential climatic amplifier? *Glob. Biogeochem. Cycles* **22,** GB2021 (2008).
15. Lanoil, B. *et al.* Bacteria beneath the West Antarctic Ice Sheet. *Environ. Microbiol.* **11,** 609–615 (2009).
16. Boyd, E. S., Skidmore, M., Mitchell, A. C., Bakermans, C. & Peters, J. W. Methanogenesis in subglacial sediments. *Environ. Microbiol. Rep.* **2,** 685–692 (2010).
17. Pattyn, F. Antarctic subglacial conditions inferred from a hybrid ice sheet/ice stream model. *Earth Planet. Sci. Lett.* **295,** 451–461 (2010).
18. Maule, C. F., Purucker, M. E., Olsen, N. & Mosegaard, K. Heat flux anomalies in Antarctica revealed by satellite magnetic data. *Science* **309,** 464–467 (2005).
19. Claypool, G. F., Lorensen, T. D. & Johnson, C. A. Authigenic carbonates, methane generation, and oxidation in continental rise and shelf sediments, ODP leg 188, sites 1165 and 1166, offshore Antarctica (Prydz Bay). *Proc. ODP Sci. Results* **188,** 1–15 (2004).
20. Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23,** GB2023 (2009).
21. Houghton, R. A. Balancing the global carbon budget. *Annu. Rev. Earth Planet. Sci.* **35,** 313–347 (2007).
22. Meyers, P. A. & Ishiwatari, R. Lacustrine organic geochemistry—an overview of indicators of organic matter sources and diagenesis in lake sediments. *Org. Geochem.* **20,** 867–900 (1993).
23. Boudreau, B. P. & Ruddick, B. R. On a reactive continuum representation of organic-matter diagenesis. *Am. J. Sci.* **291,** 507–538 (1991).
24. Reeburgh, W. S., Whalen, S. C. & Alperin, M. J. In *Microbial Growth on C1 Compounds* (eds Murrell, J. C. & Kelly, D. P.) 1–14 (Intercept Ltd, 1993).
25. Etiope, G. & Klusman, R. W. Geologic emissions of methane to the atmosphere. *Chemosphere* **49,** 777–789 (2002).
26. Torres, M. E. *et al.* Fluid and chemical fluxes in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR. I: Hydrological provinces. *Earth Planet. Sci. Lett.* **201,** 525–540 (2002).
27. Wallmann, K., Drews, M., Aloisi, G. & Bohrmann, G. Methane discharge into the Black Sea and the global ocean via fluid flow through submarine mud volcanoes. *Earth Planet. Sci. Lett.* **248,** 545–560 (2006).
28. Boswell, R. & Collett, T. S. Current perspectives on gas hydrate resources. *Energy Environ. Sci.* **4,** 1206–1215 (2011).
29. Pollard, D. & DeConto, R. M. Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* **458,** 329–332 (2009).
30. Dyke, A. S., Prest, V. K. & Narraway, J. D. in *Map/Geological Survey of Canada 1702A* (Geological Survey of Canada, 1987).
31. Dickens, G. R., Paull, C. K. & Wallace, P. Direct measurement of in situ methane quantities in a large gas-hydrate reservoir. *Nature* **385,** 426–428 (1997).
32. Barker, P. F., Camerlenghi, A. & Acton, G. D. Leg 178 Summary. *Proc ODP Init. Rep.* **178,** 60 (1999).

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

Acknowledgements This research was funded by the Natural Environment Research Council (UK—NERC grant NE/E004016/1) and the National Science Foundation WISSARD project (NSF-AISS 0839142). Support to J.L.W. was also provided by the Leverhulme Trust via a Phillip Leverhulme award and to S.A. by the Netherlands Organisation for Scientific Research (NWO). We acknowledge NSERC and Antarctica New Zealand for financial and logistic support for sampling in Antarctica and the Polar Continental Shelf Project for financial and logistic support for sampling in Arctic Canada. We thank S. Fitzsimons for assistance with sampling in Antarctica. A.D. was funded by an NSERC Undergraduate Student Research Award. This research used data provided by the Ocean Drilling Program. The Ocean Drilling Program is sponsored by the US National Science Foundation and participating countries under the management of the Joint Oceanographic Institutions, Inc. We thank C. Ruppel for comments on this manuscript.

Author Contributions J.L.W. wrote the paper and directed the work, and led the sample collection in Greenland. S.A. did the numerical modelling and contributed to manuscript preparation. S.T. assisted with the modelling and contributed to manuscript preparation. M.S. contributed to the writing of the manuscript and did experimental work. J.T. did the initial design of incubation experiments, laboratory analysis of incubation experiments, and sample collection. G.P.L. performed laboratory analysis of the incubation experiments, and did sample collection. E.L. performed laboratory analysis of incubation experiments. A.D. performed laboratory analysis of the incubation experiments. M.T. assisted with the manuscript and modelling calculations. M.J.S. added input to the incubation experiments, and did sample collection of Antarctic subglacial material. A.M.A. assisted with writing the manuscript and advised upon incubation experiments. A.R. assisted with manuscript preparation and numerical modelling. C.B. assisted with the laboratory analysis of the incubation experiments.

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 J.L.W. (j.l.wadham@bris.ac.uk).

LETTER

doi:10.1038/nature11370

'Melt welt' mechanism of extreme weakening of gabbro at seismic slip rates

Kevin M. Brown[1] & Yuri Fialko[1]

Laboratory studies of frictional properties of rocks at slip velocities approaching the seismic range (~0.1–1 m s^{-1}), and at moderate normal stresses (1–10 MPa), have revealed a complex evolution of the dynamic shear strength, with at least two phases of weakening separated by strengthening at the onset of wholesale melting[1–4]. The second post-melting weakening phase is governed by viscous properties of the melt layer and is reasonably well understood[5,6]. The initial phase of extreme weakening, however, remains a subject of much debate. Here we show that the initial weakening of gabbro is associated with the formation of hotspots and macroscopic streaks of melt ('melt welts'), which partially unload the rest of the slip interface. Melt welts begin to form when the average rate of frictional heating exceeds 0.1–0.4 MW m^{-2}, while the average temperature of the shear zone is well below the solidus (250–450 °C). Similar heterogeneities in stress and temperature are likely to occur on natural fault surfaces during rapid slip, and to be important for earthquake rupture dynamics.

Several mechanisms have been proposed to explain the pre-melting weakening, including flash melting[7,8], lubrication by hydrous films coating the gouge particles[9,10], silica gel formation[2], nano-powder lubrication, thermal decomposition and degassing[11]. Of the proposed mechanisms, only the first two could potentially explain extreme weakening in crystalline mafic rocks[1,4]. To investigate the nature of extreme dynamic weakening before the onset of wholesale melting, we performed a series of medium- to high-speed friction experiments on diabase (fine-grained gabbro). The tests were conducted using a rotary shear apparatus (see Methods for details of the experimental setup). Figure 1 shows a typical experimental record. At velocities well below the critical value (defined below) of ~0.1 m s^{-1}, the coefficient of friction drops by 10–20% from a peak static value to a nearly constant kinetic value of ~0.6–0.8—consistent with Byerlee's law[12]—within a few tens of seconds, or a slip distance of several metres. Our focus in this study, however, is a subsequent dramatic weakening that reduces the effective coefficient of friction by a factor of 2 or more (Fig. 1 and Supplementary Fig. 2). This extreme weakening is observed above a critical velocity V_c of ~0.06–0.2 m s^{-1}, in agreement with previously reported high-speed friction data[1–4]. Most of our tests were conducted near the critical velocity V_c to explore relationships between V_c, the normal stress, the slip distance, and the average temperature of the shear zone. The tests revealed a systematic decrease in V_c with increasing normal stress (Supplementary Fig. 3). The boundary between weakened and unweakened regimes is fairly sharp; near this boundary, a velocity increment of as little as 10^{-2} m s^{-1} is often sufficient to trigger weakening, and the critical increment appears to decrease at higher normal stress (Supplementary Fig. 3). The observed dependence of the onset of weakening on normal stress (Supplementary Fig. 3) and slip distance (Fig. 1 and Supplementary Fig. 2) is suggestive of a thermally activated mechanism. Figure 2 shows the dissipative power density (work done against friction per unit slip area per unit time, τV, where τ is the average shear stress, and V is the slip velocity) plotted against the time to the onset of weakening, t_c. The data reveal a critical power density, demarcating the boundary between the weakened and unweakened states, of ~0.1–0.4 MW m^{-2}, in agreement with previous estimates[13]. The power density associated with seismic ruptures, $\tau V \approx (10^1–10^2)$ MW m^{-2}, greatly exceeds the critical power density inferred from laboratory data (Fig. 2), so the experimentally deduced conditions for the onset of weakening are certainly met during natural earthquakes, provided that seismic slip is sufficiently localized[14,15]. The time until the onset of weakening, t_c, is greatly reduced if frictional losses τV exceed the critical power density τV_c.

To investigate the nature of the weakening mechanism, we evaluated temperature on the slip interface using a combination of direct temperature measurements inside the sample and numerical modelling (see Methods). Figure 3 shows the observed critical velocity, V_c, as a function of the background temperature of the shear zone, T_b, at the onset of weakening. As can be seen in Fig. 3, weakening occurs when the average temperature of the shear zone varies from ~270–350 °C (at normal stress of ~1 MPa and slip rate greater than 0.1 m s^{-1}) to ~450–550 °C (at normal stress of ~4–5 MPa and slip rate less than 0.1 m s^{-1}). These average temperatures are well above the boiling point of aqueous fluids, thus ruling out processes associated with nano-layers of absorbed water on fine gouge particles[9,10]. The data shown in Fig. 3 can be also used to test the flash melting hypothesis[7,8]. Flash melting is believed to occur when the heat generated from rubbing of microscopic, highly stressed asperities is sufficient to thermally weaken the asperities on a timescale of a transient contact. The critical weakening velocity V_c at the onset of flash melting is given by[7]

$$V_c = \frac{\pi\kappa}{a}\left(\frac{\rho c(T_w - T_b)}{\tau_c}\right)^2 \quad (1)$$

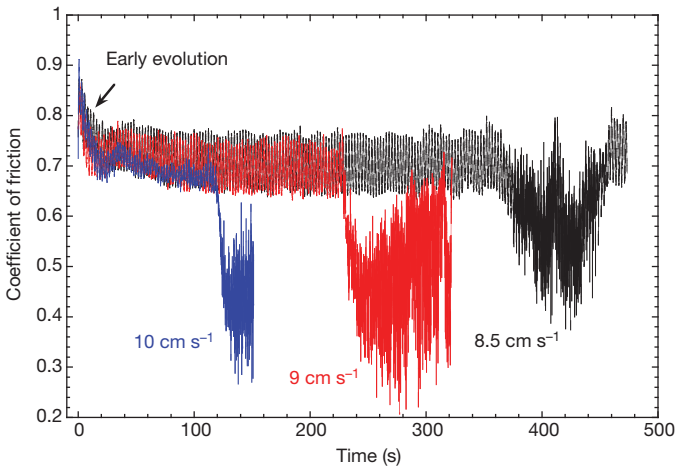


Figure 1 | Experimentally measured evolution of the coefficient of friction at three different 'near-critical' velocities. Blue, red and black lines show data from tests 90.31, 90.32 and 90.33, respectively (see Supplementary Table 1), with velocities as labelled. Normal stress equals 3 MPa. Note the onset of strong dynamic weakening at ~120, 230 and 360 s, respectively.

[1]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

where a is the characteristic size of contacts, κ is thermal diffusivity, c is specific heat, ρ is density, τ_c is the characteristic strength of contacts, and T_w is the weakening temperature, at which contacts lose strength. The flash melting hypothesis predicts that the critical weakening velocity should be independent of normal stress, in apparent contradiction with observations shown in Supplementary Fig. 3. Note, however, that higher normal stress can affect the onset of weakening indirectly, through increases in ambient temperature T_b. We used equation (1) to calculate the critical weakening velocity predicted by the flash melting model, assuming that the background temperature equals the estimated temperature of the shear zone at the onset of weakening. The respective model predictions are shown in Fig. 3. The best-fitting values of the asperity size a (2–5 μm) are within the observed range of gouge particle sizes in our scanning electron microscope (SEM) images (1–10 μm; see Fig. 4b—note that there is a poorly resolved nanoscale component in the wear products), so the flash melting hypothesis can in principle explain the inverse scaling of weakening velocity with normal stress (Supplementary Fig. 3). However, additional observations suggest that a different physical mechanism may be responsible for the strong dynamic weakening in our experiments. Stopping slip just after the onset of weakening revealed dark glassy streaks of sub-millimetre width (hereafter referred to as 'melt welts'; Fig. 4a and Supplementary Fig. 7a), indicating the onset of macroscopic melting. Further increases in velocity, normal stress or slip duration resulted in coalescence and expansion of melt streaks into broader rings (Supplementary Figs 7b,c). We propose that the observed melt welts (Fig. 4, Supplementary Fig. 7) result from stress heterogeneities that spontaneously develop and grow on the slip interface, possibly owing to a feedback between local temperature increases and thermal expansion. We verified this hypothesis by placing temperature sensors 2 mm away from the shear zone, spanning the ring-shaped sample both across and along the slip direction (Fig. 5). These measurements showed that below the critical velocity, or before weakening, the temperature of the slip interface increases relatively uniformly. A sharp transition to a highly heterogeneous temperature distribution then occurs at time t_c, corresponding to the onset of weakening (Fig. 5). This behaviour may be similar to 'hotspotting' in brakes and clutches, well known in engineering tribology[16–18]. If so, this mechanism is different from flash melting[7,19,20] in that regions of high temperature do not need to be microscopic, highly transient and uniformly distributed on the slip interface. By comparing temperature records from thermocouples to the location of glassy streaks in experiments stopped shortly after the onset of weakening, we confirmed that, on weakening, elevated temperatures were localized near melt welts, while the surrounding areas of the slip interface cooled. Anti-correlated variations in temperature across the ring sample in the post-weakening regime over a period of several seconds (Fig. 5) indicate that hotspots can migrate across the slip interface, presumably marking the birth and destruction of incipient melt welts.

Melt welts initially occupy a small fraction of the total contact area (Fig. 4), and cannot appreciably reduce the effective coefficient of friction (Fig. 1) if shear stress on the rest of the contact area remains high. We therefore suggest that the observed dynamic weakening results primarily from volumetric expansion in melt welts, which reduces the effective normal stress on the rest of the slip interface. Unloading of the slip interface adjacent to melt welts is evident from dramatically reduced rates of frictional heating and reversals in temperature trends (Fig. 5). Furthermore, inspection of normal stress at the onset of weakening revealed small increases in compression (Supplementary Fig. 8), consistent with net dilation in the shear zone. We confirmed a concomitant dilation of several micrometres by precise measurements of the rate of shortening across the sample (Supplementary Fig. 9).

We speculate that hotspots may initially develop as a result of a positive feedback between heating and dilatancy (for example, thermoelastic instability[17,18]) or at the site of local stress concentrators such as ploughing gouge particles (much like galling in metals[21]; see Fig. 4) that are sufficiently long-lived to raise local temperature by several hundred degrees and initiate melting. Thus, the onset of weakening is controlled by both the average temperature of the slip zone (which increases during slip) and the transient heating of hotspots. The resulting melt welts may undergo further dilation owing to a volume increase associated with the phase change. A net reduction in the effective coefficient of friction may occur if normal stress on the unmelted part of the slip interface decreases while shear stress within melt welts does not increase in proportion to the locally increased normal stress. An important question is whether melt welts can maintain excess pressure over time. In the case of shear between two rigid surfaces, pressurized patches of melt should eventually be smeared across the slip interface. However, in the presence of gouge, lateral

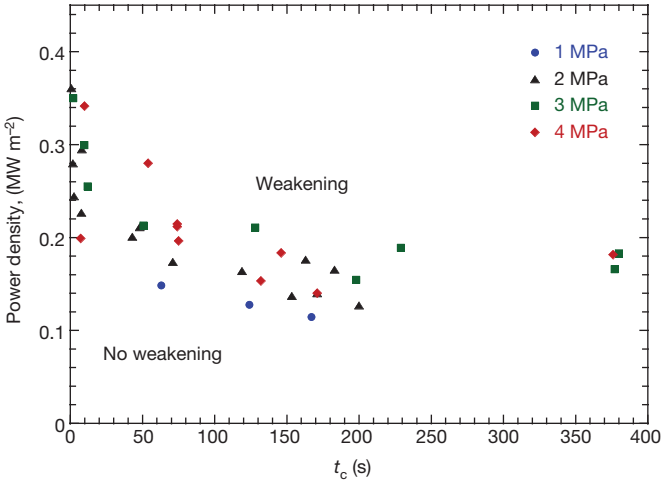


Figure 2 | Power density versus time until the onset of weakening, for experiments at four different values of normal stress. The normal stress (coloured symbols) is held constant during each experiment. Delayed weakening (larger t_c) is associated with lower power density, possibly owing to elevated temperature of the slip interface and interplay between rate- and temperature-dependent weakening mechanisms. Rapid weakening occurs when power density exceeds the threshold value denoted by coloured symbols, whereas no weakening is expected at lower values of power density.

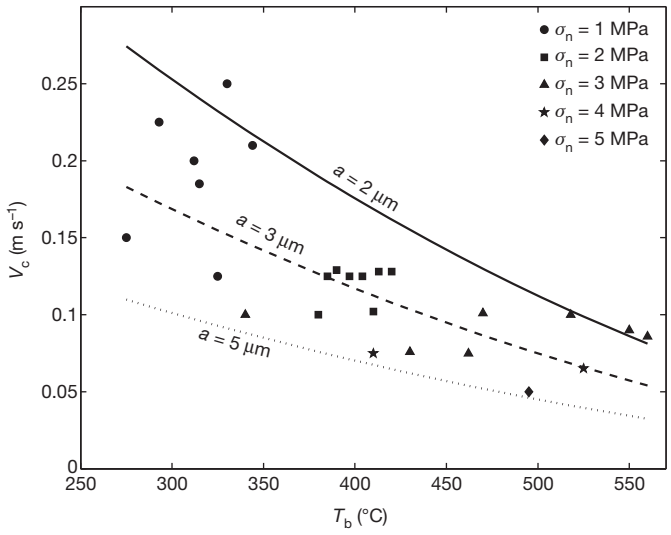


Figure 3 | Observed critical slip velocity as a function of the estimated average temperature of the slip interface at the time of weakening. Symbols denote normal stress; lines show predictions of the flash melting model (equation (1)) for different assumed values of the asperity size a. Other parameters used in the calculations include: $\tau_c = 3$ GPa, $T_w = 900$ °C (ref. 7), and a thermal diffusivity of 5.5×10^{-7} m s^{-2} (see Supplementary Table 2).

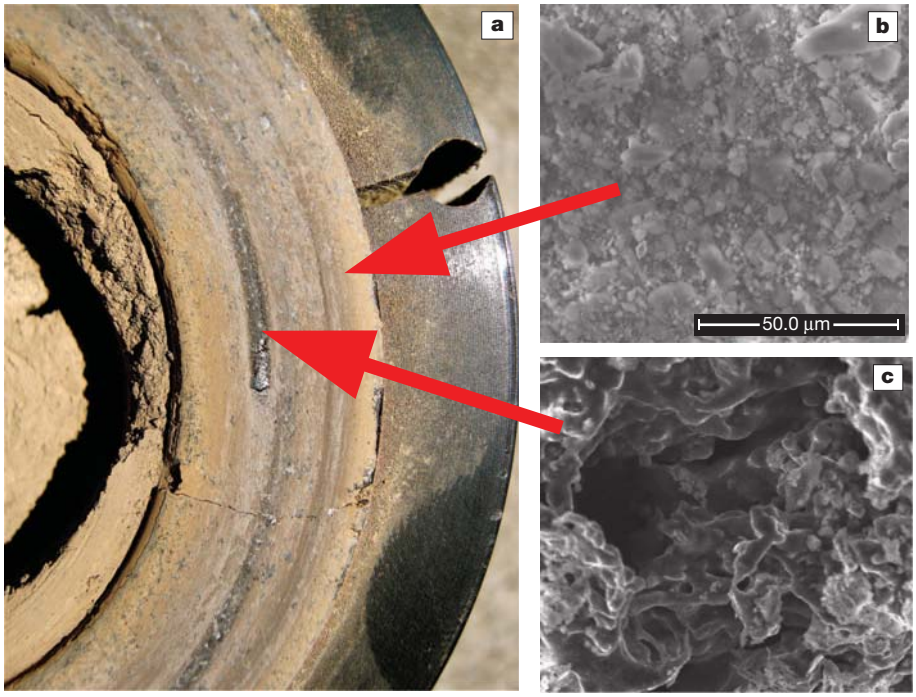


Figure 4 | Initiation of melt welts at the slip interface during the onset of dynamic weakening. Data from test 97.4. **a**, Photograph of the slip interface. **b**, **c**, SEM images of the areas shown by red arrows in **a**. **b**, Regular, unmolten surface. Note gouge particles. **c**, Glassy streak area. Dark, smooth surfaces, a scarcity of clasts, and 'ropy' flow structures are indicators of local melting. Scale bar, 50 μm; same scale applies to panel **c**.

propagation of melt blobs may be limited by the strength of the gouge layer. The melt extrusion would require displacement (or absorption) of the gouge particles ahead of the advancing melt front, thus allowing melt welts to maintain excess pressure and push the shear zone apart. Increases in slip velocity, normal stress or total slip distance lead to widening and coalescence of melt welts (Supplementary Fig. 7), and ultimately to wholesale melting. As growing melt welts occupy a progressively larger fraction of the slip interface, thereby increasing the true contact area and reducing the contribution of unmelted (low-normal-stress) contact, the average shear stress may increase. The onset of strengthening in our experiments is manifested by visible flashes at the edges of the slip zone, and the peak strength is associated with formation of a continuous, through-going layer of melt that covers the entire contact, in agreement with previously published results[1,3,4]. Note that transient strengthening before wholesale melting observed in high-speed, low-normal-stress friction experiments cannot be readily explained by the flash melting model. We also point out that the melt-welt mechanism does not require an extreme strain localization in the gouge layer, nor special assumptions about the size and strength of gouge particles.

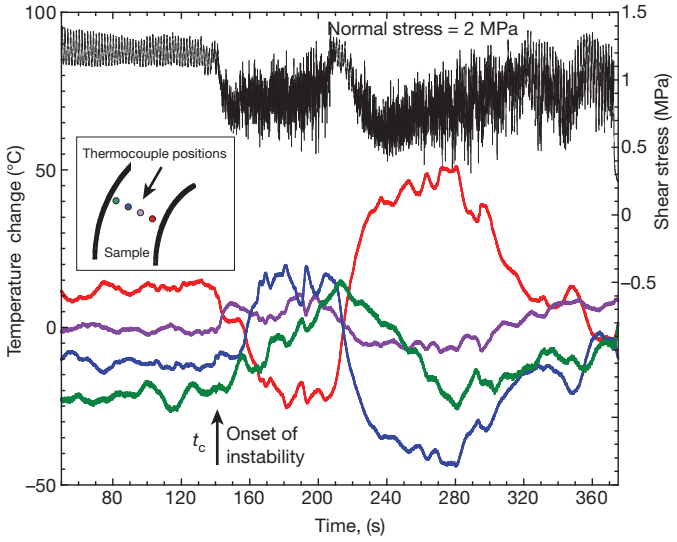


Figure 5 | Evolution of shear stress and temperature through the onset of weakening. Black line, right axis: measured shear stress. Coloured lines, left axis: temperature deviations recorded by individual thermocouples with respect to the mean value for the four thermocouples (~280 °C), at 2 mm distance from the slip interface. Thermocouple positions are shown in the inset. Note anticorrelated variations in temperature (for example, blue and red lines), probably indicating initiation and termination of melt welts at different locations on the interface. Data from test 98.20; slip velocity $V = 0.125\ \mathrm{m\ s^{-1}}$.

It is reasonable to assume that if extreme stress and temperature heterogeneities are developing and growing spontaneously on a nominally flat frictional interface in carefully controlled laboratory experiments, they are also present on rough fault surfaces during seismic slip. Extrapolation of critical velocities measured in our experiments to conditions at the base of the brittle layer, where most large earthquakes nucleate, implies that strong dynamic weakening may occur at much earlier stages of seismic rupture than commonly believed. Assuming that the slip instability is highly localized and that the scalings inferred from low-normal-stress data (Fig. 2, Supplementary Fig. 3) hold, strong dynamic weakening at seismogenic depths may initiate at slip velocities as small as several millimetres per second. At low normal stress, the melt-welt mechanism may induce strong dynamic weakening on shallow parts of seismogenic faults that are nominally velocity-strengthening and prone to creep in the interseismic period. The onset of melt-welt weakening requires the rate of energy dissipation on a slip interface to be greater than ~0.1–0.35 MW $\mathrm{m^{-2}}$ (Fig. 2). These values are similar to the critical power density reported in a recent compilation of high-speed laboratory measurements of rock friction[13]. Further experimental studies (in particular, at higher normal stress) are required to verify if the data shown in Fig. 2 and Supplementary Fig. 3 can indeed be extrapolated. The proposed mechanism does not preclude the occurrence of other potential mechanisms of strong dynamic weakening (for example, aquathermal pressurization). In fact, such additional mechanisms may be required

to reconcile the inferred low dynamic strength of earthquake ruptures with the scarcity of pseudotachylites on exhumed natural faults[5,7,22,23].

METHODS SUMMARY

Our high-speed friction experiments were conducted on a modified industrial lathe (Supplementary Fig. 1) at normal stresses between 0.5 and 5 MPa and slip velocities of 10^{-2} to 0.5 m s^{-1}. The rock samples consisted of a ring with internal diameter $D_1 = 54$ mm and external diameter $D_2 = 80$ mm on the rotating side of the apparatus, and a square plate on the stationary side embedded within the non-rotating flywheel (Supplementary Fig. 1). Both the ring and the plate are 12 mm thick. The ring geometry was chosen to minimize variations in slip rate and temperature across the slip interface. To monitor frictional heating and temperature variations inside the sample, we placed thermocouples in small holes drilled laterally into the side of the non-rotating plate at three different distances from the slip interface, as well as at four positions equally spaced across the sample in the radial direction at 2 mm distance from the slip interface (Fig. 5 inset). Numerical simulations were performed using the finite-element program Abaqus/Simulia[24].

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

Received 14 December 2011; accepted 29 June 2012.

1. Tsutsumi, A. & Shimamoto, T. High-velocity frictional properties of gabbro. *Geophys. Res. Lett.* **24,** 699–702 (1997).
2. Goldsby, D. & Tullis, T. Low frictional strength of quartz rocks at subseismic slip rates. *Geophys. Res. Lett.* **29** (2002).
3. Di Toro, G., Goldsby, D. L. & Tullis, T. E. Friction falls towards zero in quartz rock as slip velocity approaches seismic rates. *Nature* **427,** 436–439 (2004).
4. Hirose, T. & Shimamoto, T. Fractal dimension of molten surfaces as a possible parameter to infer the slip-weakening distance of faults from natural pseudotachylytes. *J. Struct. Geol.* **25,** 1569–1574 (2003).
5. Fialko, Y. & Khazan, Y. Fusion by earthquake fault friction: Stick or slip? *J. Geophys. Res.* **110,** B12407 (2005).
6. Nielsen, S., Di Toro, G., Hirose, T. & Shimamoto, T. Frictional melt and seismic slip. *J. Geophys. Res.* **113,** B01308 (2008).
7. Rice, J. R. Heating and weakening of faults during earthquake slip. *J. Geophys. Res.* **111,** B05311 (2006).
8. Goldsby, D. & Tullis, T. Flash heating leads to low frictional strength of crustal rocks at earthquake slip rates. *Science* **334,** 216–218 (2011).
9. Reches, Z. & Lockner, D. A. Fault weakening and earthquake instability by powder lubrication. *Nature* **467,** 452–455 (2010).
10. Sammis, C., Lockner, D. & Reches, Z. The role of adsorbed water on the friction of a layer of submicron particles. *Pure Appl. Geophys.* **168,** 2325–2334 (2011).
11. Han, R., Shimamoto, T., Hirose, T., Ree, J.-H. & Ando, J. Ultralow friction of carbonate faults caused by thermal decomposition. *Science* **316,** 878–881 (2007).
12. Byerlee, J. Friction of rock. *Pure Appl. Geophys.* **116,** 615–626 (1978).
13. Di Toro, G. *et al.* Fault lubrication during earthquakes. *Nature* **471,** 494–498 (2011).
14. Chester, F. M. & Chester, J. S. Ultracataclasite structure and friction processes of the Punchbowl fault, San Andreas system, California. *Tectonophysics* **295,** 199–221 (1998).
15. Fialko, Y. Temperature fields generated by the elastodynamic propagation of shear cracks in the Earth. *J. Geophys. Res.* **109,** B01303 (2004).
16. Dow, T. Thermoelastic effects in a thin sliding seal — a review. *Wear* **59,** 31–52 (1980).
17. Anderson, A. & Knapp, R. Hot spotting in automotive friction systems. *Wear* **135,** 319–337 (1990).
18. Lee, K. & Barber, J. Frictionally excited thermoelastic instability in automotive disk brakes. *J. Tribol.* **115,** 607–614 (1993).
19. Bowden, F. B. & Persson, P. A. Deformation heating and melting of solids in high speed friction. *Proc. R. Soc. Lond. A* **260,** 433–458 (1960).
20. Molinari, A., Estrin, Y. & Mercier, S. Dependence of the coefficient of friction on the sliding conditions in the high velocity range. *J. Tribol.* **121,** 35–41 (1999).
21. Schedin, E. & Lehtinen, B. Galling mechanisms in lubricated systems: a study of sheet metal forming. *Wear* **170,** 119–130 (1993).
22. Brune, J. N., Henyey, T. & Roy, R. Heat flow, stress, and rate of slip along San Andreas fault, California. *J. Geophys. Res.* **74,** 3821–3827 (1969).
23. Lapusta, N., Rice, J. R., Ben-Zion, Y. & Zheng, G. Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction. *J. Geophys. Res.* **105,** 23765–23789 (2000).
24. Abaqus/Simulia v.6.11; available at http://www.3ds.com/products/simulia/overview/ (Dassault Systèmes, 2012).

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

Acknowledgements We thank D. Lockner for comments that improved this manuscript. The SIO Marine Science Development Center provided the lathe used in our experiments. This work was supported by NSF (grant EAR-0838255).

Author Contributions K.M.B. built the apparatus, K.M.B. and Y.F. designed and conducted experiments, Y.F. performed numerical modelling.

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 K.M.B. (kmbrown@ucsd.edu) or Y.F. (yfialko@ucsd.edu).

METHODS

Our high-speed rotary friction apparatus was developed on a modified industrial lathe chassis (Supplementary Fig. 1). The apparatus allows measurements of friction at slip velocities between 0.003 and 2 m s^{-1}, normal stress up to 5 MPa, and unlimited total displacement. Modifications to the lathe chassis included removing excess components, tightening the clutch and attaching a stronger, digitally controlled drive motor. The rock samples consisted of a ring with internal diameter $D_1 = 54$ mm and external diameter $D_2 = 80$ mm on the rotating side of the apparatus, and a square plate on the stationary side embedded within the non-rotating flywheel (Supplementary Fig. 1). Both the ring and the plate are 12 mm thick. The ring geometry was chosen to minimize variations in slip rate and temperature across the slip interface.

The samples were ground to an initial surface roughness of 60–80 grit and had an initial parallelism of the mating surfaces of the order of micrometres over tens of centimetres. Before each experiment, we ran the samples in for several tens of metres to develop a gouge layer (typically having thickness of a fraction of a millimetre throughout the experiment). In each experiment, the sample was first put under normal load, and then accelerated to the target speed. The reported slip velocity corresponds to that in the middle of the ring sample, $V = \pi(D_1 + D_2)\omega/2$, where ω is the revolution speed of the sample holder, in revolutions per second. Shear and normal stress were calculated by dividing the tangential and axial forces acting on a sample by the nominal contact area, $\pi(D_2^2 - D_1^2)/4$.

Considerable efforts were made to increase the stiffness and inertial dampening of the machine, to prevent chatter and high-frequency vibrations during the tests. A significant chatter was observed in the initial low-stiffness configuration of the apparatus. Sampling at high frequency (>5,000 Hz) showed that the chatter in fact resulted from stick–slip behaviour, even at nominal slip rates in excess of 1 m s^{-1}. Chatter resulted in instantaneous peak slip rates an order of magnitude higher than the nominal slip speeds. After modification of the apparatus, involving stiffening of the loading shaft and adding a massive fly-wheel, most of the high-frequency chatter was suppressed. Still remaining are periodic oscillations in torque and normal stress, most likely owing to imperfectly parallel contacts (see, for example, Fig. 1 and Supplementary Fig. 2). The magnitude of the oscillations typically changes with time, increasing and decreasing as wear modifies the geometry of the slip interface. The onset of weakening in our experiments is manifested in a transition to irregular, high-frequency oscillations (Supplementary Fig. 4), and sometimes a barely audible rumble.

Because temperature on the slip interface could not be measured directly, we evaluated the average surface temperature using a numerical model constrained by temperature measurements inside the sample. Thermocouples were placed in small holes drilled laterally into the side of the non-rotating plate at three different distances from the slip interface, equally spaced along the contact with the ring sample, as well as at four positions equally spaced across the sample in the radial direction at 2 mm distance from the slip interface (Fig. 5 inset).

Our model explicitly accounted for frictional heating on the slip interface, unsteady heat transfer inside the apparatus, and convective cooling on the external surfaces. We assumed axial symmetry and built a finite-element mesh of the rock sample, sample holder and adjacent parts of the apparatus (Supplementary Fig. 5). The model used experimentally measured time-dependent shear stress τ and slip velocity V as input to calculate the rate of mechanical energy done against friction (τV). The latter was applied as a heat-flux boundary condition on the slip interface, assuming that all mechanical energy is ultimately transferred into heat. A good agreement between the modelled and observed temperatures inside the sample (Supplementary Fig. 6) confirmed that frictional losses are converted primarily to thermal energy, and that surface energy associated with production of gouge particles is negligible. Convective cooling at the contacts with air was taken into account using the Robin boundary condition, $q = h(T - T_a)$, where q is the heat flux across the surface, $h = 30$ W m^{-2} °C^{-1} is the heat transfer coefficient, and $T_a = 22$ °C is the ambient air temperature. We used the 'gap conductance' boundary condition at the contacts between the sample and sample holder, with the heat transfer coefficient proportional to the applied normal stress. Supplementary Table 2 summarizes the thermophysical properties of gabbro and steel used in our simulations. The model parameters were calibrated using data from low-speed experiments, in which no strong weakening occurred. Supplementary Fig. 6 shows a validation example in which temperatures observed in one of the experimental runs (not used to calibrate the model parameters) are compared to model predictions at the respective positions of temperature sensors inside the sample. A good agreement between the observed and calculated temperatures implies that inferences of average temperature on the shear interface are robust.

To measure the rate of wear and to verify the inferred dilatancy of the shear zone at the onset of weakening, we attached three linear variable differential transformers (LVDTs), two equidistant on either side of the sample (see Supplementary Fig. 1), and the third, with somewhat lower resolution, at a right angle to the first two, so that we could average the result to get the net axial displacement. The precision LVDTs have a maximum resolution of 0.5–1 μm, depending on integration time (typically 0.0167 Hz), and on the full voltage range setting chosen. LVDT measurements indicated that the periodic stress oscillations before the onset of weakening are caused by a sample misalignment of 5–10 μm. This is equivalent to changes of one or two grain sizes in the gouge layer generated in the shear zone during the course of the test. During the onset of strong dynamic weakening, LVDT data reveal that the shear zone is dilating by about 10 μm (Supplementary Fig. 9), consistent with the observed transient increases in normal stress (Supplementary Figs 8 and 9).

Numerical simulations were performed using the finite-element program Abaqus/Simulia[24].

LETTER

doi:10.1038/nature11399

Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice

Lisa S. Andersson[1]*, Martin Larhammar[2]*, Fatima Memic[2]*, Hanna Wootz[2]*, Doreen Schwochow[1], Carl-Johan Rubin[3], Kalicharan Patra[2], Thorvaldur Arnason[4], Lisbeth Wellbring[1], Göran Hjälm[3], Freyja Imsland[3], Jessica L. Petersen[5], Molly E. McCue[5], James R. Mickelson[5], Gus Cothran[6], Nadav Ahituv[7,8], Lars Roepstorff[9], Sofia Mikko[1], Anna Vallstedt[2], Gabriella Lindgren[1], Leif Andersson[1,3]* & Klas Kullander[2]*

Locomotion in mammals relies on a central pattern-generating circuitry of spinal interneurons established during development that coordinates limb movement[1]. These networks produce left–right alternation of limbs as well as coordinated activation of flexor and extensor muscles[2]. Here we show that a premature stop codon in the *DMRT3* gene has a major effect on the pattern of locomotion in horses. The mutation is permissive for the ability to perform alternate gaits and has a favourable effect on harness racing performance. Examination of wild-type and *Dmrt3*-null mice demonstrates that Dmrt3 is expressed in the dI6 subdivision of spinal cord neurons, takes part in neuronal specification within this subdivision, and is critical for the normal development of a coordinated locomotor network controlling limb movements. Our discovery positions *Dmrt3* in a pivotal role for configuring the spinal circuits controlling stride in vertebrates. The *DMRT3* mutation has had a major effect on the diversification of the domestic horse, as the altered gait characteristics of a number of breeds apparently require this mutation.

Horses show considerable variation in the pattern of locomotion. The three naturally occurring gaits in all equids are, in order of increasing speed, walk, trot and canter/gallop. Some horses can use alternate gaits, typically at intermediate speed, and 'gaitedness' is a trait upon which many specialized breeds have been developed. Based on variation in footfall pattern, timing and cadence, these alternate gaits can be generally divided into four categories: pace, regular rhythm ambling, lateral ambling and diagonal ambling (Supplementary Notes and Supplementary Table 1). Pace is a two-beat gait in which the horse moves the two legs on the same side of the body in a synchronized, lateral movement (Fig. 1a) in contrast to the trot, where the diagonal front and hind legs move forward and backward together (Fig. 1b). Ambling gaits are four-beat gaits in which footfall pattern, foot placement and timing are often unique to specific breeds (Supplementary Notes and Supplementary Table 1). Tölt is a regular ambling gait characteristic of the Icelandic horse. Many Icelandic horses also have the ability to pace and test scores for pace show a bimodal distribution (Fig. 1c) and high heritability, in the range 0.60–0.73 (ref. 3).

A genome-wide association analysis using 30 Icelandic horses classified as four-gaited (walk, tölt, trot and gallop) and 40 classified as five-gaited (walk, tölt, trot, gallop and pace) revealed a highly significant association between the ability to pace and a single nucleotide polymorphism (SNP; BIEC2_620109) at nucleotide position 22967656 on chromosome 23 (Fig. 1d). The two flanking markers showed only weak association to the pacing phenotype, indicating that the mutation(s) underlying the association is located within the 684-kilobases interval chr23:22628976–23315071. We resequenced selected regions of the 684-kb interval in a panel of four- and five-gaited Icelandic horses. The five-gaited horses were all homozygous for a minimal 438-kb haplotype (chr23:22877015–23315071), that was inferred to be identical-by-descent (IBD; Supplementary Table 2). This region contains only three genes encoding different isoforms of the doublesex and mab-3 related transcription factors, DMRT1-3 (Fig. 1e). The DMRT family of transcription factors carry a DM (dsx and mab-3) DNA-binding domain, conferring sequence-specific DNA binding distinct from a classical zinc-finger[4].

We performed whole-genome resequencing of one four-gaited and one five-gaited Icelandic horse, homozygous for opposite alleles at the SNP associated with the ability to pace. Average sequence coverage of 30× was obtained and polymorphisms identified in the critical 438-kb interval were compiled (Supplementary Table 3). Homozygosity mapping using the sequenced five-gaited horse confirmed an IBD region of about 438 kb. In this interval, we identified 65 sequence differences (60 SNPs and five small insertions/deletions) unique to the five-gaited horse when comparing data for the two horses and the reference genome (Supplementary Table 4); no structural rearrangements were detected. We found five intronic or intragenic SNPs at sites showing some degree of evolutionary conservation, and a single base change at nucleotide position chr23:22999655 causing a premature stop at codon 301 in *DMRT3* (*DMRT3_Ser301STOP*; Fig. 1f). The allele is expected to encode a truncated protein lacking 174 amino acid residues of the full-length protein (Fig. 1g), of which 161 (92.5%) are identical between human and horse Dmrt3. *DMRT3_Ser301STOP* was evaluated as the candidate causative mutation.

We genotyped 352 additional Icelandic horses and found that all but one of the five-gaited horses were homozygous *A/A* for the *DMRT3* nonsense mutation (Table 1); further investigation of competition records revealed that this single discordant horse was most likely phenotypically misclassified. In contrast, only 31% of the four-gaited horses were homozygous *A/A* ($P = 2.4 \times 10^{-14}$). Thus, homozygosity for the *DMRT3* nonsense mutation is required for the ability to pace in this breed. The observation that a considerable number of homozygous mutant horses are considered four-gaited may reflect phenotype misclassifications, but more likely incomplete penetrance due to other genetic factors, maturity and environmental effects, in particular training.

The *DMRT3* genotype distribution across breeds was markedly dichotomous, with the mutation occurring at high frequency in all gaited breeds, whereas all tested non-gaited horses were homozygous wild type (Table 1), with the exception of horses used for harness

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-75124 Uppsala, Sweden. [2]Department of Neuroscience, Uppsala University, SE-75124 Uppsala, Sweden. [3]Department of Medical Biochemistry and Microbiology, Uppsala University, SE-75123 Uppsala, Sweden. [4]Faculty of Land and Animal Resources, The Agricultural University of Iceland, IS-311 Borgarnes, Iceland. [5]College of Veterinary Medicine, University of Minnesota, St Paul, Minnesota 55108, USA. [6]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas 77483, USA. [7]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California 94143, USA. [8]Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA. [9]Unit of Equine Studies, Swedish University of Agricultural Sciences, Uppsala, SE-75007 Sweden.

*These authors contributed equally to this work.

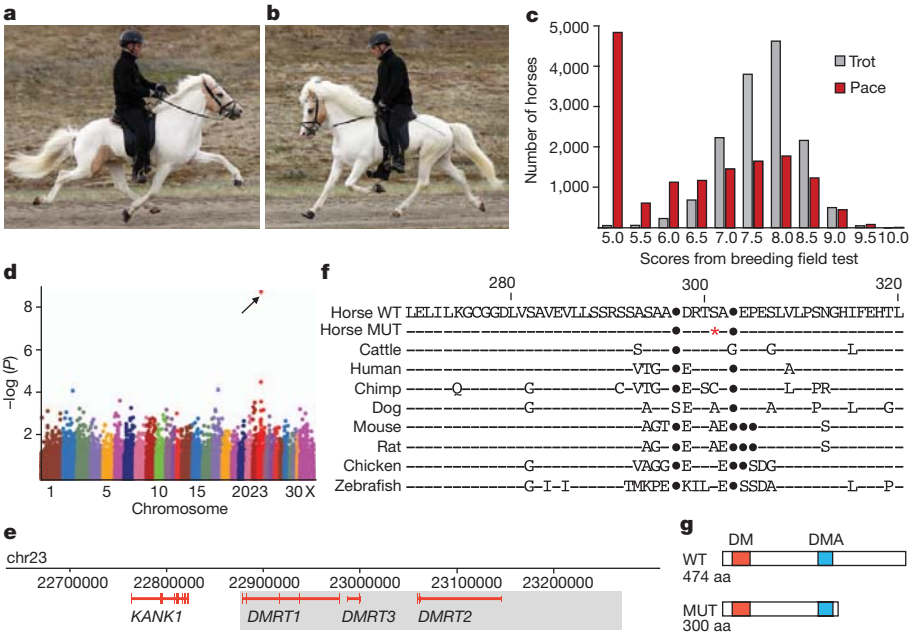


Figure 1 | Identification of a *DMRT3* mutation in horses. **a**, A pacing Icelandic horse, fore- and hindlegs on the same side of the body are synchronized. **b**, A trotting Icelandic horse, the diagonal fore- and hindlegs are synchronized. **c**, Distribution of breeding evaluation test scores for pace and trot in Icelandic horses. Score 5.0 indicates 'gait not shown'. **d**, Genome-wide association analysis revealed a highly significant association between the ability to pace and SNP BIEC2_620109 on chromosome 23 ($P_{raw} = 1.7 \times 10^{-9}$, corrected empirical *P*-value (EMP2) $= 2.0 \times 10^{-4}$, genome-wide significance). **e**, The 684 kb genomic interval associated with the *Gait* locus; the minimum *Gait* IBD region (438 kb) is shaded. **f**, Partial amino acid alignment of the predicted Dmrt3 protein in wild-type (WT) and mutant (MUT) horses and in other vertebrates. Horse nonsense mutation (red asterisk), sequence identities (dashes), insertions/deletions (dots). **g**, Schematics of wild-type and mutant Dmrt3. DM, zinc-finger like DNA binding module; DMA, protein domain of unknown function present in DMRT proteins.

racing (see below). Nearly all individuals from other gaited breeds were homozygous mutant, regardless of whether their four-beat alternate gait is characterized by lateral or diagonal couplets (Supplementary Table 1). Thus, the *DMRT3* mutation is permissive for the ability to perform alternate gaits, which can be either pace or four-beat ambling gaits. Although this mutation must be advantageous for gaited horses, it may be disadvantageous for others. In fact, Icelandic horse homozygous mutants had inferior scores for gallop and trot (Supplementary Table 5). Thus, there may be selection against the mutation in non-gaited horses bred for dressage, show jumping or high-speed gallop.

Table 1 | Allele frequency of the *DMRT3* nonsense mutation among horse populations

Breed	*n*	*p*(A)
Icelandic horses*		
Four-gaited†	124	0.65
Five-gaited	66	0.99
Random sample	162	0.89
Other gaited horses		
Kentucky mountain saddle horse	22	0.95
Missouri fox trotter	40	1.00
Paso fino	45	1.00
Peruvian paso	19	1.00
Rocky mountain horse	17	1.00
Tennessee walking horse	33	0.98
Non-gaited horses		
Arabian horse	18	0.00
Gotland pony	28	0.00
North-Swedish draft horse	31	0.00
Przewalski's horse	6	0.00
Shetland pony	20	0.00
Swedish ardennes	22	0.00
Swedish warmblood	64	0.00
Thoroughbred	29	0.00
Horses bred for harness racing		
Standardbred, trotter (Sweden)	270	0.97
Standardbred, trotter (USA)	57	1.00
Standardbred, pacer (USA)	40	1.00
French trotter (France)	47	0.77

* These do not include the horses used in the initial genome-wide association and therefore provide a replication of the highly significant association.
† Thirty-eight of the 124 four-gaited horses were homozygous *A/A*.
n, number of horses; *p*(A)=allele frequency of the *DMRT3* nonsense mutation.

We found a high frequency of the *DMRT3* mutation in horses bred for harness racing (Table 1). These horses have the ability to trot or pace at high speed without breaking into a gallop, the natural gait at high speed for horses. The American Standardbred was established in the 19th century and bred for harness racing. Competitions are held separately in trot or pace and assortative mating based on preferred gait has subdivided the breed into two populations, pacers and trotters. In contrast to the pattern in Icelandic horses, where homozygosity for *DMRT3_Ser301STOP* was associated with the ability to pace, both Standardbred pacers and trotters are homozygous for the mutation. Thus, the mutation may promote the ability to trot or pace at high speed and genetic modifiers determine the gait to which the horse is best suited.

The Swedish Standardbred is largely developed from the American Standardbred but is not completely fixed for the *DMRT3* mutation, probably owing to the import of French trotters, a breed with a fairly high frequency of the wild-type allele (Table 1). The segregation of the two alleles in the Swedish Standardbred provided an opportunity to examine the effect of the mutation on racing performance. The *DMRT3* mutation was associated with superior breeding values (BV) for racing performance ($BV_{CA} = 95.7 \pm 1.7$, $n = 17$; $BV_{AA} = 109.0 \pm 0.8$, $n = 206$; $P < 0.0001$) and increased earned prize money ($X_{CA} = 48{,}000 \pm$ US$35,000, $n = 17$; $X_{AA} = 161{,}000 \pm$ US$24,000, $n = 206$; $P_{one\text{-}sided} = 0.007$). We also genotyped 61 horses from one racing camp in a blind test; two of these had major difficulties in sustaining trot at high speed (Supplementary Fig. 1a, b and Supplementary Movie 1) and were heterozygous *C/A*, whereas all others were homozygous *A/A* ($P = 0.0005$).

Whereas the horse discovery demonstrates that *DMRT3* has an effect on gait coordination, studies of its possible role in locomotor

circuitry are more tractable in mice. We used *Dmrt3*-null mice[5] (Supplementary Fig. 2) and evaluated locomotion performances. Motor coordination and balance were largely normal in $Dmrt3^{-/-}$ mice (Supplementary Fig. 3a–c). In water, $Dmrt3^{-/-}$ mice spent less time swimming and showed frequent twitching limb movements rarely observed in controls (Fig. 2a and Supplementary Movie 2). Next, mice were placed on a TreadScan apparatus, which performs an automated and unbiased analysis to collect multiple gait parameters (Supplementary Table 6). $Dmrt3^{-/-}$ mice, but not control mice (wild-type littermates), had major difficulties running at higher velocities (Fig. 2b and Supplementary Movie 3). Gait analysis (Fig. 2c) revealed significantly increased stride length in all limbs of $Dmrt3^{-/-}$ mice (Fig. 2d). Swing times (flexion) were increased in all limbs, whereas stance time (extension) was increased in forelimbs (Fig. 2e, f). Moreover, propulsion time increased in all limbs, whereas brake time was decreased in hindlimbs, indicating that $Dmrt3^{-/-}$ mice may emphasize extension movements, resulting in a longer stride (Supplementary Table 6). Heterozygotes did not differ significantly from controls.

Development of weight-supported walking was similar in wild-type and $Dmrt3^{-/-}$ mice (Supplementary Fig. 3d, e), whereas limb coordination in neonatal mice, scored during air-stepping, was markedly different (Fig. 2g). At postnatal (P) day 4, we observed similar numbers of alternating step movements in all wild-type limbs as well as in $Dmrt3^{-/-}$ mice forelimbs. However, alternating hindlimb movements were almost absent in $Dmrt3^{-/-}$ mice (Fig. 2h), accompanied by increased uncoordinated step movements (Supplementary Fig. 3f). This early effect was emphasized by a similar phenotype at P1 (Supplementary Fig. 3g–i). Next, we analysed central pattern generator output in the isolated neonatal spinal cord using drug-induced fictive locomotion. Cords from wild-type mice generated a stable rhythm, whereas cords from $Dmrt3^{-/-}$ mice had uncoordinated and irregular firing rhythms as well as increased burst and interburst durations (Fig. 2i–k). Moreover, the coefficient of variation, a normalized measure of variability, increased two- to threefold (Supplementary Fig. 4). We analysed the rhythm relationship between the left-right (l/r) and flexion-extension (f/e) outputs using a continuous wavelet transform that measures the coherence between the spinal ventral roots with high resolution[6,7]. Wild-type and heterozygous mice had coherence frequencies around 0.4 Hz with clear coordinated left-right and flexion-extension alternation (coherence in wild-type l/r 95%, f/e 92%). In contrast, $Dmrt3^{-/-}$ mice showed more variable and lower frequency values (≈0.1 Hz), and bursts were non-coherent between left-right and flexion-extension (Fig. 2l, m). The coherence values were lower (l/r 23%, f/e 25%) and significantly reduced compared to wild-types ($P = 0.01$, l/r; $P = 0.01$, f/e). The remaining coherent activity showed no clear direction towards synchrony or alternation (Supplementary Fig. 4). Excitation-inhibition balance can be influenced by the glycine

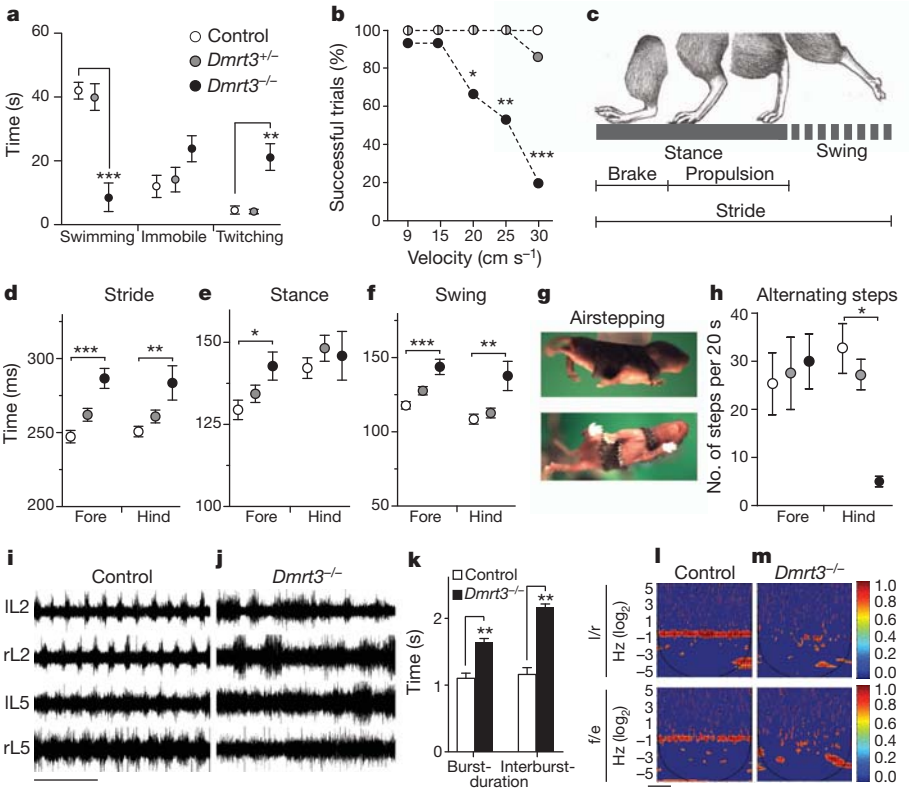


Figure 2 | Characterization of motor coordination in mice lacking *Dmrt3*. **a**, $Dmrt3^{-/-}$ mice showed decreased swimming duration ($P < 0.0001$), and an increase in twitching movements compared to control and $Dmrt3^{+/-}$ ($P = 0.0002$); $n = 5$ control and $Dmrt3^{-/-}$, $n = 7$ $Dmrt3^{+/-}$. Time spent immobile was similar between genotypes ($P = 0.13$). **b**, Mice were tested for their ability to run (>5 step cycles) at different treadmill speeds (9, 15, 20, 25, 30 cm s^{-1}) ($n = 15$ trials per genotype, five animals). $Dmrt3^{-/-}$ mice had difficulties at 20 cm s^{-1} ($P = 0.04$) with markedly reduced number of successful trials at 25 cm s^{-1} ($P = 0.006$) and 30 cm s^{-1} ($P < 0.0001$) compared to controls. **c**, Schematic drawing of the gait parameters analysed in treadmill locomotion. **d**–**f**, At 20 cm s^{-1} $Dmrt3^{-/-}$ showed increased stride (**d**, forelimb: $P < 0.0001$, hindlimb: $P = 0.006$), hind limb stance (**e**, $P = 0.02$) and swing (**f**, forelimb: $P < 0.0001$, hindlimb: $P = 0.001$) time duration compared to control ($n = 11$–12 trials per genotype, five animals). **g**, Images of a P4 mouse airstepping. **h**, $Dmrt3^{-/-}$ mice showed a decreased number of alternating hindlimb steps during airstepping ($P = 0.02$ compared to controls; $n = 4$ per genotype, except controls $n = 3$). Wild-type littermate controls (white), $Dmrt3^{+/-}$ (grey), $Dmrt3^{-/-}$ (black). Mean ± s.e.m. **i**–**m**, Fictive locomotion was recorded from ventral left (l) and right (r) lumbar (L) root 2 and 5 from neonatal spinal cords. **i**, **j**, Representative traces from control (**i**) and $Dmrt3^{-/-}$ (**j**) spinal cords. Time scale 10 s. **k**, $Dmrt3^{-/-}$ mice displayed increased burst and interburst durations compared to control animals, analysis made on L2s ($n = 6$ per genotype, $P = 0.02$, burst; $P = 0.001$, interburst). **l**, **m**, Coherence power spectra analysis of left/right (l/r) and flexor/extensor (f/e) recordings (colour-graded scale; $n = 3$–5 per genotype). Time scale 100 s.

re-uptake inhibitor sarcosine[8]; however, the collapsed coordination observed here was unaffected by such treatment. Moreover, strychnine-induced synchronous bursting was similar between genotypes, indicating that excitatory cross coupling was present in *Dmrt3*$^{-/-}$ mice (Supplementary Fig. 4).

Post-natally, *Dmrt3* messenger RNA was expressed in the ventral spinal cord at all levels, whereas prenatal expression was evident in more dorsally located cells, indicative of dorsoventral migration (Fig. 3a, b and Supplementary Fig. 5). Already at embryonic day (E) 12.5, *Dmrt3* cells were found in the medioventral domain, a pattern that persisted in adults. To pinpoint the origin of *Dmrt3* cells, we used markers for dorsal and ventral progenitors giving rise to different subclasses of interneurons[9] (Fig. 3c). Immunostainings at E11.5 for Dmrt3 and Pax7, a marker for the dorsal progenitor domain[10], demonstrated that Dmrt3 immunopositive ($^+$) cells are generated near the ventral Pax7$^+$ domain border (Fig. 3d). Moreover, Dmrt3$^+$ cells were found within the Pax2 domain marking dI4, dI5 and V0d progenitors[11], while they were negative for Evx1, a post-mitotic marker for V0v interneurons[12]. Dmrt3 expression did not coincide with the dI5 marker Lmx1b[13] whereas labelling with the dI4-6 postmitotic marker Lbx1 (ref. 14) indicated that the Dmrt3$^+$ population arise from the ventral-most Lbx1 domain (Fig. 3d). These data indicate that Dmrt3 marks a dI6 interneuron subpopulation, further confirmed by *in situ* hybridization analysis and additional markers (Supplementary Fig. 5). At E14.5, we found a partial overlap with Wt1 (Fig. 3e), proposed to label the dI6 population[15]. Dmrt3$^+$ interneurons in P11 spinal sections were positive for *Viaat* (also known as *Slc32a1*), marking inhibitory neurons, but negative for *Vglut2* (also known as *Slc17a6*), marking the majority of spinal excitatory neurons[16] (Fig. 3f). Fluorescein-dextran retrograde tracings in E15 spinal cords[17] revealed Dmrt3$^+$ interneurons that extended projections ipsilaterally (22%) and contralaterally (39%). Contralateral fibres were not observed at E12.5, indicating that midline crossing occurs between E12.5 and E15.5 (Supplementary Fig. 6). Trans-synaptic pseudorabies-virus tracing in hindlimb muscles was used to examine whether Dmrt3 interneurons contact motor neurons (Fig. 3g). Forty hours post-infection, we found virus$^+$/Dmrt3$^+$ cells both ipsi- and contralateral to the injected muscle, indicating direct connections to motor neurons[18] and corroborating the presence of commissural fibres from Dmrt3$^+$ cells (Fig. 3h). Although their character may change in the adult mouse, our developmental analysis suggests that Dmrt3-expressing cells originate from dI6 progenitors at around E11.5, develop into inhibitory interneurons with projecting axons ipsi- and contralateral, and make synaptic connections to motor neurons.

Because loss of Dmrt3 may affect interneuron development in mice, we next analysed the dI6 population, and the flanking dI5 and V0d populations, by Pax2, Brn3a and Lbx1 immunostainings. All three populations remained of similar sizes in wild-type and *Dmrt3*$^{-/-}$ mice (Fig. 3i). In contrast, we found a 58% increase in the number of Wt1$^+$ neurons in *Dmrt3*$^{-/-}$ mice (Fig. 3j–l, $P < 0.0001$), suggesting a fate change within a specific subset of dI6 neurons. Moreover, retrograde tracing in E15 and P0 animals revealed significant decreases in commissural interneuron numbers in *Dmrt3*$^{-/-}$ mice compared to wild-type at E15 ($P = 0.001$) and P0 ($P = 0.005$) (Supplementary Fig. 6). Thus, loss of Dmrt3 resulted in an increased number of Wt1$^+$ cells and fewer commissural interneurons, probably explained by an altered fate of the Dmrt3$^+$ population. In general, loss of transcription factor expression within progenitor domains results in neuron specification defects, presumably by suppression of differentiation programs operating in adjacent domains[15,19–21]. Our results suggest an early reprogramming of spinal interneurons; however, circuit reorganization, compensation issues and the direct role of Dmrt3$^+$ interneurons require further investigation.

As the horse *DMRT3* mutation occurs in the last exon, the mRNA is not expected to be subject to nonsense-mediated RNA decay[22] and the mutant allele is probably translated into a truncated form (Fig. 1g).

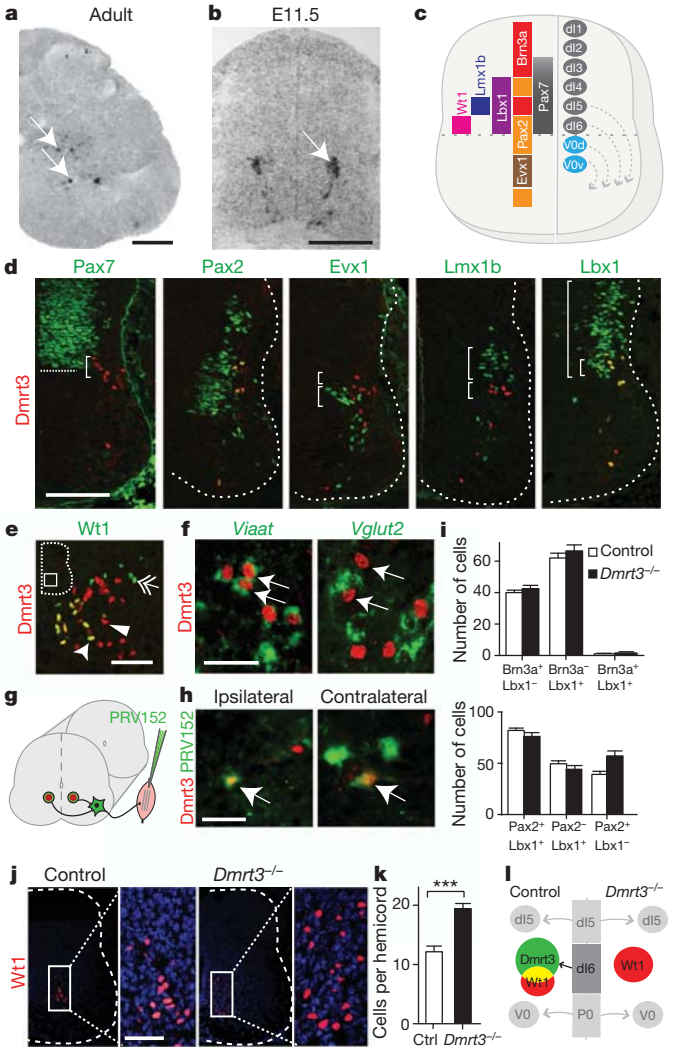


Figure 3 | Characterization of Dmrt3-expressing cells in the mouse spinal cord. **a**, *Dmrt3* mRNA expression pattern in adult spinal cord (P60). **b**, *Dmrt3* mRNA expression in a restricted population of neurons migrating ventrally in the developing spinal cord at E11.5. **c**, A schematic spinal cord cross-section showing progenitor and transcription factor domains. **d**, Double immunolabelling of Dmrt3 and Pax7 shows that Dmrt3$^+$ cells originate from the ventral-most part (bracket) of the dorsal domain (border indicated by line). Dmrt3$^+$ cells overlap with the dI4/dI6/V0d marker Pax2, but not with the $V0_V$/$V0_C$/$V0_G$ marker Evx1 or the dI5 marker Lmx1b (compare brackets). Dmrt3$^+$ cells overlap with the dI4/dI5/dI6 marker Lbx1. **e**, Double immunolabelling with Dmrt3 (arrowhead) and Wt1 (double arrow) show a partial overlap (arrow). **f**, Dmrt3$^+$ interneurons (arrows) co-labelled with *Viaat* mRNA (green) but not with *Vglut2* mRNA (green). **g**, Schematic of trans-synaptic muscle tracing of Dmrt3 neurons in the spinal cord ($n = 6$, ipsilateral; $n = 5$, contralateral). **h**, Double immunolabelling of Dmrt3 and green fluorescent protein (PRV152) show that both ipsi- and contralateral premotor interneurons overlap with Dmrt3$^+$ cells (arrows). **i**, Quantification of Brn3a/Lbx1-positive neurons (control $n_{section} = 16$, *Dmrt3*$^{-/-}$ $n_{section} = 21$) and of Lbx1/Pax2-positive neurons (control $n_{section} = 11$, *Dmrt3*$^{-/-}$ $n_{section} = 19$). **j**, Immunolabelling of Wt1$^+$ cells (red) in spinal cord sections from control and *Dmrt3*$^{-/-}$ E15.5 embryos. **k**, Quantification demonstrated that loss of Dmrt3 leads to an expanded Wt1$^+$ cell subpopulation (control $n = 43$, *Dmrt3*$^{+/-}$ $n = 36$, *Dmrt3*$^{-/-}$ $n = 46$, $^{***}P < 0.0001$). **l**, Schematic illustration of fate change in the Dmrt3 dI6 population of neurons. Mean ± s.e.m. Scale bars: 400 μm (**a**), 70 μm (**b**, **d**, **j**), 50 μm (**e**, **f**, **h**).

Expression levels between mutant and wild-type homozygotes were similar and *DMRT3* mRNA was found in a small population of neurons located in the ventral horn and around the central canal in both wild-type and mutant horses (Supplementary Fig. 1c–e). Furthermore,

transfection experiments and an electrophoretic mobility shift assay indicated that the mutant Dmrt3 protein maintain its cellular localization and DNA-binding profile (Supplementary Fig. 1f, g). It may therefore be a dominant negative form with normal DNA binding but defective protein interactions.

The remarkable association between the *DMRT3* nonsense mutation and gaitedness across horse breeds, combined with the demonstration that mouse *Dmrt3* is required for normal development of a coordinated locomotor network in the spinal cord, allow us to conclude that *DMRT3_Ser301STOP* is a causative mutation affecting the pattern of locomotion in horses. The horse phenotype indicates that Dmrt3 neurons not only have a critical role for left/right coordination but also for coordinating the movement of the fore- and hindlegs. The mutation facilitates lateral gaits, ambling and pace, and inhibits the transition from trot or pace to gallop. Homozygosity for the mutation is required, but not sufficient for pacing, as many Standardbred trotters and some Icelandic horses that are homozygous mutant do not pace. In mice, a complete loss of *Dmrt3* on one allele does not lead to any detectable phenotype, whereas in Icelandic horses, heterozygosity for the *DMRT3_Ser301STOP* mutation promotes tölt, supporting our hypothesis that the mutant protein in horses acts as a dominant negative form. Dmrt3 neurons are present in the horse and mouse spinal cord, and in the mouse, they develop into premotor inhibitory interneurons projecting ipsi- and contralaterally. Inhibitory commissural connections have been suggested as major constituents of left-right phasing during locomotion[23–25], and in cat, such interneurons have been implicated in mediation of the crossed reflexes and in the selection of different motor patterns[26,27]. Thus, Dmrt3 neurons have a character and position in spinal cord circuitry that concurs with gait coordination.

METHODS SUMMARY

A summary of the methods can be found in the Supplementary Information and includes detailed information on study populations, genotyping methods and genome-wide association analysis, genome resequencing and calling of genetic variants, *Dmrt3*-null mice, immunohistochemistry, *in situ* hybridization of mouse and horse tissue, spinal cord and muscle tracing, extracellular physiology, behaviour recordings and statistical analyses, expression analysis using mouse and horse tissue, transfection experiments and electrophoretic mobility shift assays.

Received 18 April; accepted 5 July 2012.

1. Grillner, S. Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron* **52,** 751–766 (2006).
2. Kullander, K. Genetics moving to neuronal networks. *Trends Neurosci.* **28,** 239–247 (2005).
3. Albertsdóttir, E., Eriksson, S., Sigurdsson, A. & Arnason, T. Genetic analysis of 'breeding field test status' in Icelandic horses. *J. Anim. Breed. Genet.* **128,** 124–132 (2011).
4. Hong, C. S., Park, B. Y. & Saint-Jeannet, J. P. The function of *Dmrt* genes in vertebrate development: it is not just about sex. *Dev. Biol.* **310,** 1–9 (2007).
5. Ahituv, N. *et al.* Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5,** e234 (2007).
6. Gallarda, B. W., Sharpee, T. O., Pfaff, S. L. & Alaynick, W. A. Defining rhythmic locomotor burst patterns using a continuous wavelet transform. *Ann. NY Acad. Sci.* **1198,** 133–139 (2010).
7. Mor, Y. & Lev-Tov, A. Analysis of rhythmic patterns produced by spinal neural networks. *J. Neurophysiol.* **98,** 2807–2817 (2007).
8. Kullander, K. *et al.* Role of EphA4 and EphrinB3 in local neuronal circuits that control walking. *Science* **299,** 1889–1892 (2003).
9. Alaynick, W. A., Jessell, T. M. & Pfaff, S. L. SnapShot: spinal cord development. *Cell* **146,** 178e1 (2011).
10. Jostes, B., Walther, C. & Gruss, P. The murine paired box gene, *Pax7*, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* **33,** 27–37 (1990).
11. Burrill, J. D., Moran, L., Goulding, M. D. & Saueressig, H. PAX2 is expressed in multiple spinal cord interneurons, including a population of EN1$^+$ interneurons that require PAX6 for their development. *Development* **124,** 4493–4503 (1997).
12. Moran-Rivard, L. *et al. Evx1* is a postmitotic determinant of V0 interneuron identity in the spinal cord. *Neuron* **29,** 385–399 (2001).
13. Müller, T. *et al.* The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* **34,** 551–562 (2002).
14. Gross, M. K., Dottori, M. & Goulding, M. *Lbx1* specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34,** 535–549 (2002).
15. Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Rev. Neurosci.* **10,** 507–518 (2009).
16. Gezelius, H., Wallen-Mackenzie, A., Enjin, A., Lagerstrom, M. & Kullander, K. Role of glutamate in locomotor rhythm generating neuronal circuitry. *J. Physiol. Paris* **100,** 297–303 (2006).
17. Rabe, N., Gezelius, H., Vallstedt, A., Memic, F. & Kullander, K. Netrin-1-dependent spinal interneuron subtypes are required for the formation of left-right alternating locomotor circuitry. *J. Neurosci.* **29,** 15642–15649 (2009).
18. Jovanovic, K., Pastor, A. M. & O'Donovan, M. J. The use of PRV-Bartha to define premotor inputs to lumbar motoneurons in the neonatal spinal cord of the mouse. *PLoS ONE* **5,** e11743 (2010).
19. Lanuza, G. M., Gosgnach, S., Pierani, A., Jessell, T. M. & Goulding, M. Genetic identification of spinal interneurons that coordinate left-right locomotor activity necessary for walking movements. *Neuron* **42,** 375–386 (2004).
20. Pierani, A. *et al.* Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1. *Neuron* **29,** 367–384 (2001).
21. Vallstedt, A. *et al.* Different levels of repressor activity assign redundant and specific roles to *Nkx6* genes in motor neuron and interneuron specification. *Neuron* **31,** 743–755 (2001).
22. Culbertson, M. R. & Leeds, P. F. Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.* **13,** 207–214 (2003).
23. Cowley, K. C. & Schmidt, B. J. Effects of inhibitory amino acid antagonists on reciprocal inhibitory interactions during rhythmic motor activity in the *in vitro* neonatal rat spinal cord. *J. Neurophysiol.* **74,** 1109–1117 (1995).
24. Grillner, S. & Wallen, P. Does the central pattern generation for locomotion in lamprey depend on glycine inhibition? *Acta Physiol. Scand.* **110,** 103–105 (1980).
25. Jankowska, E. & Noga, B. R. Contralaterally projecting lamina VIII interneurones in middle lumbar segments in the cat. *Brain Res.* **535,** 327–330 (1990).
26. Harrison, P. J., Jankowska, E. & Zytnicki, D. Lamina VIII interneurones interposed in crossed reflex pathways in the cat. *J. Physiol. (Lond.)* **371,** 147–166 (1986).
27. Jankowska, E., Edgley, S. A., Krutki, P. & Hammar, I. Functional differentiation and organization of feline midlumbar commissural interneurones. *J. Physiol. (Lond.)* **565,** 645–658 (2005).

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

Acknowledgements Thanks to S. Mikulovic and E. Restrepo for valuable input, C. Birchmeier for Lbx1 antibody, L. Enquist and J. Martin for PRV152, S. Ewart for horse samples, and B. Ågerup for access to race horses. The work was supported by grants from the Swedish Foundation for Strategic Research, the Swedish Research Council Formas (221-2009-1631), Swedish Research Council Medicine and Health (2007-3630/4479, 2010-4394), Swedish Society for Medical Research (H.W.), National Institute of Child Health & Human Development R01HD059862 (N.A.), and the Swedish Brain Foundation. Sequencing was performed by the SNP&SEQ Technology Platform, supported by Uppsala University and Hospital, SciLife Lab – Uppsala and the Swedish Research Council (80576801 and 70374401). Computer resources were supplied by UPPMAX. K.K. is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Author Contributions L.S.A., S.M., G.L. and L.W. collected the horse material and/or performed the genome-wide association analysis. L.S.A., D.S., M.L., G.H. and L.A. planned, designed, performed and/or analysed horse experiments. M.L., F.M., H.W., K.P., A.V. and K.K. planned, designed performed and/or analysed mouse experiments. C.-J.R. performed bioinformatic analysis. T.A. analysed horse performance data. N.A., F.I., J.L.P., M.E.M., J.R.M. and G.C. contributed with materials. L.R. recorded horse gaits. L.A. led positional cloning and characterisation of horse *DMRT3*. K.K. led the mouse studies. K.K. and L.A. wrote the paper with contributions from all authors.

Author Information The Illumina reads have been submitted to the short reads archive (http://www.ncbi.nlm.nih.gov/sra); the accession number for the study is SRP012260 and accession numbers for individual data are: four-gaited horse, SRS309533; five-gaited horse, SRS309532. Sanger sequencing data have been submitted to GenBank (accession numbers JQ922365–JQ922395). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.A. (leif.andersson@imbim.uu.se).

LETTER

doi:10.1038/nature11310

Autistic-like behaviour and cerebellar dysfunction in Purkinje cell *Tsc1* mutant mice

Peter T. Tsai[1], Court Hull[2]*, YunXiang Chu[2]*, Emily Greene-Colozzi[1], Abbey R. Sadowski[1], Jarrett M. Leech[1], Jason Steinberg[1], Jacqueline N. Crawley[3], Wade G. Regehr[2] & Mustafa Sahin[1]

Autism spectrum disorders (ASDs) are highly prevalent neurodevelopmental disorders[1], but the underlying pathogenesis remains poorly understood. Recent studies have implicated the cerebellum in these disorders, with post-mortem studies in ASD patients showing cerebellar Purkinje cell (PC) loss[2,3], and isolated cerebellar injury has been associated with a higher incidence of ASDs[4]. However, the extent of cerebellar contribution to the pathogenesis of ASDs remains unclear. Tuberous sclerosis complex (TSC) is a genetic disorder with high rates of comorbid ASDs[5] that result from mutation of either *TSC1* or *TSC2*, whose protein products dimerize and negatively regulate mammalian target of rapamycin (mTOR) signalling. TSC is an intriguing model to investigate the cerebellar contribution to the underlying pathogenesis of ASDs, as recent studies in TSC patients demonstrate cerebellar pathology[6] and correlate cerebellar pathology with increased ASD symptomatology[7,8]. Functional imaging also shows that TSC patients with ASDs display hypermetabolism in deep cerebellar structures, compared to TSC patients without ASDs[9]. However, the roles of *Tsc1* and the sequelae of *Tsc1* dysfunction in the cerebellum have not been investigated so far. Here we show that both heterozygous and homozygous loss of *Tsc1* in mouse cerebellar PCs results in autistic-like behaviours, including abnormal social interaction, repetitive behaviour and vocalizations, in addition to decreased PC excitability. Treatment of mutant mice with the mTOR inhibitor, rapamycin, prevented the pathological and behavioural deficits. These findings demonstrate new roles for *Tsc1* in PC function and define a molecular basis for a cerebellar contribution to cognitive disorders such as autism.

To evaluate the role of *Tsc1* in cerebellar PCs, we generated mice with *Tsc1* deleted in cerebellar PCs ($L7^{Cre}$;$Tsc1^{flox/+}$ (heterozygous) and $L7^{Cre}$;$Tsc1^{flox/flox}$(mutant))[10]. Cre expression is high in PCs, and expression is evident by postnatal day 6 (P6)[11]. $Tsc1^{+/+}$ (wild-type), $L7^{Cre}$;$Tsc1^{+/+}$ (L7Cre), $Tsc1^{flox/flox}$ (Flox), heterozygous and mutant mice did not show reduced survival, and weights were comparable across genotypes, unlike the severe phenotype of neuronal or glial *Tsc* mutants (Supplementary Fig. 1)[12].

To ensure that TSC1 function was impaired in PCs, we evaluated staining of phosphorylated S6 (pS6), a downstream effector of mTOR signalling. We expected TSC1 dysfunction to result in increased mTOR activity, and increased pS6 staining was detected in heterozygous and mutant PCs (Supplementary Figs 2–4). To assess the specificity of Cre-mediated recombination, we crossed $L7^{Cre}$ and *Rosa26* reporter mice and found only infrequent, scattered recombination in non-cerebellar areas, as described previously (Supplementary Fig. 5)[11]. We also examined pS6 staining in other brain regions but found no differences between mutants and controls, except in cerebellar PCs (Supplementary Fig. 6).

One of the most consistent pathologic findings in post-mortem studies of ASD patients is reduced cerebellar PC numbers[2]. In the mutant cerebellum, although basic cellular architecture was maintained in adult mice, the PC layer was abnormal with increased soma area and reduced PC numbers compared to control or heterozygous littermates (Fig. 1a, Supplementary Fig. 7). To investigate why PCs were decreased

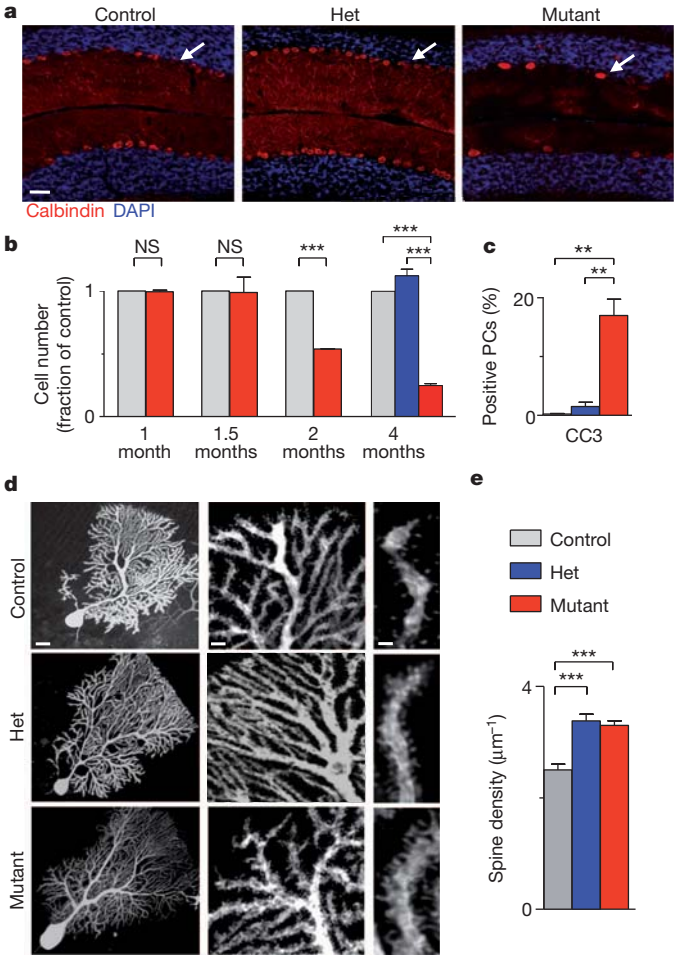


Figure 1 | PC *Tsc1* mutants display reduced PC numbers and abnormal PC morphology. **a**–**c**, Calbindin staining showed that mutants had reduced PC numbers (**a**; Het, heterozygote). Arrows, PCs. PC loss occurred by 2 months of age in mutants, but no PC loss was seen in heterozygotes at 4 months of age (**b**). NS, not significant. Increased cleaved caspase 3 (CC3) staining was found in mutants (**c**). In **b** and **c**, $n = 3$ mice for controls and mutants; $n = 2$ mice for heterozygotes; >500 cells per group. **d**, **e**, Mutant PCs displayed increased spine density ($n = 10$ cells, 3 mice for controls; $n = 10$ cells, 4 mice for heterozygotes; $n = 11$ cells, 3 mice for mutants). Middle and right panels, ×5 and ×20 magnifications of the left panel. ***$P < 0.001$, two-way analysis of variance (ANOVA), Bonferroni's post-hoc analysis. Scale bars in **a**.100μm; in **d**, 20 μm (left), 5 μm (middle) and 2 μm (right). Error bars, means ± s.e.m.

[1]The F.M. Kirby Neurobiology Center, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Laboratory of Behavioral Neuroscience, Intramural Research Program, National Institute of Mental Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

in mutants, we quantified PC numbers throughout development. Decreased cell numbers were first noted at 2 months of age, with a further reduction by 4 months of age, a reduction that was not seen in heterozygotes (Fig. 1b). As these findings suggested cell loss, we investigated markers of apoptosis and found increased TUNEL (TdT-mediated dUTP nick end labelling) and cleaved caspase 3 staining in mutant PCs at 7–8 weeks of age (Fig. 1c, and Supplementary Figs 8 and 9). Recently, neuronal stress in the cerebellum has been implicated in ASD pathogenesis[13], and studies have demonstrated critical roles for the TSC–mTOR pathway in mediating neuronal stress responses[14,15]. To investigate whether similar mechanisms were involved in *Tsc1*-mutant PC death, we evaluated markers for both endoplasmic reticulum (GRP78) and oxidative (haem oxygenase 1 (HO1)) stress and found significantly elevated levels of both markers (Supplementary Fig. 9).

As TSC–mTOR signalling has important roles in neuronal morphology and function[16,17], we also investigated whether *Tsc1* loss resulted in morphological changes in PCs at 4 weeks of age. TSC has known roles in the regulation of cell size[17,18], and the PC soma area was significantly increased in mutant, but not heterozygous, mice (Fig. 1d, Supplementary Fig. 10). TSC has also been implicated in regulating dendritic spine numbers[19], and we found increased spine density on heterozygous and mutant PC dendrites (Fig. 1d, e). Interestingly, decreased spine density has been reported in hippocampal and cortical neurons with *Tsc* loss[17,19,20], indicating that the mechanisms underlying the regulation of dendritic spines by TSC1 and TSC2 are diverse. We also found many axonal varicosities and abnormal axonal collaterals in mutants (Supplementary Fig. 11), consistent with known roles for TSC in regulating axonal morphology[16,21].

To investigate whether PC *Tsc1* mutants might show abnormal behaviours that are found in ASDs, we first evaluated social interaction, using a three-chambered assay of social approach and preference for social novelty. We found social impairments in both heterozygous and mutant animals, with no significant differences found between the time that they spent in the chamber or in interaction with the novel mouse versus the novel object (Fig. 2a). Subsequently, in a social novelty paradigm, control animals spent significantly more time in the chamber and in close interaction with the novel animal than with the familiar animal, whereas heterozygous and mutant animals displayed no significant preference for social novelty by either measure (Fig. 2b). We tested further whether mutants had impaired male–female social interactions and observed significant reductions in the time that the mutant mice spent interacting, compared with controls (Supplementary Fig. 12).

Taking into account the role of the cerebellum in motor functions, we investigated whether motor deficits contributed to social impairment. The motor activity of mutant mice was indistinguishable from littermates until approximately 7–8 weeks of age when mutants displayed initial signs of ataxia. Ataxia progressed, and by 4 months of age there were marked changes in gait parameters (Supplementary Fig. 13). However, heterozygotes displayed no ataxia (Supplementary Fig. 13), and locomotion during social testing and open-field testing was not significantly different between genotypes (Supplementary Figs 14 and15), indicating that motor impairments were not responsible for observed social deficits.

In rodents, social interaction depends largely on olfactory cues. We observed that comparable time was spent investigating three non-social olfactory cues—water, almond extract and banana extract (Supplementary Fig. 16)—indicating that olfactory function in mutants is intact. However, consistent with observed social impairment phenotypes, heterozygous and mutant mice demonstrated reduced investigation of social odours compared with controls, suggesting that impaired discrimination of social olfactory cues contributed to social deficits in mutants.

ASD patients also show repetitive behaviours and cognitive, behavioural inflexibility. To model the perseverative thinking and cognitive inflexibility exhibited by patients with ASDs, we tested animals in a reversal learning paradigm using a water T-maze. Two

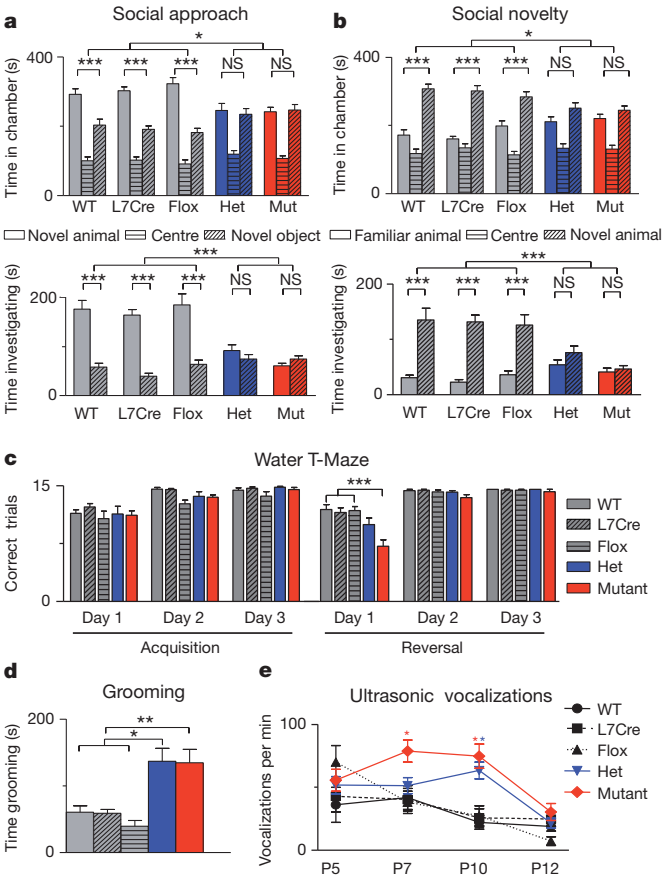


Figure 2 | PC *Tsc1* heterozygotes and mutants display autistic-like behaviours. **a**, In assays of social approach, unlike all control genotypes (WT, L7Cre and Flox), heterozygotes and mutants (Mut) showed no significant preference for a novel mouse over a novel object, as measured by time spent in a chamber (top panel) and time spent in close interaction with a novel mouse (bottom panel). **b**, Unlike controls, heterozygotes and mutants also failed to display preference for social novelty, as measured by time spent in the chamber (top panel) and in close interaction with either a familiar or a novel mouse (bottom panel) ($n \geq 11$ for each group). **c**, Mutants displayed normal acquisition of escape-platform location but impairments on day 1 of reversal learning in a water T-maze (total trials = 15; $n \geq 13$ for each group). **d**, Heterozygotes and mutants spent significantly more time self-grooming ($n \geq 11$ for each group). **e**, Heterozygote (P10) and mutant (P7 and P10) pups emitted significantly more ultrasonic vocalizations than controls ($n \geq 8$ for each time point and group). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ (two-way ANOVA, Bonferroni's post-hoc analysis). Error bars, means ± s.e.m.

measures of learning performance were used—the number of correct trials and the number of trials needed before five consecutive correct trials were achieved—and showed that mutant animals had similar acquisition learning of a submerged escape-platform location (days 1–3) compared to control littermates (Fig. 2c and Supplementary Fig. 17). However, when the escape platform location was reversed, mutant animals demonstrated significantly impaired learning of the new platform location. We also examined repetitive behaviour in a repetitive grooming task and found significantly increased self-grooming rates in heterozygotes and mutants (Fig. 2d).

ASD-like behaviours also include deficits in communication. Murine pups use ultrasonic vocalizations to communicate with their mothers, and abnormal mother–pup communication has recently been shown in *Tsc2*$^{+/-}$ mice[22]. We evaluated ultrasonic vocalizations from P5–12 and, similar to reported ASD mouse models[23], we found increased vocalizations in both heterozygotes and mutants (Fig. 2e). Consistent with roles for *Tsc1* in regulating these early phenotypes, pS6 levels were elevated by P7 in mutant PCs (Supplementary Fig. 3). Motor deficits are also found in over 50% of patients with ASDs. To

evaluate whether mutants have impaired motor learning, we evaluated mutant animals before ataxia onset, using the accelerating rotarod test, and found significantly impaired motor learning in mutants (Supplementary Fig. 18).

The changes in PC morphology, combined with previous reports that *Tsc1* loss can alter synaptic properties[17,20], suggested that synaptic inputs to PCs might also be affected. PCs receive a single, strong climbing-fibre input and many weak granule-cell parallel-fibre inputs (Fig. 3a). However, we found no difference in the amplitude of single fibre climbing-fibre inputs between mutant and littermate controls at P28 (Fig. 3b). In control animals, when synapses are stimulated twice in rapid succession, climbing-fibre synapses depress, whereas parallel synapses facilitate, consistent with the high and low release probabilities of these synapses, respectively (Fig. 3b). The same characteristic plasticity was observed in mutants (Fig. 3b). We also stimulated parallel fibres, which produce both a direct excitatory short-latency parallel-fibre excitatory postsynaptic current (EPSC) and a disynaptic inhibitory postsynaptic current (IPSC) that arises from parallel-fibre activation of molecular layer interneurons (Fig. 3c). There was a trend towards a reduction in the ratio of the amplitudes of the EPSCs and IPSCs recorded in PCs, but this was not statistically significant (Fig. 3c). Although it is difficult to exclude a subtle effect on synaptic properties, these results indicate that in spite of morphological differences, significant alterations in synaptic transmission are not apparent in mutants.

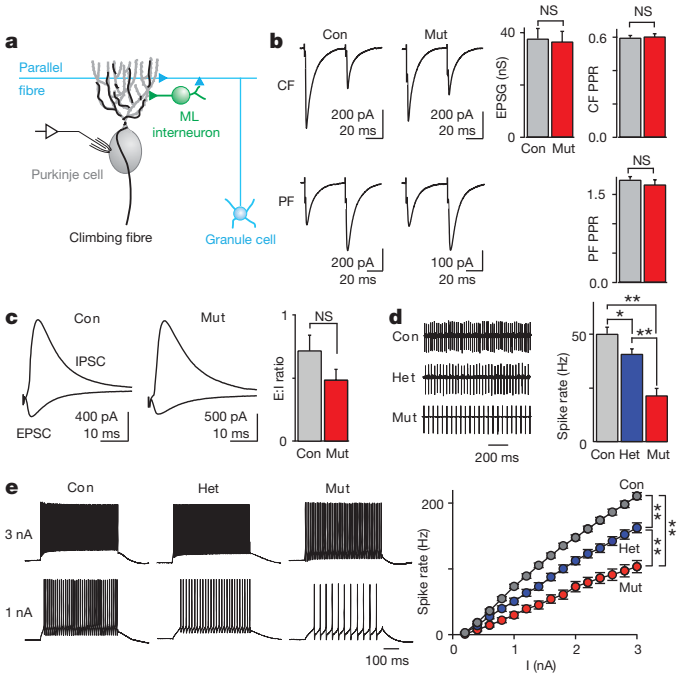


Figure 3 | PC excitability is reduced in PC *Tsc1* heterozygotes and mutants, but no significant difference is found in synaptic inputs to PCs. **a**, Schematic of the electrophysiological recording configuration and synaptic inputs onto cerebellar PCs. **b**, Electrical stimulation of climbing fibres (CF, top panel) and granule-cell-mediated parallel fibres (PF, bottom panel). Whole-cell voltage-clamp recordings showed no difference in the conductance (G) of single-fibre CF inputs ($n = 20$ cells, 5 mice for controls (Con); $n = 16$ cells, 5 mice for mutants; $P = 0.34$), or the paired-pulse ratio (PPR) of either CF or PF inputs (for CF PPR, $n = 8$ for controls, $n = 5$ for mutants; $P = 0.51$). EPSG, excitatory post synaptic conductance; ML, molecular layer. **c**, Electrical stimulation revealed no differences in the ratio of evoked synaptic excitation to inhibition (E:I ratio; $n = 29$ for controls, $n = 17$ for mutants; $P = 0.19$). **d**, Extracelluar recording of spontaneous PC spiking (left panel) revealed a significantly lower spike rate in heterozygotes and mutants (right panel; $n = 77$ cells, 12 mice for controls, $n = 62$ cells, 7 mice for heterozygotes, $n = 42$ cells, 9 mice for mutants). **e**, Whole-cell current-clamp recordings showed reduced excitability in PCs from heterozygotes and mutants compared to control PCs. Left, current injections of 1 and 3 nA produced fewer spikes in PCs from heterozygotes and mutants than in control PCs. Right, average data showed a reduced spike output for mutant PCs ($n = 77$ cells, 12 mice for controls, $n = 48$ cells, 7 mice for heterozygotes; $n = 43$ cells, 9 mice for mutants). $^*P < 0.05$; $^{**}P < 0.01$ (two-way ANOVA, Bonferroni's post-hoc analysis). Error bars, means ± s.e.m.

Previous studies of *Tsc1* have also focused on neurons that are quiescent in the absence of excitatory input, whereas PCs fire spontaneous action potentials even in the absence of synaptic inputs. Because PC firing rate is thought to be critical for encoding cerebellar output in deep cerebellar nuclei[24], we examined the intrinsic excitability of PCs using extracellular recordings, and found a significantly lower, graded spontaneous spiking rate in heterozygotes and mutants compared with controls (Fig. 3d). Moreover, current injection evoked fewer action potentials in heterozygous and mutant PCs, also in a graded fashion (Fig. 3e). A plot of firing frequency versus injected current shows that heterozygous and mutant PCs were significantly less excitable than controls (Fig. 3e). Injection of small hyperpolarizing currents resulted in smaller voltage changes in mutant and heterozygous PCs, indicating a decrease in the effective input resistance (Supplementary Fig. 19A), which has been described previously for hippocampal neurons[17]. This decrease in input resistance likely contributed to the reduced excitability of PCs in mutant and heterozygous animals. By 6 weeks of age there was an even more marked reduction in excitability in mutant mice (Supplementary Fig. 19B). Hence, despite receiving seemingly normal functioning synaptic inputs, the output of the cerebellar cortex of heterozygous and mutant animals seems to be strongly reduced, both tonically and in response to incoming excitatory drive. Our findings implicate reduced PC excitability as a potential mechanism underlying the abnormal behaviours in PC *Tsc1* mutant mice, consistent with clinical observations of impaired cerebellar function in ASD patients[9,25].

To evaluate whether the abnormal phenotypes seen in PC *Tsc1* mutant mice were modifiable as shown in other models of increased mTOR signalling[12,19,26], we treated animals with the mTOR inhibitor, rapamycin, starting at P7. Whereas vehicle treatment resulted in identical phenotypes to untreated cohorts, rapamycin treatment prevented the development of pathologic deficits in mutant animals, and mutant soma size and PC numbers were indistinguishable from controls (Fig. 4a and Supplementary Fig. 20).

We subsequently examined whether the abnormal behaviours could also be rescued with rapamycin treatment. In vehicle-treated mice, behavioural phenotypes were identical to untreated cohorts (Fig. 4b–c and Supplementary Figs 21–23); however, rapamycin treatment ameliorated the motor phenotypes seen in mutant animals in gait testing and the rotarod (Supplementary Figs 21 and 24). Rapamycin treatment also prevented deficits in the water T-Maze, with no significant differences seen between rapamycin-treated mutants and controls in both acquisition and reversal learning (Fig. 4b and Supplementary Fig. 22). In addition, following rapamycin treatment, mutants displayed social behaviour comparable to that of controls in both social approach and social novelty assays (Fig. 4c and Supplementary Fig. 23). Thus, rapamycin prevented both pathologic and behavioural phenotypes in *Tsc1* PC mutants, supporting the possibility of a therapeutic role for mTOR inhibition.

Our study shows that the TSC–mTOR pathway may have critical, novel roles in cerebellar PCs. We find that mice with homozygous and heterozygous loss of *Tsc1* in PCs demonstrated social impairment, restrictive behaviour and abnormal vocalizations; representative of the three core deficits in ASDs. Mutant mice, but not heterozygous mice, also displayed pathologic features found in ASD postmortem studies with reduced PC numbers and evidence of increased neuronal stress. Although PC loss has been reported in postmortem studies of ASD patients, several lines of evidence suggest that PC death cannot fully explain the abnormal behaviours seen in PC *Tsc1* mice. Prior to PC death, mutant mice displayed abnormal vocalizations and motor-learning impairments. In addition, mice with heterozygous loss of *Tsc1* showed no evidence of PC loss, but displayed autistic-like behaviours.

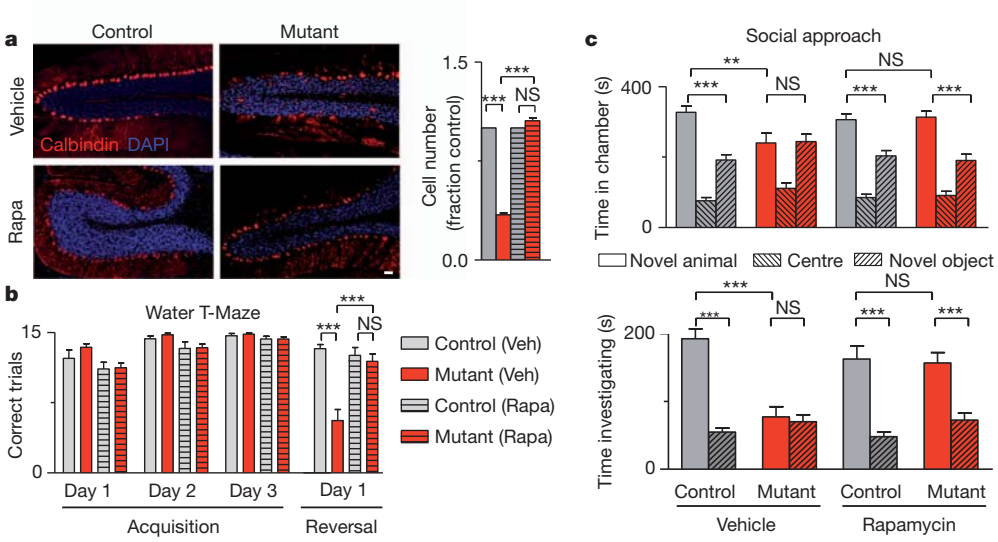


Figure 4 | Rapamycin treatment prevents pathological and behavioural abnormalities in PC *Tsc1* mutant mice. **a**, Treatment with rapamycin (Rapa) prevented cell loss seen in vehicle (Veh)-treated PC *Tsc1* mutants (left panel; scale bar, 100 μm), and quantification of cell numbers (right panel; $n > 500$ cells, 2 mice per group) **b**, **c**, Vehicle-treated mutants displayed behavioural deficits, whereas rapamycin-treated mutants displayed amelioration of these deficits, as shown using assays of reversal learning in the water T-maze ($n \geq 9$ in each group) (**b**) and social approach in the three-chambered apparatus ($n \geq 10$) (**c**). As behavioural phenotypes were not significantly different between the three control genotypes in untreated mice, control genotypes were pooled into a single group for these studies. $^{*}P < 0.05$; $^{***}P < 0.001$ (two-way ANOVA, Bonferroni's post-hoc analysis). Error bars, means ± s.e.m.

In this study, we also show that loss of *Tsc1* from cerebellar PCs is sufficient to result in abnormal autistic-like behaviours. These findings implicate the cerebellum in the neural circuitry mediating core features of autism. The cerebellum has been suggested previously to have roles in social interaction[27], and cerebellar abnormalities are associated with ASDs as well as cognitive and behavioural disturbances[28]. How the cerebellum modulates the abnormal behaviours of autism remains a topic of intense investigation. Autism has been proposed to be a disorder of abnormally distributed networks[29]. The cerebellum is connected, through the DCN, to these networks and cortical areas implicated in ASDs. Akin to its role in motor coordination, the cerebellum has been proposed to modulate these cognitive networks, with dysfunction resulting in abnormally regulated behaviours comparable to cognitive, behavioural dysmetria[30]. Our data displayed markedly impaired PC excitability in both heterozygotes and mutants. As PC firing rates are critical determinants of DCN output, by affecting DCN activity, PC dysfunction could alter these downstream neuronal networks, thereby contributing to abnormal autistic-like behaviours. Therefore, PC *Tsc1* mutants should provide a valuable experimental system to investigate the effects of PC dysfunction on these neuronal networks and other mechanisms contributing to the pathogenesis of ASDs.

METHODS SUMMARY

All behavioural procedures were reviewed and approved by the Animal Research Committee of Boston Children's Hospital. Descriptions of behavioural assays are provided in greater detail in the Methods. Experiments were performed with the examiner blind to genotype. Vehicle or rapamycin treatments (6 mg kg^{-1}) were given by intraperitoneal injection starting at P7 (a summary of the statistical analysis and the number of animals used in behavioural tests is provided in Supplementary Figs 25 and 26, and Supplementary Tables 1–3).

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

Received 22 August 2011; accepted 11 June 2012.
Published online 1 July 2012.

1. Prevalence. of autism spectrum disorders — Autism and Developmental Disabilities Monitoring Network, United States, 2006. *MMWR Surveill. Summ.* **58,** 1–20 (2009).
2. Bauman, M. L. & Kemper, T. L. Neuroanatomic observations of the brain in autism: a review and future directions. *Int. J. Dev. Neurosci.* **23,** 183–187 (2005).
3. Amaral, D. G., Schumann, C. M. & Nordahl, C. W. Neuroanatomy of autism. *Trends Neurosci.* **31,** 137–145 (2008).
4. Limperopoulos, C. *et al.* Does cerebellar injury in premature infants contribute to the high prevalence of long-term cognitive, learning, and behavioral disability in survivors? *Pediatrics* **120,** 584–593 (2007).
5. Jeste, S. S., Sahin, M., Bolton, P., Ploubidis, G. B. & Humphrey, A. Characterization of autism in young children with tuberous sclerosis complex. *J. Child Neurol.* **23,** 520–525 (2008).
6. Ertan, G., Arulrajah, S., Tekes, A., Jordan, L. & Huisman, T. A. Cerebellar abnormality in children and young adults with tuberous sclerosis complex: MR and diffusion weighted imaging findings. *J. Neuroradiol.* **37,** 231–238 (2010).
7. Weber, A. M., Egelhoff, J. C., McKellop, J. M. & Franz, D. N. Autism and the cerebellum: evidence from tuberous sclerosis. *J. Autism Dev. Disord.* **30,** 511–517 (2000).
8. Eluvathingal, T. J. *et al.* Cerebellar lesions in tuberous sclerosis complex: neurobehavioral and neuroimaging correlates. *J. Child Neurol.* **21,** 846–851 (2006).
9. Asano, E. *et al.* Autism in tuberous sclerosis complex is related to both cortical and subcortical dysfunction. *Neurology* **57,** 1269–1277 (2001).
10. Kwiatkowski, D. J. *et al.* A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. *Hum. Mol. Genet.* **11,** 525–534 (2002).
11. Barski, J. J., Dethleffsen, K. & Meyer, M. Cre recombinase expression in cerebellar Purkinje cells. *Genesis* **28,** 93–98 (2000).
12. Tsai, P. & Sahin, M. Mechanisms of neurocognitive dysfunction and therapeutic considerations in tuberous sclerosis complex. *Curr. Opin. Neurol.* **24,** 106–113 (2011).
13. Sajdel-Sulkowska, E. M., Xu, M. & Koibuchi, N. Increase in cerebellar neurotrophin-3 and oxidative stress markers in autism. *Cerebellum* **8,** 366–372 (2009).
14. Di Nardo, A. *et al.* Tuberous sclerosis complex activity is required to control neuronal stress responses in an mTOR-dependent manner. *J. Neurosci.* **29,** 5926–5937 (2009).
15. Reith, R. M., Way, S., McKenna, J. III, Haines, K. & Gambello, M. J. Loss of the tuberous sclerosis complex protein tuberin causes Purkinje cell degeneration. *Neurobiol. Dis.* **43,** 113–122 (2011).
16. Nie, D. *et al.* Tsc2-Rheb signaling regulates EphA-mediated axon guidance. *Nature Neurosci.* **13,** 163–172 (2010).
17. Tavazoie, S. F., Alvarez, V. A., Ridenour, D. A., Kwiatkowski, D. J. & Sabatini, B. L. Regulation of neuronal morphology and function by the tumor suppressors Tsc1 and Tsc2. *Nature Neurosci.* **8,** 1727–1734 (2005).
18. Meikle, L. *et al.* A mouse model of tuberous sclerosis: neuronal loss of Tsc1 causes dysplastic and ectopic neurons, reduced myelination, seizure activity, and limited survival. *J. Neurosci.* **27,** 5546–5558 (2007).
19. Meikle, L. *et al.* Response of a neuronal model of tuberous sclerosis to mammalian target of rapamycin (mTOR) inhibitors: effects on mTORC1 and Akt signaling lead to improved survival and function. *J. Neurosci.* **28,** 5422–5432 (2008).
20. Bateup, H. S., Takasaki, K. T., Saulnier, J. L., Denefrio, C. L. & Sabatini, B. L. Loss of Tsc1 *In vivo* impairs hippocampal mGluR-LTD and increases excitatory synaptic function. *J. Neurosci.* **31,** 8862–8869 (2011).

21. Choi, Y. J. *et al.* Tuberous sclerosis complex proteins control axon formation. *Genes Dev.* **22,** 2485–2495 (2008).
22. Young, D. M., Schenk, A. K., Yang, S. B., Jan, Y. N. & Jan, L. Y. Altered ultrasonic vocalizations in a tuberous sclerosis mouse model of autism. *Proc. Natl Acad. Sci. USA* **107,** 11074–11079 (2010).
23. Scattoni, M. L., Crawley, J. & Ricceri, L. Ultrasonic vocalizations: a tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neurosci. Biobehav. Rev.* **33,** 508–515 (2009).
24. De Zeeuw, C. I. *et al.* Spatiotemporal firing patterns in the cerebellum. *Nature Rev. Neurosci.* **12,** 327–344 (2011).
25. Ryu, Y. H. *et al.* Perfusion impairments in infantile autism on technetium-99m ethyl cysteinate dimer brain single-photon emission tomography: comparison with findings on magnetic resonance imaging. *Eur. J. Nucl. Med.* **26,** 253–259 (1999).
26. Zhou, J. *et al.* Pharmacological inhibition of mTORC1 suppresses anatomical, cellular, and behavioral abnormalities in neural-specific *Pten* knock-out mice. *J. Neurosci.* **29,** 1773–1783 (2009).
27. Insel, T. R. & Fernald, R. D. How the brain processes social information: searching for the social brain. *Annu. Rev. Neurosci.* **27,** 697–722 (2004).
28. Steinlin, M. Cerebellar disorders in childhood: cognitive problems. *Cerebellum* **7,** 607–610 (2008).
29. Müller, R. A. The study of autism as a distributed disorder. *Ment. Retard. Dev. Disabil. Res. Rev.* **13,** 85–95 (2007).
30. Schmahmann, J. D. Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *J. Neuropsychiatry Clin. Neurosci.* **16,** 367–378 (2004).

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

Acknowledgements We thank G. Corfas, M. Fagiolini, P. Rosenberg, S. Goldman and the Neurodevelopmental Behavioral Core of Boston Children's Hospital for assistance with behavioural experiments. We are grateful to C. Walsh, L. Benowitz and members of the Sahin laboratory for critical reading of the manuscript, and to M. Gregas for advice regarding statistical analysis. P.T.T. received support from the Developmental Neurology Training grant (T32 NS007473), American Academy of Neurology, and the Nancy Lurie Marks Family Foundation. This work and M.S. are supported in part by the National Institutes of Health (NIH; grant R01 NS58956), the John Merck Scholars Fund, Autism Speaks, the Nancy Lurie Marks Family Foundation, Boston Children's Hospital Translational Research Program, Manton Center for Orphan Disease Research and Boston Children's Hospital Intellectual and Developmental Disabilities Research Center (grant P30 HD18655). J.N.C. is supported by the Intramural Research Program, National Institute of Mental Health. W.G.R. is supported by the NIH (grant R01NS032405) and the Simons Foundation (grant SFARI 232304). Y.X.C. is supported by the Howard Hughes Medical Institute Medical Research Fellowship.

Author Contributions P.T.T. and M.S. conceived and designed the experimental approach, performed experiments and prepared the manuscript. C.H. and Y.X.C. designed and performed cerebellar slice experiments and assisted in manuscript preparation. E.G.-C., A.R.S. and J.M.L. assisted with behavioural experiments and statistical analysis. J.S. contributed to experimental quantification. J.N.C. provided training for behavioural phenotyping and contributed to behavioural experimental design and analysis. W.G.R. and M.S. supervised the project, and contributed to experimental design and analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of this article at www.nature.com/nature. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (mustafa.sahin@childrens.harvard.edu) or P.T.T. (peter.tsai@childrens.harvard.edu).

METHODS

Mice. $L7^{cre}$;$Tsc1^{flox/flox}$ (mutant) animals were generated by crossing *L7/Pcp2-Cre* ($L7^{Cre}$) transgenic mice[11] with floxed *Tsc1* mice ($Tsc1^{flox/flox}$)[10] to yield $L7^{Cre}$;$Tsc1^{flox/+}$ progeny. These progeny were then crossed with one another or with $Tsc1^{flox/flox}$ animals to give litters consisting of control ($Tsc1^{+/+}$ (wild-type), $Tsc1^{flox/+}$, $L7^{Cre}$;$Tsc1^{+/+}$ (L7Cre), or $Tsc1^{flox/flox}$ (Flox)) mice, heterozygous ($L7^{Cre}$;$Tsc1^{flox/+}$) mice or mutant ($L7^{Cre}$;$Tsc1^{flox/flox}$) mice. Only male animals were used for behavioural experiments except for analysis of ultrasonic vocalizations, in which both male and female pups were used. Mice were of mixed genetic backgrounds (C57Bl/6J, 129 SvJae and BALB/cJ). During analysis, germline deletion was discovered to occur in the mouse colony at a frequency of ~5% (more than 200 animals were tested). Genotyping for heterozygous mice would exclude these animals from this cohort, but it is possible that a small percentage of $L7^{Cre}$;$Tsc1^{flox/-}$ mice were included in the mutant cohort. We therefore repeated behavioural analysis with a cohort of mutant ($L7^{Cre}$;$Tsc1^{flox/flox}$) mice and littermate controls and found no significant differences with the previous mutant cohort (Supplementary Figs 25 and 26), and thus cohorts were combined for analysis. As behavioural data revealed comparable behavioural phenotypes between all untreated control ($Tsc1^{+/+}$, $Tsc1^{flox/flox}$ and $L7^{Cre}$;$Tsc^{+/+}$) genotypes, these genotypes were pooled for behavioural studies involving rapamycin treatment. All experimental protocols were approved by the Animal Research Committee of the Children's Hospital, Boston.

Social interaction. Animals were tested for social interaction in the three-chambered apparatus (Dold Laboratories) as described previously[31]. Time spent in chambers and the number of crossings between chambers was recorded in an automated manner (National Instruments). Time spent interacting with the novel animal and object was recorded by an examiner with a stopwatch. Animals were tested between 7 and 9 weeks of age. For male–female interactions, tested males were placed into an open field with control females and evaluated for male-initiated interaction over a 5-minute period. All behavioural assays (including social interaction) were performed with the examiner blind to the genotypes.

Gait analysis. Animals were placed at the end of an apparatus that is 5 cm in width (preventing animals from turning around) and 56 cm in length, with paper placed along the floor of the entire apparatus. Apparatus walls consisted of opaque material and were 15 cm in height, preventing animals from looking over or escaping beyond walls. Paws were painted with red ink (forepaws) and black ink (hindpaws). The longest stride for each trial was taken as the stride length for that trial. Stride width was measured as the distance between the hind paws at time of the longest stride. Measurements were taken from 14–16-week-old animals.

Open field. Open-field testing was performed as described for a 15-min period[32]. Movement and time spent in centre quadrants were recorded by video camera and analysed by Noldus (Virginia) analysis software. Measurements were taken from animals aged 7–10 weeks.

Olfaction. Olfaction was tested as described previously[33]. Animals were presented sequentially with odours on cotton-tipped applicators: first, non-social odours, then social odours. Odours were presented in three consecutive trials per odorant stimulus (2 min per trial) in the following order: water, almond extract, banana extract, social odour 1, social odour 2. Social odours were swipes from cages containing unfamiliar gender- (male) and age-matched animals. Measurements were taken from animals aged 8–12 weeks.

Grooming. After habituation, animals were observed for 10 min, and time spent grooming was recorded as described[34]. Measurements were taken from animals aged 8–12 weeks.

Water T-Maze. Reversal learning was tested using the water T-maze as described[35]. On days 1–3, mice were given 15 trials and tasked to locate a submerged platform placed in one of the maze arms. After 15 trials on day 3, the platform was changed to the other maze arm. Mice were then tested for 15 additional trials (reversal day 1). Then for 2 subsequent days (reversal days 2 and 3), mice were given 15 trials per day. The number of correct trials and the number of trials required to achieve five consecutive correct trials were recorded. Measurements were taken from animals aged 8–12 weeks.

Ultrasonic vocalizations. Ultrasonic vocalizations were tested as described on postnatal days 5–12 (ref. 23). Pups were individually removed from their mother and placed inside a soundproof container in which three detectors were used to monitor vocalizations for 5 min. Calls were recorded using Ultravox recording software (Noldus)[36]. The maternal genotype in all experiments was $L7^{Cre}$;$Tsc1^{flox/+}$.

Accelerating rotarod. Animals were tested using the accelerating rotarod, as described previously[37], over 5 consecutive days. Animals were tested before overt ataxia, between 5 and 7 weeks of age.

Slices. Acute sagittal slices (250–300-μm thick) were prepared from the cerebellar vermis of 4- and 6-week-old mutant and control littermates. Slices were cut in an ice-cold artificial cerebrospinal fluid (ACSF) solution consisting of (mM): 125 NaCl, 26 $NaHCO_3$, 1.25 NaH_2PO_4, 2.5 KCl, 1 $MgCl_2$, 2 $CaCl_2$ and 25 glucose (pH 7.3, osmolarity 310), equilibrated with 95% O_2 and 5% CO_2. Slices were initially incubated at 34 °C for 25 min, and then at room temperature before recording in the same ACSF.

Recordings. Visually guided (infrared DIC videomicroscopy and water-immersion ×40 objective) whole-cell recordings were obtained with patch pipettes (2–4 MΩ) pulled from borosilicate capillary glass (World Precision Instruments) with a Sutter P-97 horizontal puller. Electrophysiological recordings were performed at 31–33 °C.

IPSCs were recorded at the EPSC reversal potential, and parallel-fibre EPSCs were recorded at the IPSC-reversal potential. To measure climbing-fibre synaptic inputs, 500 nM NBQX (2,3-Dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide) was used to reduce the size of synaptic currents, and picrotoxin (20 μM) was used to block GABAergic inhibition. For experiments recorded at the EPSC-reversal potential and for climbing-fibre EPSCs, the internal pipette solution contained (in mM): 140 Cs-gluconate, 15 HEPES, 0.5 EGTA, 2 TEA-Cl, 2 MgATP, 0.3 NaGTP, 10 phosphocreatine-$Tris_2$, 2 QX 314-Cl. The pH was adjusted to 7.2 with CsOH. Membrane potentials were not corrected for the liquid junction potential. The EPSC and IPSC reversal potentials were determined in each experiment by adjusting the membrane potential until no EPSC or IPSC was evident; these potentials were approximately +15 mV for the EPSC reversal and −65 mV for the IPSC reversal. For current-clamp recordings, the internal solution contained (in mM): 150 K-gluconate, 3 KCl, 10 HEPES, 0.5 EGTA, 3 MgATP, 0.5 GTP, 5 phosphocreatine-$Tris_2$ and 5 phosphocreatine-Na_2. The pH was adjusted to 7.2 with NaOH. Current-clamp and extracellular recordings were performed in NBQX (5 μM), R-CPP (2.5 μM) and picrotoxin (20 μM) to block AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors, NMDA (*N*-methyl-D-aspartate) receptors, and GABA subtype A ($GABA_A$) receptors, respectively. All drugs were purchased from Sigma-Aldrich or Tocris. Electrophysiolgical data were acquired as described previously[38].

Rapamycin treatment. Rapamycin was dissolved in 0.25% polyethylene glycol and 0.25% Tween before usage. Vehicle or rapamycin was administered intraperitoneally every Monday, Wednesday and Friday, with rapamycin dosed at 6 mg kg^{-1} per injection starting at P7. As behavioural data revealed comparable behavioural phenotypes between all untreated control ($Tsc1^{+/+}$, $Tsc1^{flox/flox}$ and $L7^{Cre}$;$Tsc^{+/+}$) genotypes, these genotypes were pooled for behavioural studies involving rapamycin treatment.

Immunohistochemistry. Mice were perfused and post-fixed with 4% paraformaldehyde. Sections were prepared by cryostat sectioning and were stained with the following antibodies: PhosphoS6 (Cell Signaling), calbindin (Sigma, Cell Signaling), GRP78 (Stressgen), Haem Oxygenase-1 (Stressgen), cleaved caspase 3 (Cell Signaling). TUNEL staining was performed according to the manufacturer's instructions (Millipore).

Microscopy. Intracellular labelling of Purkinje cells was accomplished using recording pipettes filled with 0.05% biocytin (Tocris). Neurons in deeper portions of the Purkinje-cell layer were targeted and filled for 3 min, and then the pipette was slowly withdrawn so that the cell membrane could reseal. Slices (250-μm thick) were then fixed in 4% paraformaldehyde in 0.1 M phosphate buffer for 24 h, rinsed thoroughly in phosphate buffer solution (PBS) and incubated for 90 min in a PBS solution containing 0.5% Triton-X, 10% normal goat serum and streptavidin Alexa Fluor 488 conjugate (1:500, Life Technologies). Slices were then rinsed in PBS, mounted on Superfrost Plus slides (VWR International), air-dried, and coverslipped in Vectashield mounting media (Vector Labs). Immunohistochemical studies were captured using Zeiss Confocal LSM710. Images were processed and morphology quantified using ImageJ software with studies performed by examiner blinded to genotypes.

Statistics. Data are reported as mean ± s.e.m., and statistical analysis was carried out with GraphPad Prism software using one- and two-way ANOVA with Bonferroni's multiple comparison tests for post-hoc analysis unless otherwise specified.

31. Yang, M., Silverman, J. L. & Crawley, J. N. Automated three-chambered social approach task for mice. *Curr Protoc Neurosci* Ch 8, Unit 8 26 (2011).
32. Holmes, A. *et al.* Behavioral characterization of dopamine D5 receptor null mutant mice. *Behav. Neurosci.* **115,** 1129–1144 (2001).
33. Yang, M. & Crawley, J. N. Simple behavioral assessment of mouse olfaction. *Curr Protoc. Neurosci.* Ch 8, Unit 8 24 (2009).
34. Silverman, J. L. *et al.* Sociability and motor functions in *Shank1* mutant mice. *Brain Res.* **1380,** 120–137 (2011).
35. Bednar, I. *et al.* Selective nicotinic receptor consequences in APP_{SWE} transgenic mice. *Mol. Cell. Neurosci.* **20,** 354–365 (2002).
36. Kurejova, M. *et al.* An improved behavioural assay demonstrates that ultrasound vocalizations constitute a reliable indicator of chronic cancer pain and neuropathic pain. *Mol. Pain* **6,** 18 (2010).
37. Buitrago, M. M., Schulz, J. B., Dichgans, J. & Luft, A. R. Short and long-term motor skill learning in an accelerated rotarod training paradigm. *Neurobiol. Learn. Mem.* **81,** 211–216 (2004).
38. Hull, C. & Regehr, W. G. Identification of an inhibitory circuit that regulates cerebellar Golgi cell activity. *Neuron* **73,** 149–158 (2012).

LETTER

doi:10.1038/nature11333

Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2

Claudia A. Doege[1], Keiichi Inoue[1], Toru Yamashita[1], David B. Rhee[1], Skylar Travis[1], Ryousuke Fujita[1], Paolo Guarnieri[2,3], Govind Bhagat[1,3], William B. Vanti[1], Alan Shih[4], Ross L. Levine[4], Sara Nik[5], Emily I. Chen[5,6] & Asa Abeliovich[1]

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) by using the pluripotency factors Oct4, Sox2, Klf4 and c-Myc (together referred to as OSKM)[1]. iPSC reprogramming erases somatic epigenetic signatures—as typified by DNA methylation or histone modification at silent pluripotency loci—and establishes alternative epigenetic marks of embryonic stem cells (ESCs)[2]. Here we describe an early and essential stage of somatic cell reprogramming, preceding the induction of transcription at endogenous pluripotency loci such as *Nanog* and *Esrrb*. By day 4 after transduction with OSKM, two epigenetic modification factors necessary for iPSC generation, namely poly(ADP-ribose) polymerase-1 (Parp1) and ten-eleven translocation-2 (Tet2), are recruited to the *Nanog* and *Esrrb* loci. These epigenetic modification factors seem to have complementary roles in the establishment of early epigenetic marks during somatic cell reprogramming: Parp1 functions in the regulation of 5-methylcytosine (5mC) modification, whereas Tet2 is essential for the early generation of 5-hydroxymethylcytosine (5hmC) by the oxidation of 5mC (refs 3,4). Although 5hmC has been proposed to serve primarily as an intermediate in 5mC demethylation to cytosine in certain contexts[5–7], our data, and also studies of Tet2-mutant human tumour cells[8], argue in favour of a role for 5hmC as an epigenetic mark distinct from 5mC. Consistent with this, Parp1 and Tet2 are each needed for the early establishment of histone modifications that typify an activated chromatin state at pluripotency loci, whereas Parp1 induction further promotes accessibility to the Oct4 reprogramming factor. These findings suggest that Parp1 and Tet2 contribute to an epigenetic program that directs subsequent transcriptional induction at pluripotency loci during somatic cell reprogramming.

We performed a functional screen for epigenetic modification factors that promote OSKM-mediated somatic cell reprogramming. Overexpression of a single pool of 29 candidate epigenetic modification factors, selected on the basis of a proteomic analysis of iPSCs (Fig. 1a–c and Supplementary Tables 1 and 2), promoted iPSC colony production in mouse embryonic fibroblast (MEF) cultures transduced with OSKM (OSKM-MEFs; relative to green fluorescent protein vector control transduced MEF; Fig. 1d). Candidate epigenetic modification factors were retested in successive subpools, and Parp1 was identified as a potent inducer of OSKM-MEF reprogramming (Fig. 1d). iPSCs generated by OSKM-MEFs with Parp1 were confirmed as pluripotent by immunocytochemistry (Supplementary Fig. 2a), gene expression (Supplementary Fig. 2b–f and Supplementary Table 3) and pyrosequencing (Supplementary Fig. 2g–i and Supplementary Table 4) analyses at multiple pluripotency loci. Parp1 overexpression did not alter the proliferation rate of transduced cultures (as determined by bromodeoxyuridine (BrdU) incorporation; Supplementary Fig. 2j).

Parp1 is a broadly expressed nuclear protein involved in the detection and repair of DNA damage, the remodelling of chromatin and the regulation of transcription[9]. A time course of endogenous *Parp1* expression revealed detectable levels even in wild-type (WT) MEFs and a peak at day 4 (d4) after transduction with OSKM (Fig. 1e, f and Supplementary Figs 2k and 3a), thus preceding expression of the endogenous pluripotency loci including *Nanog*, *Oct4* and *Esrrb* (Fig. 1f and Supplementary Fig. 3b). Immunocytochemistry analysis at d4 after transduction of WT MEFs with OSKM (WT d4-OSKM-MEFs) demonstrated Parp1 accumulation in a majority of nuclei in comparison with d4 control vector-transduced MEFs (WT d4-CONT-MEFs; Fig. 1g). The increased accumulation of Parp1 holoprotein during reprogramming was not paralleled by a corresponding accumulation of cleaved Parp1, a marker of apoptosis (Supplementary Fig. 3a).

In view of the effect of Parp1 overexpression in somatic cell reprogramming, we next tested the impact of Parp1 deficiency. Reprogramming of iPSCs was suppressed in the context of *Parp1*$^{-/-}$ OSKM-MEFs relative to WT OSKM-MEFs (Fig. 2a and Supplementary Fig. 2k, l). Resupplying WT Parp1 partly rescued iPSC generation in *Parp1*$^{-/-}$ OSKM-MEFs. In contrast, expression of Parp1 mutants[10,11], compromising either the catalytic activity or the DNA-binding activity, failed to rescue iPSC generation, indicating that both activities are required for iPSC reprogramming (Fig. 2a and Supplementary Fig. 2l).

The early expression and functional role of endogenous *Parp1* in the reprogramming process suggested the possibility of a direct interaction at pluripotency loci. Consistent with this notion, chromatin immunoprecipitation (ChIP) at the *Nanog* and *Esrrb* pluripotency loci, in the transcription start site regions (Supplementary Table 5), demonstrated increased Parp1 binding in d4-OSKM-MEFs and iPSCs (compared with d4-CONT-MEFs; Fig. 2b, c and Supplementary Fig. 3g, h, k).

Parp1 has been broadly implicated in the regulation of epigenetic remodelling events[7,12], but its role during the reprogramming of pluripotency loci is unclear. We therefore investigated the impact of modified Parp1 expression on two distinct forms of cytosine methylation at pluripotency loci, 5mC and 5hmC, during somatic cell reprogramming. 5hmC is a more recently described DNA modification that has been suggested to participate in the maintenance of pluripotency at regulatory elements[5,13–17] in ESCs, but its role during somatic cell reprogramming has not previously been described. We quantified total cytosine methylation (5mC plus 5hmC; either by *Hpa*II digestion sensitivity assays or pyrosequencing of bisulphite-treated DNA) or 5hmC alone (either by *Msp*I sensitivity assay (glucosylation-coupled methylation-sensitive quantitative polymerase chain reaction; GlucMS-qPCR)[16,18] or hydroxymethylated DNA immunoprecipitation (hMeDIP); see Supplementary Tables 6 and 7 and Supplementary Methods) at regulatory regions of the *Nanog* or *Esrrb* loci. Both d4-OSKM-MEFs and iPSCs showed a significant and consistent

[1]Departments of Pathology and Cell Biology and Neurology, Taub Institute for Aging, Columbia University, New York, New York 10032, USA. [2]Biomedical Informatics Shared Resources, Bioinformatics Division, Columbia University, New York, New York 10032, USA. [3]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York 10032, USA. [4]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York 10016, USA. [5]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, New York 11794, USA. [6]Stony Brook University Proteomics Center, School Of Medicine, Stony Brook, New York 11794, USA.

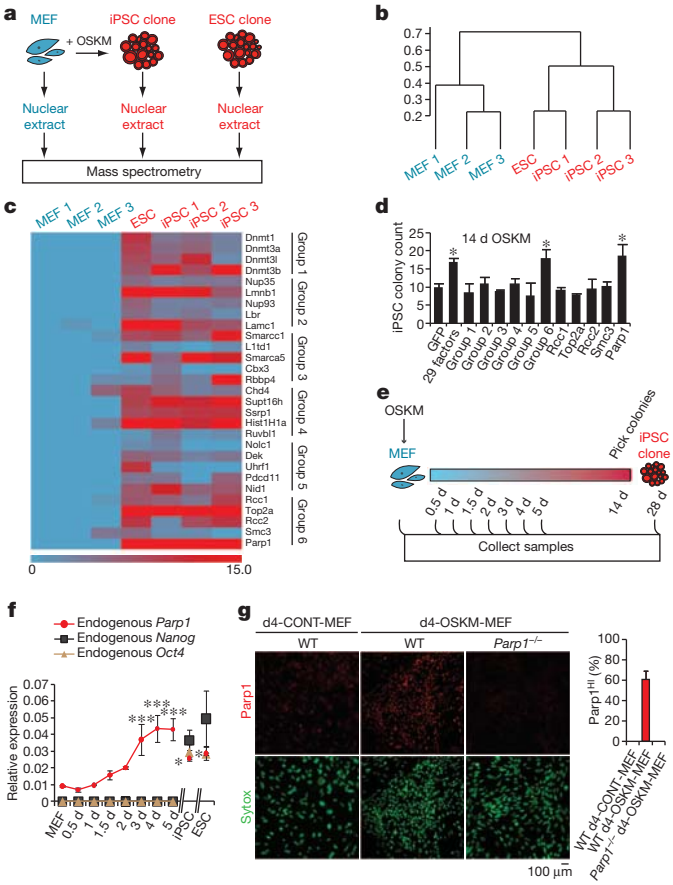


Figure 1 | Parp1 promotes OSKM-mediated iPSC generation. **a**, Diagram of proteomic strategy to identify candidate epigenetic modification (EM) factors. **b**, Unsupervised hierarchical clustering analysis (Spearman rank correlation) of mass spectrometry data from nuclear extracts of MEFs ($n = 3$), iPSCs ($n = 3$) and ESCs ($n = 1$). The scale bar represents the correlation height ($= 1 - \text{Abs}[\text{correlation}]$). **c**, Dual-colour heat map for expression levels of 29 proteins highly enriched in both the iPSC and ESC samples (relative to MEFs). The colour scale bar represents the spectral count. Candidate EM factors were divided into six groups for further functional testing. **d**, Functional screen of candidate EM factors for promotion of somatic cell reprogramming in OSKM-MEFs. EM candidates were transduced together as a single pool of 29 genes, as 6 subpools, or as individual factors from group 6 (as in **c**). Alkaline phosphatase-positive (AP^+) iPSC colonies were counted at day 14 after transduction with OSKM. **e**, Diagram of time-course analyses of iPSC reprogramming. **f**, Gene expression time course of endogenous *Parp1*, *Nanog* and *Oct4*. **g**, Immunocytochemistry analysis of WT or *Parp1*$^{-/-}$ d4-OSKM-MEFs and d4-CONT-MEFs with an antibody against Parp1 (upper panels; red), and counterstained with Sytox nuclear marker (lower panels; green). Increased Parp1 expression in d4-OSKM-MEFs is quantified on the right ($Parp1^{HI}$; defined as mean plus 2 s.d. or greater than the expression level in d4-CONT-MEFs); modified nuclear morphology apparent in d4-OSKM-MEFs is as described previously[30]. Results in **d**, **f** and **g** are shown as means and s.d. for three independent experiments. Asterisk, $P < 0.05$; three asterisks, $P < 0.001$.

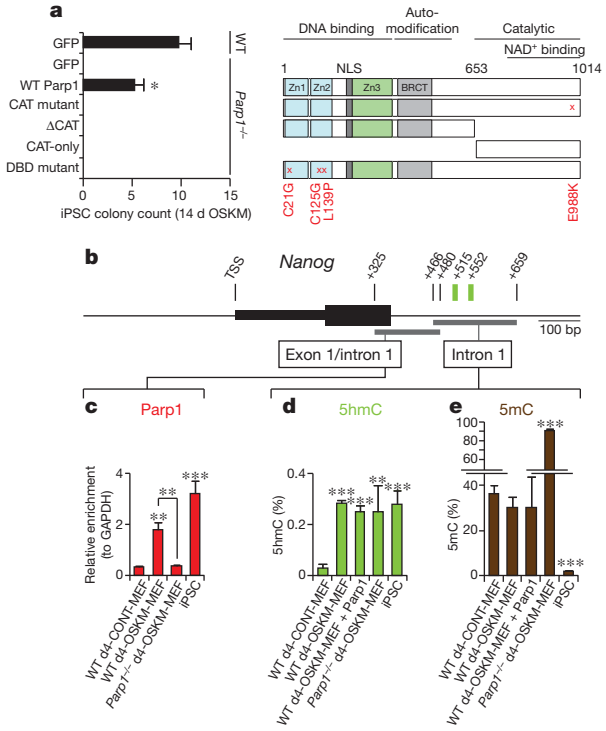


Figure 2 | Parp1 activities during iPSC reprogramming. **a**, Functional analysis of Parp1 mutants for rescue of iPSC colony formation in *Parp1*$^{-/-}$ OSKM-MEFs. Cultures were transduced with green fluorescent protein (GFP), WT Parp1, or Parp1 mutants encoding a catalytic domain missense mutation (CAT mutant; E988K), deletion of the catalytic domain (ΔCAT), deletion of the DNA-binding and automodification domains (CAT-only), or triple missense mutation of the DNA-binding domain (DBD mutant; C21G/C125G/L139P). Zn, zinc fingers; BRCT, BRCA1 carboxy terminus. **b**, Schematic representation of the *Nanog* locus transcription start site (TSS) region. Indicated are *Hpa*II/*Msp*I sites (green bars) and amplicons for 'exon 1/intron 1' and 'intron 1' regions (thick grey lines). bp, base pairs. **c**, Parp1 ChIP analyses of the cultures as indicated, presented as the relative enrichment to glyceraldehyde-3-phosphate dehydrogenase (GAPDH). **d**, Content of 5hmC assessed by GlucMS-qPCR (as a percentage of total cytosine). **e**, Content of 5mC, quantified by subtraction of 5hmC content (as in **d**) from the total methylated cytosine (5mC + 5hmC, as determined by *Hpa*II digestion insensitivity; see Supplementary Fig. 3f). Results in **a** and **c**–**e** are shown as means and s.d. for three independent experiments. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

increase in 5hmC relative to d4-CONT-MEFs (Fig. 2d and Supplementary Fig. 3c, d, n–q) at the pluripotency loci. In contrast to 5hmC, 5mC was not accumulated at the early time point (in d4-OSKM-MEFs, relative to d4-CONT-MEFs) at either locus (Fig. 2e and Supplementary Fig. 3j, m). The two pluripotency loci differed with respect to the overall pattern of 5mC during reprogramming, as previously described[2,19]: the *Nanog* locus showed a canonical pattern of hypermethylation (elevated 5mC) in MEFs and became demethylated (low 5mC) in reprogrammed iPSCs (Fig. 2e and Supplementary Fig. 3e, f), whereas the *Esrrb* locus showed a relatively low level of methylation even in MEFs, and remained hypomethylated in fully reprogrammed iPSCs (Supplementary Fig. 3i, j, l, m). Thus among these pluripotency loci, 5hmC rather than 5mC seemed to be an early and predictive epigenetic mark for subsequent activation. Immunocytochemistry analysis with an antibody against 5hmC showed increased nuclear staining in most cells in d4-OSKM-MEFs (relative to d4-CONT-MEF cultures; Fig. 3a).

We next sought to determine the role of Parp1 in the regulation of 5hmC and 5mC epigenetic marks at the pluripotency loci. Parp1 deficiency (which suppressed iPSC reprogramming) led to a consistent, large increase in 5mC accumulation in *Parp1*$^{-/-}$ d4-OSKM-MEFs, relative to WT d4-OSKM-MEF cultures at both the *Nanog* and *Esrrb* loci (Fig. 2e and Supplementary Fig. 3j, m). The increased 5hmC in WT d4-OSKM-MEFs was not suppressed in *Parp1*$^{-/-}$ d4-OSKM-MEFs; rather, in the context of Parp1 deficiency, 5hmC induction seemed similar to that of WT cells, for example at the *Nanog* locus, or modestly further increased, for example at the *Esrrb* locus (Fig. 2d and Supplementary Fig. 3d, n–q). Parp1 overexpression (which promoted iPSC reprogramming) did not consistently modify 5mC or 5hmC in d4-OSKM-MEFs, although a modest increase in 5hmC levels was observed at the *Esrrb* locus but not at the *Nanog* locus (Fig. 2d and Supplementary Fig. 3o, q). Taken together, these data implicate Parp1 in the regulation of 5mC; in contrast, we speculate that the variable impact of Parp1 on 5hmC may be indirect.

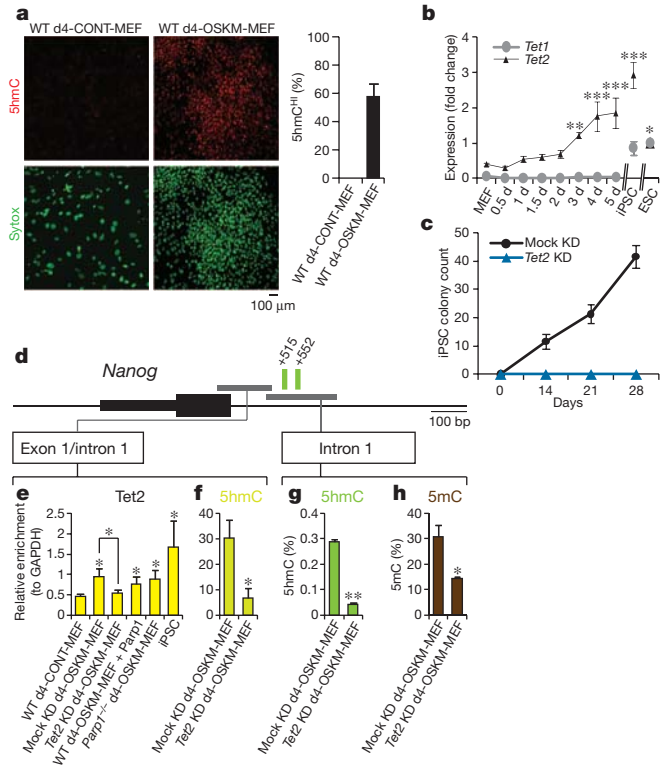


Figure 3 | Tet2 is required for 5hmC formation at the *Nanog* locus. **a**, Immunocytochemistry of d4-OSKM-MEFs and d4-CONT-MEFs with an antibody against 5hmC (ref. 5) (upper panels; red) and counterstained with Sytox nuclear marker (lower panels; green). Representative images show increased 5hmC in d4-OSKM-MEFs, as quantified on the right ($5hmC^{HI}$; defined as mean plus 2 s.d. or greater above the level in d4-CONT-MEFs). **b**, Time course of *Tet1* and *Tet2* gene expression assessed by qPCR (relative to ESC level). **c**, OSKM-mediated iPSC colony formation assay (AP^+) in shRNA-mediated *Tet2* knockdown (*Tet2* KD; blue) and non-silencing control shRNA (mock KD; black)-treated MEFs. **d**, Diagram of the *Nanog* locus; regions are the same as in Fig. 2b. **e**, Tet2 ChIP-qPCR at the exon 1/intron 1 amplicon. **f**, **g**, Content of 5hmC in the cultures indicated, assessed by hMeDIP of the exon 1/intron 1 region (**f**, relative to GAPDH) or GlucMS-qPCR intron 1 amplicon (**g**, as a percentage of total cytosine). **h**, Content of 5mC at the intron 1 amplicon, quantified by subtraction of 5hmC (as in **g**) from the total methylated cytosine levels (as in Supplementary Fig. 4k; determined by *Hpa*II sensitivity assay; as a percentage of total cytosine). Results in **a**–**c** and **e**–**h** are shown as means and s.d. for three independent experiments. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

Given the potential role of 5hmC early in somatic cell reprogramming, we obtained an expression time-course analysis of the TET enzymes, a family of Fe II and 2-oxoglutarate-dependent enzymes that generate 5hmC from 5mC (ref. 4). Expression of *Tet2*, but not *Tet1* or *Tet3*, was significantly induced in WT d4-OSKM-MEFs and remained elevated in iPSCs (Fig. 3b and Supplementary Fig. 4a). Consistent with a functional role for Tet2, short hairpin RNA (shRNA)-mediated *Tet2* knockdown (KD; Supplementary Table 8) abolished iPSC colony formation (Fig. 3c and Supplementary Fig. 4b, c). ChIP analysis with an antibody against Tet2 showed a direct interaction with the *Nanog* and *Esrrb* pluripotency loci in WT d4-OSKM-MEFs; this was not altered in *Parp1*$^{-/-}$ d4-OSKM-MEFs (Fig. 3d, e and Supplementary Fig. 4d, e, l, m). Furthermore, *Tet2* KD in d4-OSKM-MEFs suppressed the typical induction of 5hmC at both the *Nanog* and *Esrrb* pluripotency loci (Fig. 3f, g and Supplementary Fig. 4f, g). In contrast, the effect of *Tet2* KD in d4-OSKM-MEFs on 5mC seemed variable: 5mC seemed to be decreased at the *Nanog* locus but mildly increased at the *Esrrb* locus (which is typically hypomethylated even in MEFs[19]; Fig. 3h and Supplementary Fig. 4h–k). Given the early induction of 5hmC but not 5mC at the pluripotency loci, as well as the consequences of Tet2 deficiency in preventing transcriptional activation and 5hmC induction even in the absence of 5mC reduction, these data support a primary role for 5hmC as a distinct epigenetic mark in the somatic cell reprogramming process, and argue against an alternative model in which 5hmC functions simply as an intermediate in the 5mC demethylation process[20]. Epigenetic marks with 5hmC may recruit select chromatin modification factors to the pluripotency loci[21].

To further probe the roles of Tet2 and Parp1 in early epigenetic events, we evaluated the chromatin state of the *Nanog* and *Esrrb* loci in d4-OSKM-MEFs. ChIP analysis revealed an enrichment in the occupancy of these loci by the activation-associated marker histone H3 lysine 4 dimethylation (H3K4me2)[22,23] and a parallel decrease in the transcriptional-silencing-associated marker histone H3 lysine 27 trimethylation (relative to d4-CONT-MEFs; H3K27me3 (refs 24–26); Fig. 4a–d and Supplementary Fig. 5a, b). Deficiency of either Parp1 or Tet2 diminished the H3K4me2 chromatin mark at the pluripotency loci of d4-OSKM-MEFs (Fig. 4a, c and Supplementary Fig. 5a). The effects on H3K27me3 were variable: Parp1 deficiency did not significantly alter H3K27me3 at either locus, whereas *Tet2* KD led to a decrease at the *Nanog* locus but not at the *Esrrb* locus (Fig. 4b, d and Supplementary Fig. 5b). We speculated that altered chromatin states in d4-OSKM-MEFs may affect chromatin accessibility at the pluripotency loci, for example to the transduced Oct4 pluripotency factor. Oct4 occupancy, as quantified by ChIP of d4-OSKM-MEFs, was significantly diminished in the context of Parp1 deficiency at both pluripotency loci, whereas *Tet2* KD did not diminish Oct4 occupancy (Fig. 4e, g and Supplementary Fig. 5c). Consistent with this, Parp1 overexpression potentiated Oct4 binding at both pluripotency loci of d4-OSKM-MEFs (Fig. 4e, g and Supplementary Fig. 5c). Furthermore,

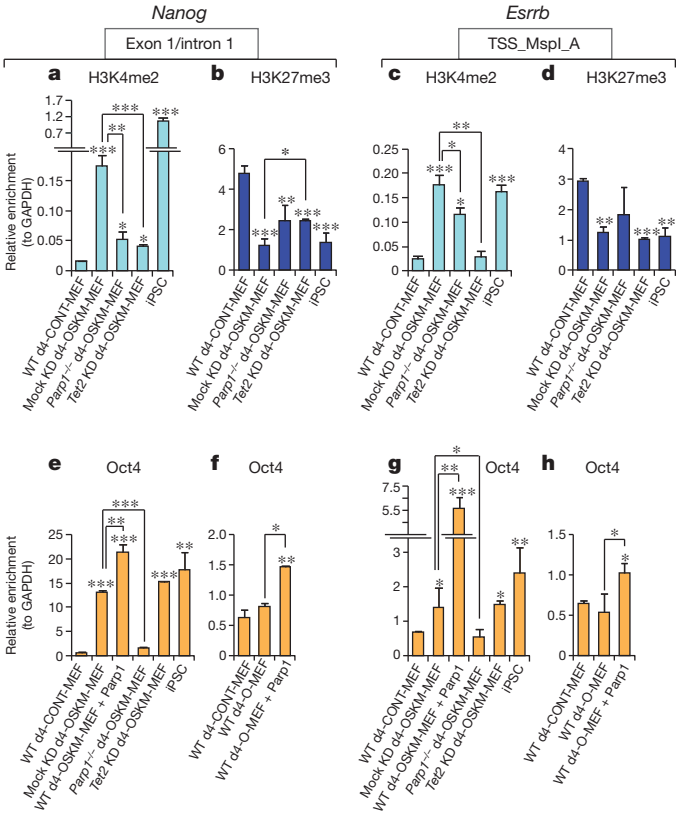


Figure 4 | Impact of Parp1 and Tet2 on chromatin state and Oct4 accessibility at the *Nanog* and *Esrrb* loci. **a**–**d**, H3K4me2 (**a**, **c**) and H3K27me3 (**b**, **d**) ChIP-qPCR at *Nanog* or *Esrrb* amplicons in cultures as indicated. **e**–**h**, Oct4 ChIP-qPCR. Results are shown as means and s.d. for three independent experiments. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

Parp1 overexpression robustly promoted exogenous Oct4 binding to the pluripotency loci even in the absence of transduction of the other pluripotency factors necessary for somatic cell reprogramming (d4-O-MEFs without SKM; Fig. 4f, h and Supplementary Fig. 5d).

Taken together, these data support necessary but distinct roles for Tet2 and Parp1 in the regulation of epigenetic marks and local chromatin structure at pluripotency loci during an early stage of somatic cell reprogramming that precedes their transcriptional activation (Supplementary Fig. 1a, b). The data further suggest that 5hmC, generated by Tet2, does not simply represent an intermediate in the 5mC demethylation process, but functions as an epigenetic mark, possibly recruiting *trans*-acting factors that promote chromatin remodelling[4,21]. The induction of *Parp1* and *Tet2* gene expression early in the course of reprogramming may reflect their direct activation by OSKM factors (Supplementary Fig. 5e, f)[27]. Finally, loss of Tet2 function is strongly implicated in human malignancies[8], and thus Tet2-mediated chromatin remodelling may affect tumour risk associated with potential iPSC therapies.

METHODS SUMMARY

Cell culture and generation of iPSCs. MEF were prepared from WT or *Parp1*$^{-/-}$ embryonic day 13.5 embryos (129S-*Parp1*^{tm1Zqw}/J; Jackson Laboratories)[28]. Tail-tip fibroblasts were prepared from WT and *Tet2*$^{-/-}$ mice[29]. For iPSC generation, MEFs were transduced by incubation with OSKM retroviral supernatants for 24 h as described[1]. Subsequently, cells were cultured in ESC medium and samples were collected at the time points indicated. Epigenetic modification factors were cloned into lentiviral vectors (pLenti6.3; Invitrogen) for lentiviral production and subsequent transduction.

***Tet2* knockdown.** *Tet2* knockdown was achieved by using a cocktail of five shRNA lentiviral vectors (Open Biosystems) specific for *Tet2*, or control non-silencing shRNA.

Mass spectrometry. Multidimensional protein identification technology (MudPIT) was performed to analyse nuclear fractions of MEFs, iPSC clones and ESC clones as detailed in Supplementary Methods.

Chromatin immunoprecipitation. ChIP was performed with the Magnify kit (492024; Invitrogen) and the following antibodies: anti-dimethyl K4 of H3 (2 μg; 07-030; Millipore), anti-trimethyl K27 of H3 (2 μg; ab6002; Abcam), anti-Oct4 (2 μg; sc-8628 X; Santa Cruz), anti-Parp1 (2 μg; sc-74469 X; Santa Cruz), anti-Tet2 (2 μg; sc-136926; Santa Cruz), normal goat IgG (2 μg; 005-000-003; Jackson ImmunoResearch), normal rabbit IgG (1 μg; Magnify; Invitrogen) and normal mouse IgG (1 μg; Magnify; Invitrogen).

Digestion with *Hpa*II and *Msp*I. Detection of 5hmC and 5mC as percentages of total cytosine species was performed with the EpiMark Kit (E3317S; NEB). The technique has been described in detail in ref. 16.

Pyrosequencing. Genomic DNA (1 μg) was bisulphite-converted with the EZ DNA Methylation Kit (D5001; Zymo Research), followed by amplification by polymerase chain reaction with the PyroMark PCR Kit (978703; Qiagen). PCR products were sequenced with the PyroMark Q24 instrument (Qiagen).

Received 3 August 2011; accepted 18 June 2012.
Published online 19 August 2012.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
3. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).
4. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
5. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).
6. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18,** 553–567 (2010).
7. Guo, J. U. *et al.* Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).
8. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468,** 839–843 (2011).
9. Krishnakumar, R. & Kraus, W. L. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell* **39,** 8–24 (2010).
10. Wacker, D. A. *et al.* The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription. *Mol. Cell. Biol.* **27,** 7475–7485 (2007).
11. Langelier, M. F. *et al.* The Zn3 domain of human poly(ADP-ribose) polymerase-1 (PARP-1) functions in both DNA-dependent poly(ADP-ribose) synthesis activity and chromatin compaction. *J. Biol. Chem.* **285,** 18877–18887 (2010).
12. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329,** 78–82 (2010).
13. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
14. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473,** 394–397 (2011).
15. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25,** 679–684 (2011).
16. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
17. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473,** 389–393 (2011).
18. Davis, T. & Vaisvila, R. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *J. Vis. Exp.* (48), e2661, http://dx.doi.org/10.3791/2661 (2011).
19. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
20. Bhutani, N., Burns, D. M. & Blau, H. M. DNA demethylation dynamics. *Cell* **146,** 866–872 (2011).
21. Yildirim, O. *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147,** 1498–1510 (2011).
22. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–181 (2005).
23. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
24. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
25. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).
26. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
27. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8,** 200–213 (2011).
28. Wang, Z. Q. *et al.* Mice lacking ADPRT and poly(ADP-ribosyl)ation develop normally but are susceptible to skin disease. *Genes Dev.* **9,** 509–520 (1995).
29. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20,** 11–24 (2011).
30. Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nature Biotechnol.* **28,** 521–526 (2010).

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

Acknowledgements We thank G. Q. Daley, A. P. Feinberg, A. Doi, R. M. Santella and M. A. Kappil for reagents and for technical assistance with pyrosequencing; A. Califano and A. Lachmann for assistance with the bioinformatics analyses; E. O. Mazzoni for assistance with the ChIP analyses; and O. Hobert for critical reading of the manuscript. This work was supported by New York State Stem Cell Science (NYSTEM) grants C024402 and C024403 and National Institutes of Health (NIH) grant RO1 NS064433 to A.A., NYSTEM Institution Development Grant N08G-071 to E.I.C., NIH grant RO1 138424 to R.L.L, and a shared NIH/National Center for Research Resources instrument grant for mass spectrometry, 1 S10 RR023680-1.

Author Contributions C.A.D. and A.A. designed the experiments and analysed data. C.A.D., D.B.R., S.T., R.F. and W.B.V. conducted molecular and cellular experiments. T.Y., G.B. and K.I. performed and analysed murine *in vivo* studies. R.L.L. and A.S. supplied essential reagents. P.G. performed bioinformatics analyses. S.N. and E.I.C. conducted proteomics. C.A.D. and A.A. wrote 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 A.A. (aa900@columbia.edu).

LETTER

doi:10.1038/nature11323

IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics

Masato Sasaki[1]*, Christiane B. Knobbe[1,2]*, Joshua C. Munger[3], Evan F. Lind[1], Dirk Brenner[1], Anne Brüstle[1], Isaac S. Harris[1,4], Roxanne Holmes[5], Andrew Wakeham[1], Jillian Haight[1], Annick You-Ten[1], Wanda Y. Li[1], Stefanie Schalm[9], Shinsan M. Su[9], Carl Virtanen[6], Guido Reifenberger[2], Pamela S. Ohashi[1], Dwayne L. Barber[4], Maria E. Figueroa[7], Ari Melnick[8], Juan-Carlos Zúñiga-Pflücker[5] & Tak W. Mak[1,4]

Mutations in the *IDH1* and *IDH2* genes encoding isocitrate dehydrogenases are frequently found in human glioblastomas[1] and cytogenetically normal acute myeloid leukaemias (AML)[2]. These alterations are gain-of-function mutations in that they drive the synthesis of the 'oncometabolite' R-2-hydroxyglutarate (2HG)[3]. It remains unclear how *IDH1* and *IDH2* mutations modify myeloid cell development and promote leukaemogenesis. Here we report the characterization of conditional knock-in (KI) mice in which the most common *IDH1* mutation, IDH1(R132H), is inserted into the endogenous murine *Idh1* locus and is expressed in all haematopoietic cells (Vav-KI mice) or specifically in cells of the myeloid lineage (LysM-KI mice). These mutants show increased numbers of early haematopoietic progenitors and develop splenomegaly and anaemia with extramedullary haematopoiesis, suggesting a dysfunctional bone marrow niche. Furthermore, LysM-KI cells have hypermethylated histones and changes to DNA methylation similar to those observed in human *IDH1*- or *IDH2*-mutant AML. To our knowledge, our study is the first to describe the generation and characterization of conditional IDH1(R132H)-KI mice, and also the first report to demonstrate the induction of a leukaemic DNA methylation signature in a mouse model. Our report thus sheds light on the mechanistic links between *IDH1* mutation and human AML.

IDH1/*IDH2* mutations typically produce mutant enzymes with aberrant activity. Whereas wild-type (WT) Idh proteins metabolize isocitrate and $NADP^+$ to yield α-ketoglutarate (αKG) and NADPH, mutant Idh proteins convert αKG into 2HG while consuming NADPH[3]. 2HG competitively inhibits tet methylcytosine dioxygenases (Tet2), which regulate DNA methylation, as well as JmjC domain-containing histone demethylases[4–6]. Accordingly, human AML cells with *IDH1*/*IDH2* mutation show global DNA hypermethylation[5].

To create a murine model of the IDH1(R132H) mutation, we used the lox-stop-lox (LSL) system to generate a conditional *Idh1* knock-in (KI) mouse ($Idh1^{LSL/WT}$) (Supplementary Fig. 2). Mutant mice were crossed with LysMCre mice[7] ($Idh1^{LSL/WT}LysMCre^{+/WT}$, LsyM-KI), because the LysM promoter is activated very early during myeloid development[8] and human AML leukaemic stem cells show similarities with very early myeloid primed progenitors (LMPP)[9].

LysM-KI mice were born at the expected Mendelian ratio, were viable and fertile, and had normal lifespans. However, serum 2HG levels were elevated approximately tenfold in both young (7–16 weeks old) and older (47–56 weeks old) LysM-KI mice (Supplementary Fig. 3a–c). Whereas peripheral blood cell counts of young LysM-KI mice were normal, older (42–46 weeks old) mutants developed anaemia (Supplementary Fig. 3d, e). In the OP9/OP9-DL1 *in vitro* differentiation system[10], haematopoietic cell lineages developed normally from several IDH1(R132H)-KI embryonic stem cell clones (Supplementary Fig. 4). Macroscopic analysis of mice revealed mild splenic enlargement in young LysM-KI mice that progressed to overt splenomegaly in all older mutants. Histologically, splenic architecture became increasingly disorganized, with age-dependent expansion of spaces between lymphoid follicles and obvious extramedullary haematopoiesis (Fig. 1a, b).

We next evaluated the haematopoietic stem cell (HSC) and haematopoietic progenitor cell (HPC) compartments in LysM-KI mice. Bone marrow (BM) cellularity was normal in young LysM-KI mice but reduced in older mutants (Fig. 1c), where it correlated inversely with splenomegaly and the degree of anaemia. Histologically, the BM of young mutants appeared normal but the BM of older mutants showed fewer mature cells and more immature cells (Fig. 1d). Flow cytometric analyses of BM and spleen revealed altered numbers of mature cells ($CD11b^+$, $Gr1^+$, $B220^+$, $CD4^+$, $CD8^+$) in LysM-KI mice (Fig. 1e, f).

More refined flow cytometric analyses confirmed that lineage-negative (Lin^-) cells underwent an age-dependent expansion in LysM-KI BM (Fig. 2a, b and Supplementary Table 1). Moreover, LSK ($Lin^-Sca1^+cKit^+$) cells were increased by approximately 1.8-fold in young LysM-KI mice and by approximately 5.4-fold in older mutants (Fig. 2a, b and Supplementary Table 1). The LSK population contains long-term HSCs (LT-HSCs; $CD150^+CD48^-$), short-term HSCs (ST-HSCs; $CD150^+CD48^+$), and lineage-restricted progenitors (LRPs; $CD150^-CD48^+$). We found that LysM-KI mice demonstrated an age-dependent increase in LRPs (Fig. 2b and Supplementary Table 1). There were no significant alterations to the LK population ($Lin^-Sca1^-cKit^+$), which contains common myeloid progenitors (CMPs; $CD34^+CD16/32^{med}$LK), granulocyte-macrophage progenitors (GMPs; $CD34^+CD16/32^{high}$LK), and megakaryocyte-erythroid progenitors (MEPs; $CD34^-CD16/32^{low}$LK), or to the common lymphoid progenitor population (CLPs; $Lin^-Il7Ra^+Sca1^{med}cKit^{med}$) (Fig. 2b and Supplementary Table 1). In colony-forming cell (CFC) assays, BM cells from young or older LysM-KI mice showed statistically normal production of granulocyte (G), granulocyte-macrophage (GM), granulocyte-erythrocyte-macrophage-megakaryocyte (GEMM), and macrophage (M) colonies (Fig. 2c). Flow cytometric analysis of nucleated splenic cells from older LysM-KI mice revealed significantly elevated numbers of LSK and LRP cells, recapitulating the pattern observed in the bone marrow of these mice (Fig. 2a, d). CFC assays of nucleated splenic cells showed increased numbers of all four myeloid colony types in LysM-KI mice (Fig. 2e). These data confirm that extramedullary haematopoiesis occurs in older LysM-KI mice.

We speculated that the LRP accumulation in LysM-KI BM (and spleen), despite the near-normal peripheral blood counts of these

[1]The Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, University Health Network, Toronto, Ontario M5G 2C1, Canada. [2]Department of Neuropathology, Heinrich Heine University, 40225 Düsseldorf, Germany. [3]School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642, USA. [4]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2C1, Canada. [5]Department of Immunology, Sunnybrook Research Institute, University of Toronto, Toronto, Ontario M4N 3M5, Canada. [6]University Health Network Microarray Center, Toronto, Ontario M5S 1A8, Canada. [7]Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. [8]Weill Cornell Medical College, Cornell University, New York, New York 10021, USA. [9]Agios Pharmaceuticals Incorporated, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

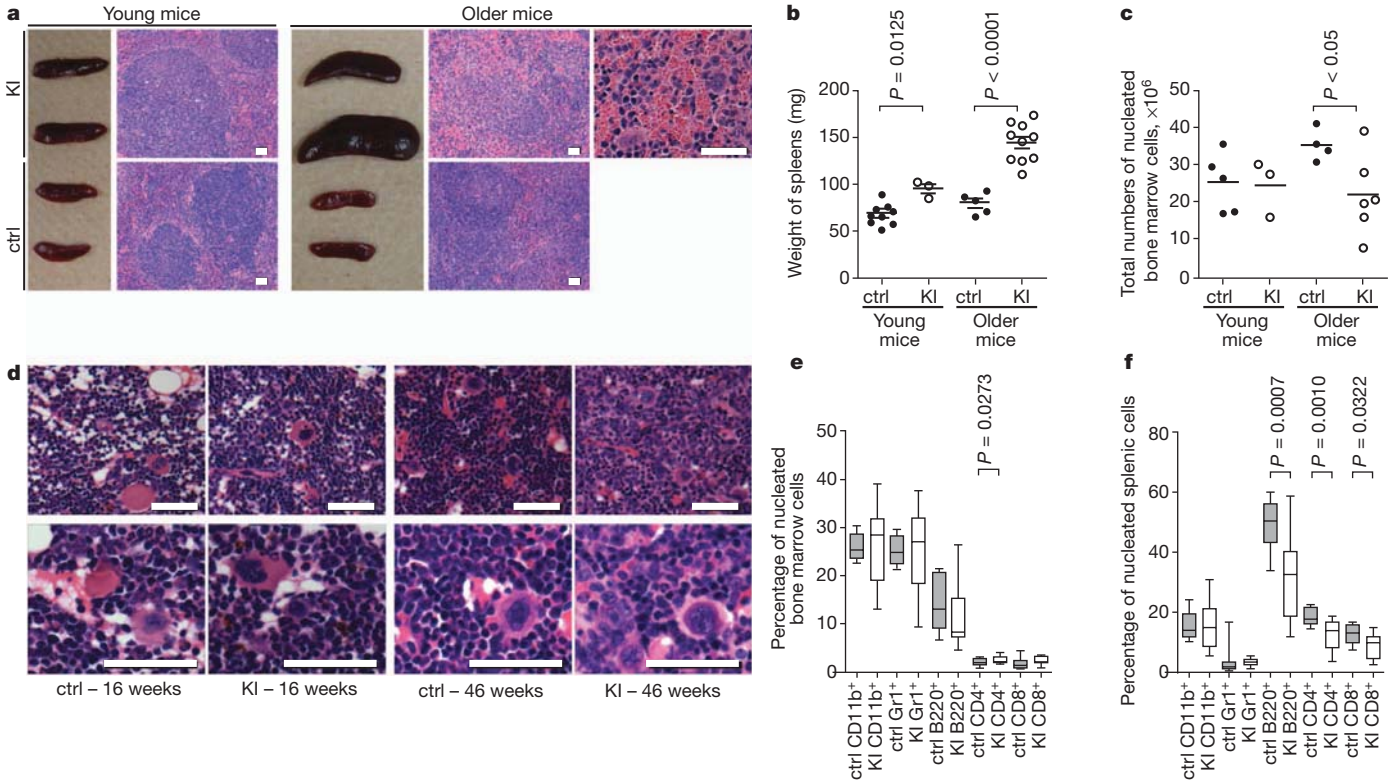


Figure 1 | LysM-KI mice show age-dependent splenomegaly and decreased bone marrow cellularity. **a**, Macroscopic appearance and haematoxylin-and-eosin-stained sections of spleens from young and older control (ctrl) and LysM-KI (KI) mice. **b**, **c**, Spleen weights (**b**) and total numbers of nucleated BM cells (**c**) from young and older control and LysM-KI mice ($n = 3$–5 per group). Horizontal line, mean. **d**, Representative haematoxylin-and-eosin-stained sections of BM from femurs of young and older control and LysM-KI mice ($n = 4$ per group). **e**, **f**, Quantifications of flow cytometric analyses of mature haematopoietic cell populations among total nucleated BM (**e**) and splenic cells (**f**) from older control and LysM-KI ($n = 10$–14 per group) mice. *P* values were determined using the unpaired *t*-test with Welch's correction. Scale bars, 50 μm.

animals and their normal BM cell CFC activity, might be due to increased symmetric cell division and proliferation, and/or a partial block in differentiation. Serial plating experiments showed that, whereas control BM cells stopped proliferating after three rounds of plating (24 days), LysM-KI BM cells continued to grow at an exponential rate for six rounds of plating (45 days) (Fig. 2f).

To extend our findings, we crossed $Idh1^{LSL/WT}$ mice with VavCre mice[11] to generate Vav-KI mutants expressing IDH1(R132H) in all haematopoietic cells, from LT-HSCs to terminally differentiated cells. Lin-negative, LSK, LRP and CLP cells were all significantly increased in BM of young Vav-KI mice (Supplementary Fig. 5a), as were splenic LSK, LK, LRP, MEP and CLP cells (Supplementary Fig. 5b). Like young LysM-KI mice, young Vav-KI mice had normal total BM cell numbers and showed no overt haematological changes (Supplementary Fig. 5c–e).

Next, we performed competitive BM repopulation assays in which equal numbers (10^5) of nucleated BM cells from donor mice (control or LysM-KI; $CD45.2^+$) and from recipient mice (control; $CD45.1^+$) were injected into lethally-irradiated recipients ($CD45.1^+$). LysM-KI BM cells showed no defects in short-term or long-term repopulation capacity for at least 170 days after transplantation, and relative peripheral blood cell counts were not altered in recipients of LysM-KI BM (Supplementary Fig. 6a, b).

The homeostasis of HSCs and HPCs is influenced by reactive oxygen species (ROS)[12,13]. Because high ROS reduces HSC longevity, these cells strive to keep ROS low[13]. In contrast, higher levels of ROS in developing myeloid cells may be necessary for their differentiation[13]. ROS elevation in AML cells triggers their differentiation[14]. It has been assumed that *IDH1*/*IDH2* mutations increase ROS[3]. However, when using CM-H_2DCFDA (5-(and-6)-chloromethyl-2′,7′-dichlorodihydrofluorescein diacetate, acetyl ester; a membrane-permeable indicator of reactive oxygen species) to measure total ROS in granulocytes ($CD11b^+Gr1^+$), macrophages/monocytes ($CD11b^+Gr1^{low}$), LSK and LK cells in BM of young and older LysM-KI mice, no abnormalities were observed (Supplementary Fig. 7a, b). Liquid chromatography-mass spectrometry (LC-MS) analysis of 2HG as well as of NAD^+, NADH, $NADP^+$ and NADPH in $CD11b^+$ cells isolated from BM and spleen of control and LysM-KI mice, and in BM-derived macrophages (BMDMs) revealed highly elevated 2HG in all LysM-KI samples, whereas NAD^+, NADH, $NADP^+$ and NADPH levels were not altered (Supplementary Fig. 7c–e). Taken together, these data indicate that the phenotype of LysM-KI mice is not due to alterations in ROS or NADPH levels, nor to 2HG-mediated complete abolition of myeloid differentiation.

The hypoxia-inducible transcription factors Hif1α and Hif2α are critical for HSC survival and maintenance[15] but their involvement in HPC differentiation and beyond is unclear. *IDH1* mutations have been shown to alter Hif1α stability[6,16]. We compared messenger RNA levels of several Hif1α target genes in sorted control and LysM-KI LSK cells but observed no differences (Supplementary Fig. 8). Thus, it is unlikely that Hif1α signalling is altered in LysM-KI LSK cells.

To analyse global mRNA expression, we performed gene expression microarray analyses of sorted LSK cells from young LysM-KI and control mice. Two-way hierarchical clustering of significantly altered genes (*P* value < 0.05, >1.5-fold change) revealed a clear separation of control and LysM-KI samples (Supplementary Fig. 9). Gene ontology analysis of the identified gene signature (Supplementary Table 3) revealed a significant enrichment of genes in several categories related to cellular growth and proliferation (Supplementary Table 4).

Human gliomas[17] and AMLs with *IDH1*/*IDH2* mutations have DNA hypermethylation that, in AML, is caused by 2HG-mediated inhibition of Tet2 (ref. 5). 2HG also inhibits histone demethylases, resulting in hypermethylation of histone lysine residues (especially in H3)[4,6]. To examine DNA methylation, we performed high-throughput

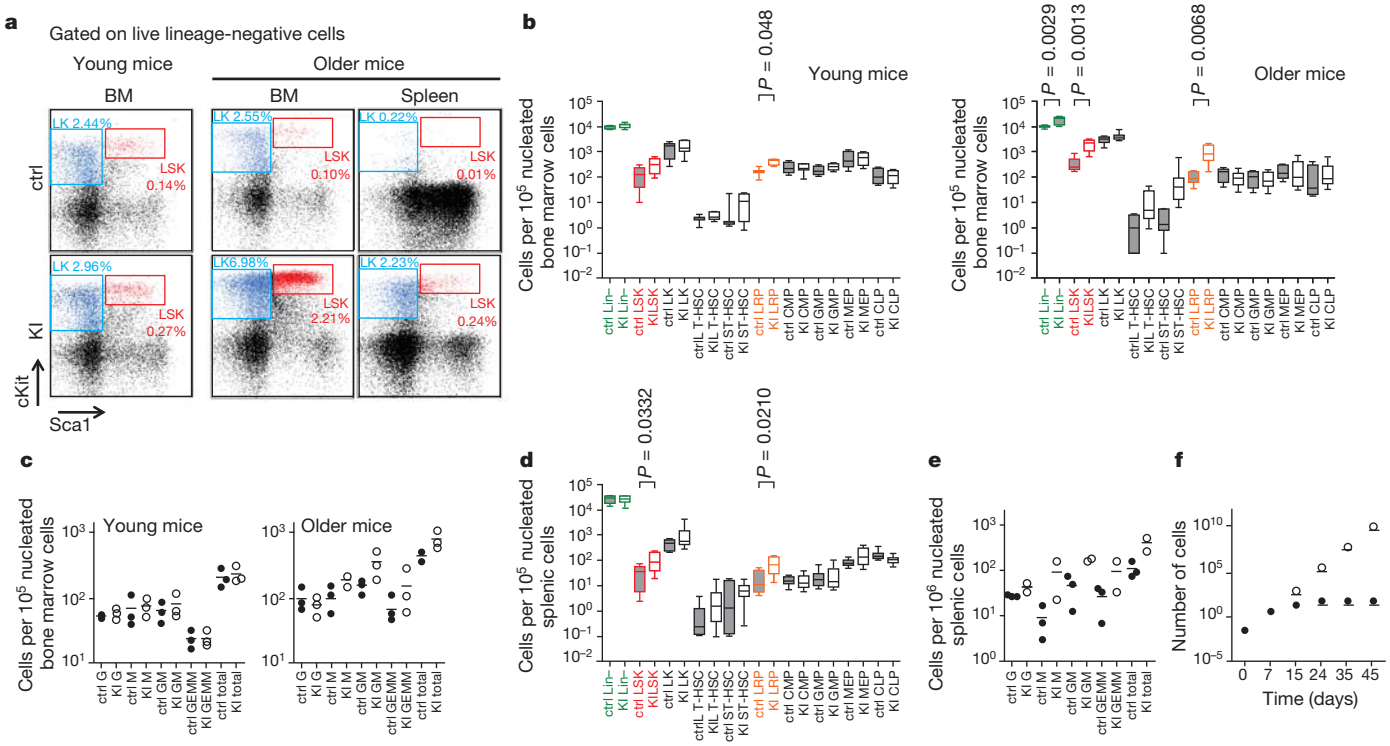


Figure 2 | LysM-KI mice show age-dependent increases in lineage-restricted progenitors and extramedullary haematopoiesis. **a**, Flow cytometric analysis of live lineage-negative cells for cKit and Sca1 in cells from the BM of young (left), and from the BM and spleen of older control and LysM-KI mice (right). LSK (red) and LK (blue) populations are highlighted. The percentage of a given population among total viable BM cells is indicated. **b**, Quantification of flow cytometric analyses for the indicated HSC and HPC populations in BM of young ($n = 3$–14 per group) and older ($n = 4$–8 per group) control and LysM-KI mice. Horizontal line, median; box, median ± 25%; end of bars, minimum and maximum. **c**, CFC assays of BM cells from young and older control and LysM-KI mice ($n = 3$ per group). Data points are the mean number of colonies of the indicated lineage for individual mice. Horizontal line, mean of values for all three mice in a group. **d**, Quantification of flow cytometric analyses for the indicated HSC and HPC populations in the spleens of older ($n = 7$–8 per group) control and LysM-KI mice, determined as for **b**. **e**, CFC assays of nucleated splenic cells from older control ($n = 3$) and LysM-KI ($n = 2$) mice determined as for **c**. **f**, Serial replating in methylcellulose of BM cells from control and LysM-KI mice ($n = 2$ per group; control, filled circles; LysM-KI, empty circles). Data are the mean total cell number for both mice in a group (in units of 10^6).

sequencing of bisulphite-treated DNA from sorted LSK cells from young LysM-KI and control mice. DNA from LysM-KI LSK cells showed a significantly greater proportion of highly methylated CpG sites, with a marked increase in CpG sites showing greater than 80% methylation (Fig. 3a). A more detailed analysis of the differential distribution pattern of DNA methylation revealed wide-spread global changes, with all chromosomes equally affected (Supplementary Fig. 10a, b and Supplementary Table 5). A clear preference for promoters and intragenic regions was observed that paralleled the changes in DNA methylation observed in human *IDH1*/*IDH2*-mutant AMLs (Fig. 3b, Supplementary Fig. 10a, b and Supplementary Table 5)[18]. Comparison of the 1,559 hypermethylated murine genes in LysM-KI LSK cells to the 8195 hypermethylated human genes in *IDH1*/*IDH2*-mutant AMLs revealed an overlap of 784 genes ($P = 1.63 \times 10^{-36}$; Supplementary Table 6). Ingenuity pathway analysis (IPA) implicated several signalling pathways involved in haematopoietic cell proliferation and differentiation, leukaemogenesis and leukaemic stem cell maintenance (Supplementary Table 7 and Supplementary Fig. 11)[19–22] including the WNT, NOTCH and TGF-β pathways, which are also targeted by aberrant methylation in human *IDH1*/*IDH2*-mutant AMLs[18]. Immunoblotting of lysates from control and LysM-KI BMDMs revealed increased methylation of multiple H3 lysine residues (Fig. 3c).

In conclusion, we show that the IDH1(R132H) mutation in all haematopoietic cells (Vav-KI), and in the myeloid lineage in particular (LysM-KI), results in the accumulation of LSK and LRP cells first in the bone marrow and later in the spleen. We believe that the accumulations of these cells are most likely due to 2HG-induced DNA and histone hypermethylation that affect LSK cell division and/or differentiation (Supplementary Fig. 1). Our finding that IDH1(R132H) does

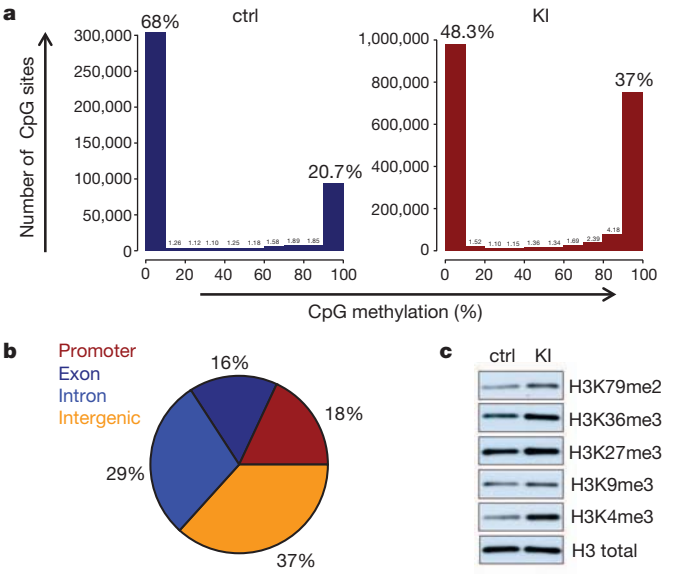


Figure 3 | Altered methylation of DNA and histones in LysM-KI cells. **a**, Sequencing data from bisulphite-treated DNA from sorted BM LSK cells of young control ($n = 2$) and LysM-KI ($n = 3$) mice. Numbers of fragments showing the indicated percentage of CpG methylation were calculated as detailed in Methods. Data shown are from one LysM-KI and one control mouse and are representative of all animals in a group. **b**, Graphic representation of the differentially methylated genomic regions with respect to their genomic location. **c**, Immunoblot analysis of methylation of the indicated H3 lysine residues in lysates of BMDMs derived from young control and LysM-KI mice. Results are representative of two independent experiments.

not completely block myeloid differentiation may explain why *IDH1*/*IDH2* mutations are found in almost all FAB (French-American-British classification) AML subtypes[23]. Moreover, our mutants show extramedullary haematopoiesis, suggesting a dysfunctional bone marrow niche. Extramedullary haematopoiesis occurs in myelodysplastic syndromes, which frequently progress to AML and harbour *IDH1*/*IDH2* mutations in 10–15% of cases[24]. Our study is, to our knowledge, the first to model epigenetic changes of human *IDH1*/*IDH2*-mutant AML in mice and thereby advances and will advance our understanding of the links between *IDH1*/*IDH2* mutations and leukaemogenesis.

METHODS SUMMARY

Methods used for the generation of the conditional IDH1(R132H) KI mouse, for the generation of KI embryonic stem cell clones expressing IDH1(R132H), the OP9 *in vitro* differentiation system, for histology, flow cytometric analyses, cell sorting, CFC assays, competitive BM repopulation assays, generation of BMDMs, mRNA extraction, cDNA synthesis, real-time RT–PCR analysis, immunoblotting, mass spectrometry, DNA methylation analysis, and mRNA expression profiling are detailed in the Methods.

For all experiments, LysM-KI mice were heterozygous for both the Idh^{LSL} allele and the LysMCre allele ($Idh1^{LSL/WT}LysMCre^{+/WT}$), and Vav-KI mice were heterozygous for both the Idh^{LSL} allele and the VavCre allele ($Idh1^{LSL/WT}VavCre^{+/WT}$). Control mice were wild-type animals or animals heterozygous for either the $Idh1^{LSL}$ allele or the LysMCre allele, if LysM-KI mice were used, or the VavCre allele, if Vav-KI mice were used ($Idh1^{WT/WT}LysMCre^{WT/WT}$, $Idh1^{WT/WT}LysMCre^{+/WT}$, $Idh1^{LSL/WT}VavCre^{WT/WT}$, or $Idh1^{WT/WT}VavCre^{+/WT}$). All animal procedures were approved by the Animal Care and Use Committee of the University Health Network (Toronto, Ontario, Canada).

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

Received 18 November 2011; accepted 18 June 2012.

Published online 4 July; corrected online 29 August 2012 (see full-text HTML for details).

1. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321,** 1807–1812 (2008).
2. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
3. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462,** 739–744 (2009).
4. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19,** 17–30 (2011).
5. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18,** 553–567 (2010).
6. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12,** 463–469 (2011).
7. Clausen, B. E. *et al.* Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8,** 265–277 (1999).
8. Ye, M. *et al.* Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. *Immunity* **19,** 689–699 (2003).
9. Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19,** 138–152 (2011).
10. Holmes R. & Zúñiga-Pflücker J.-C. The OP9–DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells *in vitro*. *Cold Spring Harb. Protoc.* http://dx.doi.org/10.1101/pdb.prot5156 (2009).
11. de Boer, J. *et al.* Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33,** 314–325 (2003).
12. Ito, K. *et al.* Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nature Med.* **12,** 446–451 (2006).
13. Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128,** 325–339 (2007).
14. Callens, C. *et al.* Targeting iron homeostasis induces cellular differentiation and synergizes with differentiating agents in acute myeloid leukemia. *J. Exp. Med.* **207,** 731–750 (2010).
15. Eliasson, P. & Jönsson, J. I. The hematopoietic stem cell niche: low in oxygen but a nice place to be. *J. Cell. Physiol.* **222,** 17–22 (2010).
16. Zhao, S. *et al.* Glioma-derived mutations in *IDH1* dominantly inhibit IDH1 catalytic activity and induce HIF-1α. *Science* **324,** 261–265 (2009).
17. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17,** 510–522 (2010).
18. Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* (in the press).
19. Campbell, C. *et al.* Signal control of hematopoietic stem cell fate: Wnt, Notch, and Hedgehog as the usual suspects. *Curr. Opin. Hematol.* **15,** 319–325 (2008).
20. Heidel, F. H. *et al.* Self-renewal related signaling in myeloid leukemia stem cells. *Int. J. Hematol.* **94,** 109–117 (2011).
21. Klinakis, A. *et al.* A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature* **473,** 230–233 (2011).
22. Söderberg, S. S. *et al.* Complex and context dependent regulation of hematopoiesis by TGF-β superfamily signaling. *Ann. NY Acad. Sci.* **1176,** 55–69 (2009).
23. Patel, K. P. *et al.* Acute myeloid leukemia with *IDH1* or *IDH2* mutation: frequency and clinicopathologic features. *Am. J. Clin. Pathol.* **135,** 35–45 (2011).
24. Patnaik, M. M. *et al.* Differential prognostic effect of *IDH1* versus *IDH2* mutations in myelodysplastic syndromes: a Mayo Clinic Study of 277 patients. *Leukemia* **26,** 101–105 (2012).

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

Acknowledgements We thank the Animal Research Colony (ARC) at the Ontario Cancer Institute for mouse care; I. Ng, A. Shahinian, J. Sylvester and S. McCracken for administrative and organizational expertise; M. Bailey and J. Tsao for technical assistance; F. Tong and R. Nayyar for assistance with flow cytometric analysis and sorting; the Weill Cornell Medical College (WCMC) Epigenomics Core Facility for technical help and expertise; G. Melino, D. Green, M. Minden, H. Chang and P. Lang for helpful discussions; J. Thomsen for figure layout and M. Saunders for scientific editing. C.B.K. and D.B. were supported in part by a Feodor-Lynen Postdoctoral Research Fellowship from the Alexander-von-Humboldt-Foundation, Germany. D.B. and A.B. were supported in part by a Fellowship from the German Research Foundation (DFG). J.C.M. is supported by a National Institute of Health grant (NIH R01AI081773) and is a Damon Runyon-Rachleff Innovation Awardee supported by the Damon Runyon Cancer Research Foundation (DRR-09-10). P.S.O. holds a Canada Research Chair in Autoimmunity and Tumor Immunity. M.E.F. is supported by the Leukemia & Lymphoma Society Special Fellow Award and a Doris Duke Clinical Scientist Development Award. A.M. is supported by an LLS SCOR grant (7132-08), a Burroughs Wellcome Clinical Translational Scientist Award and a Starr Cancer Consortium grant (I4-A442). J.-C.Z.-P. is supported by a Canada Research Chair in Developmental Immunology. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) and the Ontario Ministry of Health and Long Term Care to T.W.M., and a program grant from the Terry Fox Foundation to P.S.O., J.-C.Z.-P. and T.W.M. Please note that the views expressed do not necessarily reflect those of the OMOHLTC.

Author Contributions C.B.K., M.S. and T.W.M. initiated the project. M.S. designed the conditional *IDH1(R132H)*-KI mouse and generated it, as well as the IDH1(R132H)-KI embryonic stem cell clones with the help of S.S., S.M.S., A.W., J.H. and A.Y.-T.; C.B.K. designed and organized experiments involving the hematopoietic system in consultation with T.W.M., M.E.F., A.M., J.-C.Z.-P., G.R., P.S.O., C.V. and D.L.B.; C.B.K. performed the flow cytometry analyses, bone marrow transfers, cell sorting, RNA extraction and real-time PCR analyses, immunoblotting, colony formation assays and serial plating experiments. J.C.M. performed the LC-MS analyses. E.F.L., D.B., A.B. and I.S.H. helped with flow cytometric analyses and BM transfers. R.H. performed the OP9/OP9-DL1 embryonic stem cell *in vitro* differentiation experiments. W.Y.L. helped with flow cytometric analyses and cell culture experiments. M.E.F. and A.M. performed the DNA methylation experiments and analysed the data. C.V. performed the mRNA expression microarray experiments and analysed data. C.B.K. and T.W.M. wrote the manuscript with the help of A.B. and D.B. All authors discussed the results extensively and agree with the conclusions presented in the manuscript.

Author Information The microarray and sequencing data have been deposited in the Gene Expression Omnibus of the National Center for Biotechnical Information with the accession numbers GSE38589 and GSE38687. 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.W.M. (tmak@uhnres.utoronto.ca).

METHODS

Generation of LSL-Idh1R132H mice and breeding. The targeting vector containing both the loxP-flanked STOP (LSL) cassette and the IDH1(R132H) mutation is shown in Supplementary Fig. 2. The LSL cassette (Addgene) is composed of the puromycin resistance gene, a splicing acceptor sequence that stops transcription at the insertion point, and four repeats of the SV40 polyA sequence. Thus, the presence of LSL inhibits expression of the IDH1(R132H) protein, and Idh1$^{LSL/WT}$ mice are therefore heterozygous for the Idh1 wild-type allele. Cre-mediated excision of LSL allows expression of the IDH1(R132H) protein (Idh1$^{LSL/WT}$ Cre^{+}).

The targeting vector was assembled by cloning the LSL cassette plus PCR-generated genomic fragments of the *Idh1* gene representing short, middle and long arms of homology into the DT-A vector where the middle fragment contains murine *Idh1* exon 4 bearing the R132H mutation, which was created by PCR-based mutagenesis. The linearized targeting vector was electroporated into E14K or Bruce4 embryonic stem cells. Homologous recombination was confirmed by Southern blotting using 5′-flanking, 3′-flanking, and puromycin-specific probes. Chimaeric mice were produced by microinjection of targeted embryonic stem cells into E3.5 blastocysts. Chimaeras derived from E14K or Bruce4 embryonic stem cells were bred to C57BL/6J or B6-Tyrc-2J/J mice, respectively. Germline transmission of the mutant *Idh1* allele was confirmed by genomic Southern blotting. Both E14K-derived and Bruce4-derived strains were crossed to C57Bl/6J, and F2–F5 generations were used for analysis. The phenotypes of the resulting progeny were identical. For transplantation studies only offspring derived from the Bruce4-derived chimaeras was used. Pups were genotyped by PCR analysis using specific primer pairs to detect the wild-type and mutant *Idh1* alleles. Sense and antisense primers for the KI, wild-type and LSL alleles were 5′-ACCAGCACCTCCC AACTTGTAT-3′, 5′-AGGTTAGCTCTTGCCGATCCGT-3′, and 5′-CAGCA GCCTCTGTTCCACATAC-3′, which yielded PCR products of the predicted sizes of 394 base pairs (bp), 307 bp, and 223 bp, respectively (Supplementary Fig. 2c). All animals were treated in accordance with the NIH Guide for Care and Use of Laboratory Animals as approved by the Ontario Cancer Institute Animal Care Committee (Toronto, Ontario, Canada).

LysMCre mice (B6.129P2-Lyz2tm1(cre)Ifo/J; catalogue no. 004781), VavCre mice (B6.Cg-Tg(Vav1-cre)A2Kio/J; catalogue no. 008610) and CD45.1 mice (B6.SJL-Ptprca Pepcb/BoyJ; catalogue no. 002014) were purchased from the Jackson Laboratory. LysMCre mice and VavCre mice were backcrossed into the C57Bl/6J background for at least 10 generations. LysM-KI mice were generated by breeding Idh1$^{LSL/WT}$ mice with LysMCre$^{+/WT}$ mice as well as by crossing LysM-KI mice with wild-type C57BL/6 mice. Vav-KI mice were generated by breeding Idh1$^{LSL/WT}$ mice with VavCre$^{+/WT}$ mice as well as by crossing Vav-KI mice with wild-type C57BL/6 mice.

For all experiments, LysM-KI mice were heterozygous for both the *Idh*^{LSL} allele and the LysMCre allele (Idh1$^{LSL/WT}$LysMCre$^{+/WT}$), and Vav-KI mice were heterozygous for both the *Idh*^{LSL} allele and the VavCre allele (Idh1$^{LSL/WT}$VavCre$^{+/WT}$). Control mice were wild-type animals or animals heterozygous for either the Idh1^{LSL} allele or the LysMCre allele, if LysM-KI mice were used, or the VavCre allele, if Vav-KI mice were used (Idh1$^{WT/WT}$LysMCre$^{WT/WT}$, Idh1$^{WT/WT}$LysMCre$^{+/WT}$, Idh1$^{LSL/WT}$VavCre$^{WT/WT}$, or Idh1$^{WT/WT}$VavCre$^{+/WT}$). All animal procedures were approved by the Animal Care and Use Committee of the University Health Network (Toronto, Ontario, Canada).

Generation of IDH1(R132H)-KI embryonic stem cell clones. KI embryonic stem cells which had been generated while making the KI mouse were expanded and transiently transfected with pCAGGS-nlsCre using Lipofectamine2000 (Invitrogen). After transfection, cells were trypsinized and plated on gelatin-coated cell culture plates. Single colonies were picked and plated in duplicates into 96-well plates to test for puromycin sensitivity as the Cre-mediated recombination of the LSL cassette results in loss of the puromycin-resistance cassette. Puromycin-sensitive clones were expanded, tested by PCR-genotyping, and then used for *in vitro* differentiation experiments.

OP9/OP9-DL1 embryonic stem cell *in vitro* differentiation system. Experiments including co-culture and differentiation of embryonic stem cells in the presence of OP9 or OP9-DL1 cells were carried out as described elsewhere[9].

Histology. For histological analyses, mouse tissues were fixed in 10% buffered formalin. After fixation, femurs were decalcified in formic acid. Fixed tissues were paraffin-embedded, sectioned, and stained with haematoxylin and eosin according to standard laboratory protocols.

Flow cytometric analyses. Flow cytometric analyses were performed according to standard protocols. In brief, mice were euthanized, single-cell suspensions from bone marrow (BM) were generated by flushing out the bone marrow from tibia and femur in Iscove's modified Dulbecco's medium supplemented with 2% heat-inactivated fetal bovine serum (HI-FBS) (IMDM2). Single-cell suspensions of spleens were generated by mashing the spleen in IMDM2. Cell suspensions were passed through 70 μm cell strainers, centrifuged and resuspended in ammonium-based red cell lysis buffer (Sigma). Nucleated cells were washed in IMDM2, centrifuged and resuspended in flow cytometric buffer (PBS without Mg^{2+} or Ca^{2+} ($PBS^{-/-}$), 2% HI-FBS, 5 mM EDTA, pH 8.0). Viable cell numbers were determined by cell counting after trypan blue staining.

For flow cytometric analyses, cells were incubated in Fc block (BD Biosciences) and then immunostained with fluorophore-linked antibodies (BD Biosciences, eBioscience). For immunostaining of lineage-positive cells, a cocktail containing biotin-labelled primary antibodies against CD5, B220, CD11b, 7-4, Gr-1 and Ter-119 (Miltenyi Biotec) was used, followed by staining with streptavidin-linked fluorophore-labelled secondary antibodies.

To detect total reactive oxygen species (ROS), immunostained cells were incubated with the live cell dye CM-H_2DCFDA (100 μM) at 37 °C for 15 min followed by two washes in flow cytometric buffer. flow cytometric data were acquired on a BD CantoII and analysed using FlowJo (TreeStar). For flow cytometric analyses of differentiated cells, 10^4–10^5 events were acquired. For HSC and HPC analyses, up to 2.5×10^6 events were acquired. Statistical analyses of flow cytometric data were performed using the GraphPad Prism software.

LSK cell sorting. Single-cell suspensions from BM were generated as described above from young mice (8 to 12 weeks of age) and cells were resuspended in flow cytometric buffer and counted. Lineage-negative cells were isolated using the Miltenyi lineage cell depletion kit. Cells were counted and immunostained as described above with anti-lineage-biotin/streptavidin-APC-Cy7, anti-cKit-PE-Cy7 and anti-Sca-1-APC antibodies. Cells were then sorted on a BD FACSAria cell sorter, collected in IMDM2, centrifuged and pellets shock-frozen on dry ice and stored at −80 °C until further processing.

Real-time RT–PCR analysis. RNA was purified from sorted LSK cells using TRIzol (Invitrogen), resuspended in water and quantified on a NanoDrop spectrophotometer before transcription into cDNA using the iScript complementary DNA synthesis kit (Bio-Rad). Primers used for RT–PCR are listed in Supplementary Table 2. Real-time RT–PCR analyses were performed using Power SybrGreen on a 7900HT Fast-Real Time PCR system (Applied Biosystems). C_t values were normalized to the housekeeping gene *Hprt* (ΔC_t) and then to the average of the Δ C_t values of all samples ($\Delta\Delta C_t$). Relative mRNA expression levels were calculated ($2^{-\Delta\Delta C_t}$) and normalized to the average of the relative control mRNA expression levels.

Colony-forming cell (CFC) assays. Single-cell suspensions from BM or spleen from control and LysM-KI mice were generated as described above. Cells were then plated at a density of 10^4 ml^{-1} for BM and 10^5 ml^{-1} for spleen in triplicates in complete methylcellulose media containing stem cell factor, Il-3, Il-6 and erythropoietin (Stem Cell Technologies, M3434). Colonies were counted after 7 days at 37 °C and 5% CO_2. For serial plating cells were collected from methylcellulose media, washed once in IMDM2, counted and replated in complete methylcellulose media at a density of 10^4 ml^{-1}.

Bone marrow competitive repopulation experiments. For competitive BM transfer, donor BM cells from control or LysM-KI mice (CD45.2^{+}), and competitor BM cells from WT mice (CD45.1^{+}), were isolated as described above and then washed in $PBS^{-/-}$ twice. Cell numbers were adjusted to 10^6 ml^{-1}. Lethally-irradiated (10.5 Gy) recipient mice (CD45.1^{+}) were injected with 10^5 donor BM cells plus 10^5 competitor BM cells (2×10^5 total). For flow cytometric analyses of peripheral blood, mice were bled from the tail vein and immunostaining was carried out as described above. Immunostained blood samples were fixed in FoxP3 fixation buffer (eBioscience) to lyse red blood cells, washed in flow cytometric buffer, and analysed by flow cytometry.

Generation of bone marrow-derived macrophages (BMDMs). BMDMs were generated from single-cell BM suspensions of young control or LysM-KI mice using a standard *in vitro* M-CSF-based differentiation protocol. In brief, 5×10^6 cells were plated on a 10-cm tissue culture plate in 5 ml of RPMI1640 media supplemented with 10% HI-FBS and 50 ng ml^{-1} M-CSF. On days 2, 3, and 4 an additional 1 ml of the complete media was added to every tissue culture plate. Cells were collected on day 5 of the *in vitro* differentiation procedure and washed in PBS containing Mg^{2+} and Ca^{2+} ($PBS^{+/+}$) before being used in experiments.

Immunoblot analysis. For immunoblot analysis of histone-3 methylation, equal numbers of BMDMs were generated as described above and lysed in 1× cell lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche). SDS (1%) was added to lysates and incubation was continued for another 30 min. Lysates were centrifuged, loading buffer was added to supernatants, and samples were subjected to electrophoresis on 12% Bis-Tris gels using MES running buffer (Invitrogen). Separated proteins were transferred onto nitrocellulose membranes (Invitrogen). Proteins were then detected using standard laboratory procedures. Primary antibodies used in this study recognized total histone-3 (Abcam ab10799), histone-3 K4me3 (Millipore 07-473), histone-3 K9me3 (Abcam ab8898), histone-3 K27me3 (Millipore 07-449), histone-3

K36me3 (Abcam ab9050), or histone-3 K79me2 (Cell Signaling Technology 9757).

Mass spectrometry. $CD11b^+$ cells were isolated from single-cell suspensions of BM or spleen cells using a CD11b-magnetic bead purification kit (Miltenyi Biotec). BMDMs were generated by *in vitro* differentiation as described above. For serum, blood was collected from mice and separated by centrifugation. Metabolites were extracted from an equivalent number of $CD11b^+$ cells or BMDMs using dry-ice-temperature 80% methanol or equal volumes of serum diluted to 80% methanol using dry-ice-temperature 100% methanol. Cells and serum were centrifuged at 14,000 r.p.m. (20,800*g*) for 10 min at 4 °C. Supernatants were dried down under nitrogen gas, resuspended in 50% methanol, and centrifuged at 13,000*g* for 5 min at 4 °C. The supernatant was transferred to HPLC sample vials and each extract was analysed by reverse-phase liquid chromatography (LC) with tributylamine as an ion pairing reagent. The LC was coupled to a triple-quadrupole mass spectrometer running in negative mode (Thermo Quantum Ultra). Specific chromatography conditions and mass spectrometry parameters were as described[25]. Peak heights of specific metabolites were measured by metabolite-specific multiple reaction monitoring (MRM) scans as previously reported[26], with the exception of hydroxyglutarate, which was monitored using an MRM scan consisting of a 147 to 129 *m/z* transition and a collision energy of 13 eV. The mzrock mass spectrometry tool kit (http://code.google.com/p/mzrock/) was used for data analysis and visualization. To minimize the impact of inter-day instrument variation, extracted chromatogram peak heights for a given metabolite were normalized by the maximum sample value for that specific metabolite measured in the same experiment. These relative values were then scaled by 10,000.

DNA methylation analysis. High molecular weight genomic DNA (gDNA) was isolated from sorted LSK cells from two control and three LysM-KI mice (7 to 15 weeks of age) using the Puregene kit (Qiagen). Genomic DNA from two *IDH1*/*IDH2*-mutant AML samples was obtained from previously reported, de-identified patient samples[5]. Two normal $CD34^+$ bone marrow control samples were purchased from AllCells. Institutional review board approval was obtained at Weill Cornell Medical Center and this study was performed in accordance with the Helsinki protocols. DNA was isolated from each primary sample using the Puregene kit (Qiagen). The gDNA was then characterized using a modified reduced representation bisulphite sequencing (RRBS)[27] approach called enhanced reduced representation bisulphite sequencing (ERRBS)[18].

Generation of libraries. In brief 2.5-10 ng gDNA were digested with MspI. Digested DNA was isolated using a standard phenol-chloroform extraction followed by ethanol precipitation and resuspended in 10 mM Tris pH 8.0. End-repair of digested DNA was performed using T4 DNA polymerase, Klenow DNA polymerase, T4 polynucleotide kinase, dNTPs in T4 DNA ligase buffer (New England Biolabs). After incubation the products were purified using QIAquick PCR purification columns (Qiagen). Adenylation was performed using Klenow fragment 3′ to 5′ exo minus (New England Biolabs) and dATP in Klenow buffer. Products were purified using MinElute PCR purification columns (Qiagen). Adenylated DNA fragments were ligated with pre-annealed 5-methylcytosine-containing Illumina adapters using T4 DNA ligase (New England Biolabs). Products were isolated using MinElute columns (Qiagen). Library fragments of 150–250 bp and 250–400 bp were gel-isolated using the QIAquick Gel Extraction kit (Qiagen). Bisulphite treatment was performed using the EZ DNA Methylation Kit (Zymo Research) with the incubation after the addition of CT conversion reagent being conducted in a thermocycler (Eppendorf MasterCycler). Purified products were subjected to PCR amplification using the FastStart High Fidelity PCR System (Roche). PCR products were isolated using AMPure XP beads. All amplified libraries underwent quality control using a Qubit 1.0 fluorometer and a Quant-iT dsDNA HS Assay Kit for quantification (Invitrogen) as well as Bioanalyzer visualization (Agilent 2100 Bioanalyzer). The amplified libraries were sequenced on an Illumina Genome Analyzer II or HiSeq2000 per manufacturer's recommended protocol for 50 bp single-end read runs. Image capture, analysis and base calling were performed using Illumina's CASAVA 1.7.

Primary data analysis. For alignment of bisulphite treated reads and methylation calls, reads were filtered from the adaptor sequences using CUTADAPT software (http://code.google.com/p/cutadapt/). Adaptor sequence contamination usually occurs towards 3′ ends of some reads. The adaptor matching part of the read was removed if it aligned with the adaptor sequence at least 6 bp and had at most a 0.2 mismatch error rate. Reads were aligned to whole genome using the bismark alignment[28] with a maximum of 2 mismatches in a directional manner and only uniquely aligning reads were retained. To call methylation score for a base position, we required that read bases aligning to that position have at least 20 phred quality score and the base position should have at least 10× coverage. Only CpG dinucleotides that satisfied these coverage and quality criteria were retained for subsequent analysis. Percentage of bisulphite converted Cs (representing unmethylated Cs) and non-converted Cs (representing methylated Cs) were recorded for each C position in a CpG context.

Downstream data analysis. CpG islands, RefSeq genes and repeat sequences for the MM9 genome were downloaded from the UCSC genome browser[29]. CpG shores were defined as 1,000-bp flanking regions on upstream and downstream of a given CpG island. If a 1,000 bp shore overlapped with another island, then the shore was clipped so that its last base falls before the start of the overlapping CpG island. Similarly, if shores were overlapping they were merged into a single shore. In addition, the genome was partitioned into intergenic, intronic, exonic and promoter regions. Promoter regions were defined as the 2-kilobases window centred around the transcription start sites (TSS) of RefSeq genes. We classified CpG dinucleotides as promoter, intronic, exonic or intergenic based on their overlap with these predefined regions. In addition, we classified CpG dinucleotides as CpG island or shore overlapping.

Calculating differential methylation. Methylation values for genomic regions (intergenic, intronic, exonic and promoters, CpG islands and island shores) between different samples were compared by taking the mean methylation percentage of CpG dinucleotides overlapping those regions. To calculate the correlation between different samples and generate the appropriate scatter plots we required that in any given region at least 3 CpG dinucleotides were covered by reads in both control and KI samples. Testing for differential methylation was performed at both the single-base and predefined region levels. For the base-level comparison, coverage on base positions for each sample was required whereas for the region-level comparison we required at least 3 covered bases on all samples. The number of methylated and unmethylated Cs aligning to each base/region were counted and compared across samples. To determine significant differential methylation between two groups of samples we applied logistic regression and the likelihood ratio test was used. Observed *P* values were adjusted with the *q* value method[30].

Pathway analysis of differentially methylated regions. Differentially methylated cytosines were annotated to the nearest RefSeq gene and pathway analysis performed using the Ingenuity Pathway Analysis software (Ingenuity Systems) and the Ingenuity knowledge database as reference.

mRNA expression profiling. RNA was extracted from LSK cell pellets from young mice using TRIzol (Invitrogen) followed by on-column DNase digest and purification using the RNeasy Micro Kit (Qiagen). RNA concentration and quality for all samples was assessed using the RNA 6000 Pico Kit (Agilent). Five nanograms of RNA from each of the samples as well as one Universal Mouse Reference RNA (Stratagene/Agilent) were amplified using the WT-Ovation Pico RNA amplification System Version 1.0 to generate cDNA (Nugen). cDNA samples were biotin labelled according to the Nugen Illumina protocol and hybridized to Mouse WG-6 V2.0 BeadChips which were hybridized, washed and stained according to the manufacturer's protocol (Illumina) and scanned on the iScan (Illumina). The data files were quantified in GenomeStudio Version 2010.2 (Illumina).

Array processing and data analysis. Microarray data were quality-checked for technical outliers before analysis using the LUMI package[31] and quality control metrics in the R/Bioconductor statistical software package (v2.14.1). All samples were of sufficient quality to proceed. Data were imported into the Genespring (v11.5.1, Agilent) software framework for statistical testing and visualization and normalized using a quantile normalization function followed by median centring and logging (log base 2). Of the 45,281 probes, 9,437 did not meet a minimum threshold of expression across both sample groups as defined by being higher than the 20th percentile of measured expression in at least 80 percent of the samples in either the wild-type or KI groups. To assess statistical differences between control and KI sample groups a straightforward *t*-test ($P < 0.05$) followed by a fold-change cut-off of 1.5 was used (Supplementary Table 3). False discovery rate (FDR)-corrected[32] *t*-statistics as well as FDR-corrected moderated *t*-statistics[33] were assessed but did not yield any significant results, probably as a result of the increased variability between samples due to the RNA amplification and the small number of replicates in each group. In total, 240 probes were found to be upregulated in KI samples along with 108 downregulated ones. Corrected hypergeometric test statistic following ref. 34 (FDR $q < 0.03$) was used to find genes in each list that were enriched for gene ontology categories (Supplementary Table 4). To visualize the differences that were found between sample groups, we performed a two-way hierarchical clustering using Pearson-centred distance metrics under average-linkage tree building rules (Supplementary Fig. 8).

25. Munger, J. *et al.* Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nature Biotechnol.* **26,** 1179–1186 (2008).
26. Bajad, S. U. *et al.* Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1125,** 76–88 (2006).

27. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33,** 5868–5877 (2005).
28. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
29. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39,** D876–D882 (2011).
30. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).
31. Du, P. *et al.* lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547–1548 (2008).
32. 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).
33. Smyth G. K.. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** 3 (2004).
34. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29,** 1165–1188 (2001).

LETTER

doi:10.1038/nature11282

Recurrent R-spondin fusions in colon cancer

Somasekar Seshagiri[1]*, Eric W. Stawiski[1]*, Steffen Durinck[1]*, Zora Modrusan[1]*, Elaine E. Storm[1], Caitlin B. Conboy[2], Subhra Chaudhuri[1], Yinghui Guan[1], Vasantharajan Janakiraman[1], Bijay S. Jaiswal[1], Joseph Guillory[1], Connie Ha[1], Gerrit J. P. Dijkgraaf[1], Jeremy Stinson[1], Florian Gnad[3], Melanie A. Huntley[3], Jeremiah D. Degenhardt[3], Peter M. Haverty[3], Richard Bourgon[3], Weiru Wang[4], Hartmut Koeppen[5], Robert Gentleman[3], Timothy K. Starr[6], Zemin Zhang[3], David A. Largaespada[2], Thomas D. Wu[3] & Frederic J. de Sauvage[1]

Identifying and understanding changes in cancer genomes is essential for the development of targeted therapeutics[1]. Here we analyse systematically more than 70 pairs of primary human colon tumours by applying next-generation sequencing to characterize their exomes, transcriptomes and copy-number alterations. We have identified 36,303 protein-altering somatic changes that include several new recurrent mutations in the Wnt pathway gene *TCF7L2*, chromatin-remodelling genes such as *TET2* and *TET3* and receptor tyrosine kinases including *ERBB3*. Our analysis for significantly mutated cancer genes identified 23 candidates, including the cell cycle checkpoint kinase *ATM*. Copy-number and RNA-seq data analysis identified amplifications and corresponding overexpression of *IGF2* in a subset of colon tumours. Furthermore, using RNA-seq data we identified multiple fusion transcripts including recurrent gene fusions involving R-spondin family members *RSPO2* and *RSPO3* that together occur in 10% of colon tumours. The *RSPO* fusions were mutually exclusive with *APC* mutations, indicating that they probably have a role in the activation of Wnt signalling and tumorigenesis. Consistent with this we show that the RSPO fusion proteins were capable of potentiating Wnt signalling. The R-spondin gene fusions and several other gene mutations identified in this study provide new potential opportunities for therapeutic intervention in colon cancer.

Colorectal cancer (CRC) is the fourth most prevalent cancer, accounting for over 50,000 deaths per year in the United States[2]. Approximately 15% of CRCs have microsatellite instability (MSI) arising from defects in the DNA mismatch-repair (MMR) system[3], whereas the other ~85% of microsatellite-stable (MSS) CRCs are the result of chromosomal instability[3]. Genomic studies have identified mutations in genes, chromosomal structural variants and pathway alterations that probably contribute to CRC development[3–6].

In an effort to understand better the pathogenesis of human CRCs we generated exome sequences (72 tumour–normal pairs, 15 MSI and 57 MSS; Supplementary Fig. 1 and Supplementary Table 2), RNA-seq (68 tumour–normal pairs), Illumina 2.5 M single-nucleotide-polymorphism (SNP) array data (74 tumour–normal pairs) and whole-genome sequences (1 MSI and 1 MSS tumour–normal pair) for a set of 74 colon tumour–normal pairs (Supplementary Table 1).

Exome sequencing identified 95,075 somatic mutations, of which 36,303 were protein altering (Fig. 1a). Two MSS samples were hypermutated (24,830 and 5,780 mutations, of which 9,479 and 2,332 were protein-altering mutations, respectively; Fig. 1a) and were not included in background mutation-rate calculations. We found 12,153 somatic mutations in 55 MSS samples studied (3,922 missense, 289 nonsense, 6 stop lost, 69 essential splice site, 20 protein-altering insertions/deletions (indels), 1,584 synonymous, 4,375 intronic and 1,888 others) and 52,312 somatic mutations in the 15 MSI samples (18,436 missense, 929 nonsense, 22 stop lost, 436 essential splice site, 363 protein-altering indels, 8,065 synonymous, 16,675 intronic and 7,386 others) (Fig. 1a and Supplementary Tables 3, 4 and 5). About 98% (35,524 out of 36,303) of the protein-altering single-nucleotide variants identified are new and are not reported in COSMIC v58 (ref. 7). We further confirmed a subset of the mutations reported to be somatic using RNA-seq data and/or mass spectrometry genotyping (Supplementary Methods, Supplementary Fig. 2 and Supplementary Table 3). In addition, we confirmed the presence of aberrant/alternate transcripts for 112 genes with canonical splice-site mutations using RNA-seq data (Supplementary Table 4). We observed a mean non-synonymous mutation frequency of 2.8 mutations per 10^6 bases (31–149 coding-region mutations in 55 samples) for the MSS samples and 47 mutations per 10^6 bases (764–3,113 coding-region mutations in 15 samples) for the MSI samples, consistent with the MMR defect in the latter (Fig. 1a) and the presence of alterations in one or more of the MMR genes in ~80% of the samples (Supplementary Fig. 3 and Supplementary Table 6). Consistent with previous mutation reports[4,6,8], analysis of base-level transitions and transversions at mutated sites using both exome and whole-genome sequencing (WGS) showed that C to T transitions were predominant in CRCs, regardless of MMR status (Fig. 1b, c), except for the two hypermutated tumour samples, which also showed a higher proportion of C to A and T to G transversions (Fig. 1b, c).

In an effort to understand the impact of the mutations on gene function we applied SIFT[9], PolyPhen[10] and mCluster[11] and found, based on at least two of the three methods, that 34% of the somatic mutations probably have a functional consequence (Supplementary Table 3). We applied a previously described *Q* score metric to rank significantly mutated cancer genes[12]. In MSS samples, we identified 23 significantly mutated cancer genes (*Q* score, ≥ 1; false discovery rate, $\leq 5\%$; Supplementary Table 7). The MSS CRC genes identified from this analysis included previously reported genes such as *KRAS*, *APC*, *TP53*, *SMAD4*, *FBXW7* and *PIK3CA* and several new genes including the cell cycle checkpoint gene *ATM* (Fig. 2a and Supplementary Table 7). *KRAS* and *TP53* were among the top mutated MSI CRC genes; however, owing to the limited number of samples analysed, none of the genes achieved statistical significance.

Of the 432 candidate CRC genes identified in colon cancer mouse model screens[13,14], 356 were mutated in this study (Supplementary Table 8). Frequently mutated genes that overlapped included *KRAS*, *APC*, *SMAD4*, *FBXW7* and *EP400*, as well as genes involved in chromatin remodelling such as *SIN3A*, *SMARCA5*, *NCOR1* and histone-modifying enzyme *JARID2* (Supplementary Table 8). *TCF12*, a CRC gene identified in a mouse model screen, was mutated in five of the MSI samples. A hotspot mutation in *TCF12* at Arg 603 (three out of five mutations) was found within the *TCF12* helix–loop–helix domain and probably abolishes its DNA-binding ability (Fig. 2b).

[1]Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [2]Department of Genetics, Cell Biology and Development, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. [3]Department of Bioinformatics and Computational Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [4]Department of Structural Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [5]Department of Pathology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [6]Department of Obstetrics, Gynecology & Women's Health, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA.
*These authors contributed equally to this work.

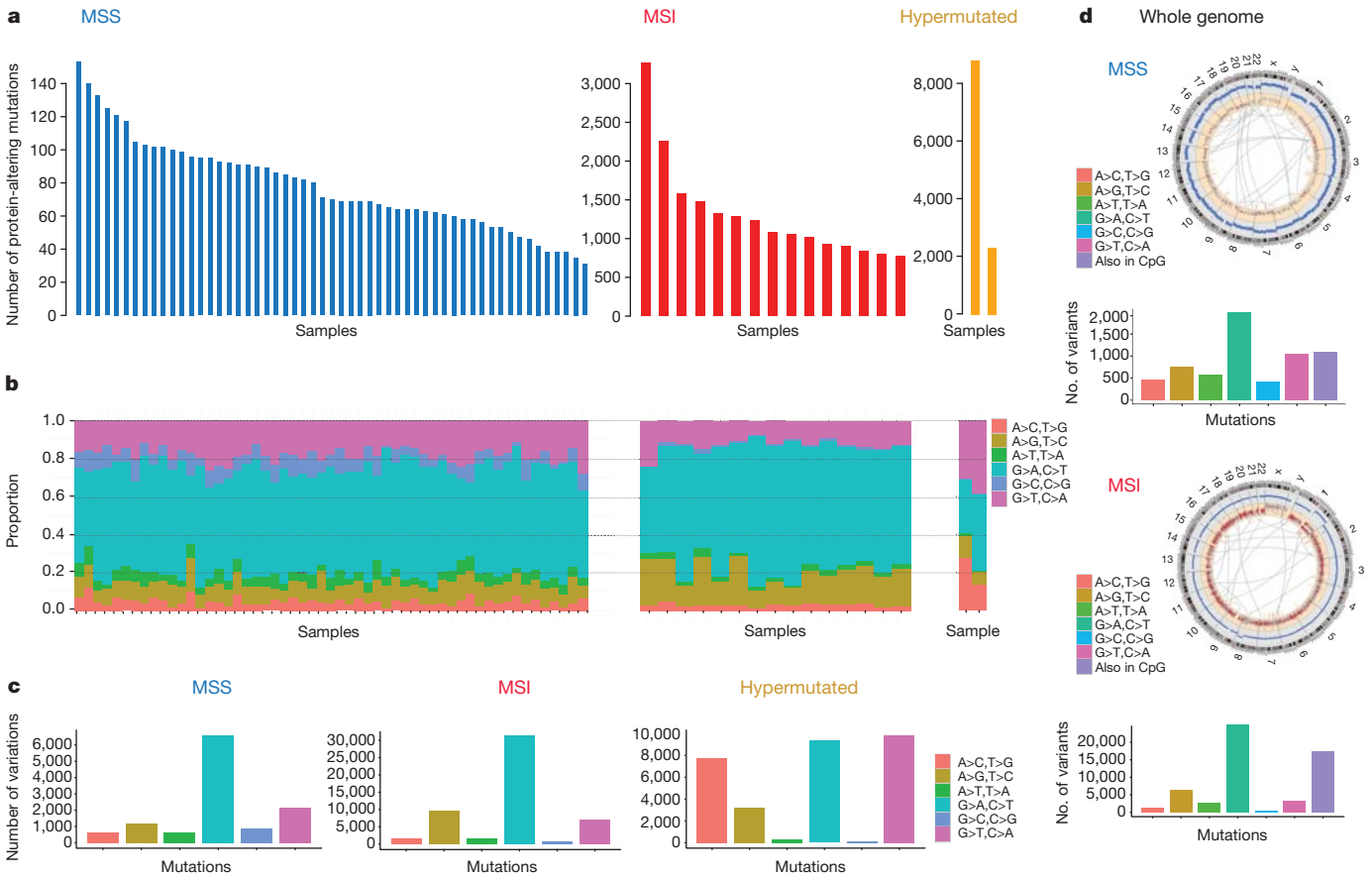


Figure 1 | Somatic mutations in colon cancer. **a**, Number of mutations in each of the MSS, MSI and hypermutated subtypes. **b**, Transitions and transversions at the individual sample level for the three groups in the same order as shown in **a**. **c**, Transitions and transversions averaged across each of the three sample groups. **d**, Circos plot representation of the whole genome of an MSS and an MSI colon tumour. The outer blue band shows copy number as measured using WGS reads. Red dots indicate somatic non-synonymous, splice-site and stop-gained mutations (MSS genome, 17,651 mutations, average of 6.2 mutations per 10^6 bases, 3,257 mutated genes; MSI genome, 97,968 mutations, average of 34.5 mutations per 10^6 bases, 9,851 mutated genes). Grey dots indicate all other somatic mutation types. Orange and grey lines show predicted intra- and interchromosomal rearrangements, respectively.

Mutational hotspots are indicative of functionally relevant driver genes. In this study we have identified 274 genes with hotspot mutations (Supplementary Table 9 and Supplementary Figs 4 and 5), 40 of which were not previously reported in COSMIC[7]. Comparison with mutations reported in COSMIC[7] identified an additional 435 hotspot mutations in 361 genes (Supplementary Table 10), 18 of which were also identified in mouse CRC screens[13,14] (Supplementary Table 8). Mutations in the *TET* family of genes that encode methylcytosine dioxygenases have not been reported in CRC[4,15]. We found that all three family members—*TET1*, *TET2* and *TET3*—were mutated in our study (Fig. 2c). In addition, *ERBB3*, which was mutated in 8% (6 out of 72) of the samples, contained multiple hotspot mutations that were oncogenic, as reported in a separate study (B.S.J. *et al.*, submitted).

The top differentially expressed genes identified using RNA-seq (Supporting Methods, Supplementary Figs 6 and 7 and Supplementary Table 11) included *FOXQ1* and *CLDN1*, both of which have been implicated in colon tumorigenesis[16]. Notably, we also identified *IGF2* upregulation in 12% (8 out of 68) of the samples (Supplementary Fig. 7). Most (7 out of 8) of the tumours with *IGF2* overexpression also showed focal amplification of the *IGF2* locus as measured by Illumina 2.5 M arrays (Supplementary Fig. 8). Overall, the differentially expressed genes affect multiple signalling pathways, including Wnt and glutamate receptor signalling (Supplementary Fig. 26 and Supplementary Table 12).

We assessed copy-number alterations in 74 tumour–normal pairs by applying GISTIC[17] to the circular binary segmented (CBS)[18] copy-number data. In addition to the *IGF2* amplifications, we found known amplifications involving *KRAS* (13%; 10 out of 74) and *MYC* (23%; 17 out of 74) located in a broad amplicon on chromosome 8q (Supplementary Fig. 9 and Supplementary Table 13). Deletion involving *FHIT*, a tumour-suppressor gene[19], was observed in 30% (22 out of 74) of the samples (Supplementary Table 14 and Supplementary Fig. 9). We also detected deletion of *APC* (8%; 6 out of 74), *PTEN* (4%; 3 out of 74) and *SMAD3* (9%; 10 out of 74). *SMAD4* and *SMAD2* are both altered in 27% (20 out of 74) of the samples and are located within 3 megabases (Mb) of each other on chromosome 18q, which is frequently lost (Supplementary Fig. 9). Finally, we found chromosome 20q to be frequently gained (Supplementary Fig. 9).

Using RNA-seq data we identified 36 rearrangements that result in gene fusions[20], including two recurrent ones (Supplementary Fig. 10 and Supplementary Table 15). The recurrent fusions found in MSS samples involve the R-spondin family members, *RSPO2* (3%; 2 out of 68) and *RSPO3* (8%; 5 out of 68). R-spondins are secreted proteins that can potentiate canonical Wnt signalling[21]. The recurrent *RSPO2* fusion identified in two tumour samples involves *EIF3E* exon 1 and *RSPO2* exon 2 (Fig. 3b). This fusion transcript is expected to produce a functional RSPO2 protein driven by the *EIF3E* promoter (Fig. 3d). A second *RSPO2* fusion that was not expected to produce a functional protein, *EIF3E(e1)–RSPO2(e3)*, was also detected in the same samples (Supplementary Table 15). To confirm the nature of the alteration at the genome level, we performed WGS of the tumours containing *RSPO2* fusions. Analysis of junction-spanning reads, mate-pair reads and copy-number data derived from the WGS data identified a 158 kilobases (kb) deletion in one sample and a 113 kb deletion in

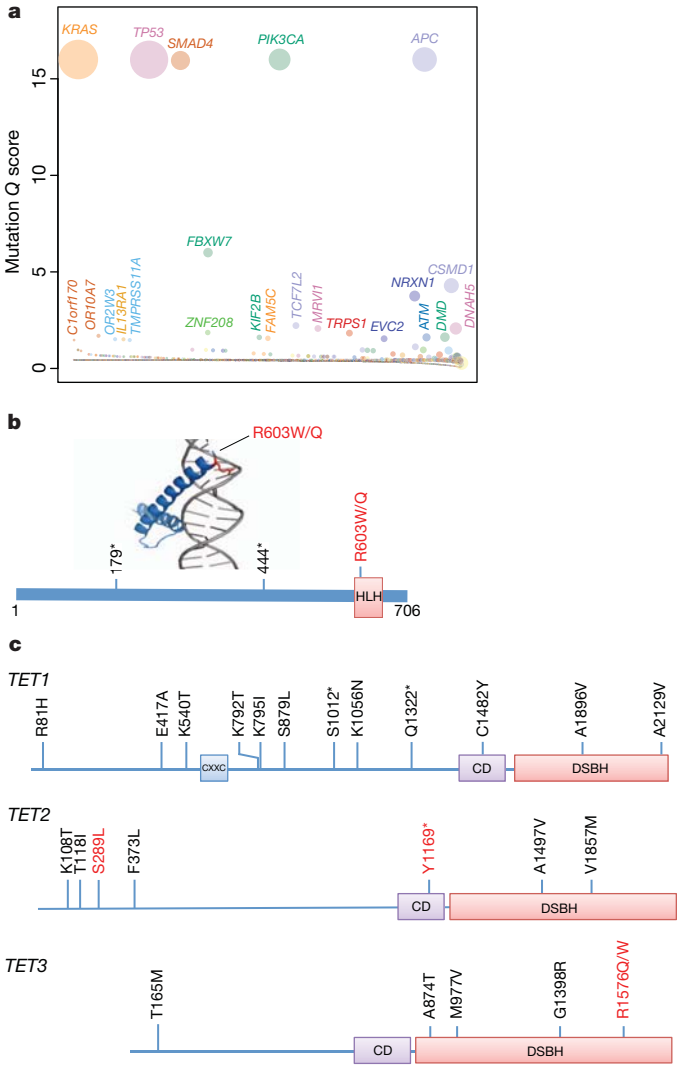


Figure 2 | Significantly mutated colon cancer genes. **a**, Genes evaluated for significance using *Q* score criteria for MSS samples are shown. Each circle represents a gene and the size of the circle is proportional to the mutation count for that gene. The genes are represented in order of increasing number of expected mutations from left to right on the *x* axis. Genes with a statistically significant *Q* score are labelled. **b**, *TCF12* helix–loop–helix (HLH) domain structure bound to DNA showing the Arg 603 hotspot and a schematic diagram depicting the various *TCF12* mutations identified in this study. **c**, Somatic mutations in the *TET* gene family including *TET1*, *TET2* and *TET3* shown on the domain architecture of the *TET* genes. Recurrent mutations found in this study for *TET3* and for *TET2* by comparison to COSMIC data are shown in red. CD, Cysteine-rich domain; CXXC, Cys-X-X-Cys-type domain; DSBH, double-stranded β-helix 2OG-Fe(II)-dependent dioxygenase domain. Asterisks in **b** and **c** denote nonsense mutations that result in premature stop codons.

the second sample, both of which places exon 1 of *EIF3E* in close proximity to the 5′ end of *RSPO2* (Supplementary Figs 11 and 12), further supporting the fusions observed.

The *RSPO3* gene fusions were observed in 5 out of 68 tumours and involve *PTPRK* as the 5′ partner. We identified two different *RSPO3* fusion variants consisting of either exon 1 or exon 7 of *PTPRK* and exon 2 of *RSPO3* (Fig. 3f–h and Supplementary Fig. 13). WGS reads from the five tumours expressing the *RSPO3* fusions showed rearrangements involving a simple (three samples) or a complex (two samples) inversion that places *RSPO3* in proximity to *PTPRK* on the same strand as *PTPRK* on chromosome 6q (Supplementary Figs 14–18). The *PTPRK(e1)–RSPO3(e2)* transcript found in four samples is an in-frame fusion that preserves the entire coding sequence of *RSPO3* and replaces its secretion signal sequence with that of *PTPRK*

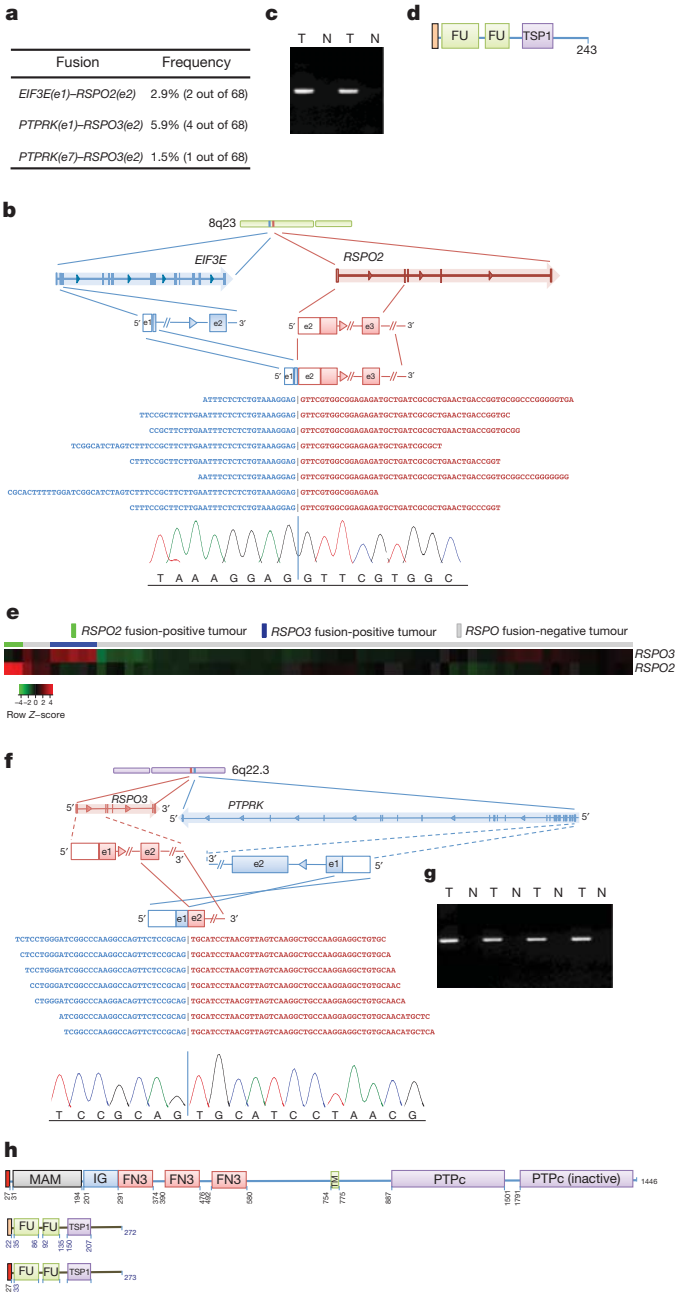


Figure 3 | Recurrent R-spondin gene fusions. **a**, List of type and frequency of R-spondin gene fusions in colon cancer. **b**, Cartoon depicting the location, orientation and exon–intron architecture of *EIF3E–RSPO2* fusion on the genome. The read evidence for the *EIF3E(e1)–RSPO2(e2)* fusion identified using RNA-seq data is shown. **c**, Independent reverse transcriptase (RT)–PCR-derived products confirming the *EIF3E(e1)–RSPO2(e2)* somatic fusion resolved on an agarose gel. The RT–PCR product was Sanger-sequenced confirming the fusion junction and a representative chromatogram is shown in **b**. **d**, Schematic of the resulting EIF3E(e1)–RSPO2(e2) fusion protein. **e**, Tumours harbouring R-spondin fusions show elevated expression of the corresponding *RSPO* gene as depicted by the heat-map. **f**, Cartoon depicting the location, orientation and exon–intron architecture of the *PTPRK–RSPO3* gene fusion on the genome. The read evidence for *PTPRK(e1)–RSPO3(e2)* fusion identified using RNA-seq data is shown. **g**, Independent RT–PCR-derived products confirming the *PTPRK(e1)–RSPO3(e2)* somatic fusion resolved on an agarose gel. The RT–PCR product was Sanger-sequenced confirming the fusion junction, and a representative chromatogram is show in panel **f**. **h**, Schematic of PTPRK, RSPO3 and the resulting PTPRK(e1)–RSPO3(e2) fusion proteins. FN3, fibronectin type 3 domain; FU, furin-like repeats; IG, immunoglobulin; MAM, meprin A5 receptor protein tyrosine phosphatase mu domain; N, normal; PTPc, protein tyrosine phosphatase catalytic domain; T, tumour; TM, transmembrane domain; TSP1, thrombospondin type 1 repeats.

(Fig. 3h). The *PTPRK(e7)–RSPO3(e2)* fusion, detected in one sample, is also an in-frame fusion that encodes a ~70 kDa protein consisting of the first 387 amino acids of PTPRK, including its secretion signal sequence, and the RSPO3 amino acids 34–272 lacking its native signal peptide (Supplementary Fig. 13c). Interestingly, RNA-seq data showed that the expression of *RSPO2* and *RSPO3* in colon tumour samples containing the fusions was elevated compared with the tumour samples lacking R-spondin fusions (Fig. 3e). Furthermore, all of the *RSPO*-positive fusion tumours expressed the potential R-spondin receptors *LGR4*, *LGR5* and *LGR6* (refs 22–24), although *LGR6* expression was lower compared with *LGR4* and *LGR5* (Supplementary Fig. 19).

To confirm the functional relevance of the R-spondin fusion proteins, we first expressed R-spondin fusion constructs in 293T cells and confirmed that the fusion proteins were expressed and secreted as predicted (Fig. 4a). We then examined the biological activity of the R-spondin fusions using a Wnt-responsive luciferase reporter, TOPbrite[22–25]. As observed with wild-type *RSPO2* and *RSPO3*, conditioned media from cells transfected with *RSPO* fusion expression constructs led to activation of the Wnt pathway (Fig. 4b). The observed activation, although apparent in the absence of exogenous Wnt, was further potentiated in the presence of recombinant Wnt, consistent with the known role of R-spondins in Wnt signalling[21–24]. In addition to 293T cells, we confirmed the ability of RSPO fusion proteins to activate Wnt signalling in HT-29, a human colon cancer cell line (Supplementary Fig. 20).

To characterize further the *RSPO* gene fusions we analysed them in the context of other mutations (Fig. 4c, d, Supplementary Figs 21 and 25 and Supplementary Table 16). We found that most of the colon tumours carried alterations in Wnt, TGF-β, Rtk/Ras and TP53 pathways (Supplementary Figs 22–26). Interestingly, the *RSPO2* and *RSPO3* fusions were mutually exclusive and occurred in tumours that did not contain *APC* mutations or copy loss (Fig. 4d; $P = 0.038$, Fisher's exact test), with the exception of one tumour that contained an *RSPO2* fusion and a single *APC* allele (Fig. 4d). In addition, the *RSPO* gene fusions were mutually exclusive with mutations in *CTNNB1*, another Wnt pathway gene that is mutated in CRC. However, all the samples with *RSPO* gene fusions also carried a mutation in either *KRAS* or *BRAF* (Supplementary Fig. 21). Most *APC* mutant samples also have Ras pathway gene mutations, indicating that the *RSPO* gene fusions may function analogously to *APC* mutations in promoting Wnt signalling during colon tumour development. This is further supported by the fact that the *RSPO* fusion-positive tumours showed upregulation of Wnt pathway target genes similar to tumours carrying *APC* mutations (Supplementary Fig. 27). Interestingly, transposon-based insertional mutagenesis screens in mouse models of CRC identified *Rspo2* as a top-candidate cancer gene[14] (Supplementary Table 8). Furthermore, analysis of these mouse colon tumours showed that the transposon insertions were in most cases mutually exclusive between *Apc* and *Rspo2* (Supplementary Fig. 28a), confirming the mutual exclusivity observed in human samples. Most of the transposon insertions in the mouse *Rspo2* gene occur near the transcriptional start site in the same orientation as the *Rspo2* promoter, and probably lead to elevated *Rspo2* expression (Supplementary Fig. 28b). Consistent with this and with the elevated expression of *RSPO* genes observed in human tumours, we found a ~20-fold increase in *Rspo2* messenger RNA expression in a mouse tumour carrying a transposon insertion near the *Rspo2* transcription start site, compared with adjacent normal tissue (Supplementary Fig. 28c). Taken together, our findings indicate that the R-spondins probably function as drivers in human CRCs. Although further studies will be required to fully

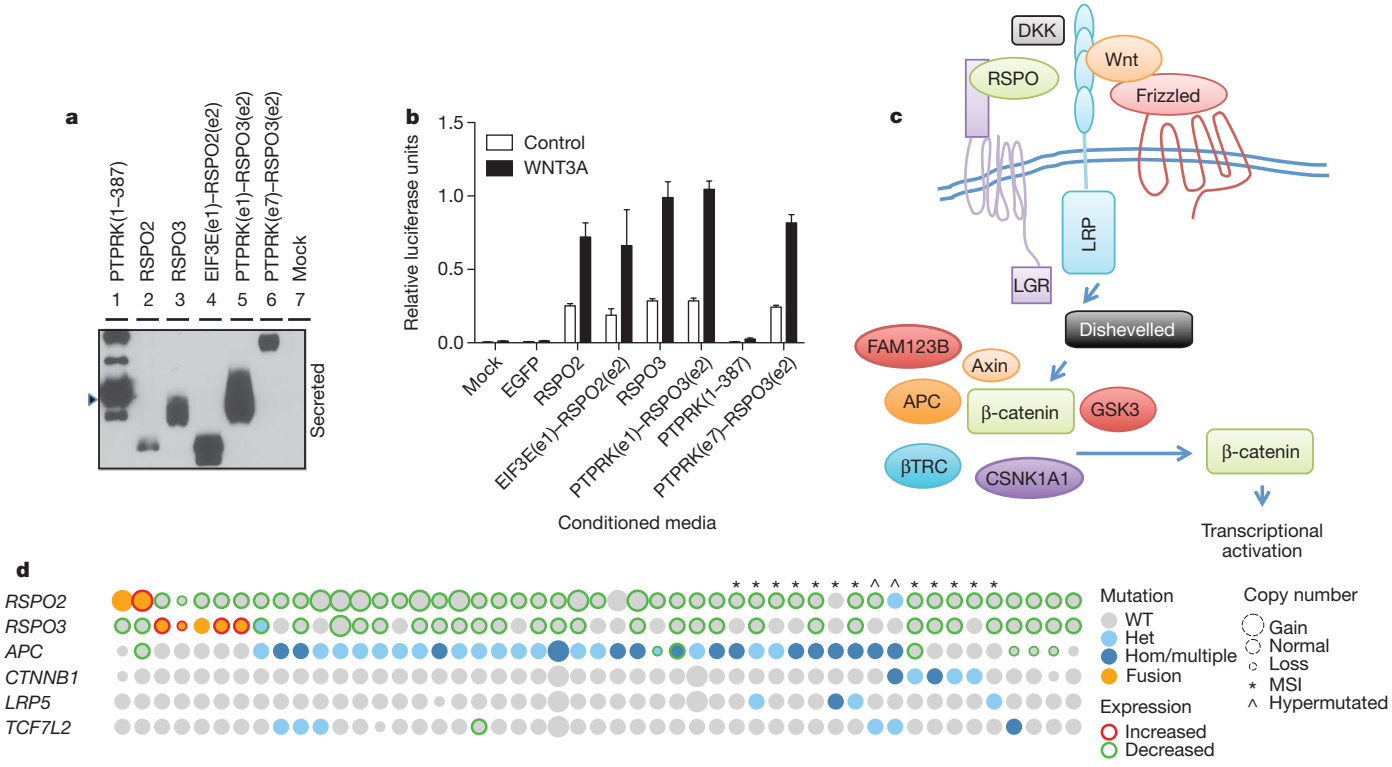


Figure 4 | RSPO fusion products activate Wnt signalling. **a**, Secreted RSPO fusion proteins detected by western blot in media from 293T cells transfected with expression constructs encoding the fusion proteins. The arrowhead indicates the expected product for RSPO 1–387. **b**, RSPO fusion proteins activate and potentiate Wnt signalling as measured using a luciferase reporter assay (see Supplementary Methods). Data shown are from conditioned media derived from cells transfected with the fusion constructs. Representative data from at least three experiments are shown. Error bars represent mean ± s.d. from three replicate experiments. EGFP, enhanced green fluorescent protein. **c**, Cartoon representing R-spondin-mediated Wnt signalling pathway activation. **d**, Quilt plot depicting *RSPO* fusions, somatic mutations, copy-number alteration and expression changes across a select set of Wnt-signalling pathway genes. Each column represents a tumour sample. Het, heterozygous; Hom, homozygous; WT, wild type.

understand the role of R-spondin fusions in CRC development, they represent attractive targets for antibody-based therapy in CRC patients positive for R-spondin fusions. In addition, other therapeutic strategies that target downstream components of the Wnt signalling cascade will probably be effective against tumours positive for R-spondin fusions.

METHODS SUMMARY

Fresh–frozen primary human colon tumours and their matched normal tissues were obtained from commercial sources with appropriate institutional approval. DNA and RNA isolated from tumours with >70% tumour content and their corresponding matched normal tissue were subject to exome sequencing, RNA-seq and SNP array analysis. A complete description of the materials and methods is provided in the Supplementary Information.

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

Received 23 December 2011; accepted 6 June 2012.
Published online 15 August 2012.

1. Chin, L., Andersen, J. N. & Futreal, P. A. Cancer genomics: from discovery science to personalized medicine. *Nature Med.* **17,** 297–303 (2011).
2. Siegel, R., Ward, E., Brawley, O. & Jemal, A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J. Clin.* **61,** 212–236 (2011).
3. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6,** 479–507 (2011).
4. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).
5. Timmermann, B. *et al.* Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS ONE* **5,** e15661 (2010).
6. Bass, A. J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent *VTI1A–TCF7L2* fusion. *Nature Genet.* **43,** 964–968 (2011).
7. Forbes, S. A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38,** D652–D657 (2010).
8. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314,** 268–274 (2006).
9. Ng, P. C. & Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **12,** 436–446 (2002).
10. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30,** 3894–3900 (2002).
11. Yue, P. *et al.* Inferring the functional effects of mutation through clusters of mutations in homologous proteins. *Hum. Mutat.* **31,** 264–271 (2010).
12. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466,** 869–873 (2010).
13. March, H. N. *et al.* Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nature Genet.* **43,** 1202–1209 (2011).
14. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323,** 1747–1750 (2009).
15. Mohr, F., Dohner, K., Buske, C. & Rawat, V. P. TET genes: new players in DNA demethylation and important determinants for stemness. *Exp. Hematol.* **39,** 272–281 (2011).
16. Kaneda, H. *et al.* FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth. *Cancer Res.* **70,** 2053–2063 (2010).
17. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).
18. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).
19. Pichiorri, F. *et al.* Fhit tumor suppressor: guardian of the preneoplastic genome. *Future Oncol.* **4,** 815–824 (2008).
20. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature Rev. Genet.* **12,** 87–98 (2011).
21. Yoon, J. K. & Lee, J. S. Cellular signaling and biological functions of R-spondins. *Cell. Signal.* **24,** 369–377 (2012).
22. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proc. Natl Acad. Sci. USA* **108,** 11452–11457 (2011).
23. de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476,** 293–297 (2011).
24. Glinka, A. *et al.* LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling. *EMBO Rep.* **12,** 1055–1061 (2011).
25. Zhang, Y. *et al.* Inhibition of Wnt signaling by Dishevelled PDZ peptides. *Nature Chem. Biol.* **5,** 217–219 (2009).

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

Acknowledgements The authors would like to acknowledge Genentech DNA Sequencing, Oligo, Microarray and gCell laboratories for their help with the project. We thank the Genentech Bioinformatics group for informatics infrastructure support and the Pathology Core Labs for providing histology, immunohistochemistry and tissue-management support. We also thank M. Costa, M. Callow, P. Polakis and the de Sauvage and Seshagiri laboratories for comments and suggestions. Work in the Largaespada laboratory was supported by National Institutes of Health grant R01-CA134759 (to D.A.L.).

Author Contributions S.S. and F.J.d.S. conceived the study and designed experiments. E.W.S. and S.D. performed the exome, WGS, RNA-seq and copy data analysis. Z.M., C.H. and Y.G. performed the validation of the fusions. Z.M. managed exome capture. S.C., V.J., B.S.J. and E.E.S. validated and tested the fusions for activity. C.B.C., T.K.S. and D.A.L. performed the mouse model analysis. G.J.P.D., J.S. and Z.M. performed tissue management and nucleic acid preparation. J.S. and J.G. collected and performed mutation validation. F.G. and Z.Z. predicted the functional effect of mutations. E.W.S., M.A.H., J.D.D., T.D.W. and R.G. provided bioinformatics support, including an algorithm for variant calling, fusion detection and copy number calling. R.B. and P.M.H. analysed SNP array data. W.W. performed structural biology support and H.K. performed pathology support. S.S., E.W.S., S.D., Z.M. and F.J.d.S. wrote the paper, which was reviewed and edited by the other co-authors.

Author Information Sequencing and genotype data has been deposited at the European Genome-Phenome Archive (http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under accession number EGAS00001000288. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. 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.S. (sekar@gene.com) or F.J.d.S. (sauvage@gene.com).

METHODS

Samples, DNA and RNA preparation and MSI testing. Patient-matched fresh–frozen primary colon tumours and normal tissue samples with appropriate Institutional Review Board approval and patient-informed consent were obtained from commercial sources (Supplementary Table 1) and subjected to genomic analysis described below. The human tissue samples used in the study were de-identified (double-coded) before their use and hence the study using these samples is not considered human subject research under the US Department of Human and Health Services regulations and related guidance (45 CFR Part 46). All tumour and normal tissues were subject to pathology review. From a set of 90 paired tumour–normal samples we identified 74 sample pairs for further analysis. Tumour DNA and RNA were extracted using the Qiagen AllPrep DNA/RNA kit (Qiagen). Tumour samples were assessed for microsatellite instability using an MSI detection kit (Promega).

Exome capture and sequencing. In total, 72 tumour samples and matched normal tissues were analysed by exome sequencing (Supplementary Table 1). Exome capture was performed using SeqCap EZ human exome library v2.0 (NimbleGen) consisting of 2.1 million empirically optimized long oligonucleotides that target ~30,000 coding genes (~300,000 exons, total size 36.5 Mb). The library is capable of capturing a total of 44.1 Mb of the genome, including genes and exons represented in RefSeq (Jan 2010), Consensus CDS Project (CCDS; Sept 2009) and miRBase (v.14, Sept 2009). Exome-capture libraries generated were sequenced on HiSeq 2000 (Illumina). One lane of 2×75 base pair (bp) paired-end data were collected for each sample. The targeted regions had a mean coverage of 179× with 97.4% bases covered at ≥ 10 times (Supplementary Fig. 1 and Supplementary Table 2).

RNA-seq. RNA from 68 colon tumour and matched normal sample pairs was used to generate RNA-seq libraries using TruSeq RNA Sample Preparation kit (Illumina). RNA-seq libraries were multiplexed (two per lane) and sequenced on HiSeq 2000 as per manufacturer's recommendation. We generated ~30 million paired-end (2×75 bp) sequencing reads per sample (Supplementary Fig. 5).

Sequence data processing. All short-read data were evaluated for quality control using the Bioconductor ShortRead package[26]. To confirm that all samples were identified correctly, all exome and RNA-seq data variants that overlapped with the Illumina 2.5 M array data were compared and checked for consistency. An all-by-all sample comparison was done on germline variants to check that all pairs were matched correctly between the respective patient tumour and matched normal sample and correspondingly did not match with any other patient pair above a cutoff of 90%.

Variant calling. Sequencing reads were mapped to the University of California, Santa Cruz (UCSC) human genome (GRCh37/hg19) using Burrows–Wheeler alignment (BWA) software[27] set to default parameters. Local realignment, duplicate marking and raw variant calling were performed as described previously[28]. Known germline variations represented in dbSNP Build 131[29], but not represented in COSMIC v54[8], were filtered out. In addition, variants that were present in both the tumour and normal samples were removed as germline variations. Remaining variations present in the tumour sample but absent in the matched normal were predicted to be somatic. Predicted somatic variations were additionally filtered to include only positions with a minimum of 10× coverage in both the tumour and matched normal as well as an observed variant allele frequency of <3% in the matched normal and a significant difference in variant allele counts using Fisher's exact test. To evaluate the performance of this algorithm we randomly selected 807 protein-altering variants and validated them using Sequenom nucleic acid technology as described previously[13]. Of these, 93% (753) validated as cancer specific, with the invalidated variants being equally split between not being seen in the tumour and also being seen in the adjacent normal (germ line). In addition, we also targeted and validated 164 mutations belonging to significant Q score genes and 212 hotspot mutations using Sequenom mass spectrometry technology. All variants that invalidated were removed from the final set. Indels were called using the GATK Indel Genotyper Version 2, which reads both the tumour and normal BAM file for a given pair[28].

To identify additional variants that had a lower mean allele frequency or were affected by a specific mapper, we included Sequenom validated variants using the following algorithm. Reads were mapped to the UCSC human genome (GRCh37/hg19) using GSNAP[30]. Variants seen at least twice at a given position and greater than 10% allele frequency were selected. These variants were additionally filtered for significant biases in strand and position using Fisher's exact test. In addition, variants that did not have adequate coverage in the adjacent normal as determined as at least a 1% chance of being missed using a binomial distribution at a normal allele frequency of 12.5% were excluded. All new protein-altering variants identified by the second algorithm were validated by Sequenom, which resulted in a total of 515 additional variants. The effect of all non-synonymous somatic mutations on gene function was predicted using SIFT[10] and PolyPhen[11]. All variants were annotated using Ensembl (release 59, http://www.ensembl.org).

Validation of somatic mutations and indels. We used single base pair extension followed by nucleic acid mass spectrometry (Sequenom) to validate the predicted somatic mutations as described previously[13]. Tumour and matched normal DNA were whole-genome amplified using the REPLI-g Whole Genome Amplification Midi kit (Qiagen) and cleaned up as per manufacturer's recommendations. Variants found as expected in the tumour but absent in the normal DNA were designated somatic. Those that were present in both tumour and normal samples were classified as germline. Variants that could not be validated in tumour or normal samples were designated as failed. For indel validation we designed primers for PCR that generated an amplicon of ~300 bp that contained the indel region. We PCR-amplified the region in both tumour and matched normal samples using Phusion (New England Biolabs) as per manufacturer's instructions. We then purified the PCR products on a gel, isolated the relevant fragments and Sanger-sequenced them. The sequencing trace files were analysed using Mutation Surveyor (SoftGenetics). Indels that were present in the tumour and absent in the normal sample were designated somatic and are reported in Supplementary Table 4.

Mutational significance and pathway analysis. We evaluated the mutational significance of genes using a previously described method[13]. In brief, this method can identify genes that have more statistically significant protein-altering mutations than what would be expected based on a calculated background-mutation rate. The background-mutation rate is calculated for six different nucleotide mutation categories (A, C, G, T, CG1, CG2), in which there was sufficient coverage (≥10×) in both the tumour and matched normal sample. A non-synonymous to synonymous ratio, r_i, was calculated using a simulation of mutating all protein-coding nucleotides and seeing if the resulting change would result in a synonymous or non-synonymous change. The background mutation rate, f_i, was determined by multiplying the number of synonymous somatic variants by r_i and normalizing by the total number of protein-coding nucleotides. The number of expected mutations for a given gene was determined as the number of protein-coding bases multiplied by f_i and integrated across all mutation categories. A P value was calculated using a Poisson probability function given the expected and observed number of mutations for each gene. P values were corrected for multiple testing using the Benjamini–Hochberg method and the resulting Q values were converted to Q scores by taking the negative $\log_{10}$ of the Q values with a maximum possible value of 16. Given that different mutation rates existed for the MSI and MSS samples, Q scores were calculated separately for each, with the two hypermutated samples being removed completely. To avoid underestimating the background mutation rates, samples with less than 50% tumour content were excluded from the analysis. Pathway mutational significance was also calculated as previously described[13], with the exception that the BioCarta Pathway database was used, which was downloaded as part of the MSigDB[31]. For quilt plots a fourfold increase or decrease compared with the matched normal was used as the cutoff (Fig. 4d and Supplementary Figs 20–24). A $\log_2$ tumour to normal ratio of 0.585 and −0.7 was used for copy-number gain and loss, respectively, for these plots. Samples where exome data were not collected were excluded from the plot (Fig. 4d and Supplementary Figs 20–24), Only samples with either a copy number change and/or a mutation are depicted. To help differentiate single-copy from double-copy loss in MMR genes a cutoff of −1 was used for copy-loss events (Supplementary Fig. 3)

WGS and analysis. Paired-end sequence reads were aligned to GRCh37 using BWA. Further processing of the alignments to obtain mutation calls was done in a similar manner to the exome-sequencing analysis using the GATK pipeline. Copy number was calculated by computing the number of reads in 10-kb non-overlapping bins and taking the tumour–normal ratio of these counts. Chromosomal breakpoints were predicted using BreakDancer[32]. Genome plots were created using Circos[33]. For detection of genomic alteration in tumours with *RSPO* fusion, reads aligning to the genomic region containing the *RSPO*, *EIF3E* and *PTPRK* genes were analysed further to reveal the deletions and insertions.

RNA-seq data analysis. RNA-seq reads were aligned to the human genome version GRCh37 using GSNAP[30]. Expression counts per gene were obtained by counting the number of reads aligning concordantly within a pair and uniquely to each gene locus as defined by CCDS. The gene counts were then normalized for library size and subsequently variance stabilized using the DESeq Bioconductor software package[34]. Differential gene expression was computed by pairwise t-tests on the variance-stabilized counts followed by correction for multiple testing using the Benjamini–Hochberg method.

SNP array data generation and analysis. Illumina HumanOmni2.5_4v1 arrays were used to assay 74 colon tumours and matched normal for genotype, DNA copy and loss of heterozygosity at ~2.5 million SNP positions. These samples all passed our quality-control metrics for sample identity and data quality (see below). A subset of 2,295,239 high-quality SNPs was selected for all analyses.

After making modifications to permit use with Illumina array data, we applied the PICNIC[35] algorithm to estimate total copy-number and allele-specific copy number/loss of heterozygosity. Modifications included replacement of the segment-initialization component with the CBS algorithm[36], and adjustment of the prior distribution for background raw copy-number signal (adjusted mean of 0.7393 and s.d. of 0.05). For the pre-processing required by PICNIC's hidden Markov model, we used a Bayesian model to estimate cluster centroids for each SNP. For SNP k and genotype g, observed data in normal samples were modelled as following a bivariate Gaussian distribution. Cluster centres for the three diploid genotypes were modelled jointly by a six-dimensional Gaussian distribution, with mean treated as a hyperparameter and set empirically based on a training set of 156 normal samples. Cluster centre and within-genotype covariance matrices were modelled as inverse Wishart, with scale matrix hyperparameters also set empirically, and with degrees of freedom manually tuned to provide satisfactory results for a wide range of probe behaviour and minor allele frequencies. Finally, signal for SNP k (for the A and B alleles separately) was transformed with a nonlinear function $y = \alpha_k x^{\gamma_k} + \beta_k$, with parameters selected on the basis of the posterior distributions computed above.

Sample identity was verified using genotype concordance: pairs of tumours from the same patient were expected to have >90% concordance, and all other pairs <80% concordance. Samples failing these criteria were excluded. Quality of the overall HMM fit was assessed by root mean squared (r.m.s.) error with respect to fitted values; samples with a r.m.s. error of >1.5 were excluded. Finally, to account for two commonly observed artefacts, fitted copy number values were set to 'NA' for singletons with a fitted total copy number of 0, or when the observed and fitted means differed by more than 2 for regions of inferred copy gain.

Recurrent DNA copy-number gain and loss. Genomic regions with recurrent DNA copy gain and loss were identified using GISTIC, version 2.0[17]. Probe integer total copy-number values obtained from PICNIC, c, were converted to $\log_2$ ratio values, y, as $y = \log_2(c_{\text{tumour}} + 0.1) - \log_2(c_{\text{control}} + 0.1)$. These ratio values were then segmented using CBS[18]. Cut-offs of +0.584 and −0.7 were used to categorize $\log_2$ ratio values as gain or loss, respectively. A minimum segment length of 20 SNPs and a $\log_2$ ratio 'cap' value of 3 were used. The same CBS segmented data and thresholds were used to compute a frequency of gain or loss for each gene found to be significant by GISTIC, which is reported in Supplementary Tables 13 and 14.

Fusion detection and validation. Putative fusions were identified using a computational pipeline we have developed called GSTRUCT-fusions. Our pipeline is based on a generate-and-test strategy that is fundamentally similar to our methodology reported previously for finding read-through fusions[37]. Paired-end reads were aligned using our alignment program GSNAP[37]. GSNAP has the ability to detect splices representing translocations, inversions and other distant fusions within a single read end.

These distant splices provided one set of candidate fusions for the subsequent testing stage. The other set of candidate fusions derived from unpaired unique alignments, in which each end of the paired-end read aligned uniquely to a different chromosome, and also from paired, but discordant unique alignments, in which each end aligned uniquely to the same chromosome, but with an apparent genomic distance that exceeded 200,000 bp or with genomic orientations that suggested an inversion or scrambling event.

Candidate fusions were then filtered against known transcripts from RefSeq and aligned to the genome using GMAP[38]. We required that both fragments flanking a distant splice, or both ends of an unpaired or discordant paired-end alignment, map to known exon regions. This filtering step eliminated approximately 90% of the candidates. We further eliminated candidate inversions and deletions that suggested rearrangements of the same gene, as well as apparent read-through fusion events involving adjacent genes in the genome, which our previous research indicated were likely to have a transcriptional rather than a genomic origin.

For the remaining candidate fusion events, we constructed artificial exon–exon junctions consisting of the exons distal to the supported donor exon and the exons proximal to the supported acceptor exon. The exons included in the proximal and distal computations were limited so that the cumulative length along each gene was within an estimated maximum insert length of 200 bp. As a control, we also constructed all exon–exon junctions consisting of combinations of exons within the same gene, for all genes contributing to a candidate fusion event.

In the testing stage of our pipeline, we constructed a genomic index from the artificial exon–exon junctions and controls using the GMAP_BUILD program included as part of the GMAP and GSNAP package. We used this genomic index and the GSNAP program with splice detection turned off to re-align the original read ends that were not concordant to the genome. We extracted reads that aligned to an intergenic junction corresponding to a candidate fusion, but not to a control intragenic junction.

We filtered the results of the re-alignment to require that each candidate fusion have at least one read with an overhang of 20 bp. We also required that each candidate fusion have at least ten supporting reads. For each remaining candidate fusion, we aligned the two component genes against each other using GMAP and eliminated the fusion if the alignment had any region containing 60 matches in a window of 75 bp. We also aligned the exon–exon junction against each of the component genes using GMAP and eliminated the fusion if the alignment had coverage greater than 90% of the junction and identity greater than 95%.

Validation of gene fusions was done using an RT–PCR approach using both colon tumour and matched normal samples. 500 ng of total RNA was reverse transcribed to complementary DNA with a High Capacity cDNA Reverse Transcription kit (Life Technologies) following manufacturer's instructions. 50 ng of cDNA was amplified in a 25 µl reaction containing 400 pM of each primer, 300 µM of each deoxynucleoside triphosphate and 2.5 U of LongAmp Taq DNA polymerase (New England Biolabs). PCR was performed with an initial denaturation at 95 °C for 3 min followed by 35 cycles of 95 °C for 10 s, 56 °C for 1 min and 68 °C for 30 s, and a final extension step at 68 °C for 10 min. 3 µl of PCR product was run on 1.2% agarose gel to identify samples containing the gene fusion. Specific PCR products were purified with either a QIAquick PCR Purification kit or Gel Extraction kit (Qiagen). The purified DNA was either sequenced directly with PCR primers specific to each fusion (Supplementary Table 14) or cloned into TOPO cloning vector pCR2.1 (Life Technologies) before Sanger sequencing. The clones were sequenced using Sanger sequencing on an ABI 3730 XL (Life Technologies) as per manufacturer's instructions. The Sanger sequencing trace files were analysed using Sequencher (Gene Codes Corp.).

RSPO fusion activity testing. A eukaryotic expression plasmid pRK5E driving the expression of carboxy-terminal Flag-tagged EIF3E, PTPKR (amino acids 1–387), RSPO2, RSPO3, EIF3E(e1)–RSPO2(e2), PTPRK(e1)–RSPO3(e2) and PTPRK(e7)–RSPO3(e2) was generated using standard PCR and cloning strategies.

Cells, conditioned media, immunoprecipitation and western blotting. HEK293T and HT-29 cells were obtained from Genentech's cell banking facility. HEK 293T cells were maintained in DMEM buffer supplemented with 10% FBS. HT-29 cells were maintained in RPMI 1640 supplemented with 10% FBS. For expression analysis and condition media generation 3×10^5 HEK29T3 cells were plated in six-well plates in 1.5 ml DMEM containing 10% FBS. Cells were transfected with 1 µg DNA using Fugene 6 (Roche) according to manufacturer's instructions. Media was conditioned for 48 h, collected, centrifuged and used to stimulate the luciferase reporter assay (final concentration, ×0.1–0.4). For expression analysis, media was collected, centrifuged to remove debris and used for immunoprecipitation.

Luciferase reporter assays. HEK293T (50,000 cells per ml) or HT-29 cells (110,000 cells per ml) were plated in 90 µl of media containing 2.5% FBS per well of a 96-well plate. After 24 h, cells were transfected using Fugene 6 according to manufacturer's instructions (Roche) with the following DNA per well: 0.04 µg (HEK293T) or 0.08 µg (HT-29) TOPbrite Firefly reporter[25] and 0.02 µg (HEK293T) and 0.04 µg (HT-29) pRL-SV40 *Renilla* (Promega) constructs. Cells were stimulated with 25 µl of either fresh or conditioned media (described above) containing 10% FBS with or without recombinant murine Wnt3a (20–100 ng per ml (final), R&D Systems). After 24 h stimulation, 50 µl of media was removed and replaced with Dual-Glo luciferase detection reagents (Promega) according to manufacturer's instructions. An Envision Luminometer (Perkin–Elmer) was used to detect luminescence. To control for transfection efficiency, firefly luciferase levels were normalized to *Renilla* luciferase levels to generate the measure of relative luciferase units. Experimental data are presented as mean ± s.d. from three independent wells.

Immunoprecipitation and western blotting. To confirm that the RSPO wild-type and RSPO fusion proteins were secreted, Flag-tagged proteins were immunoprecipitated from the media using anti-Flag–M2 antibody-coupled beads (Sigma), boiled in SDS–PAGE loading buffer, resolved on 4–20% SDS–PAGE (Invitrogen) and transferred onto a nitrocellulose membrane. RSPO and other Flag-tagged proteins expressed in cells were detected from cell lysates using western blot as described previously. In brief, immunoprecipitated proteins and proteins from cell lysates were detected by western blot using Flag–horseradish peroxidase-conjugated antibody and Super signal West Dura chemiluminescence detection substrate (Thermo Fisher Scientific).

Transposon-based genetic screen. In brief, a cohort of triple-transgenic mice (*Rosa26–LsL–Sb11*, *T2/Onc*, *Villin–Cre*) were generated to allow tissue-restricted mobilization of a mutagenic *Sleeping Beauty* transposon in the gastrointestinal epithelium[6]. Triple-transgenic mice developed intestinal intraepithelial neoplasias, adenomas and adenocarcinomas. In total, 135 tumours from 47 mice were analysed for transposon-insertion sites by DNA extraction, linker-mediated PCR to amplify genomic DNA-transposon junctions, and 454 pyrosequencing. 16,000 non-redundant transposon insertion sites were mapped, and a Monte Carlo method was used to define common insertion sites (CISs) that were recurrently targeted by transposon insertions at a higher frequency than predicted by chance.

Quantitative RT–PCR. Quantitative RT–PCR reactions were performed twice in duplicate using Fast Start Universal SYBR Master (Roche) and a

Mastercycler ep realplex thermal cycler (Eppendorf). Primers sequences are listed below (5′ to 3′):
Mouse Actb_F1: TCCAGCCTTCCTTCTTGGGTATGGA
Mouse Actb_R1: CGCAGCTCAGTAACAGTCCGCC
Human RSPO2_F1: TTATTTGCATAGAGGCCGTTGCTTT
Human RSPO2_R1: GCTCCAATGACCAACTTCACATCCTT

26. Morgan, M. *et al.* ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25,** 2607–2608 (2009).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
28. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
29. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
30. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26,** 873–881 (2010).
31. 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).
32. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6,** 677–681 (2009) CrossRef.
33. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
34. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
35. Greenman, C. D. *et al.* PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. *Biostatistics* **11,** 164–175 (2010).
36. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (2007).
37. Nacu, S. *et al.* Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med. Genomics* **4,** 11 (2011).
38. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21,** 1859–1875 (2005).

LETTER

doi:10.1038/nature11308

Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors

Bon-Kyoung Koo[1], Maureen Spit[2]*, Ingrid Jordens[2]*, Teck Y. Low[3,4], Daniel E. Stange[1], Marc van de Wetering[1], Johan H. van Es[1], Shabaz Mohammed[3,4], Albert J. R. Heck[3,4], Madelon M. Maurice[2] & Hans Clevers[1]

LGR5$^+$ stem cells reside at crypt bottoms, intermingled with Paneth cells that provide Wnt, Notch and epidermal growth factor signals[1]. Here we find that the related RNF43 and ZNRF3 transmembrane E3 ubiquitin ligases are uniquely expressed in LGR5$^+$ stem cells. Simultaneous deletion of the two genes encoding these proteins in the intestinal epithelium of mice induces rapidly growing adenomas containing high numbers of Paneth and LGR5$^+$ stem cells. *In vitro*, growth of organoids derived from these adenomas is arrested when Wnt secretion is inhibited, indicating a dependence of the adenoma stem cells on Wnt produced by adenoma Paneth cells. In the HEK293T human cancer cell line, expression of RNF43 blocks Wnt responses and targets surface-expressed frizzled receptors to lysosomes. In the *RNF43*-mutant colorectal cancer cell line HCT116, reconstitution of RNF43 expression removes its response to exogenous Wnt. We conclude that RNF43 and ZNRF3 reduce Wnt signals by selectively ubiquitinating frizzled receptors, thereby targeting these Wnt receptors for degradation.

The small intestinal epithelium self renews every 4–5 days, making it the fastest proliferating mammalian tissue. Crypt base columnar (CBC) cells[2] are small cycling cells residing between the terminally differentiated Paneth cells at crypt bottoms. We identified *Lgr5* as a marker for these cells and integrated an enhanced green fluorescent protein (*EGFP*)–internal ribosome entry site (*IRES*)–*creERT2* cassette into the locus. As shown by lineage tracing, LGR5$^+$ cells generate all cell types of the epithelium over the lifetime of the animal[3]. Conditions were then established in which single LGR5$^+$ cells generate long-lived organoids that retain all characteristics of the epithelium[4]. Organoids derived from a single adult LGR5$^+$ cell were transplanted into recipient mice with chemically induced colitis, where they integrated to generate normal epithelium for at least 6 months[5]. Paneth cells provide Wnt, Notch and epidermal growth factor (EGF) signals to LGR5$^+$ stem cells[1]. Gene-expression profiling of sorted LGR5–GFP cells allowed us to define a short list of enriched genes, such as *Ascl2* and *Olfm4* (ref. 6). Here we address the role of two related genes that occur in this list: *Rnf43* and *Znrf3*.

We arbitrarily sorted crypt cells on the basis of LGR5–GFP expression into five fractions and performed microarray expression profiling (Supplementary Fig. 2a, b). As expected, *Lgr5* and other stem-cell genes (*Olfm4*, *Ascl2* and *Tnfrsf19*) showed an expression gradient that was highest in fraction 5+ (Fig. 1a). From this microarray experiment (Gene Expression Omnibus accession number GSE36497), we identified additional genes enriched in the 5+ fraction. Among these were two related RING-type E3 ubiquitin ligases, *Rnf43* and *Znrf3*. The encoded proteins contain an ectodomain, a single transmembrane region and a cytoplasmic RING domain (Fig. 1b and Supplementary Fig. 2c). *In situ* hybridization and quantitative reverse transcriptase (RT)–PCR confirmed expression of *Rnf43* and *Znrf3* in CBC cells (Fig. 1c, red arrows; Supplementary Fig. 2d).

We generated conditional alleles of the genes by inserting two *loxP* sites flanking the exon encoding the RING domain (Supplementary Fig. 3). The alleles were individually crossed with intestinal epithelium-specific Cre lines (*Cyp1a1-cre* or villin (*Vil*)-*creERT2*), which allow gene deletion upon β-naphthoflavone and tamoxifen administration, respectively[7,8]. Homozygous mutants did not yield any discernible phenotype (Supplementary Fig. 4). We then generated *Rnf43 Znrf3*

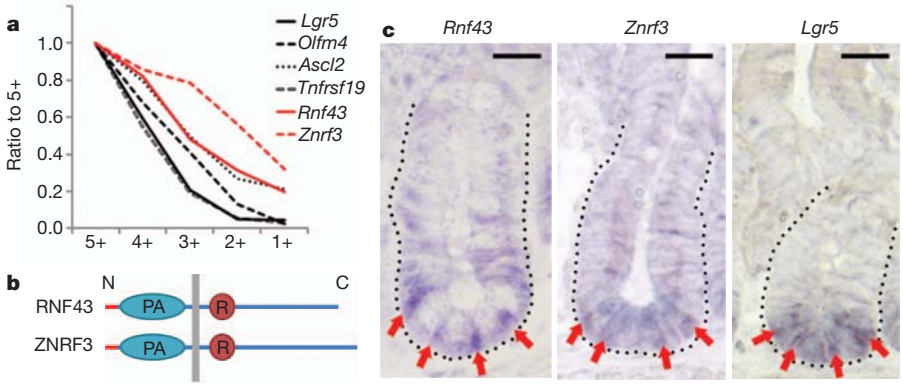


Figure 1 | LGR5$^+$ stem cell genes: *Rnf43* and *Znrf3*. **a**, Graph of the expression pattern of stem-cell genes (*Lgr5*, *Olfm4*, *Ascl2* and *Tnfrsf19*) and two RING-type E3 ubiquitin ligases (*Rnf43* and *Znrf3*). Every fraction is compared to fraction 5+. **b**, Schematic overview of RNF43 and ZNRF3. The red line denotes the signal peptide and the grey line denotes the membrane. PA, protease-associated domain; R, RING domain. **c**, *In situ* hybridization of *Rnf43*, *Znrf3* and *Lgr5* in the mouse small intestinal crypt. CBC cells in between Paneth cells are indicated with red arrows. Dotted line indicates the intestinal epithelial lining of the crypt. Scale bars, 20 μm.

[1]Hubrecht Institute, KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [2]Department of Cell Biology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands. [3]Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands. [4]The Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands.
*These authors contributed equally to this work.

compound mutant mice with *Vil-creERT2*. When both E3 ligases were deleted simultaneously, the mutant intestine underwent a marked expansion of the proliferative compartment (Fig. 2a, b), resembling the effects of acute deletion of *Apc*[9]. Ki67$^+$ cells not only occupied the enlarged crypts but also appeared on the flanks of villi. The hyperproliferative cells contained high levels of β-catenin (Fig. 2d, inset), indicating a strong activation of Wnt/β-catenin signalling such as seen upon *Apc* deletion[9]. Indeed, Wnt target genes such as *Cd44*, *Ephb2* and *Axin2* were strongly upregulated (Supplementary Fig. 5). In wild-type intestine, *Olfm4*-positive stem cells reside at crypt bottoms (Fig. 2e). In the compound mutant intestine, numbers of *Olfm4*-positive cells were strongly increased (Fig. 2f), as were the numbers of *Cryptdin6* (also known as *Defa6*)-expressing Paneth cells[10] (Fig. 2h). The expansion of the stem–Paneth cell zone was confirmed by using other stem-cell (*Msi1* and *Sox9*) and Paneth cell (lysozyme) markers (Supplementary Fig. 6). Clonal deletion of *Rnf43* and *Znrf3* using a low dose of tamoxifen resulted in adenoma formation within 1 month (Fig. 2i). In the adenoma, intestinal stem cells continuously expanded while generating Paneth cells (Fig. 2j, k). We also deleted *Rnf43* and *Znrf3* stochastically in LGR5$^+$ stem cells using *Lgr5-creERT2*. Again, these mutant stem cells generated adenomas that contained *Olfm4*-positive stem cells and lysozyme-positive Paneth cells, but no other differentiated cell types (Supplementary Fig. 7).

We cultured intestinal organoids from wild-type and *Rnf43 Znrf3* compound mutant mice. *Rnf43 Znrf3* compound mutant organoids grew faster than controls (Supplementary Fig. 8). R-spondin1 (RSPO1), an essential component of the medium, is a ligand of LGR5, which itself resides in the frizzled–LRP complex. Through these interactions, RSPO1 potently enhances Wnt signals[11–13]. Whereas control organoids died within 5 days after the removal of RSPO1,

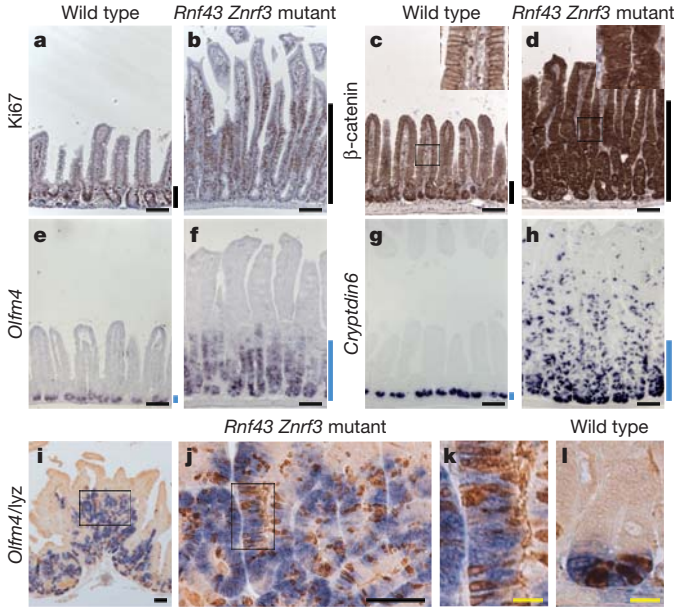


Figure 2 | Strong proliferation of the *Rnf43 Znrf3* compound mutant intestine is accompanied by Wnt/β-catenin activation as well as stem cell and Paneth cell metaplasia. **a**–**h**, Wild-type (**a**, **c**, **e** and **g**) and *Rnf43 Znrf3* compound mutant (**b**, **d**, **f** and **h**) intestines are stained for Ki67 (**a**, **b**; for proliferating cells), β-catenin (**c**, **d**), *Olfm4* (**e** and **f**; for stem cells) and *Cryptdin6* (**g**, **h**; for Paneth cells). These intestines were analysed 1 week after tamoxifen induction (5 mg). Side bars indicate Ki67-positive proliferative zone (black) and *Olfm4*-positive stem cell zone (blue), respectively. **i**–**l**, Adenoma formation in the *Rnf43 Znrf3* compound mutant intestine. Formation of adenoma was observed 1 month after tamoxifen induction (1 mg) in *Rnf43 Znrf3* compound mutant (**i**–**k**) and wild-type mice (**l**). **j** and **k** are high-magnification images of boxed regions in **i** and **j**, respectively. Note that stem-cell (*Olfm4*, in blue) expansion accompanies Paneth cell metaplasia (Lyz, in brown). Lyz, lysozyme. Scale bars, 100 μm (black) and 10 μm (yellow).

the mutant organoids could be maintained for at least 4 weeks in the absence of RSPO1 (Supplementary Fig. 9f, i). We proposed that mutant stem cells were hypersensitive to Paneth-cell-derived WNT3, rendering them RSPO1-independent. To test this, we treated the cultures with a small molecule inhibitor of porcupine[14], thus blocking the secretion of functional WNT3 by Paneth cells[1]. Inhibitor treatment resulted in the death of both the wild-type and *Rnf43 Znrf3* mutant organoid cultures within 5 days, illustrating that Paneth-cell-derived WNT3 was indeed essential (Supplementary Fig. 9g, i). The effect of the inhibitor was completely rescued by exogenous WNT3A (Supplementary Fig. 9d, h). Taken together, we conclude that mutation of *Rnf43* and *Znrf3* rendered LGR5$^+$ stem cells hypersensitive to locally secreted Wnt, resulting in adenoma formation *in vivo* and RSPO1-independent growth *in vitro*. Unlike *Apc*-mutant adenoma cells that activate the Wnt pathway cell autonomously, *Rnf43 Znrf3* mutant cells depend on paracrine Wnt signals for their growth.

Because RNF43 and ZNRF3 acted as negative Wnt regulators, we tested the effect of RNF43 overexpression in mouse intestinal *Vil-creERT2*-derived organoids by retroviral expression using a vector that harbours a floxed *dsRed* complementary DNA upstream of *RNF43* (ref. 15). After tamoxifen treatment, the induced RNF43 expression inhibited organoid growth immediately (Supplementary Fig. 9m). Growth could be restored with a GSK3-β inhibitor, but not by exogenous WNT3A (Supplementary Fig. 9n–p), indicating that the negative effect of RNF43 on Wnt/β-catenin signalling occurred upstream of GSK3-β.

We also performed Wnt reporter assays in HEK293T cells transfected with either wild-type *RNF43* or *RNF43* mutants with inactivating point mutations in their RING domain (MT). A small amount of RNF43 completely abolished the response to exogenous WNT3A (Fig. 3a). A similar effect was observed after transfection of *Znrf3* (Supplementary Fig. 10). The RNF43 RING mutants showed no inhibition but instead acted in a dominant-negative manner and enhanced Wnt signalling activity (Fig. 3a). The inhibitory effect of RNF43 was not observed when the Wnt pathway was activated by dominant-active LRP6 with an amino-terminal deletion (ΔN-LRP6)[16], or by a GSK3-β inhibitor (Fig. 3b), suggesting that the effect occurred at the level of the frizzled receptors for Wnt.

RNF43 mutations have been identified in two colorectal cancer cell lines containing an activating β-catenin (*CTNNB1*) mutation rather than an *APC* mutation[17]. HCT116 cells harbour one wild-type and one mutant *CTNNB1* allele. As a consequence, the Wnt pathway is modestly active in these cells, and can be strongly enhanced by exogenous Wnt[18]. We confirmed the homozygous *RNF43* mutation in HCT116 cells (Supplementary Fig. 11a). Re-expression of RNF43 in these cells blocked the effect of WNT3A (Fig. 3c). These results indicate that mutations of *RNF43* synergize with mutations in *CTNNB1* in colon cancer cells.

To assess whether RNF43 selectively controls Wnt receptor expression at the cell surface, we generated stable transfectants of HEK293 cells in which RNF43 expression could be induced by doxycycline (Dox). Cells were surface-labelled with biotin both before and 24 h after the addition of Dox, and their surface proteomes were compared by mass spectrometry. We confirmed RNF43 as the most upregulated protein on the cell surface, whereas FZD1, FZD2, FZD3 and LRP5 were among the eight most strongly downregulated surface proteins (Fig. 3d).

We next investigated how expression of RNF43 affects the subcellular distribution of frizzled proteins. To this end, we expressed a functional version of FZD5 that carried a SNAP tag in its extracellular domain. SNAP–FZD5 labelling with membrane-impermeable SNAP–Alexa549 showed its stable localization at the cell surface (Fig. 3e). After co-expression with RNF43, FZD5 was rapidly internalized (Fig. 3e). A similar effect was noted after transfection of *Znrf3* (Supplementary Fig. 10b). This was not observed when various mutants of RNF43 were co-expressed with FZD5, indicating that both RING

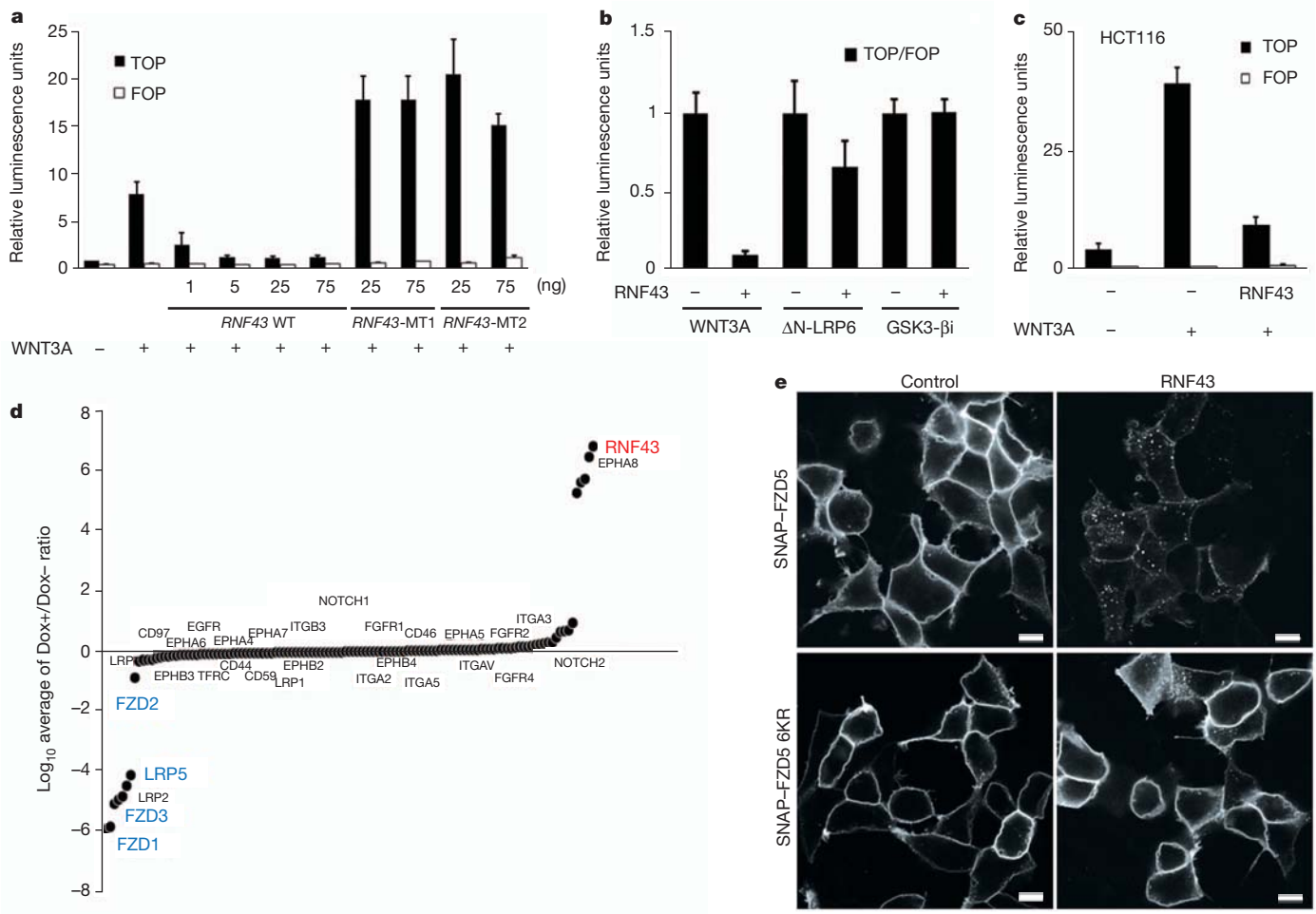


Figure 3 | RNF43 suppresses the Wnt/β-catenin pathway by reducing surface levels of frizzled receptors. **a**, Luciferase assay of TCF4–β-catenin activity in the presence of RNF43. RNF43 abolishes WNT3A-activated luciferase activity in a dose-dependent manner. RING domain *RNF43* point mutants (*RNF43*-MT1, C290S/H292S; *RNF43*-MT2, H295S/C298S) do not downregulate luciferase activity. WT, wild type. **b**, Luciferase assay of TCF4–β-catenin activity with various stimuli (WNT3A, ligand; ΔN-LRP6, active form of co-receptor; GSK3-βi, inhibitor of GSK3-β that activates signalling downstream of the Wnt receptors). RNF43 overexpression strongly suppresses WNT3A-mediated signalling whereas signalling downstream of frizzled (ΔN-LRP6 and GSK3-βi) remains unaffected. Luciferase activities are normalized to control cells. **c**, Luciferase assay of TCF4–β-catenin activity in HCT116 cells. Re-expression of RNF43 in HCT116 cells blocks the effect of exogenous WNT3A. **d**, Relative abundance of surface proteins detected by mass spectrometry in the presence or absence of RNF43. After induction of RNF43, surface levels of FZD1, FZD3 and LRP5 are largely reduced. **e**, Subcellular localization of SNAP–FZD5 and lysine mutant SNAP–FZD5 6KR in HEK293T cells co-transfected with control or RNF43. Surface SNAP–FZD5 was labelled with SNAP–Alexa549 for 15 min and chased for 5 min. Note that RNF43 induces rapid internalization of SNAP–FZD5. SNAP–FZD5 6KR is resistant to RNF43-induced internalization. Scale bars, 10 μm. Error bars denote s.d., $n = 3$. Three independent biological experiments were performed for the luciferase assays (**a**–**c**). FOP, mutant negative control luciferase reporter; TOP, TCF optimal luciferase reporter.

activity and the ectodomain are important to promote FZD5 internalization (Supplementary Fig. 11b). A non-ubiquitinatable FZD5 6KR variant (in which all cytosolic lysine residues are substituted by arginine residues) was resistant to RNF43-mediated internalization, indicating a direct role for ubiquitination (Fig. 3e). Flow cytometric analyses confirmed that RNF43 reduced surface expression of FZD5 (Supplementary Fig. 11c).

Ubiquitin modification commonly drives cell-surface receptor internalization and lysosomal degradation[19]. De-ubiquitination controls the recycling of receptors, including frizzled receptors[20]. Co-expression of RNF43 strongly and selectively decreased levels of the mature, complex glycosylated form of FZD5, whereas the immature form remained unaffected (Fig. 4a and Supplementary Fig. 12a). Both the RNF43 RING domain and FZD5 lysine acceptor sites for ubiquitin were essential for this effect. Indeed, RNF43 strongly promoted ubiquitination of FZD5 (Fig. 4b). Notably, lysosomal inhibition (bafilomycin A1) counteracted the RNF43-mediated degradation of FZD5, whereas proteasomal inhibition (MG132) had no effect (Fig. 4b and Supplementary Fig. 12b). RNF43 interacted with FZD5 as assessed by co-immunoprecipitation, and both proteins co-localized in internal vesicles (Fig. 4c, d). On expression of RNF43, surface FZD5 was efficiently targeted to RAB5$^+$ early endosomes and CD63$^+$ lysosomes (Supplementary Fig. 12c, d). We conclude that RNF43 negatively regulates Wnt/β-catenin signalling by ubiquitinating frizzled receptors, thereby targeting this Wnt receptor to the lysosomal pathway for degradation.

Previous studies have identified *Rnf43* and *Znrf3* as target genes of the Wnt pathway[6,21,22]. The current study demonstrates that they act as inhibitors of Wnt signalling by negatively regulating frizzled receptor surface expression. The same notion was put forward in a study[23] that appeared while the current manuscript was under revision. Thus, these genes join the class of negative-feedback Wnt pathway regulators, which includes DKK1 (ref. 24), APCDD1 (ref. 25) and AXIN2 (ref. 26). The inverse mechanism has been proposed for the Wnt receptor tyrosine kinase RYK: the ubiquitin ligase MIB1 triggers its internalization and thereby activates Wnt signalling[27].

Homeostasis of the intestinal stem-cell compartment is governed by a number of positive- and negative-feedback regulators of

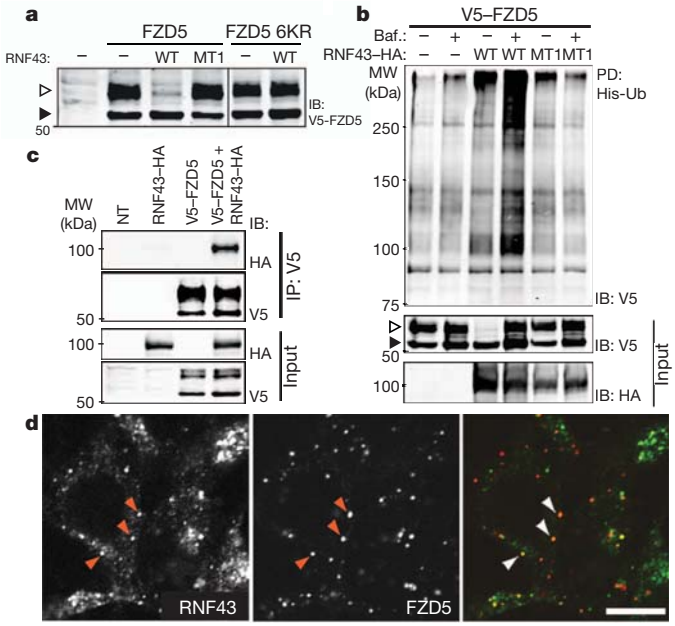


Figure 4 | RNF43 promotes ubiquitin-mediated endocytosis of frizzled receptors. **a**, RNF43 regulates FZD5 turnover. Levels of mature, complex glycosylated V5-conjugated FZD5 (V5–FZD5) are selectively downregulated in HEK293T cells co-expressing RNF43, but not mutant RNF43 (MT1). V5–FZD5 6KR levels remain unaffected by RNF43 co-expression. Arrowheads denote immature (filled) and mature (open) V5–FZD5. IB, immunoblot. **b**, RNF43 mediates ubiquitination of FZD5. HEK293T cells were transfected with V5–FZD5, His-ubiquitin and wild-type or mutant (M1) haemagglutinin-conjugated RNF43 (RNF43–HA), as indicated. Treatment of cells with bafilomycin A1 (Baf.), an inhibitor of lysosomal acidification, rescues FZD5 from RNF43-mediated downregulation (middle). Levels of ubiquitinated FZD5 (top) significantly accumulate upon RNF43 co-expression and lysosomal inhibition, but not with RNF43-MT1. Arrowheads denote immature (filled) and mature (open) V5–FZD5. PD, pull-down; Ub, ubiquitin. **c**, RNF43 interacts with FZD5. V5-FZD5 was immunoprecipitated (IP) from bafilomycin A1-treated HEK293T cells expressing V5–FZD5 and RNF43–HA, as indicated. NT, non-transfected. **d**, RNF43–GFP co-localizes with SNAP–FZD5 in internal vesicles. Surface SNAP–FZD5 was labelled with SNAP–Alexa549 for 15 min and chased for 30 min in HEK293T cells. Arrowheads indicate co-localization. Scale bars, 10 μm.

Wnt/β-catenin signalling[6,12], indicating the importance of balancing adequate levels of Wnt activity. We propose that RNF43 and ZNRF3 serve to restrict the size of the intestinal stem-cell zone. Indeed, their genetic removal leads to unrestricted expansion of the stem-cell zone in a fashion that is reminiscent of the removal of another negative Wnt pathway regulator, APC[9]. Unlike *APC*-mutant adenomas however, *Rnf43 Znrf3* compound mutant adenomas depend on paracrine Wnt factors from Paneth cells within the mutant clone. RNF43 is a genuine tumour suppressor: it was recently found to be mutated in human intraductal papillary mucinous neoplasm, a class of pancreatic neoplasia[28] and in cholangiocarcinoma[29]. *RNF43* was also found to be mutated in a transposon-based screen for intestinal adenoma formation in mice[30] and in two colon cancer cell lines[17]. Deep-sequencing efforts have shown further cases of *RNF43* mutations in colorectal and other cancers (http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=mutations&ln=RNF43&start=1&end=784&coords=AA:AA&neg=off&page=1). Our findings indicate that human tumours that harbour mutations in *RNF43* may be treatable by Wnt pathway inhibitors that act at the level of Wnt secretion or receptor activation.

METHODS SUMMARY

Reagents. Murine recombinant EGF was from PeproTech. Noggin- and RSPO1-conditioned media were generated from HEK293 cells. WNT3A-conditioned media was generated from L cells. Y-27632 was from Sigma. CHIR99021 was from Stemgent. Porcupine inhibitor IWP1 was provided by L. Lum[14]. Endoribonuclease-prepared short interfering RNAs for *RNF43* (EHU058181) and *ZNRF3* (EHU136411) were from Sigma. N-terminal SNAP-tagged mouse FZD5 was designed on the basis of previously described V5–Fz5 (ref. 31).

Mice. *Rnf43* and *Znrf3* floxed alleles were generated through homologous recombination in embryonic stem cells as described[12]. Details of these constructs are available on request. Experiments were performed according to guidelines and reviewed by the Dier Experimenten Commissie (DEC) of the KNAW.

Crypt isolation, cell dissociation and culture. Crypt isolation, cell dissociation and culture have been described previously[4]. For the analysis of organoid growth under various conditions, established organoid cultures were dissociated into small pieces and seeded in Matrigel (BD Biosciences) under the appropriate growth condition. Growth morphology of organoids was imaged daily and viability of each organoid was scored.

Microarray analysis. LGR5–GFP-positive cells from *Lgr5–GFP–IRES–creERT2* knock-in intestines were sorted into five different fractions based on their GFP intensity. RNeasy Micro kit (Qiagen) was used to isolate RNA, and microarray analysis (Agilent) was performed as previously described[1]. The data were deposited to the Gene Expression Omnibus (accession number GSE36497).

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

Received 23 March; accepted 11 June 2012.
Published online 15 August 2012.

1. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469,** 415–418 (2011).
2. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. *Am. J. Anat.* **141,** 461–479 (1974).
3. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
4. Sato, T. *et al.* Single Lgr5 stem cells build crypt–villus structures *in vitro* without a mesenchymal niche. *Nature* **459,** 262–265 (2009).
5. Yui, S. *et al.* Functional engraftment of colon epithelium expanded *in vitro* from a single adult $Lgr5^{+}$ stem cell. *Nature Med.* **18,** 618–623 (2012).
6. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136,** 903–912 (2009).
7. El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39,** 186–193 (2004).
8. Ireland, H. *et al.* Inducible cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of β-catenin. *Gastroenterology* **126,** 1236–1246 (2004).
9. Sansom, O. J. *et al. Myc* deletion rescues Apc deficiency in the small intestine. *Nature* **446,** 676–679 (2007).
10. van Es, J. H. *et al.* Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nature Cell Biol.* **7,** 381–386 (2005).
11. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proc. Natl Acad. Sci. USA* **108,** 11452–11457 (2011).
12. de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476,** 293–297 (2011).
13. Kazanskaya, O. *et al.* The Wnt signaling regulator R-spondin 3 promotes angioblast and vascular development. *Development* **135,** 3655–3664 (2008).
14. Chen, B. *et al.* Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nature Chem. Biol.* **5,** 100–107 (2009).
15. Koo, B. K. *et al.* Controlled gene expression in primary Lgr5 organoid cultures. *Nature Methods* **9,** 81–83 (2012).
16. Liu, G., Bafico, A., Harris, V. K. & Aaronson, S. A. A novel mechanism for Wnt activation of canonical signaling through the LRP6 receptor. *Mol. Cell. Biol.* **23,** 5825–5835 (2003).
17. Ivanov, I., Lo, K. C., Hawthorn, L., Cowell, J. K. & Ionov, Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. *Oncogene* **26,** 2873–2884 (2007).
18. Li, V. S. *et al.* Wnt signaling through inhibition of β-catenin degradation in an intact Axin1 complex. *Cell* **149,** 1245–1256 (2012).
19. Haglund, K. & Dikic, I. The role of ubiquitylation in receptor endocytosis and endosomal sorting. *J. Cell Sci.* **125,** 265–275 (2012).
20. Mukai, A. *et al.* Balanced ubiquitylation and deubiquitylation of Frizzled regulate cellular responsiveness to Wg/Wnt. *EMBO J.* **29,** 2114–2125 (2010).
21. Van der Flier, L. G. *et al.* The intestinal Wnt/TCF signature. *Gastroenterology* **132,** 628–632 (2007).
22. Yagyu, R. *et al.* A novel oncoprotein RNF43 functions in an autocrine manner in colorectal cancer. *Int. J. Oncol.* **25,** 1343–1348 (2004).
23. Hao, H. X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485,** 195–200 (2012).
24. Niida, A. *et al.* DKK1, a negative regulator of Wnt signaling, is a target of the β-catenin/TCF pathway. *Oncogene* **23,** 8520–8526 (2004).
25. Shimomura, Y. *et al.* APCDD1 is a novel Wnt inhibitor mutated in hereditary hypotrichosis simplex. *Nature* **464,** 1043–1047 (2010).

26. Jho, E. H. *et al.* Wnt/β-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. *Mol. Cell. Biol.* **22,** 1172–1183 (2002).
27. Berndt, J. D. *et al.* Mindbomb 1, an E3 ubiquitin ligase, forms a complex with RYK to activate Wnt/β-catenin signaling. *J. Cell Biol.* **194,** 737–750 (2011).
28. Wu, J. *et al.* Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc. Natl Acad. Sci. USA* **108,** 21188–21193 (2011).
29. Ong, C. K. e. t. a. l. Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nature Genet.* **44,** 690–693 (2012).
30. March, H. N. *et al.* Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nature Genet.* **43,** 1202–1209 (2011).
31. Tauriello, D. V. *et al.* Loss of the tumor suppressor CYLD enhances Wnt/β-catenin signaling through K63-linked ubiquitination of Dvl. *Mol. Cell* **37,** 607–619 (2010).

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

Acknowledgements We thank A. A. Rolf, I. Kuper, M. van den Born, C. Kroon-Veenboer, H. Begthel, J. Korving and S. van den Brink for technical assistance, L. Lum for providing IWP1 and S. Bartfeld for the schematic drawing. This work was funded in part by grants from the European Research Council, EU/232814-StemCeLLMark and the National Research Foundation of Korea, NRF-2011-357-C00093 (B.-K.K.); EU/Health-F4-2007-200720 (M.v.d.W.); The Centre van Biomedical Genetics (D.E.S.); Ti Pharma/T3-106 (J.H.v.E.); the European Research Council, ERC-StG no.242958 (M.M.M.) and the KNAW/3V-fund.

Author Contributions B.-K.K., M.M.M. and H.C. conceived and designed the experiments. B.-K.K., M.S., I.J., D.E.S., M.v.d.W. and J.H.v.E. performed the experiments. T.Y.L., S.M. and A.J.R.H. performed the mass spectrometry analysis. B.-K.K., M.S., I.J., M.M.M. and H.C. analysed the data. B.-K.K., M.M.M. and H.C. wrote the manuscript.

Author Information The data for the microarray analysis have been deposited to the Gene Expression Omnibus under accession number GSE36497. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. 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.M.M. (M.M.Maurice@umcutrecht.nl) or H.C. (h.clevers@hubrecht.eu).

METHODS

Plasmids. N-terminal SNAP-tagged mature mouse FZD5 (Ala27-Val559) was generated by PCR-subcloning SNAP from nSNAP (Bioke) between mouse H2-Kb signal sequence and FZD5 as described for V5–Fz5[31]. SNAP–FZD5 6KR was generated by substitution of cytoplasmic lysines 347, 434, 439, 442, 445 and 525 for arginines by site-directed mutagenesis. RNF43–2×Flag–HA and RNF43–GFP constructs were generated by PCR-subcloning. His-ubiquitin was a gift from B. Burgering.

Histology, immunohistochemistry and *in situ* hybridization. Mouse intestine was fixed in 4% paraformaldehyde and embedded in paraffin. For immunohistochemistry, the primary antibodies were mouse anti-Ki67 (1:250, Monosan), mouse anti-β-catenin (1:100, BD Transduction Laboratories) and rabbit anti-lysozyme (1:1,500, DAKO). For *in situ* hybridization, digoxigenin (DIG)-labelled complementary RNA probes were generated from cDNA-containing vectors (from IMAGE consortium or RZPD) by *in vitro* transcription.

Surface biotinylation and mass spectrometry analysis. Surface proteins from Dox-inducible RNF43-overexpressing HEK293 cells were enriched by using EZ-Link Sulfo-NHS-SS-Biotin, (Thermo Fisher) according to manufacturer's protocol, except for the following modifications. After intensive washing, the resin was rinsed with five volumes of PBS buffer to remove residual detergent. Biotinylated proteins were eluted after 1 h incubation, once with 5 mM dithiothreitol (DTT), and 0.5% RapiGest SF surfactant (Waters) in 50 mM ammonium bicarbonate, and twice with 5 mM DTT and 8 M urea in 50 mM ammonium bicarbonate. Each elution was stored separately. All samples were subjected to alkylation with iodoacetamide (in the dark at room temperature for 20 min) followed by digestion with Lys-C and then trypsin overnight. RapiGest detergent was precipitated by addition of 5% trifluoroacetic acid and removed by centrifugation. Digested samples were pooled and desalted with SEP-PAK C18 (Waters) before vacuum-drying.

Each sample was analysed in triplicates by performing nano-ultra performance liquid chromatography (LC)–tandem mass spectrometry (MS/MS) on a Proxeon EASY-nLC 1000 (Thermo Scientific) connected to a LTQ-Orbitrap Q-Exactive (Thermo Fisher Scientific). The injected sample was first trapped with a double-fritted trapping column (Dr Maisch ReproSil C18, 3 μm, 2 cm × 100 μm) before being separated in an analytical column (Agilent Zorbax SB-C18, 1.8 μm, 35 cm × 50 μm). Solvent A consisted of 0.1 M acetic acid and solvent B consisted of 0.1 M acetic acid in 80% acetonitrile. Measurement time for each sample took 240 min. Samples were first loaded at a maximum pressure of 980 bar with 100% solvent A. Subsequently, peptides were chromatographically separated by a 211 min gradient consisting of 10–40% solvent B at a flow of 100 nl min^{-1}; then ramped to 100% solvent B in 3 min and held in 100% solvent B for another 2 min. This was finally followed by a 13 min equilibration with 100% solvent A. For MS analysis, 1.7 kV was applied to the nano-spray needle. The survey scan was from 350 to 1500 *m/z* at a resolution of 30,000 and for MS/MS the resolution was set to 7500. The 20 most intense precursors were selected for fragmentation with higher energy dissociation. The target ion setting was 3E6 for MS and 5E4 for MS/MS, with a maximum fill-time of 250 ms and 120 ms, respectively.

Protein identification and extracted ion chromatogram detection. The raw files were converted to peak lists using Proteome Discoverer version 1.3 (Thermo Fisher Scientific) and searched against IPI (International Protein Index) human database version 3.37 (69,164 sequences; 29,064,824 residues) using Mascot search engine (version 2.3.02; http://www.matrixscience.com), with trypsin set as the enzyme. The database search was made with the following parameters set to consider a peptide tolerance of ± 50 p.p.m., a fragment tolerance of ± 0.02 Da allowing two missed cleavages, carbamidomethyl as a fixed modification and oxidation and protein N-terminal acetylation as variable modifications. The results were then filtered using Percolator to a false discovery rate below 1%. Filtered peptides assigned with a Mascot score lower than 20 were discarded. Using Proteome Discoverer, the selected ion chromatogram for all detected peptide precursors were computed and 'top 3' strategy was used for each protein in order to provide protein areas used for subsequent label-free quantification. These raw areas were then normalized with respect to the total intensity observed in each LC–MS analysis and against the sum of all analyses. Ultimately, a ratio was obtained by dividing the normalized area for each protein in the Dox^+ experiment by the Dox^- experiment.

Immunofluorescence and ubiquitination analysis. HEK293T cells were cultured in RPMI (Invitrogen), supplemented with 10% FBS (Sigma), 100 U ml^{-1} penicillin, and 100 μg ml^{-1} streptomycin and transfected with FuGENE 6 (Promega) according to the manufacturer's instructions. Bafilomycin A1 (10 nM; overnight) was used to block lysosomal acidification.

For microscopy analysis, HEK293T cells were grown on laminin-coated glass coverslips. Cells were transfected and after 20 h labelled with 1 μM SNAP–Alexa549 (Bioke) for 15 min at room temperature. Subsequently, cells were chased for 5 or 15 min at 37 °C and fixed in 4% paraformaldehyde. Cells were incubated with mouse anti-HA (Roche), followed by secondary antibodies conjugated to Alexa488 or -680 (Invitrogen). Cells were mounted in ProLong Gold (Invitrogen) and visualized using a Zeiss LSM510 confocal microscope.

Ubiquitination analysis was done as described[31]. In brief, HEK293T cells were transfected with His-ubiquitin and V5-tagged FZD5 for 24 h, washed and collected in ice-cold PBS with 10 mM *N*-ethyl maleimide (Sigma). Cells were lysed in a chaotropic lysis buffer containing 6 M guanidinium–HCl (Fluka), 100 mM phosphate buffer, 10 mM Tris–HCl, pH 8.0, 20 mM imidazole and 10 mM β-mercaptoethanol. After sonication, lysates were cleared by centrifugation and incubated with nickel-nitrilotriacetic acid–agarose beads (Qiagen) for 2.5 h at room temperature. Beads were washed with lysis buffer with 0.2% Triton X-100, followed by sequential washes with buffers A, B and C (A, 8 M urea, 100 mM phosphate buffer, 10 mM Tris–HCl, pH 8.0, 10 mM β-mercaptoethanol and 0.2% Triton X-100; B, as buffer A, but pH 6.3; C, as B, but with 0.1% Triton X-100). His-tagged proteins were eluted with two bead volumes of elution buffer (6.2 M urea, 10 mM Tris–HCl, pH 7.0, 100 mM P buffer and 200 mM imidazole and sample buffer). For immunoblotting, rabbit anti-V5 (Sigma) and rat anti-HA (Roche) and were used. Secondary antibodies were obtained from Invitrogen.

Immunoprecipitation. HEK293T cells were co-transfected with V5–FZD5 and mock plasmid or RNF43–HA, and treated with bafilomycin A1 (10 nM; overnight). Cells were washed and collected in ice-cold PBS and lysed in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.5% Triton X-100, 5 mM EDTA, 1 mM DTT, 50 mM sodium fluoride and protease inhibitors) for 20 min on ice followed by 30 min centrifugation at 4 °C. Immunoprecipitations were performed on the supernatant with mouse anti-V5 (Invitrogen) and protein A beads (Repligen) for 2 h at 4 °C. Beads were washed six times with lysis buffer and proteins were eluted in ×1.5 sample buffer.

Flow cytometric analysis. HEK293 cells with Dox-inducible RNF43 or RNF43-MT1 were transfected with V5–FZD5 and V5–FZD5 6KR. Murine stem cell virus–enhanced GFP was co-transfected to gate transfected cells. Experimental groups were treated with Dox to induce RNF43 or RNF43-MT1 expression 24 h after transfection. Dox-treated and untreated cells were collected after 12 h of incubation. Cells were washed with PBS and resuspended in FACS buffer (PBS with 1% FBS). Cells were labelled with mouse anti-V5 antibody followed by anti-mouse Alexa647 antibody. Cells were washed twice with PBS and analysed using a BD FACSCalibur. Levels of V5 were analysed for GFP-positive cells.

Proteasomal inhibition. HEK293T cells co-expressing V5–FZD5 and RNF43–HA were treated with vehicle (ethanol) or the proteasomal inhibitor MG132 (20 × μM) for 3 h before cells were lysed.

Pulse chase. HEK293T cells expressing V5–FZD5 were grown in depletion medium (−Met/−Cys) for 30 min at 37 °C. Next, medium was exchanged with labelling medium (100 μCi ^{35}S-Met/^{35}S-Cys per ml, Redivue, GE Healthcare) for 15 min at 37 °C. After the pulse, medium was washed away and replaced by RPMI with 10% FCS at 37 °C for the indicated time points. Subsequently, cells were washed with PBS with 10 mM *N*-ethyl maleimide and lysed in lysis buffer. Immunoprecipitation of V5–FZD5 was performed as described above. Samples were applied to SDS–PAGE, the gel was dried and a phosphorimager screen (Molecular Dynamics) was used to measure radioactivity with a Storm scanner (Amersham BioSciences).

LETTER

doi:10.1038/nature11290

Novel role of PKR in inflammasome activation and HMGB1 release

Ben Lu[1,2], Takahisa Nakamura[3], Karen Inouye[3], Jianhua Li[1], Yiting Tang[4], Peter Lundbäck[5], Sergio I. Valdes-Ferrer[1,2], Peder S. Olofsson[1], Thomas Kalb[1], Jesse Roth[1], Yongrui Zou[4], Helena Erlandsson-Harris[5], Huan Yang[1], Jenny P.-Y. Ting[7], Haichao Wang[8], Ulf Andersson[6], Daniel J. Antoine[9], Sangeeta S. Chavan[1], Gökhan S. Hotamisligil[3] & Kevin J. Tracey[1,2]

The inflammasome regulates the release of caspase activation-dependent cytokines, including interleukin (IL)-1β, IL-18 and high-mobility group box 1 (HMGB1)[1–5]. By studying HMGB1 release mechanisms, here we identify a role for double-stranded RNA-dependent protein kinase (PKR, also known as EIF2AK2) in inflammasome activation. Exposure of macrophages to inflammasome agonists induced PKR autophosphorylation. PKR inactivation by genetic deletion or pharmacological inhibition severely impaired inflammasome activation in response to double-stranded RNA, ATP, monosodium urate, adjuvant aluminium, rotenone, live *Escherichia coli*, anthrax lethal toxin, DNA transfection and *Salmonella typhimurium* infection. PKR deficiency significantly inhibited the secretion of IL-1β, IL-18 and HMGB1 in *E. coli*-induced peritonitis. PKR physically interacts with several inflammasome components, including NOD-like receptor (NLR) family pyrin domain-containing 3 (NLRP3), NLRP1, NLR family CARD domain-containing protein 4 (NLRC4), absent in melanoma 2 (AIM2), and broadly regulates inflammasome activation. PKR autophosphorylation in a cell-free system with recombinant NLRP3, apoptosis-associated speck-like protein containing a CARD (ASC, also known as PYCARD) and pro-caspase-1 reconstitutes inflammasome activity. These results show a crucial role for PKR in inflammasome activation, and indicate that it should be possible to pharmacologically target this molecule to treat inflammation.

HMGB1, a ubiquitous nuclear and cytosolic protein, is released into the extracellular space during sterile inflammation and infection[6]. Immunocompetent cells secrete HMGB1 after stimulation with microbial products and other danger signals[4–6]. Ischaemic and injured somatic cells also release HMGB1, but in this case the release mechanisms are passive[6]. Extracellular HMGB1 signals through Toll-like receptor 4 (TLR4) receptors in macrophages to induce cytokine release[6,7] and mediate cell migration by interacting with receptor for advanced glycation end-products (RAGE)[8]. The importance of these mechanisms in disease pathogenesis has been established, because neutralizing HMGB1 attenuates disease severity in infectious and sterile injury syndromes such as arthritis, colitis, sepsis and ischaemia reperfusion[6]. The mediator role of HMGB1 at the intersection of both sterile and infectious inflammation highlights the importance of mechanisms that regulate the secretion of HMGB1 from immune cells[6].

Early observations suggested that secretion of HMGB1 from macrophages requires inflammasome and caspase activity[3–5,9,10], and is highly associated with HMGB1 translocation from the nucleus to the cytoplasm[4,6]. HMGB1 is also a universal sentinel to viral invasion, and its release can be stimulated by polyinosinic:polycytidylic acid (poly(I:C)), a double-stranded RNA (dsRNA) mimetic[11,12]. Because we were interested in studying HMGB1 release mechanisms from macrophages exposed to poly(I:C), we first considered earlier observations that PKR is phosphorylated in these conditions[13,14]. As PKR is also activated by prototypical pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) known to stimulate HMGB1 release[15,16] and activate the inflammasome, we reasoned that PKR activation might participate in inflammasome activation.

Accordingly, we measured HMGB1 release by peritoneal macrophages obtained from PKR-deficient ($Pkr^{-/-}$) mice, in which the PKR kinase domain had been functionally inactivated by genetic deletion[16]. HMGB1 secretion after poly(I:C) exposure in $Pkr^{-/-}$ macrophages was significantly lower than in wild-type ($Pkr^{+/+}$) macrophages (Fig. 1a). Pharmacological inhibition of PKR in $Pkr^{+/+}$ macrophages dose-dependently, and significantly, inhibited poly(I:C)-induced HMGB1 release and PKR phosphorylation (Fig. 1b and Supplementary Fig. 1). Other prototypical danger signals, including ATP, monosodium urate (MSU), adjuvant aluminium (ALU) and live *E. coli*, also significantly increased PKR phosphorylation and HMGB1 release (Fig. 1c–e). Addition of potassium to the extracellular space significantly inhibited ATP-induced PKR activation (Fig. 1c). Inactivation of macrophage PKR by genetic deletion (Fig. 1d, e) or pharmacological inhibition (Fig. 1f and Supplementary Fig. 2) also significantly inhibited HMGB1 release.

Pyroptosis, a form of programmed, inflammatory cell death, occurs with macrophage inflammasome activation, and we observed that deletion of PKR significantly inhibited lactate dehydrogenase (LDH) release (Fig. 1g). Analysis by tandem mass spectrometry of HMGB1 released in response to ATP, MSU or ALU indicated that HMGB1 was highly acetylated in the nuclear location sequence (Fig. 1h and Supplementary Figs 3–6). By contrast, HMGB1 released from macrophages subjected to freeze/thaw cycles was not acetylated in the nuclear location sequence (Fig. 1h). Together with evidence that inflammasome activation participates in the nuclear translocation of HMGB1 (ref. 4), these results indicated that HMGB1 hyperacetylation and release, and inflammasome activation, are regulated by PKR.

To address the role of PKR in activating the NLRP3 inflammasome, we measured caspase-1 activation and IL-1β cleavage in peritoneal macrophages from $Pkr^{+/+}$ and $Pkr^{-/-}$ mice. Caspase-1 activation and IL-1β cleavage were significantly inhibited in $Pkr^{-/-}$ macrophages stimulated by exposure to ATP, MSU and ALU (Fig. 2a). Similar results were obtained in bone marrow-derived dendritic cells (BMDCs) (Supplementary Fig. 7) and bone marrow-dervied macrophages (BMDMs) (Supplementary Fig. 8). The expression of NLRP3 and pro-IL-1β did not differ significantly in $Pkr^{-/-}$ or $Pkr^{+/+}$ macrophages (Fig. 2a and Supplementary Fig. 9), but IL-1β secretion by macrophages exposed to live *E. coli*. was significantly inhibited in

[1]Laboratory of Biomedical Science, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, New York 11030, USA. [2]The Elmezzi Graduate School of Molecular Medicine, North Shore-LIJ Health system, 350 Community Drive, Manhasset, New York 11030, USA. [3]Department of Genetics & Complex Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [4]Center for Autoimmunity and Musculoskeletal Diseases, Feinstein Institute for Medical Research, Manhasset, New York, New York 11030, USA. [5]Department of Medicine, Karolinska Institute, Karolinska University Hospital, Stockholm S-17176, Sweden. [6]Women's and Children's Health, Karolinska Institute, Karolinska University Hospital, Stockholm S-17176, Sweden. [7]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [8]Department of Emergency Medicine, North Shore University Hospital, Manhasset, New York 11030, USA. [9]MRC Centre for Drug Safety Science, Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool L69 3GE, UK.

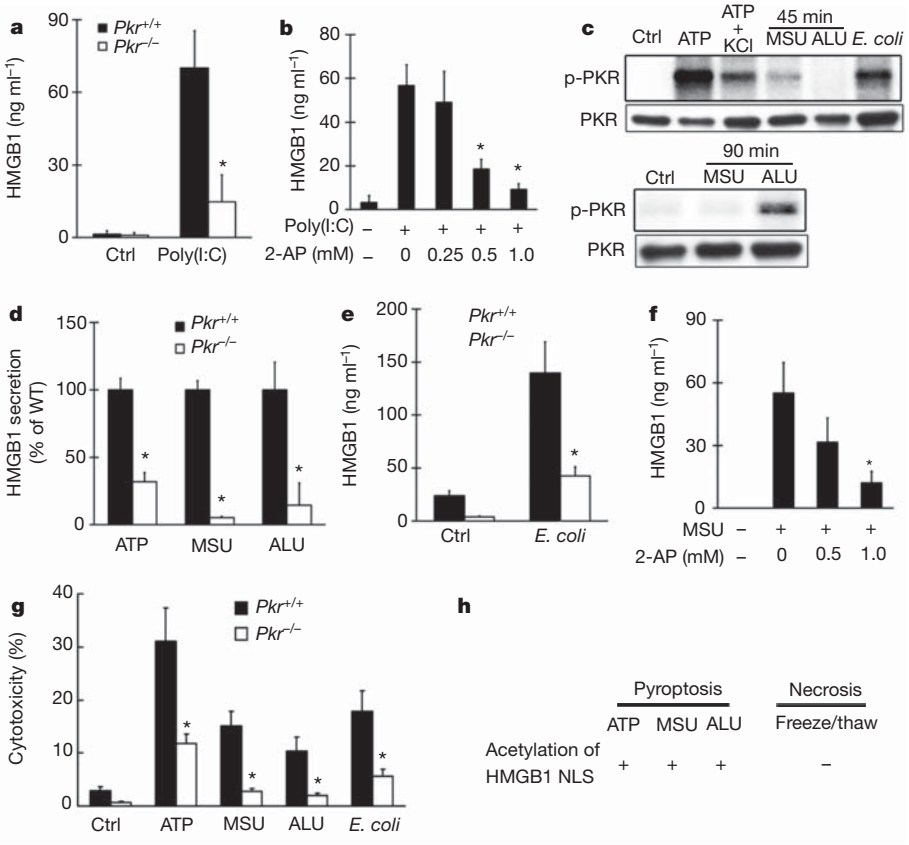


Figure 1 | Role of PKR in pyroptosis-mediated HMGB1 release. **a**, **b**, Macrophages from $Pkr^{+/+}$ or $Pkr^{-/-}$ mice (**a**) or from $Pkr^{+/+}$ mice treated with indicated doses of the PKR inhibitor 2-AP (**b**) were stimulated with poly(I:C). Ctrl, control. **c**, Lipopolysaccharide (LPS)-primed $Pkr^{+/+}$ macrophages were stimulated or treated with or without potassium-substituted medium (KCl) as indicated. Cells were lysed at indicated time points and PKR activation was monitored by autophosphorylation (denoted by p-PKR). **d**–**g**, LPS-primed $Pkr^{+/+}$ or $Pkr^{-/-}$ macrophages were stimulated or treated with 2-AP as indicated. HMGB1 levels in the supernatant were determined by western blot. WT, wild type. Cytotoxicity (**g**) was determined by LDH assay. Data are mean and s.d. of three independent experiments. $*P < 0.05$ versus wild-type-stimulated groups. **h**, Mass-spectrometric analysis of the acetylation status of nuclear location sequences (NLS) of HMGB1.

$Pkr^{-/-}$ macrophages as compared with $Pkr^{+/+}$ macrophages (Fig. 2b). Tumour necrosis factor (TNF-α) release in the $Pkr^{+/+}$ and $Pkr^{-/-}$ macrophages was comparable, indicating that the decreased release of IL-1β in the $Pkr^{-/-}$ macrophages is not attributable to a global defect in signal transduction (Supplementary Fig. 10).

Transfection of BMDCs with poly(I:C) and *E. coli* RNA significantly activated caspase-1 and stimulated IL-1β cleavage in $Pkr^{+/+}$ but not $Pkr^{-/-}$ cells (Supplementary Fig. 11). Similar observations were obtained in $Pkr^{-/-}$ and $Pkr^{+/+}$ macrophages stimulated by rotenone, which induces mitochondrial reactive oxygen species production and PKR phosphorylation (Supplementary Figs 12 and 13). Pharmacological inhibition of PKR dose-dependently suppressed MSU-induced caspase-1 activation and IL-1β cleavage. The observed half-maximum inhibitory concentrations (IC_{50} values) of 2-aminopurine (2-AP) (Fig. 2c) and $C_{13}H_8N_4OS$ (Supplementary Fig. 14) were 0.5 mM and 0.25 μM, respectively, which agree closely with their known IC_{50} against PKR. PKR inhibition significantly reduced ATP- and ALU-induced inflammasome activation in mouse macrophages (Supplementary Figs 15 and 16), and in human monocytic THP-1 cells (Supplementary Fig. 17). IL-18 release was significantly lower in $Pkr^{-/-}$ macrophages than in $Pkr^{+/+}$ macrophages stimulated with ATP, MSU or ALU, whereas TNF-α and IL-6 were not suppressed (Supplementary Fig. 18). The addition of 2-AP reduced MSU-induced IL-18 release, but not TNF-α and IL-6 (Supplementary Fig. 18). 2-AP failed to inhibit MSU-induced caspase-1 activation and IL-1β cleavage in $Pkr^{-/-}$ macrophages (Supplementary Fig. 19). Collectively, these findings establish a crucial role for PKR in activating the NLRP3 inflammasome.

$Pkr^{+/+}$ and $Pkr^{-/-}$ mice were then exposed to live *E. coli*, to activate the NLRP3 inflammasome *in vivo*[17]. Serum levels of IL-1β, IL-18 and HMGB1 were significantly reduced in $Pkr^{-/-}$ mice as compared with $Pkr^{+/+}$ mice (Fig. 2d). In agreement with previous evidence suggesting that peritoneal neutrophil infiltration requires inflammasome activation[18], we observed significantly fewer neutrophils in peritoneal lavage from $Pkr^{-/-}$ mice than in $Pkr^{+/+}$ controls (Fig. 2e). Serum IL-6, an inflammasome-independent cytokine, was comparable in both groups (Fig. 2d). Exposure of $Pkr^{-/-}$ mice to endotoxemia produced significantly lower serum IL-1β and IL-18 levels, but quantitatively similar serum levels of TNF-α and IL-6 as compared with $Pkr^{+/+}$ controls (Supplementary Fig. 20). Thus, PKR regulates inflammasome-dependent cytokine release *in vivo*.

To study the mechanisms of PKR regulation, HEK293A cells, which do not normally express the NLRP3 components, were transfected with plasmids that co-express ASC, pro-caspase-1 and NLRP3. These modified cells were then co-transfected with plasmids that express PKR, or specific short hairpin RNA (shRNA) to suppress endogenous PKR. Overexpression of PKR significantly enhanced caspase-1 activation and IL-1β cleavage, and knockdown of endogenous PKR by shRNA abolished caspase-1 cleavage in the NLRP3 inflammasome-reconstituted cells (Fig. 2f, g). IL-1β cleavage was significantly abrogated by pharmacological inhibition of PKR (Supplementary Fig. 21). PKR overexpression failed to activate caspase-1 directly in the absence of NLRP3 (Supplementary Fig. 22), and did not enhance HMGB1 release in the absence of caspase-1 (Supplementary Fig. 23). Knockdown of endogenous eIF2α, a PKR substrate[13,14], did not

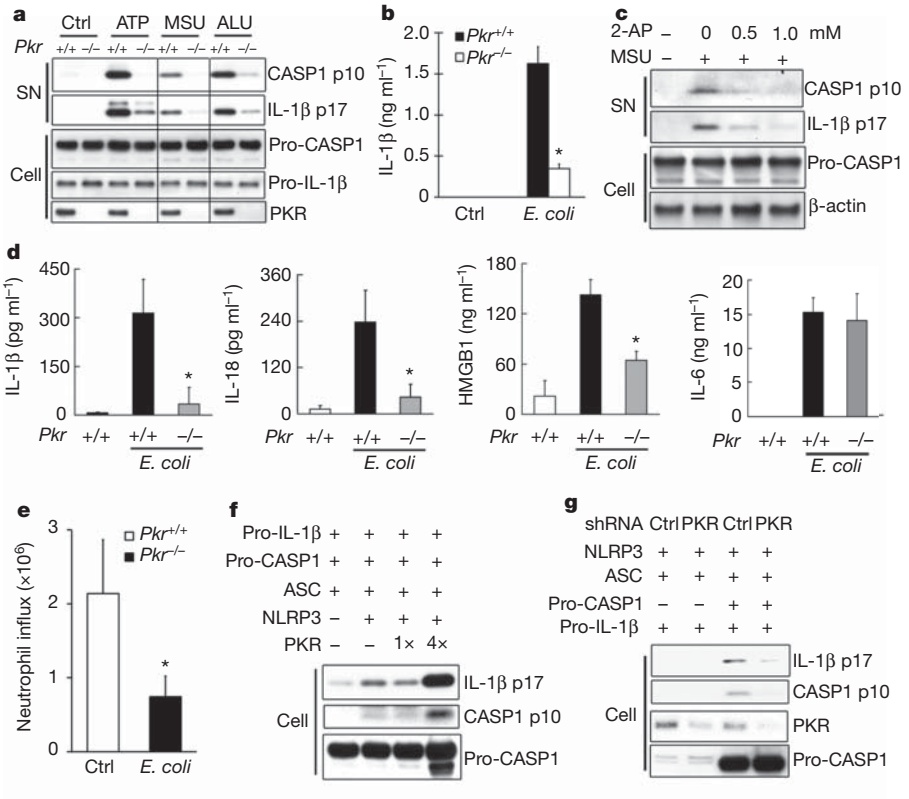


Figure 2 | PKR is important for inflammasome activation. **a**, **b**, LPS-primed $Pkr^{+/+}$ or $Pkr^{-/-}$ macrophages were stimulated as indicated. SN, supernatant. **c**, $Pkr^{+/+}$ macrophages were stimulated or treated with 2-AP as indicated. **d**, **e**, $Pkr^{+/+}$ or $Pkr^{-/-}$ mice ($n = 5$) were injected with live *E. coli*. **f**, **g**, HEK293A cells were transfected as indicated. Caspase-1 (CASP1) activation and IL-1β cleavage (as measured by the p10 and p17 subunits, respectively) were assessed by western blot. Data are representative of at least three independent experiments. Levels of IL-1β, IL-18, HMGB1 and IL-6 in the supernatant (**b**) or serum (**d**) were determined by ELISA. Peritoneal lavage fluid was collected and neutrophil content was measured by flow cytometry (**e**). Data are mean and s.d. $*P < 0.05$ versus wild-type-infected groups.

alter HMGB1 release and IL-1β cleavage in the NLRP3 inflammasome-reconstituted cells (Supplementary Fig. 24).

We therefore reasoned that PKR physically interacts with the NLRP3 inflammasome, and prepared mouse macrophage lysates for immunoprecipitation using anti-PKR antibodies. PKR pulled down NLRP3, and reciprocal immunoprecipitation of NLRP3 pulled down PKR (Supplementary Fig. 25). Similar observations were obtained in human THP-1 cells (Supplementary Fig. 26). PKR–NLRP3 complexes could not be detected in the $Pkr^{-/-}$ macrophages (Supplementary Fig. 27). Co-expression of PKR and NLRP3 in HEK293A cells, followed by immunoprecipitation, confirmed the physical interaction between PKR and NLRP3. PKR failed to associate physically with the other cytosolic receptors or inflammasome family members NOD1, NLRP12 or NLRX1 (Supplementary Fig. 28). PKR co-immunoprecipitated with truncated forms of NLRP3 containing the pyrin domain (PYD), NACHT domain and leucine-rich repeat (LRR) domain (Supplementary Fig. 29), and with the Walker A mutant form of NLRP3 (Supplementary Fig. 30), which cannot bind ATP[19]. Together with direct evidence that recombinant PKR forms a heteromeric protein complex with NLRP3 (Fig. 3a), these results indicate that PKR physically interacts with the inflammasome.

Exposure of $Pkr^{+/+}$ macrophages to ATP significantly enhanced the formation of NLRP3 and PKR complexes (Fig. 3b). Pharmacological inhibition of PKR phosphorylation by 2-AP and $C_{13}H_8N_4OS$ significantly suppressed the formation of NLRP3–PKR complexes (Fig. 3b). The addition of poly(I:C) and ATP to NLRP3 and PKR in a cell-free system significantly induced PKR phosphorylation and increased the formation of NLRP3–PKR complexes (Fig. 3a).

Next, the NLRP3 inflammasome was reconstituted in a cell-free system with recombinant NLRP3, ASC, pro-caspase-1 and PKR. Combining NLRP3 and ASC with pro-caspase-1 failed to increase caspase-1 activity (Fig. 3c). The addition of PKR and poly(I:C) or ATP significantly increased caspase-1 activity, which was inhibited by the addition of 2-AP (Fig. 3c). PKR activity is required in this cell-free system, because poly(I:C)/ATP failed to increase NLRP3 inflammasome activity in the absence of PKR. Moreover, replacing PKR with a recombinant, mutant PKR carrying the Lys296Arg mutation and thus devoid of kinase activity[13,16] failed to activate the NLRP3 inflammasome (Fig. 3c). Although PKR is an autophosphorylating kinase, it was theoretically possible that PKR might phosphorylate inflammasome components directly. Phosphorylation of NLRP3, ASC and caspase-1 could not be detected, although poly(I:C)/ATP significantly stimulated autophosphorylation of PKR (data not shown). To confirm that PKR autophosphorylation is involved in NLRP3 inflammasome activation, the PKR(K296R) mutant and NLRP3 inflammasome components were co-expressed in HEK293A cells. The PKR(K296R) mutant failed to bind NLRP3, whereas wild-type PKR did bind NLRP3 (Supplementary Fig. 31). Overexpression of the PKR(K296R) mutant failed to activate the reconstituted NLRP3 inflammasome in HEK293A cells (Supplementary Fig. 31). Together, these data indicate that PKR physically interacts with NLRP3 and is crucial for inflammasome activation.

Reasoning that PKR might associate with the inflammasome components during assembly, ATP-stimulated macrophage lysates were subjected to gel filtration chromatography. We observed that PKR and NLRP3 eluted together with ASC and caspase-1 in the high molecular mass fraction of the ATP-stimulated macrophage extract (Fig. 3d and Supplementary Fig. 32). Inhibition of PKR activation by 2-AP reduced PKR in the high molecular mass inflammasome fraction (Fig. 3d and Supplementary Fig. 33). Expression of the PKR(K296R) mutant

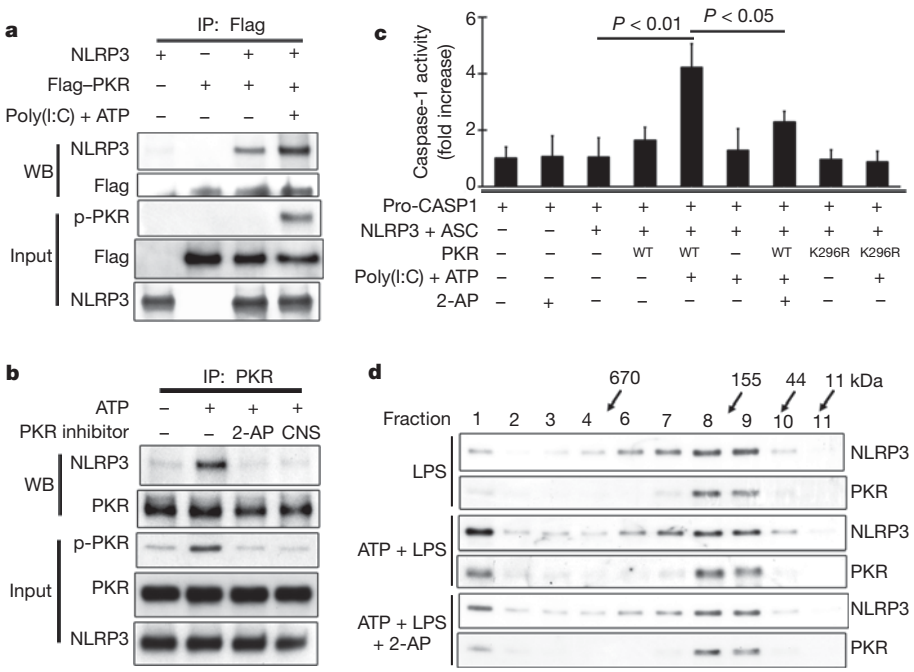


Figure 3 | PKR physically interacts with NLRP3 and facilitates inflammasome activation. **a**, **b**, Immunoprecipitation (IP) and western blot (WB) analysis of the physical interaction of PKR and NLRP3 in a cell-free system using recombinant proteins (**a**) or LPS-primed macrophages stimulated with ATP or treated with 2-AP or $C_{13}H_8N_4OS$ (CNS) as indicated (**b**). **c**, The NLRP3 inflammasome was reconstituted using recombinant proteins and ATP and poly(I:C) as indicated. Caspase-1 activity was measured by hydrolysis of the synthetic peptide WEHD-pNA. **d**, $Pkr^{+/+}$ macrophages were stimulated or treated with 2-AP as indicated. Cell lysates were subjected to gel-filtration chromatography and western blot. Results are representative of three independent experiments.

significantly reduced NLRP3 in the high molecular mass inflammasome fraction, without significantly changing total NLRP3 expression (Supplementary Fig. 34). Rendering cells deficient in PKR also significantly impaired the formation of high molecular mass protein NLRP3 inflammasome complexes, without altering total NLRP3 protein expression levels (Supplementary Fig. 35). Activation of macrophages by exposure to ATP, MSU, ALU or *E. coli* significantly shifted caspase-1 from the low molecular mass, monomeric fraction, to the high molecular mass protein complexes in $Pkr^{+/+}$ macrophages, but not in $Pkr^{-/-}$ macrophages (Supplementary Fig. 35). Thus, activated PKR is an integral component of the assembled inflammasome.

Unexpectedly, PKR also physically interacted with NLRP1, AIM2 and NLRC4 when co-expressed in HEK293A cells (Fig. 4a). To address whether PKR might also regulate NLRP1, AIM2 and NLRC4 inflammasomes, macrophages were stimulated with anthrax lethal toxin, transfected with poly(dA-dT)•poly(dA-dT) (hereafter termed poly(dA:dT)) or infected with *S. typhimurium* to activate the NLRP1, AIM2 or NLRC4 inflammasomes, respectively. This significantly enhanced PKR autophosphorylation (Supplementary Fig. 36). Inactivation of PKR by genetic deletion (Fig. 4b–d and Supplementary Figs 37 and 38) or by pharmacological inhibition (Fig. 4e and Supplementary Fig. 39) significantly inhibited caspase-1 activation, IL-1β cleavage and HMGB1 secretion. LDH release and IL-18 production were also significantly decreased in $Pkr^{-/-}$ macrophages compared with $Pkr^{+/+}$ macrophages, whereas TNF-α production was not suppressed in these conditions (Supplementary Figs 40 and 41). Overexpression of PKR significantly enhanced NLRP1, AIM2 and NLRC4 inflammasome-induced caspase-1 activation and IL-1β cleavage in HEK293A cells (Fig. 4f and Supplementary Fig. 42).

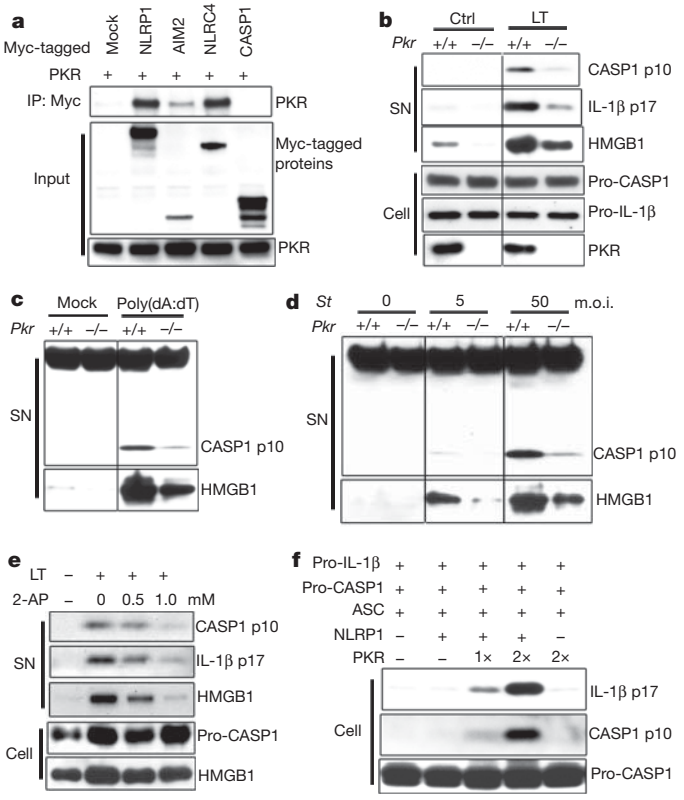


Figure 4 | PKR regulates NLRP1, AIM2 and NLRC4 inflammasome activation. **a**, HEK293A cells were transfected as indicated. The physical interaction of PKR and Myc-tagged proteins was analysed by immunoprecipitation and western blot analysis. **b**–**d**, LPS-primed (**b**) or unprimed (**c**, **d**) $Pkr^{+/+}$ or $Pkr^{-/-}$ macrophages were stimulated with anthrax lethal toxin (LT), transfected with poly(dA:dT), or infected with *S. typhimurium* (*St*). m.o.i., multiplicity of infection. **e**, $Pkr^{+/+}$ macrophages were stimulated or treated with 2-AP as indicated. **f**, HEK293A cells were transfected as indicated. Caspase-1 activation, IL-1β cleavage and HMGB1 secretion were assessed by western blot. Results are representative of at least three independent experiments.

Thus, PKR regulates the activation of the NLRP3, NLRP1, AIM2 and NLRC4 inflammasomes, but its activity does not determine which inflammasome will be activated in response to a specific stimulus.

Previous observations suggested that macrophage death by pyroptosis might initiate deleterious inflammatory responses and compromise immunity during bacterial infection[5,15], and that PKR can participate in bacteria-induced macrophage death[15]. To assess whether PKR deficiency impairs bacterial clearance, $Pkr^{+/+}$ and $Pkr^{-/-}$ mice were exposed to live *E. coli* (bearing a genetic marker) into the peritoneal cavity. Titres of the genetically labelled bacteria in the spleen and peritoneal cavity 24 h later were significantly lower in $Pkr^{-/-}$ mice than in $Pkr^{+/+}$ controls (Supplementary Fig. 43), indicating that the animals maintain the capacity to clear bacteria despite PKR deficiency.

This study establishes that PKR activity is integral to inflammasome assembly and activation. PKR physically interacts with inflammasome components, is important for caspase-1 activation, IL-1β cleavage and HMGB1 release, and mediates inflammasome activity in a cell-free system. Although originally studied as an intracellular sensor of viral dsRNA, a broader role of PKR was recently suggested as a danger-sensing molecule activated by cellular and metabolic stress[15,16,20]. Together with the established role of the inflammasome and HMGB1 in sterile and infection-associated inflammation[3,6,21–23], the identification here of PKR as a crucial regulator of inflammasome activity adds to our understanding of the mechanisms of innate immune responses to environmental, metabolic and invasive signals.

Rapidly growing evidence links the inflammasome to the pathogenesis of disease syndromes such as obesity, type 2 diabetes, atherosclerosis, gout, sepsis and colitis[4,5,10,18,24,25]. The development of therapeutics targeting the inflammasome requires knowledge of specific upstream signalling pathways that regulate inflammasome assembly and activation[25,26]. It should now be possible to design and develop therapeutics to inhibit inflammasome activity by specifically targeting PKR, without broadly impairing immunity, in the treatment of non-resolving inflammatory and meta-inflammatory syndromes.

METHODS SUMMARY

Cell stimulation. Peritoneal macrophages, BMDMs or BMDCs from $Pkr^{-/-}$ and wild-type $Pkr^{+/+}$ mice were primed with ultra-pure LPS (500 ng ml^{-1}), and then stimulated with inflammasome agonists. PKR inhibitors were added before exposure to inflammasome stimuli, where applicable. HMGB1 release, caspase-1 activation and IL-1β maturation were assessed by western blot.

Reconstituted inflammasome in HEK293A cells. Cells were transfected with plasmids expressing pro-IL-1β and inflammasome components, such as ASC, pro-caspase-1, NLRP3 and NLRP1. These modified cells were then co-transfected with plasmids that express PKR, or specific shRNA to knockdown the endogenous PKR. After 24 h, cell lysates were collected and analysed for IL-1β maturation and caspase-1 activation by western blot.

Reconstituted inflammasome in a cell-free system. Recombinant NLRP3, ASC and pro-caspase-1 were incubated with recombinant PKR and with ATP (2.5 mM) or poly(I:C) (1 ng ml^{-1}) in the PKR reaction buffer at 37 °C for 1 h. Caspase-1 activity was measured by hydrolysis of the synthetic peptide WEHD-pNA.

***E. coli*-induced peritonitis.** $Pkr^{-/-}$ and $Pkr^{+/+}$ mice were injected intraperitoneally with live *E. coli* (2×10^9 per mouse). After 6 h, blood was collected and the serum levels of IL-1β, IL-18, HMGB1 and IL-6 were measured by ELISA assay.

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

Received 10 November 2011; accepted 8 June 2012.
Published online 8 July 2012.

1. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140,** 821–832 (2010).
2. Ogura, Y., Sutterwala, F. S. & Flavell, R. A. The inflammasome: first line of the immune response to cell stress. *Cell* **126,** 659–662 (2006).
3. Lamkanfi, M. & Dixit, V. M. Modulation of inflammasome pathways by bacterial and viral pathogens. *J. Immunol.* **187,** 597–602 (2011).
4. Lamkanfi, M. *et al.* Inflammasome-dependent release of the alarmin HMGB1 in endotoxemia. *J. Immunol.* **185,** 4385–4392 (2010).
5. Kayagaki, N. *et al.* Non-canonical inflammasome activation targets caspase-11. *Nature* **479,** 117–121 (2011).
6. Andersson, U. & Tracey, K. J. HMGB1 is a therapeutic target for sterile inflammation and infection. *Annu. Rev. Immunol.* **29,** 139–162 (2011).
7. Yang, H. *et al.* A critical cysteine is required for HMGB1 binding to Toll-like receptor 4 and activation of macrophage cytokine release. *Proc. Natl Acad. Sci. USA* **107,** 11942–11947 (2010).
8. Manfredi, A. A. *et al.* Maturing dendritic cells depend on RAGE for *in vivo* homing to lymph nodes. *J. Immunol.* **180,** 2270–2275 (2008).
9. Willingham, S. B. *et al.* NLRP3 (NALP3, Cryopyrin) facilitates *in vivo* caspase-1 activation, necrosis, and HMGB1 release via inflammasome-dependent and -independent pathways. *J. Immunol.* **183,** 2008–2015 (2009).
10. Qin, S. *et al.* Role of HMGB1 in apoptosis-mediated sepsis lethality. *J. Exp. Med.* **203,** 1637–1642 (2006).
11. Jiang, W., Bell, C. W. & Pisetsky, D. S. The relationship between apoptosis and high-mobility group protein 1 release from murine macrophages stimulated with lipopolysaccharide or polyinosinic-polycytidylic acid. *J. Immunol.* **178,** 6495–6503 (2007).
12. Yanai, H. *et al.* HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* **462,** 99–103 (2009).
13. Dey, M. *et al.* Mechanistic link between PKR dimerization, autophosphorylation, and eIF2α substrate recognition. *Cell* **122,** 901–913 (2005).
14. Dar, A. C., Dever, T. E. & Sicheri, F. Higher-order substrate recognition of eIF2α by the RNA-dependent protein kinase PKR. *Cell* **122,** 887–900 (2005).
15. Hsu, L. C. *et al.* The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4. *Nature* **428,** 341–345 (2004).
16. Nakamura, T. *et al.* Double-stranded RNA-dependent protein kinase links pathogen sensing with stress and metabolic homeostasis. *Cell* **140,** 338–348 (2010).
17. Sander, L. E. *et al.* Detection of prokaryotic mRNA signifies microbial viability and promotes immunity. *Nature* **474,** 385–389 (2011).
18. Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440,** 237–241 (2006).
19. Duncan, J. A. *et al.* Cryopyrin/NALP3 binds ATP/dATP, is an ATPase, and requires ATP binding to mediate inflammatory signaling. *Proc. Natl Acad. Sci. USA* **104,** 8041–8046 (2007).
20. Bennett, R. L. *et al.* RAX, the PKR activator, sensitizes cells to inflammatory cytokines, serum withdrawal, chemotherapy, and viral infection. *Blood* **108,** 821–829 (2006).
21. Rathinam, V. A., Vanaja, S. K. & Fitzgerald, K. A. Regulation of inflammasome signaling. *Nature Immunol.* **13,** 333–342 (2012).
22. Franchi, L., Muñoz-Planillo, R. & Núñez, G. Sensing and reacting to microbes through the inflammasomes. *Nature Immunol.* **13,** 325–332 (2012).
23. Schattgen, S. A. & Fitzgerald, K. A. The PYHIN protein family as mediators of host defenses. *Immunol. Rev.* **243,** 109–118 (2011).
24. Wen, H., Ting, J. P. & O'Neill, L. A. A role for the NLRP3 inflammasome in metabolic diseases—did Warburg miss inflammation? *Nature Immunol.* **13,** 352–357 (2012).
25. Strowig, T., Henao-Mejia, J., Elinav, E. & Flavell, R. Inflammasomes in health and disease. *Nature* **481,** 278–286 (2012).
26. Bauernfeind, F. *et al.* Cutting edge: reactive oxygen species inhibitors block priming, but not activation, of the NLRP3 inflammasome. *J. Immunol.* **187,** 613–617 (2011).

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

Acknowledgements We thank H. LaQueta, M. Dancho, M. McCarty, E. Lau, D. Katz and J. Scheinerman for technical assistance. This work was supported in part by grants from the National Institutes of Health (RO1 GM62508 to K.J.T. and DK052539 to G.S.H.). B.L. and S.I.V.-F. are supported by the foundation of Elmezzi Graduate School of Molecular Medicine. T.N. is supported by fellowships from the International Human Frontier Science Program and a Career Development Award from the American Heart Association.

Author Contributions B.L. and K.J.T. designed the research; B.L., T.N., K.I., D.J.A., J.L., Y.T., P.L., S.I.V.-F. and H.E.-H. performed the experiments; B.L., S.I.V.-F., P.S.O., H.Y., S.S.C., J.R., T.K., D.J.A., G.S.H., U.A. and K.J.T. analysed the results; J.P.-Y.T. and Y.Z. provided important reagents; B.L. made the figures; B.L. and K.J.T. wrote the paper; U.A., H.W., P.S.O., S.S.C. and G.S.H. edited and commented on the manuscripts.

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 K.J.T. (kjtracey@nshs.edu) or B.L. (blu@nshs.edu).

METHODS

Mice. The $Pkr^{-/-}$ mice and their littermate wild-type controls are originally in the mixed 129Sv/BALB/C background as described previously[16]. In *E. coli* infection and endotoxemia experiments, we used the mice that had been bred into the C57Bl/6J genetic background. Genome-wide single nucleotide polymorphism analysis confirmed that about 98% of the genetic background of these mice is of C57Bl/6J. Animal care and experimental procedures were performed with approval from the Institutional Animal Care and Use Committees of the Feinstein Institute for Medical Research and Harvard University. $Pkr^{-/-}$ and $Pkr^{+/+}$ mice are genotyped with PCR primers ($Pkr^{-/-}$, forward, 5′-GGAA CTTTGGAGCAATGGA-3′, reverse, 5′-TGCCAATCACAAATCTAAAAC-3′; $Pkr^{+/+}$, forward, 5′-TGTTCTGTGGCTATCAGGG-3′, reverse, 5′-TGACGAGT TCTTCTGAGGG-3′) yielding a 240-base-pair (bp) wild-type DNA band and a 460-bp mutant DNA band after electrophoresis in 2% agarose gel in TAE buffer.

Reagents. Ultra-pure LPS and the NLRP3 inflammasome agonists ATP, MSU and ALU were obtained from Invivogen. The anti-NLRP3 (cryo-2) antibodies were from Adipogen. The rabbit monoclonal anti-PKR (phospho T446) antibodies (E120) were from Abcam. Mouse HMGB1 monoclonal antibody IgG2b 2G7 was described previously[7,10]. *eif2a* short interfering RNA (siRNA) molecules and their negative controls were obtained from Life Technologies. Recombinant NLRP3 protein was obtained from Abnova. Recombinant Flag-tagged PKR protein was purified from $Pkr^{-/-}$ mouse embryonic fibroblast cells stably expressing Flag-tagged PKR. Recombinant Flag-tagged ASC and pro-caspase-1 protein were purified from HEK293A cells overexpressing Flag-tagged ASC and pro-caspase-1 (Origene). Anti-human cleaved IL-1β (D116) was from Cell Signaling. Anti-mouse IL-1β (sc-1251), anti-mouse caspase-1 (sc-514) and anti-human caspase-1 (sc-622) antibodies were from Santa Cruz.

Cell preparation and stimulation. Peritoneal macrophages were isolated and cultured as described previously[7]. Primary BMDMs and BMDCs were cultured and differentiated *in vitro* as described previously[18]. Human THP-1 cells were differentiated for 3 h with 100 nM phorbol-12-myristate-13-acetate (PMA). One million peritoneal macrophages, BMDMs, BMDCs or PMA-differentiated THP-1 cells plated in 12-well plates were primed with ultra-pure LPS (500 ng ml^{-1}) for 4 h, and then stimulated with MSU (150 μg ml^{-1}) or ALU (200 μg ml^{-1}) for 5 h, or live *E. coli* (m.o.i. = 20) for 6 h (bacteria growth is stopped by adding antibiotics into culture medium 1 h after infection), or pulsed with ATP (5 mM) for 30 min or 1 h as indicated. Ultra-pure LPS (1 ng ml^{-1})-primed mouse peritoneal macrophages were stimulated with poly(I:C) (50 μg ml^{-1}) for 16 h. For *Salmonella* infection, wild-type *S. typhimurium* was grown overnight in Luria–Bertani (LB) broth, then reinoculated at a dilution of 1:100 and grown to mid-exponential phase (3 h) to induce expression of the *Salmonella* pathogenicity island 1 type III secretion system. To minimize the involvement of NLRP3 inflammasome activation during *Salmonella* infection, unprimed $Pkr^{+/+}$ and $Pkr^{-/-}$ macrophages were infected with wild-type *S. typhimurium* (m.o.i. is from 5 to 100). The supernatant samples were collected 1 h after infection. To study AIM2 inflammasome activation, $Pkr^{+/+}$ and $Pkr^{-/-}$ macrophages were transfected with poly(dA:dT) using Lipofectamine 2000 at a concentration of 1 μg DNA plus 3.5 μl lipofectamine 2000 per ml. The supernatant samples were collected 6 h after transfection. Caspase-1 activation, HMGB1 and LDH release were analysed. To study the DNA-transfection-induced IL-1β cleavage, LPS-primed $Pkr^{+/+}$ and $Pkr^{-/-}$ macrophages were transfected with poly(dA:dT)/lyovec (Invivogen) at 5 μg ml^{-1}. There is minimum cell death with this approach. Cell lysates and precipitated supernatants were analysed by western blot analysis.

Pharmacological inhibition of PKR. The PKR inhibitor 2-AP was purchased from Sigma and dissolved in PBS:glacial acetic acid (200:1) at 60 °C for 60 min. The PKR inhibitor $C_{13}H_8N_4OS$ was purchased from Calbiochem. Ultra-pure LPS-primed peritoneal mouse macrophages, BMDMs, BMDCs or PMA-differentiated THP-1 cells were pre-treated with 2-AP (0.5 or 1.0 mM) or $C_{13}H_8N_4OS$ (0.25 or 0.5 μM) for 1–4 h before stimulating with various inflammasome activators. Higher concentrations of the above PKR inhibitors are not recommended owing to off-target effects. Cell lysates and precipitated supernatants were analysed by western blot.

ELISA. Cell culture supernatants and mouse serum samples were assayed for mouse IL-1β, IL-6, TNF-α, IL-18 (R&D Systems) and HMGB1 (IBL International) according to the manufacturer's instructions.

Transfection. One million BMDCs from $Pkr^{+/+}$ or $Pkr^{-/-}$ mice were primed with ultra-pure LPS for 2 h, and then transfected with 5 μg of poly(I:C) by lipofectamine 2000 for 6 h. In another set of experiment, ultra-pure LPS-primed BMDCs (10^6 cells) from $Pkr^{+/+}$ or $Pkr^{-/-}$ mice were transfected with 5 μg of poly(I:C) or *E. coli* RNA (Ambion) by electroporation using Nucleofector (Program V-01) and cell line Nucleofector kit V (Amaxa AG). After 3 h, cell extracts and precipitated supernatants were obtained and analysed by western blot analysis.

Reconstituted inflammasome in HEK293A cells. For the human NLRP3 inflammasome reconstitution assays, 0.2×10^6 HEK293A cells were seeded onto 24-well plates in cell culture media lacking antibiotics. After 24 h, cells were transfected with plasmids expressing pro-IL-1β and inflammasome components (Origene), such as ASC, pro-caspase-1, NLRP3 or NLRP1, using lipofectamine 2000. These modified cells were then co-transfected with plasmids that express PKR (Invivogen) or specific shRNA (Invivogen) to knockdown the endogenous PKR. The total amount of DNA was standardized to 600 ng per well using a control plasmid. Twenty-four hours later, cell lysates were collected and analysed for IL-1β maturation and caspase-1 cleavage by western blot analysis.

Immunoprecipitation. Antibodies against the PKR, NLRP3, Flag or Myc were used to precipitate proteins from cell lysis in the presence of 20 μl protein A/G beads (Santa Cruz) overnight at 4 °C. Protein complexes were washed four times with lysis buffer, and then incubated at 95 °C for 5 min and resolved by western blot analysis.

Western blot. Proteins from cell-free supernatants were extracted by methanol/chloroform precipitation as described previously[18]. Cell extracts were prepared as described previously[7]. Samples were separated by 4–20% SDS–PAGE or 4–20% native-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes. The relative levels of HMGB1 in the culture medium were determined by western blot analysis using recombinant HMGB1 protein to make a standard curve.

PKR autophosphorylation assay. PKR phosphorylation were assessed by PKR autophosphorylation assay as described previously[15,16].

Gel filtration chromatography. As described previously[19], fresh soluble lysates from mouse macrophages were prepared by hypotonic lysis and shearing through a 27-gauge needle. Soluble lysate (0.6 mg of total protein) was run on a Biosil 400 gel filtration column using a Bio-Rad Duoflow chromatography system in PBS with 2 mM dithiothreitol (DTT).

Reconstituted inflammasome in a cell-free system. Recombinant NLRP3, ASC and pro-caspase-1 were incubated with recombinant PKR and ATP (2.5 mM) or poly(I:C) (1 ng ml^{-1}) in PKR reaction buffer (20 mM HEPES, pH 7.4, 10 mM $MgCl_2$, 40 mM KCl and 2 mM DTT) at 37 °C for 1 h. Caspase-1 activity was measured by hydrolysis of WEHD-pNA.

***E. coli*-induced peritonitis and endotoxemia model.** $Pkr^{-/-}$ and $Pkr^{+/+}$ mice were injected intraperitoneally with live *E. coli* (2×10^9 per mouse). Six hours later, blood was collected and serum levels of IL-1β, IL-18, HMGB1 and IL-6 were measured by ELISA. Meanwhile, peritoneal cavities were washed with 10 ml of PBS. The lavage fluids were analysed for neutrophil influx by flow cytometry using the neutrophil marker GR-1. For bacterial titre experiments, $Pkr^{+/+}$ and $Pkr^{-/-}$ mice received live 2×10^9 *E. coli* into the peritoneal cavity. Bacterial titres in both the peritoneal cavity and the spleen were measured 24 h after *E. coli* challenge. To prevent other bacterial contamination during sample collection process, the *E. coli* had been transformed with an ampicillin-resistant gene. Bacterial titres were determined using ampicillin-containing LB agar plates. For endotoxemia experiments, $Pkr^{-/-}$ and $Pkr^{+/+}$ mice received 20 mg kg^{-1} of LPS and blood was collected 2.5 h later. Serum IL-1β, IL-18, TNF-α and IL-6 levels were determined by ELISA.

Analysis of HMGB1 by tamdem mass spectrometry. Detailed methods are available in Supplementary Figures and Methods.

Statistical analysis. A Student's *t*-test and one-way analysis of variance (ANOVA) were used for comparison among all different groups. $P < 0.05$ was considered statistically significant.

LETTER

doi:10.1038/nature11337

T cells become licensed in the lung to enter the central nervous system

Francesca Odoardi[1], Christopher Sie[1], Kristina Streyl[2], Vijay K. Ulaganathan[2], Christian Schläger[1], Dmitri Lodygin[1], Klaus Heckelsmiller[2], Wilfried Nietfeld[3], Joachim Ellwart[4], Wolfgang E. F. Klinkert[2], Claudio Lottaz[5], Mikhail Nosov[6], Volker Brinkmann[7], Rainer Spang[5], Hans Lehrach[3], Martin Vingron[3], Hartmut Wekerle[2], Cassandra Flügel-Koch[8] & Alexander Flügel[1]

The blood–brain barrier (BBB) and the environment of the central nervous system (CNS) guard the nervous tissue from peripheral immune cells. In the autoimmune disease multiple sclerosis, myelin-reactive T-cell blasts are thought to transgress the BBB[1,2] and create a pro-inflammatory environment in the CNS, thereby making possible a second autoimmune attack that starts from the leptomeningeal vessels and progresses into the parenchyma[3–6]. Using a Lewis rat model of experimental autoimmune encephalomyelitis, we show here that contrary to the expectations of this concept, T-cell blasts do not efficiently enter the CNS and are not required to prepare the BBB for immune-cell recruitment. Instead, intravenously transferred T-cell blasts gain the capacity to enter the CNS after residing transiently within the lung tissues. Inside the lung tissues, they move along and within the airways to bronchus-associated lymphoid tissues and lung-draining mediastinal lymph nodes before they enter the blood circulation from where they reach the CNS. Effector T cells transferred directly into the airways showed a similar migratory pattern and retained their full pathogenicity. On their way the T cells fundamentally reprogrammed their gene-expression profile, characterized by downregulation of their activation program and upregulation of cellular locomotion molecules together with chemokine and adhesion receptors. The adhesion receptors include ninjurin 1, which participates in T-cell intravascular crawling on cerebral blood vessels. We detected that the lung constitutes a niche not only for activated T cells but also for resting myelin-reactive memory T cells. After local stimulation in the lung, these cells strongly proliferate and, after assuming migratory properties, enter the CNS and induce paralytic disease. The lung could therefore contribute to the activation of potentially autoaggressive T cells and their transition to a migratory mode as a prerequisite to entering their target tissues and inducing autoimmune disease.

We monitored the invasion of the CNS by green-fluorescent-protein (GFP)-expressing myelin basic protein (MBP)-reactive T cells ($T_{MBP–GFP}$ cells)[7] over the preclinical phase of Lewis rat transfer experimental autoimmune encephalomyelitis (EAE). Only very few early invading pioneer T cells[1,2] were detected in the CNS in the first 60 h after intravenous transfer of activated $T_{MBP–GFP}$ cells (<100 T cells per total spinal cord). The bulk of $T_{MBP–GFP}$ cells (>10^6 T cells per total spinal cord) started to invade the CNS at the leptomeningeal vessels of the spinal cord 60–84 h after T-cell transfer, shortly before the onset of clinical symptoms (Fig. 1a and Supplementary Fig. 1a, b)[3,4,8]. Pro-inflammatory cytokines and adhesion molecules in the CNS became detectable only during the mass T-cell invasion (Supplementary Fig. 1c)[3,9].

The $T_{MBP–GFP}$ cells nearly all disappeared from the blood almost immediately after intravenous infusion. It was not until 48 h later that a gradual increase of $T_{MBP–GFP}$ cells was observed in the blood and spleen, closely followed by the start of the main invasion of the T cells into the CNS (Fig. 1b). This coincidence of the reappearance of the $T_{MBP–GFP}$ cells in blood and spleen and their invasion into the CNS led us to test whether the re-emerged cells (termed $T_{migratory}$ cells) would enter the nervous tissue more efficiently than the initially transferred $T_{MBP–GFP}$ cell blasts (T_{blast} cells). To do this, we prepared an anastomosis between the blood circulations of two animals, one of

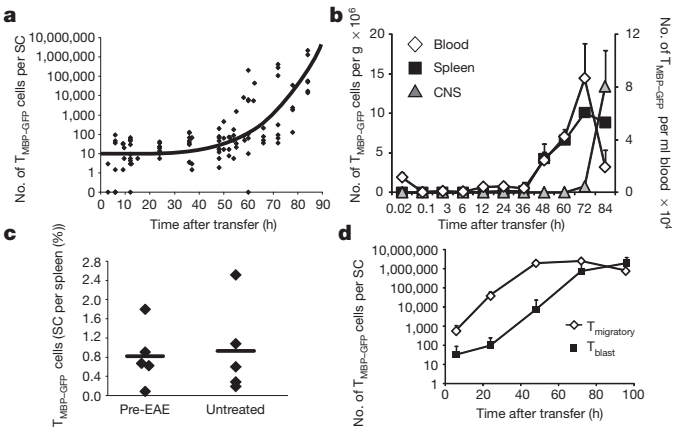


Figure 1 | $T_{migratory}$ cells but not T_{blast} cells home efficiently into the CNS. **a**, Invasion of $T_{MBP–GFP}$ cells into the CNS during transfer EAE. $T_{MBP–GFP}$ cells were quantified in the total spinal cord (SC) at the indicated time points after T-cell transfer by cytofluorometry. Each diamond represents a single animal. **b**, Transferred $T_{MBP–GFP}$ cells reappear in the blood and spleen shortly before CNS invasion. Cytofluorometric quantification of $T_{MBP–GFP}$ cells. Means ± 1 s.d. of 4 animals are shown. Representative data of eight independent experiments, including at least three animals per time point. **c**, Blood-derived $T_{migratory}$ cells of pre-EAE animals enter the CNS of naïve rats. A blood connection was established linking the tail arteries of an untreated animal with an animal that had received T_{blast} cells 48 h beforehand (pre-EAE animal). Twelve hours later the numbers of $T_{MBP–GFP}$ cells were determined by cytofluorometry. Shown are relative values of $T_{MBP–GFP}$ cells in the CNS versus spleens of five pairs of animals. **d**, Spleen-derived $T_{migratory}$ cells efficiently infiltrate the CNS. T_{blast} cells or $T_{migratory}$ cells (isolated from spleen 60 h after transfer) were transferred in naïve animals. Numbers of $T_{MBP–GFP}$ cells were quantified by cytofluorometry in total spinal cord at the indicated time points after T-cell transfer. Means ± 1 s.d. of triplicate measurements. Representative data from at least six independent experiments are shown.

[1]Institute for Multiple Sclerosis Research, Department of Neuroimmunology, Gemeinnützige Hertie-Stiftung and University Medical Centre Göttingen, 37073 Göttingen, Germany. [2]Max Planck Institute for Neurobiology, 82152 Martinsried, Germany. [3]Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. [4]Institute of Molecular Immunology, Helmholtz Zentrum München, 81377 Munich, Germany. [5]Institute for Functional Genomics, University of Regensburg, 93053 Regensburg, Germany. [6]College of Medicine, Nursing and Health Sciences, Regenerative Medicine Institute, National Centre for Biomedical Engineering Science, National University of Ireland, County Galway, Ireland. [7]Autoimmunity, Transplantation and Inflammation Novartis Institutes for BioMedical Research, 4056 Basel, Switzerland. [8]Institute for Anatomy II, Friedrich Alexander University, 91054 Erlangen, Germany.

which had received T_{blast} cells 48 h beforehand (pre-EAE animal) and the other was a naïve rat that was not pre-treated (Supplementary Fig. 2a). Interestingly, $T_{MBP-GFP}$ cells passing through the blood shunt within 12 h after anastomosis infiltrated the CNS of the animal that was not pre-treated in similar frequencies to those of the pre-EAE animal (Fig. 1c). Furthermore, in contrast to T_{blast} cells, $T_{migratory}$ cells isolated from the spleen (60 h after T-cell transfer) readily immigrated into the inflamed CNS and, more importantly, into the CNS of animals that were not pre-treated (Fig. 1d and Supplementary Fig. 2b–e). Intravital two-photon microscopy revealed that $T_{migratory}$ cells were crawling on the meningeal vessels within minutes after intravenous transfer. Twenty-four hours later, the $T_{migratory}$ cells had transgressed the vessel walls and were moving throughout the leptomeningeal surface and deep within the CNS parenchyma, like T cells usually do 3 to 4 days after transfer of T_{blast} cells (Supplementary Fig. 2e–h and Supplementary Movie 1)[6,10]. $T_{migratory}$ cells in the CNS upregulated interleukin-2 (IL-2) receptor, OX40, interferon-γ (IFN-γ) and IL-17, indicative of a recent antigen encounter and reactivation within the CNS, the trigger for generalized parenchymal inflammation and disease (Supplementary Fig. 2i–k)[3,4,6,11]. Beginning 24 h after T-cell transfer the animals developed the first clinical signs of weight loss, followed 24 h later by overt paralytic disease; approximately 48 h earlier than in the typical transfer EAE (Supplementary Fig. 2l).

To identify the homing niches of T_{blast} cells after intravenous transfer, we analysed whole body sections of animals, complemented by cytofluorometric cell quantifications. At the early time points, almost all of the T_{blast} cells were located in the lung (Fig. 2a and Supplementary Fig. 3a, b). Some cells were also found distributed in the mucosa of the small intestine. Thereafter, $T_{MBP-GFP}$ cells accumulated in lung-draining mediastinal lymph nodes, before they appeared in the blood and spleen (Fig. 1b and Supplementary Fig. 3c). Notably, $T_{MBP-GFP}$ cells extracted from the lung at 48 h after transfer and injected into naïve recipient rats accumulated in the CNS within 24 h, where they displayed signs of activation (Fig. 2b and Supplementary Fig. 3d). Thus, the cells became competent in the lung to enter and to get reactivated within the CNS. This led us to analyse $T_{MBP-GFP}$ cell behaviour within the lung in more detail. Within the first 12 h after transfer, T_{blast} cells were distributed within the peripheral lung parenchyma. Then 24–48 h after transfer, they moved along bronchial structures and accumulated in areas consisting of densely packed clusters of lymphocytes; that is, bronchus-associated lymphoid tissues (BALT) (Fig. 2c, Supplementary Fig. 3e and Supplementary Movie 2). Interestingly, $T_{MBP-GFP}$ cells significantly increased their velocity during their journey (from 3.7 μm min^{-1} to 5.2 μm min^{-1}) and at the same time the percentage of arrested cells decreased (from 75% to 50%) (Fig. 2d, Supplementary Fig. 3f–h and Supplementary Movie 3). Furthermore, $T_{MBP-GFP}$ cells were regularly located within alveoli and bronchi, and seemed to use the airways as roads (Fig. 2e and Supplementary Movie 4). T cells are known to accumulate in the lung, including the airways, during lung infections and allergic reactions,

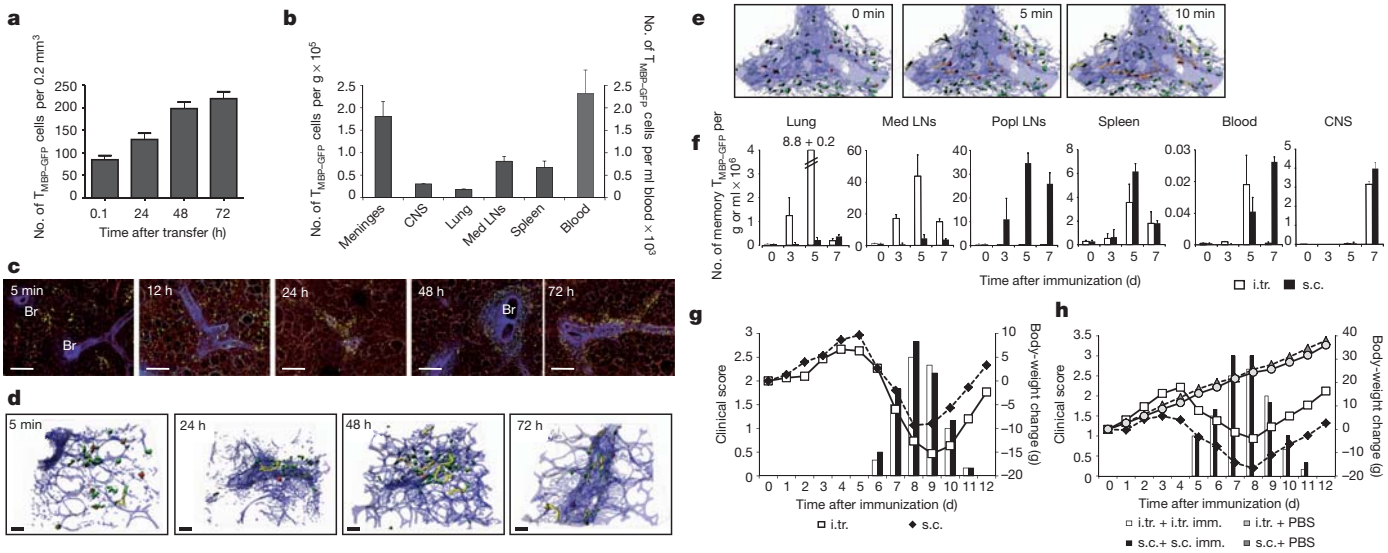


Figure 2 | The lung as a homing and reactivation site of $T_{MBP-GFP}$ cells. **a**, **b**, Quantification of lung infiltration by $T_{MBP-GFP}$ cells at the indicated time points after transfer of T_{blast} cells (**a**). Cells were counted from two-photon-microscopy overview images of acute lung slices. Shown are means and s.e.m. from at least three experiments. Lung-derived $T_{migratory}$ cells efficiently enter the naïve CNS (**b**). 1×10^6 $T_{migratory}$ cells isolated from lung 48 h after T-cell transfer were intravenously injected into naïve animals; 24 h later $T_{MBP-GFP}$ cells were quantified by cytofluorometry. Means ± 1 s.d. of three animals. Representative data from two independent experiments are shown. LNs, lymph nodes; Med LNs, mediastinal lymph nodes. **c**, Homing within the lung visualized by two-photon-microscopy. Representative two-photon-microscopy overview images of acute lung slices at the indicated time points after T-cell transfer. Green, $T_{MBP-GFP}$ cells; red, autofluorescent background and erythrocytes; blue, collagen fibres. Br, bronchial structures. Scale bars, 100 μm. **d**, Motility of $T_{MBP-GFP}$ cells within the lung over time. Three-dimensional reconstructions of representative two-photon-microscopy time-lapse videos of lung explants at the indicated time points after T-cell transfer. Scale bars, 50 μm. Green, $T_{MBP-GFP}$ cells; blue, collagen fibres; red dots, trajectories of stationary cells; yellow lines, trajectories of motile cells. **e**, $T_{MBP-GFP}$ cells use airways as roads. Snapshots of three-dimensional reconstructed time-lapse videos acquired by two-photon microscopy from lung explants 72 h after T-cell transfer. The times are in relation to the first displayed frame. Yellow, $T_{MBP-GFP}$ cells moving along alveoli and the outer wall of bronchial airways; orange, $T_{MBP-GFP}$ cells moving along the inner wall of airways; blue, collagen fibres rendered semi-transparent to reveal intraluminal cells. **f**, **g**, Memory T cells stimulated in the lung induce CNS autoimmunity as efficiently as they do after stimulation in peripheral lymph nodes. $T_{MBP-GFP}$ cells were intraperitoneally transferred into newborns and 10 to 14 weeks later, the animals were MBP-immunized either intratracheally (i.tr.) or subcutaneously (s.c.). Quantification of memory $T_{MBP-GFP}$ cells during EAE (**f**). Absolute number of memory $T_{MBP-GFP}$ cells was quantified by cytofluorometry at the indicated time points in intratracheally or subcutaneously immunized animals. Popl LNs, popliteal lymph nodes. Means ± 1 s.d. of duplicate experiments. EAE course in intratracheally or subcutaneously immunized (i.tr. imm. and s.c. imm., respectively) memory animals (**g**). Clinical score (bars) and body-weight change (lines) were measured daily. Representative data of two independent experiments, with three animals per group. **h**, Local stimulation of resting $T_{MBP-GFP}$ cells induces CNS disease. Resting $T_{MBP-GFP}$ cells were injected into the trachea or into the footpad of naïve animals. Subsequently the rats were intratracheally or subcutaneously challenged either with MBP in complete Freund's adjuvant or with PBS. Clinical score (bars) and body-weight change (lines) over the EAE course. Representative data of three independent experiments, with at least two animals per group.

although this transepithelial migratory path into the airways has been considered to be a dead end for the cells[12]. Therefore, to test whether T cells can pass through the airways, we transferred T_{blast} cells intratracheally. Similarly to intravenously transferred cells, intratracheally transferred T cells migrated along the airways and maintained their capacity to emigrate out of the lung and to induce clinical disease (Supplementary Fig. 3i).

The homing of effector T cells to the lung is not an artefact of transfer EAE. $T_{MBP-GFP}$ cells also accumulated in the lung after 'classical' antigenic stimulation in peripheral lymph nodes after subcutaneous immunization (Supplementary Fig. 4a). Moreover, memory $T_{MBP-GFP}$ cells, which had been embedded in the rat immune system over months, were also found to home and persist within the lung (Supplementary Fig. 4b). Importantly, these cells could readily be activated to become pathogenic: after intratracheal immunization with MBP the cells developed strong signs of activation and, after vigorous amplification within the lung and its draining lymph nodes, they moved into the blood and spleen from where they reached the CNS and triggered a paralytic disease (Fig. 2f, g and Supplementary Fig. 4b–e). Notably, this lung-derived autoimmune response did not differ from the one after stimulation of the memory T cells in the lymph nodes (Fig. 2f, g and Supplementary Fig. 4c–e). The stimulatory potential of the lung environment was confirmed further by local activation of resting $T_{MBP-GFP}$ cells that had been inserted into the lung by intratracheal transfer. Again, similarly to the stimulation in lymph nodes, the T cells in the lung underwent effective activation and proliferation following intratracheal administration of antigen, and they gained the capacity to invade the CNS and to evoke clinical EAE (Fig. 2h and Supplementary Fig. 4f–h).

Interestingly, $T_{MBP-GFP}$ cells changed their reactivity for chemotactic stimuli during the preclinical phase of transfer EAE. Early *ex-vivo*-isolated $T_{MBP-GFP}$ cells (5 min and 24 h after T-cell transfer) predominantly migrated towards the homeostatic chemokines CCL19 and CCL21 (Fig. 3a). These migratory preferences of the cells coincided with their migration to bronchial structures and the BALT that express CCL19 and CCL21 (Supplementary Fig. 5a)[13]. Pertussis toxin (PTX), which irreversibly blocks chemokine signalling[14], significantly reduced their accumulation within the BALT (Supplementary Fig. 5b), supporting the view that the T-cell migration to the bronchial structures and BALT was driven by CCL19 and CCL21. However, PTX pre-treatment did not interfere with the homing of the T cells to the lung and, in contrast to earlier reports about the effect of PTX on T-cell motility in lymph nodes[15], did not change their locomotion activities (Supplementary Fig. 5c). From 48 h after transfer onwards, $T_{MBP-GFP}$ cells responded mainly to gradients of inflammatory chemokines, such as CXCL11 and CCL5 (Fig. 3a). This 're-orientation' of the cells was relevant to their ability to home to their target organ: interference with CXCR3 signalling, that is, the corresponding receptor of CXCL11, significantly reduced their CNS invasion and the severity of clinical EAE (Supplementary Fig. 5d). Notably, $T_{MBP-GFP}$ cells left the lung at least partially under the control of sphingolipid signalling. Treatment with FTY720, a receptor agonist that functionally interferes with sphingosine 1-phosphate receptors (S1P1)[16], led to a substantial trapping of the T cells within the BALT, but did not hamper the entry of the cells into the lung or their locomotion behaviour there (Supplementary Fig. 5e, f). FTY720 treatment significantly reduced the numbers of $T_{MBP-GFP}$ cells in the blood and CNS, and delayed and ameliorated clinical EAE (Supplementary Fig. 5g, h), which could well be explained by the arrest of the pathogenic T cells within the lung.

These results suggest that $T_{MBP-GFP}$ cells get equipped during their journey through peripheral environments to enter their target tissue. To characterize these functional changes in more detail, we analysed genome-wide transcriptional profiles of T_{blast} cells (before intravenous transfer) and $T_{migratory}$ cells (isolated 3 days after T-cell transfer from the spleen) using Affymetrix 230.0 expression arrays. Average linkage cluster analysis (Genesis) revealed ~1800 upregulated and ~1600 downregulated transcripts (out of 30,248) for activated compared to migrating $T_{MBP-GFP}$ cells, respectively. These changes (≥2-fold) could be grouped into two major gene clusters that included reciprocally

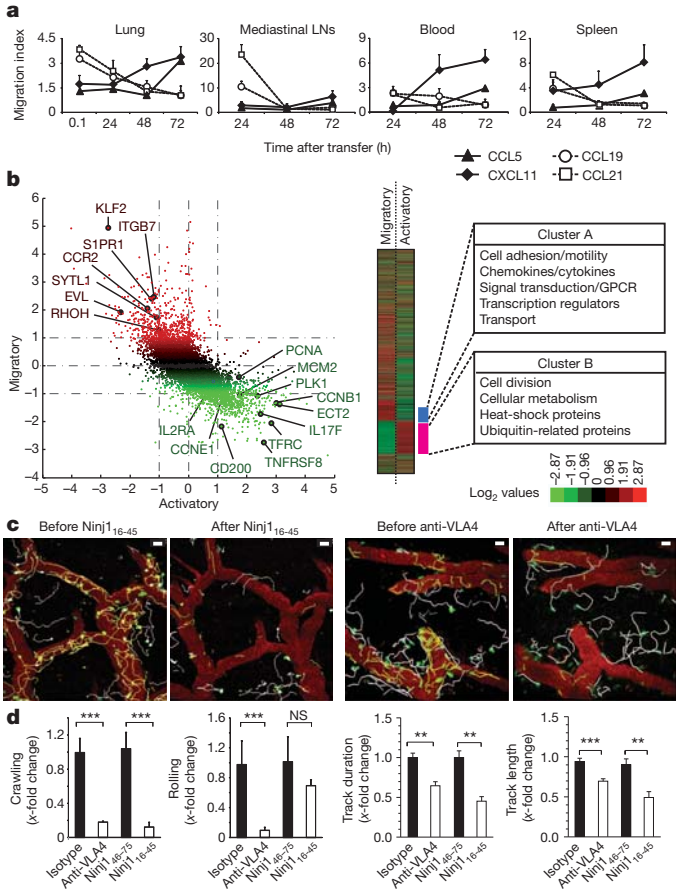


Figure 3 | Switch of T_{blast} cells into a migratory mode. a, Chemotactic reorientation of $T_{MBP-GFP}$ cells during preclinical EAE. $T_{MBP-GFP}$ cells isolated from the indicated organs were assayed for their ability to migrate towards CCL5, CXCL11, CCL19 and CCL21. Cell migration indices (ratio of transmigrated T cells with chemokine versus no chemokine) were calculated by cytofluorometry. Means ± s.e.m. of four independent experiments. **b**, Changes of transcript regulations of $T_{MBP-GFP}$ cells during their activatory (T_{blast} cells per resting $T_{MBP-GFP}$ cells) compared to their migratory state (spleen-derived $T_{migratory}$ cells per T_{blast} cells). Left, scatter plot depicting the upregulated transcripts during activation (green) or migration (red) with a selection of regulated genes. Right, global transcriptional profiles (cluster heat map). Overview of average linkage hierarchical cluster analysis of $T_{MBP-GFP}$ cells in their migratory (first column) or activatory (second column) state and annotated transcripts (rows). Gene-expression changes are depicted by a $\log_2$ pseudo-colour scale (ratio of fold-change is indicated). Two prominent clusters are marked by vertical coloured bars. The corresponding gene groups within these clusters are indicated in the boxes. GPCR, G protein-coupled receptor. **c**, **d**, Ninj1 as a novel adhesion molecule in $T_{migratory}$ cells. T cell crawling in leptomeningeal vessels is disrupted by Ninj1 or VLA4 blocking (**c**). $T_{migratory}$ cells 60 h after T-cell transfer were recorded by live two-photon microscopy before or 90 min after intravenous injection of Ninj1 blocking peptide (amino acids 16–45; $Ninj1_{16-45}$) or anti-VLA4 monoclonal antibodies. Motility tracks (30 min recording time) of $T_{MBP-GFP}$ cells crawling inside the leptomeningeal vessels (yellow tracks). Note that the peptide treatment did not interfere with the motility of extravasated $T_{MBP-GFP}$ cells (white tracks). Scale bars, 20 μm. Locomotion behaviour of $T_{migratory}$ cells after Ninj1 is compared to VLA4 blocking treatment (**d**). Numbers of intraluminally crawling and rolling $T_{migratory}$ cells, track durations and track lengths of individual $T_{migratory}$ cells in 30-min intervals were quantified in the leptomenigeal vessels before and 90 min after treatment with PBS, anti-VLA4, blocking $Ninj1_{16-45}$ and Ninj1 control peptide (amino acids 46–75; $Ninj1_{46-75}$). Fold changes after the indicated treatment are shown on the *y* axis. Cumulative data from three independent experiments per treatment. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. NS, not significant.

regulated pathways (Fig. 3b). In comparison to T_{blast} cells, $T_{migratory}$ cells downregulated their immune activation and proliferation programs (Supplementary Fig. 6a–d); that is, pro-inflammatory cytokines, costimulatory molecules, cell-cycle-associated genes, cyclin-dependent kinases and factors of the DNA replication machinery. By contrast, genes controlling cell motility and cell adhesion were strongly upregulated (Supplementary Fig. 6e). Accordingly, transcription factors that are important for regulating T-cell cycling and migration—for example, *Klf2* (ref. 17)—were among the most upregulated genes in $T_{migratory}$ cells (Fig. 3b and Supplementary Fig. 6c).

To evaluate the functional relevance of the newly assumed adhesive profile, we tested ninjurin 1 (Ninj1), a two-pass transmembrane adhesion molecule that functions by homophilic binding. Ninj1 was recently found to be expressed in myeloid cells and CNS endothelia, and was upregulated during EAE (Supplementary Fig. 7a–c)[18–20]. We found Ninj1 expression and a transient upregulation in $T_{migratory}$ cells, but not in naïve T or B cells (Supplementary Fig. 7b–d). Intriguingly, intravenous infusion of Ninj1-blocking peptide (amino acids 16–45), that blocks homophilic Ninj1 binding, strongly reduced the numbers of T cells crawling on the intraluminal side of the CNS vessels (Fig. 3c, d, Supplementary Fig. 7e and Supplementary Movie 5). This might explain the reduced $T_{migratory}$ cell entry into the CNS and the delayed and ameliorated disease course after intravenous infusion of the Ninj1-blocking peptide (Supplementary Fig. 7f, g). Overexpression of Ninj1 by retroviral gene transfer promoted the entry of T cells overexpressing GFP and *Ninj1* ($T_{MBP–GFP–Ninj1}$ cells) into the CNS and accelerated the onset of clinical EAE (Supplementary Fig. 7h–j). The integrin α_4 (very late antigen 4, VLA4), whose β1 chain was also upregulated in $T_{migratory}$ cells (Supplementary Fig. 6e), had similar effects on the intravascular T-cell motility as shown by blockage with VLA4-neutralizing antibody (Fig. 3d). It is clear that both adhesion molecules have a role in T-cell crawling by strengthening the endothelial adherence of the cells. Interestingly, Ninj1 blockage was less efficient in interfering with the initial adhesion steps of the cells to the endothelia than the anti-VLA4 treatment (Fig. 3d and Supplementary Fig. 7e).

To explore the regulation dynamics of the migratory mode in $T_{MBP–GFP}$ cells *in vivo*, we analysed the expression profiles of a selected set of the regulated genes in T cells during their migratory paths. Our results show that the reprogramming process in T cells is already initiated in the lung, and is continued during their migration through the mediastinal lymph nodes and spleen (Supplementary Fig. 8). Notably, the development of a migratory mode was not restricted to T cells during transfer EAE: memory or post-activated $T_{MBP–GFP}$ cells after intratracheal or subcutaneous immunizations assumed a very similar migratory mode before they accumulated at the CNS (Supplementary Fig. 9a–f). Furthermore, the observed homing patterns of intravenously transferred T-cell blasts and their switch to migration-competent effector cells were not restricted to T_{MBP} cells, but were also seen in neuritogenic P2-reactive T cells or T cells recognizing ovalbumin (Fig. 4 and Supplementary Fig. 9g–k).

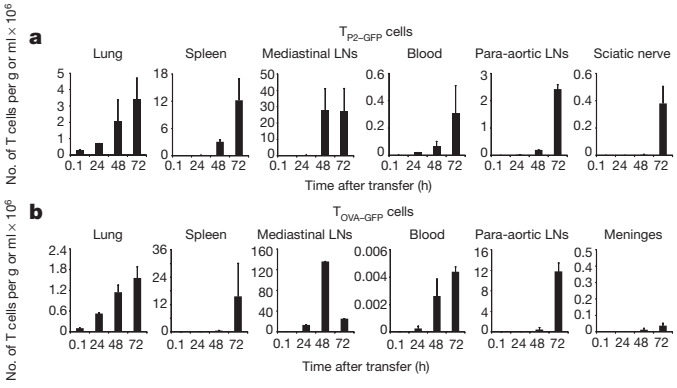


Figure 4 | Homing into the lung is not limited to encephalitogenic T cells. **a**, **b**, Homing behaviour of neuritogenic $T_{P2–GFP}$ cells and non-pathogenic $T_{OVA–GFP}$ cells resembles that of $T_{MBP–GFP}$ cells. The number of $T_{P2–GFP}$ cells (**a**) or $T_{OVA–GFP}$ cells (**b**) in the indicated organs after intravenous transfer. The cells were quantified using cytofluorometry. Means ± 1 s.d. Representative data of three independent experiments with at least three animals per group.

Our data show that the preclinical phase of transfer EAE cannot be explained solely by the integrity of the BBB or the immune hostile environment of the CNS having to be breached by activated 'pioneer' T cells[1,2] before the mass of the T cells get access into their target organ. The majority of the transferred cells during this phase of transfer EAE home to the peripheral environments of the lung from where they move to the BALT and draining lymph nodes, before they re-enter the blood circulation and finally reach the CNS. On their way, the T cells profoundly reprogram their gene-expression profile, and this includes upregulation of membrane receptors, such as the S1P1, that have a role in guiding the cells out of lymphatic structures[21]. Additionally, the T cells upregulate chemokine receptors and adhesion molecules that regulate their orientation towards inflamed tissues and simultaneously raise their capacity to transgress endothelial barriers, including those of CNS vessels[22–24]. At the same time, the T cells downregulate their proliferation and activation programs and thus prepare for antigen re-encounter and activation. A premature reactivation or hyper-reactivation would cripple the immune response by causing T-cell apoptosis or anergy[25,26]. Importantly, the switch of activated autoaggressive T cells towards a migratory mode occurs before the onset of any deleterious inflammatory processes within the attacked organs. Several of the genes recently identified in genome-wide association studies as risk genes for the development of multiple sclerosis[27] were found in the fraction of regulated genes in our transcriptome analysis (Supplementary Fig. 10). Recently, the gut environment came into focus as a potential site for the initial priming of autoreactive T cells during spontaneous EAE[28]. We demonstrate here that the lung can serve as a location where autoreactive T cells become reactivated and gain the competence to enter the CNS. It is well known that relapses in multiple sclerosis are triggered by inflammations of the respiratory tract[29]. The lung is in constant and direct contact with the outer environment and is colonized by microbiota that can actively promote local immune responses[30]. This organ represents a homing place for effector and memory cells relevant to host defence and asthma pathogenesis. At the same time, the lung represents a niche for potentially autoaggressive effector and memory T cells. Our data indicate that stimulatory environmental factors might directly elicit a pathogenic response of these cells.

METHODS SUMMARY

$CD4^+CD8^-$ $\alpha\beta TCR^+$ MBP-, ovalbumin- and P2-specific T cells retrovirally engineered to express enhanced GFP were established from lymph nodes of immunized Lewis rats as reported[7]. All cell lines were tested for cytokine profile, antigen specificity and surface phenotype. Memory animals were established by intraperitoneal transfer of $T_{MBP–GFP}$ cells into neonatal rats. If not otherwise stated, transfer EAE was induced by intravenous injection of 5×10^6 T_{blast} cells or by intratracheal administration of 2×10^6 T_{blast} cells. Active EAE was induced in memory animals by intratracheal or subcutaneous immunization. In naïve animals, aEAE was induced by subcutaneous or intratracheal transfer of resting T cells followed by immunization. Transfer of migration-competent $T_{MBP–GFP}$ cells into recipient animals, cross-circulation model and therapeutic interference are described in the Methods. Animal experiments were in compliance with regulations of the Bavarian and Lower Saxony state.

Time-lapse two-photon laser-scanning microscopy, fluorescence video microscopy time-lapse recordings and image analyses are described in Methods.

$T_{MBP–GFP}$ cells were isolated from the single organs as described[8]. Cell sorting was carried out using FACS Vantage, FACSAria (Becton Dickinson) or MoFlo (Beckman Coulter). Cytofluorometric analysis was performed with FACS-Calibur operated by Cell Quest software.

Total RNA was used to perform a genome-wide transcriptional profiling assay (Rat Genome 230, Affymetrix). Messenger RNA extraction, complementary DNA reversion and Taqman analyses are described in Methods.

Graphs were generated and statistical tests performed using Graphpad 5.0.4.

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

Received 28 April; accepted 19 June 2012.
Published online 22 August 2012.

1. Wekerle, H. *et al.* Cellular immune reactivity within the CNS. *Trends Neurosci.* **9,** 271–277 (1986).
2. Hickey, W. F., Hsu, B. L. & Kimura, H. T-lymphocyte entry into the central nervous system. *J. Neurosci. Res.* **28,** 254–260 (1991).
3. Brown, D. A. & Sawchenko, P. E. Time course and distribution of inflammatory and neurodegenerative events suggest structural bases for the pathogenesis of experimental autoimmune encephalomyelitis. *J. Comp. Neurol.* **502,** 236–260 (2007).
4. Kivisäkk, P. *et al.* Localizing central nervous system immune surveillance: meningeal antigen-presenting cells activate T cells during experimental autoimmune encephalomyelitis. *Ann. Neurol.* **65,** 457–469 (2009).
5. Karin, N. *et al.* Selective and nonselective stages in homing of T lymphocytes to the central nervous system during experimental allergic encephalomyelitis. *J. Immunol.* **150,** 4116–4124 (1993).
6. Bartholomäus, I. *et al.* Effector T cell interactions with meningeal vascular structures in nascent autoimmune CNS lesions. *Nature* **462,** 94–98 (2009).
7. Flügel, A. *et al.* Gene transfer into $CD4^+$ T lymphocytes: green fluorescent protein-engineered, encephalitogenic T cells illuminate brain autoimmune responses. *Nature Med.* **5,** 843–847 (1999).
8. Flügel, A. *et al.* Migratory activity and functional changes of green fluorescent effector cells before and during experimental autoimmune encephalomyelitis. *Immunity* **14,** 547–560 (2001).
9. Glabinski, A. R. *et al.* Central nervous system chemokine mRNA accumulation follows initial leukocyte entry at the onset of acute murine experimental autoimmune encephalomyelitis. *Brain Behav. Immun.* **9,** 315–330 (1995).
10. Kawakami, N. *et al.* Live imaging of effector cell trafficking and autoantigen recognition within the unfolding autoimmune encephalomyelitis lesion. *J. Exp. Med.* **201,** 1805–1814 (2005).
11. Kawakami, N. *et al.* The activation status of neuroantigen-specific T cells in the target organ determines the clinical outcome of autoimmune encephalomyelitis. *J. Exp. Med.* **199,** 185–197 (2004).
12. Hogan, R. J. *et al.* Protection from respiratory virus infections can be mediated by antigen-specific CD4(+) T cells that persist in the lungs. *J. Exp. Med.* **193,** 981–986 (2001).
13. Lo, J. C. *et al.* Differential regulation of CCL21 in lymphoid/nonlymphoid tissues for effectively attracting T cells to peripheral tissues. *J. Clin. Invest.* **112,** 1495–1505 (2003).
14. Pham, T. H. *et al.* S1P1 receptor signaling overrides retention mediated by G alpha i–coupled receptors to promote T cell egress. *Immunity* **28,** 122–133 (2008).
15. Asperti-Boursin, F. *et al.* CCR7 ligands control basal T cell motility within lymph node slices in a phosphoinositide 3-kinase-independent manner. *J. Exp. Med.* **204,** 1167–1179 (2007).
16. Brinkmann, V. *et al.* Fingolimod (FTY720): discovery and development of an oral drug to treat multiple sclerosis. *Nature Rev. Drug Discov.* **9,** 883–897 (2010).
17. Carlson, C. M. *et al.* Kruppel-like factor 2 regulates thymocyte and T-cell migration. *Nature* **442,** 299–302 (2006).
18. Ahn, B. J. *et al.* Ninjurin1 is expressed in myeloid cells and mediates endothelium adhesion in the brains of EAE rats. *Biochem. Biophys. Res. Commun.* **387,** 321–325 (2009).
19. Araki, T. & Milbrandt, J. Ninjurin, a novel adhesion molecule, is induced by nerve injury and promotes axonal growth. *Neuron* **17,** 353–361 (1996).
20. Ifergan, I. *et al.* Role of Ninjurin-1 in the migration of myeloid cells to central nervous system inflammatory lesions. *Ann. Neurol.* **70,** 751–763 (2011).
21. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427,** 355–360 (2004).
22. Yednock, T. A. *et al.* Prevention of experimental autoimmune encephalomyelitis by antibodies against alpha 4 beta 1 integrin. *Nature* **356,** 63–66 (1992).
23. Sporici, R. & Issekutz, T. B. CXCR3 blockade inhibits T-cell migration into the CNS during EAE and prevents development of adoptively transferred, but not actively induced, disease. *Eur. J. Immunol.* **40,** 2751–2761 (2010).
24. Zheng, H. M. *et al.* Mimic peptides bonding specifically with the first and second extracellular loops of the CC chemokine receptor 5 derived from a phage display peptide library are potent inhibitors of experimental autoimmune encephalomyelitis. *Inflamm. Res.* **60,** 759–767 (2011).
25. Lenardo, M. J. Interleukin-2 programs mouse alpha beta T lymphocytes for apoptosis. *Nature* **353,** 858–861 (1991).
26. Tham, E. L., Shrikant, P. & Mescher, M. F. Activation-induced nonresponsiveness: a Th-dependent regulatory checkpoint in the CTL response. *J. Immunol.* **168,** 1190–1197 (2002).
27. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476,** 214–219 (2011).
28. Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* **479,** 538–541 (2011).
29. Sibley, W. A., Bamford, C. R. & Clark, K. Clinical viral infections and multiple sclerosis. *Lancet* **325,** 1313–1315 (1985).
30. Ichinohe, T. *et al.* Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl Acad. Sci. USA* **108,** 5354–5359 (2011).

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

Acknowledgements The authors thank A. Stas, S. Hamann, N. Meyer and I. Haarmann for technical assistance. N. Plesnila supported this work with advice in surgical preparations; T. Issekutz provided anti-CXCR3 and anti-VLA4 monoclonal antibodies. We are grateful to D. Johnson and F. Lühder for critical reading of the manuscript. We thank W. Leibold for critical discussions. We thank C. Ludwig for text editing. Hartmut Wekerle is Senior Research Professor of the Hertie Foundation. This work was supported by the Deutsche Forschungsgemeinschaft (TRR-SFB43, FORR 1336), the Bundesministerium für Bildung und Forschung ('UNDERSTAND MS') and the Hertie Foundation (grant 1.01.1/11/004).

Author Contributions F.O. and K.S. carried out most of the cell transfer, EAE studies, cytofluorometric characterizations and cell-transfer experiments. C.S. together with F.O. performed the lung-migration studies. V.K.U. contributed by analysing the gene transcriptome data. C.Sch. contributed to the generation of imaging data and the treatment with anti-CXCR3 monoclonal antibody. D.L. and M.N. contributed with characterizations of the migratory T cells. K.H. coordinated the cell sampling for the transcriptome analyses. W.N. and H.L. contributed by carrying out the transcriptome analyses. C.L., R.S. and M.V. helped with the statistical transcriptome data analysis. V.B. provided FTY720 and gave advice about its experimental handling. W.E.F.K. and J.E. performed cell sorting. H.W. contributed with scientific advice and contributed to the design of the transcriptome analysis. C.F.-K. contributed to the morphological studies of the lung tissue. A.F. and F.O. designed the study, coordinated the experimental work and wrote the manuscript with input from co-authors.

Author Information The data discussed in this publication have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus and are accessible through GEO Series accession number GSE38645 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38645). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.F. (fluegel@med.uni-goettingen.de).

METHODS

Animals. Rats on a Lewis (LEW/Crl) genetic background were bred by the animal facilities of the Max Planck Institute for Biochemistry and the University Medical Center Göttingen. All animal experiments were performed according to the Bavarian and Lower Saxony state regulations for animal experimentation and were approved by the responsible authorities (numbers 209.1/211-2531-36/04 and 33.11.42502-04-016/09, respectively).

Antigens. T-cell clones were specific for guinea-pig MBP, ovalbumin (OVA) and P2 peptide. MBP was isolated from guinea-pig brains as described[31]. OVA was from Sigma. The synthetic peptide of bovine P2 myelin protein (amino acids 53–78) was synthesized by the Max Planck Institute for Biochemistry.

Generation and culturing of T cells. MBP-, OVA- and P2 $CD4^+$ T cells retrovirally engineered to express the enhanced GFPs (eGFPs) ($T_{MBP-GFP}$, $T_{OVA-GFP}$, T_{P2-GFP}) were generated as reported[7]. Ten days after subcutaneous immunization with MBP, OVA, P2 peptide in CFA (which was prepared by supplementing Incomplete Freund's Adjuvant (Difco Laboratories) with 2 mg ml^{-1} of inactivated *Mycobacterium tuberculosis* H37Ra (Difco Laboratories)), cells were collected from the draining lymph nodes and co-cultured in the presence of retrovirus-producing GP+E 86 packaging cells. Murine stem-cell virus (pMSCV, Invitrogen) in which the eGFP gene (Invitrogen) had been integrated (pMSCV–eGFP) was used as the retroviral vector. After 2 days of co-culture, T-lymphocyte blasts were expanded in IL-2-containing growth medium for 5 days and then re-stimulated in the presence of irradiated thymocytes pulsed with the specific antigen. Selection with G418 and amplification of the T cells was carried out as reported[32]. All the cell lines were $CD4^+$, $CD8^-$ and $\alpha\beta TCR^+$, and were extensively tested for their phenotype, cytokine profile and antigen specificity.

Memory animals were prepared as previously described[33]. Briefly, $T_{MBP-GFP}$ cells (2×10^6 cells per animal; 4 days after MBP stimulation) were transferred intraperitoneally into neonatal rats within 48 h post partum. The transferred cells (memory $T_{MBP-GFP}$ cells; $CD4^+CD8^-$) persisted lifelong within the immune repertoire of the animals maintaining a memory phenotype (L-selectin^{low} and CD45RC ^{low}, CD44^{high}). Experiments were performed in animals that were 10–14 weeks old.

EAE models. For transfer, MBP-specific T_{GFP} cells were stimulated as described above. Two days later, naïve Lewis rats (6–8 weeks old) received tail-vein injections of 5×10^6 $T_{MBP-GFP}$, T_{P2-GFP} or $T_{OVA-GFP}$ blasts (T_{blast} cells). Alternatively, 2×10^6 (in 100 µl Eagle's Hepes (EH) medium) T_{blast} cells were injected intratracheally into naïve animals previously anaesthetized and intubated. For transfer of migration-competent $T_{MBP-GFP}$ cells, 3×10^6 $T_{migratory}$ cells were isolated from the spleen 60 h after T-cell transfer and 1×10^6 from the lung 48 h after T-cell transfer and immediately transferred intravenously into recipient animals.

Active EAE was induced in naïve or memory animals by intratracheal or subcutaneous immunization.

For intratracheally induced EAE, 3×10^6 resting $T_{MBP-GFP}$ cells were intratracheally transferred into anaesthetized and intubated naïve animals. 3 h later the rats were immunized by intratracheal instillation of an emulsion (50 µl in PBS) containing CFA (20 µl per animal), MBP (1 mg ml^{-1}, 12.5 µl per animal) and China ink (1%). China ink could be detected both macroscopically and microscopically, confirming the correct intrapulmonary distribution of the antigen. The procedure was tolerated well. Note that administration of China ink alone, MBP in China ink, or OVA in CFA and China ink did not induce any clinical effect.

For subcutaneously induced active EAE, 1.5×10^6 $T_{MBP-GFP}$ cells were injected into the foot pad of naïve animals. 12 h later the animals were immunized by subcutaneous injection of an emulsion containing CFA (35 µl) and MBP (1 mg ml^{-1}, 35 µl).

The same intratracheal or subcutaneous immunization protocols were applied to induce active EAE in memory animals.

The animals were observed daily for clinical symptoms. Their disease stage was scored as follows: 0, no disease; 1, flaccid tail; 2, gait impairments; 3, complete hind-limb paralysis; 4, tetraparesis; and 5, dead.

Pre-activation of the BBB. BBB priming was induced by transfer of 5×10^6 non-labelled T_{blast} cells. Alternatively, 150 ng of LPS (25 µl of volume (LPS from *Escherichia coli* 0127:B8 Sigma-Aldrich)) were stereotactically injected into the cisterna magna of anaesthetized animals.

Two-photon laser-scanning microscopy equipment. Two-photon laser-scanning microscopy was carried out using a Zeiss Laser Scanning Microscope 710 combined with a Coherent Chameleon Vision II laser system. The excitation wavelength was set to 880 nm and the pulsed laser was routed through a Zeiss ×20 water immersion objective (NA 0.95). Fluorescent signals were detected using non-descanned detectors equipped with 525/50-nm and 630/69-nm band-pass filters (Semrock).

Surgical procedure and intravital imaging settings. Animals were anaesthetized using Ketamin (10%, Medistar) and Xylazin (Xylariem 20 mg, Riemser Arzneimittel) in equal proportions (100 µl per 100 g), tracheally intubated and ventilated with 1.5–2% isoflurane. A spinal cord imaging window was established at the level of Th12/L1 as previously described[6,34]. Animals were stabilized in a custom-made microscope stage. Body temperature was controlled by a heated pad (37 °C). Vital parameters were monitored during the imaging. Blood vessels were labelled by intravenous infusion of fluorescent dextran (Invitrogen).

Acute slicing and imaging settings. Acute spinal-cord slices were prepared as reported[10], with minor modifications. In brief, spinal cords were explanted and embedded in 7% low-melting agarose (Carl Roth GMBH). Acute slices (thickness of 300 µm) were prepared using a Vibrotome (Campden Vibroslicer).

Acute lung-slice preparation was adapted from a protocol described previously[35]. In brief, the lung was inflated with 2% low-melting agarose through the trachea. Subsequently, the lung was cut into pieces and embedded in 7% agarose gel. After hardening, slices of 1 mm in thickness were obtained using a Vibratome. Acute spinal-cord and lung slices were stored on ice in buffered cell medium and constantly oxygenized with carbogen until image acquisition. All the procedures were performed at 4 °C.

A custom-made heating chamber was supplied with humidified carbogen to visualize T-cell movement in acute slices of lung or spinal cord. High-resolution overview images were acquired using the Zeiss tile-scan module; time-lapse videos were acquired in 60-s intervals with approximately three spots imaged in parallel.

Image analysis after intravital imaging. Three-dimensional reconstructions were generated using Imaris 7.1.1 (Bitplane). To determine T-cell velocity, four-dimensional data were generated by the Imaris Track module and revised manually. Motility parameters and trajectories were calculated from the obtained position coordinates using Excel (Microsoft). Crawling and rolling cells were defined as reported previously[2].

Analysis of acute lung-slice overview images. Files were exported from the ZEN 2009 Software and analysed in ImageJ 1.45d (National Institutes of Health). Shading correction was carried out using Photoshop CS5 (Adobe Systems). For two-photon images with China-ink traces, a white-noise reduction was performed by removing signals present across all channels on a single-plane basis using ImageJ. For quantitative analysis, the green-channel image was enhanced by subtraction of autofluorescent, overlapping signals from the red and blue channels combined with noise reduction using ImageJ.

To quantify bronchiolar association of T cells in ImageJ, regions of transversally cut bronchiolar structures were subdivided into a peribronchial region (as determined by high collagen density) and a closely associated parenchymal region, defined as a 200-µm band surrounding, but not including, the peribronchial region. Mean pixel intensities (that is, the sum of all pixel values in the selection divided by the number of selected pixels) in these two regions were determined in the enhanced green-channel image and their ratio was defined as a measure for bronchiolar cell accumulation (bronchiolar association index). A non-parametric statistical analysis was performed (Dunn's multiple comparison test) using Graphpad 5.0.4 (GraphPad Software).

Overall cell counts per overview image were assessed by ImageJ's particle analyser after watershed segmentation of the enhanced green-channel image. Cell numbers from peribronchial accumulations were normalized to the relative frequency of bronchioles per analysed region to account for fluctuations in bronchiole density across multiple slices.

Analysis of acute lung-slice time-lapse videos. Three-dimensional reconstructions were created as described above. If the blue channel signal from the time-lapse data was too dim for high-quality three-dimensional reconstruction, high-resolution snap-shot acquisitions of the same area were overlaid on the time-lapse videos to provide a detailed view of the T-cell movement in the pulmonary environment.

Fluorescence-microscopy equipment and imaging set-up. A Zeiss Axio Observer Fluorescence Microscope was used in combination with an HXP120C illuminator routed through a Zeiss Plan Apochromat ×10 Air Objective. Fluorescent signals were detected using a Zeiss AxioCam HSM video camera. An atmosphere suitable for live imaging was generated by a Zeiss Heating Unit XL with constant carbogen gas supply. Glass-bottom wells (Greiner Bio-One) with custom-made compartments were used to acquire images from two treatment groups simultaneously. Multi-spot time-lapse videos were acquired with 200-ms image acquisition for the green (excitation: 470 nm, emission: 509 nm) and red (excitation: 558 nm, emission: 583 nm) channels and approximately 30-ms image acquisition for brightfield. Time-lapse image series were acquired (1-min intervals) with multiple spots in parallel.

Analysis of time-lapse videos. All videos were analysed using Imaris 7.1.1. Cell trajectories were determined using the tracking module and revised manually. All 60-frame videos (1 min per frame, 60 min per video) were split into subsets of 10 frames for analysis. The relevant statistical data (for example, track speed) were

exported to Microsoft Excel for further analysis. Cells that were visible in more than 3 of the 10 frames were included in the analysis. A cell was defined as motile if it moved more than 10 μm away from its origin during the 10-frame interval. For quantification of cell velocities, only motile cells were considered. Graphs were generated and statistical tests performed using Graphpad 5.0.4.

Cell isolation, cytofluorometry and fluorescence-activated sorting. $T_{MBP-GFP}$ cells were isolated from spleen, blood, meninges and CNS parenchyma, as described previously[6,8,34]. From lymph nodes, $T_{MBP-GFP}$ cells were retrieved by sieved homogenization. Bone-marrow $T_{MBP-GFP}$ cells were isolated by flushing the cavity of the femur. To recover $T_{MBP-GFP}$ cells from lung, the organ was mechanically minced (McIlwain Tissue Chopper) and incubated with 0.3% collagenase (Sigma-Aldrich) for 1 h at 37 °C under constant shaking. Subsequently, the tissue was forced through a metal mesh and washed in EH medium. T cells were isolated by percoll gradient. Immediately after cell preparation, T_{GFP} cells were sorted using a FACS Vantage or a FacsAria cell sorter (both by Becton Dickinson) or MoFlo (Beckman Coulter). All procedures were performed at 4 °C. Cytofluorometric analysis was carried out with a FACS-Calibur operated by Cell Quest software (Becton Dickinson). The number of $T_{MBP-GFP}$ cells was determined in relation to a known absolute number of added phycoerythrin-labelled beads (Becton Dickinson). For surface staining the following antibodies were used: CD4 (W3/25), CD8β, CD62L, αβTCR (R73), Vβ8.2 TCR, Vβ8.5 TCR, Vβ10 TCR, Vβ16 TCR, Vα4 TCR, Vα8 TCR, CD45RA (OX33), CD11b (OX42), CD25 (OX39), CD134 (OX40), transferrin receptor (CD71), basigin (CD147), ItgB1(CD29) and neuropilin (Santa Cruz). Mouse immunoglobulin-G1κ (IgG1κ) (MOPC-31 C, Sigma-Aldrich) served as isotype control. Allophycocyanin (APC)-labelled anti-mouse antibody (Dianova) was used as the secondary antibody. All antibodies were purchased from Serotec unless stated otherwise. Intracellular staining for IFN-γ and IL-17 was carried out as described[6]. For intracellular staining of Ninj1, KLF2 and KLF4, samples were fixed in paraformaldehyde (PFA; 2%) for 20 min, permeabilized by Cytofix–Cytoperm buffer (Becton Dickinson) and then incubated with anti-Ninj1 monoclonal antibodies (MABs) (Becton Dickinson), anti-KLF2 (Chemicon international) and anti-KLF4 antibodies (Affinity Bioreagents) for 45 min. APC-labelled anti-mouse antibody and APC-labelled anti-rabbit antibody (both from Dianova) were used as secondary antibodies for Ninj1, and KLF2 and KLF4, respectively.

***In vivo* proliferation and cell cycle analysis.** 5-bromodeoxyuridine (BrdU) (120 mg kg^{-1} body weight; Becton Dickinson) was injected intraperitoneally 2 h before organ recovery. BrdU-containing $T_{MBP-GFP}$ cells were isolated from the single tissues and stained using a BrdU Flow kit (Becton Dickinson) according to the manufacturer's instructions. Cells were re-suspended in 7-amino-actinomycin D (7-AAD) solution just before flow-cytometry analyses.

Cross-circulation model. Diethlyether anaesthesia followed by subcutaneous atropine (0.5 mg kg^{-1}) injection was induced in a naïve animal that was not pre-treated and in an animal that had received 5×10^6 $T_{MBP-GFP}$ cells 60 h beforehand (pre-EAE). Animals were then intubated and ventilated with 0.8% halothane in 70% N_2O and 30% O_2. A rectal probe was used to monitor body temperature. Body temperature was controlled by a heated pad (37 °C). The tail artery cannulation was used to measure the mean arterial blood pressure and for blood sampling at regular intervals to determine pH and arterial blood gases. A cannula was introduced into the tail artery in a proximal direction. Distal to the cannula, the artery was ligated and a second cannula was introduced in the distal direction of the tail artery. The two animals were connected through a heparinized catheter (Harvard Apparatus) from the proximal to the distal part and vice versa. A constant blood flow between the two animals was regularly monitored by injection of a small amount of NaCl. The arterial connection lasted 8 to 12 h. Thereafter, the animals were killed and the numbers of $T_{MBP-GFP}$ cells in the spleen and CNS of pre-EAE animals and animals that were not pre-treated were quantified by flow cytometry after the 12-h anastomosis. The efficiency of distributions across the blood shunt was measured in each experiment; as the 'transfer rates' substantially fluctuated between the different animal pairs (probably owing to variable blood flows within the shunts) relative numbers were given.

Microarray analysis. Activated T cells ($T_{MBP-GFP}$ blast cells) were obtained 2 days post re-stimulation *in vitro* using thymocytes and MBP antigen followed by FACS sorting. Resting T cells ($T_{MBP-GFP}$ resting cells) were obtained 7 days post antigen exposure followed by FACS sorting. FACS sorting of spleen-derived migratory T cells ($T_{MBP-GFP}$ migratory cells) was performed 3 days post transfer. The sorting reached over 95% of purity. Transfer EAE was induced as described above. Cells were collected in cold PBS containing 1% rat serum, re-suspended in lysis buffer containing β-mercaptoethanol and mechanically broken up with a 23G needle. RNA was isolated from $T_{MBP-GFP}$ blast cells, $T_{MBP-GFP}$ resting cells and $T_{MBP-GFP}$ migratory cells using the Qiagen RNeasy mini kit. For every RNA sample, more than 10×10^6 cells were used. RNAs from four to six independent experiments with four animals in each experiment were combined into one RNA sample. Two RNA samples from spleen, five samples from $T_{MBP-GFP}$ blast cells and four from $T_{MBP-GFP}$ resting cells were analysed further. Integrity, size range and concentration of the RNAs were evaluated using the Agilent Bioanalyser (Agilent Technologies) and agarose gel electrophoresis. Total RNA was reverted into biotinylated cRNA according to the manufacturer's protocol (Affymetrix); the resultant cRNA was fragmented and then hybridized to the GeneChip Rat Genome 230Array (Affymetrix). The arrays were washed, stained and scanned by the GeneChip System confocal scanner (Hewlett Packard). Gene-expression profiles were background corrected and normalized on probe level using the variance stabilization method described previously[36]. Normalized probe intensities were summarized into gene-expression levels using the additive model as described[37]. Comparison analysis was performed on the gene-expression data set pertaining to $T_{MBP-GFP}$ blast cells, $T_{MBP-GFP}$ resting cells and $T_{MBP-GFP}$ migratory cells, to obtain transcriptomes of activated state (T_{blast} versus $T_{resting}$) and migratory state ($T_{migratory}$ versus T_{blast}). Transcriptomes were compared using Matlab. Average linkage hierarchical cluster analysis was performed on the annotated expression data set using Genesis[38]. Pathway-based analysis was carried out using GenMAPP[39]. For Gene-ontology based analysis, recent annotations for affymetrix probesets were downloaded from the Gene Ontology consortium database.

Quantitative PCR. Total RNA was isolated using the standard Trizol method. cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen). Quantitative PCR analysis was carried out using the 7500 Real Time PCR system (PE Applied Biosystems), as described[8,34,40]. β-actin was used as the house-keeping gene. Data were obtained by independent duplicate measurements. The CT value of the individual measurements did not exceed 0.5 amplification cycles. Several rat primers and probes were recently described[6,34].

The primer probes not previously reported are: *Ccr7*; forward, 5′-GTGTAGT CCACGGTGGTGTTCTC-3′; backward, 5′-CTGGTCATTTTCCAGGTGTGCT-3′; probe, 5′-CCGATGTAGTCGTCTGTGA-3′. *Ccl21*; forward, 5′-TCCTTAGCT TGGTCCTGGTTCT-3′; backward, 5′-GAATCTTCTTCTGGCTGTACCTAA GAC-3′; probe, 5′-CCCGTGGACCCAAGGCAGTGAC-3′. *Ninj1*; forward, 5′-GGGCTGGTTTTCATCATCGT-3′; backward, 5′-GGTGCCACGTCCATGA CA-3′; probe, 5′-ACATCTTCATTACGGCCTTTGGAGTCCAG-3′. *Rhamm*; forward, 5′-GCGCAGACTCTGAGAGATGTT-3′; backward, 5′- GCCATGGC TACTTTTTCTTG-3′; probe, 5′-CTCAGTTAGAAAGCTATAAGTCATC-3′. *Txk*; forward, 5′-ACCTGGAGAAGAGCTGCTATATC-3′; backward, 5′-TCT TTACTACGCAGGCAGAACT-3′; probe, 5′-CAGGGATCTGGCTGCCAGGA ACTG-3′. *Klf4*; forward, 5′-CAGTCGCAAGTCCCCTCTCT-3′; backward, 5′-TATCAAGAGCTCATGCCAC-3′; probe, 5′-CTCTTTGGCTTGGGCTCCT-3′. *Zfp36l1*; forward, 5′-CACGACACAACAGATCCTAGTCCTT-3′; backward, 5′-GAGCATCTTGTTACCCTTGCATAA-3′; probe, 5′-CCCCGTGCGAACGC CCAC-3′. *Il7r*; forward, 5′-AATGCCCAGGATGGAGATCTAGA-3′; backward, 5′-TGAGTGCTGATTTCCATCCACTT-3′; probe, 5′-ATGATCACTCCTTCT GGTGCCACAGCC-3′. *Tnfrsf14*; forward, 5′-CTGTCCCCCACAGACCTATA CC-3′; backward, 5′-CAGGCCCATGTCTGGATCAC-3′; probe, 5′-CCCATG CGAATGGCCTGAGCAA-3′. *Myc*; forward, 5′-CCCCTAGTGCTGCATGAA GAG-3′; backward, 5′-TCCACAGACACCACATCAATTTC-3′; probe, 5′-CA CCACCAGCAGCGACTCTGAAGAAG-3′. *Batf*; forward, 5′-GACAGCAGCGA CTCCAGCTT-3′; backward, 5′-TCTCTCCTCTGAACTTTCCTCACA-3′; probe, 5′-CCCCGGCAAGCAGGACTCATCC-3′. *Plek*; forward, 5′-CACGGA CTGGGAAATGAGGATA-3′; backward, 5′-TCCTCTCTATTTTCCTAGTT GTAGACTCTTGCA-3′; probe, 5′-AGAAGCATTGACGCGAATCCGACT-3′. *Rgs1*; forward, 5′-CAGATGCTGTGAAACAAATCAATATTG-3′; backward, 5′-CGTATATGACTTTTTGTGCTTCATCAA-3′; probe, 5′-TCGACAGCCA AGAAGATCAAAACACCAACT-3′. *Arl6ip4*; forward, 5′-CTCTGGATCAGTG GCACAGATC-3′; backward, 5′-CATGGCCTGGATACGAGACTTC-3′; probe, 5′-TGGCCCAGTCCTGACAGACGAGC-3′. For all TaqMan probes, a combination of 5′-FAM and 3′-TAMRA quencher was used.

Histology. Histological analysis was performed as described previously[8]. In brief, animals were perfused with 4% paraformaldehyde in PBS. The single organs were then post-fixed in the same buffer for 48 h at 4 °C. After a 24-h incubation period in 30% sucrose in PBS, the organs were embedded in Tissue Tek O.C.T Compound (Sakura Finetek), frozen and cut transversally into slices of 15-μm thickness using a Leica CM305S cryostat (Leica Microsystems) at −20 °C. Images of the slices were acquired using a Zeiss Axio Observer Fluorescence Microscope. Full-body cut preparation was accomplished as previously described[41]. In brief, after the perfusion and postfixation steps, de-calcification was performed by incubating the total body in 14% EDTA for 2 weeks. Several tiles were assembled using the photomerge function in Adobe Photoshop CS5.

Chemotactic assays. Directional migration of effector T cells was evaluated using 96-well trans-well plates (Corning Costar, 5-μm pore size) as described previously[40]. Briefly, 1×10^5 T_{GFP} cells *ex vivo* isolated or obtained from *in vitro* culture were loaded in the upper chamber. CCL5 (0.1 ng ml^{-1}, Serotec), CXCL-11

($0.1\,ng\,ml^{-1}$, R&D), CCL19 and CCL21 (both $0.1\,ng\,ml^{-1}$, Peprotech), or no chemokines were added to the lower chamber. The plates were incubated for 4 h at 37 °C. The number of effector T cells that had migrated through the membrane was counted by flow cytometry and quantified as a percentage of the input.

Antibody and peptide treatments. Hamster anti-rat CXCR3 blocking antibody (clone XR3.2, provided by T. Issekutz) was administered intravenously at $4\,mg\,kg^{-1}$ daily for the entire EAE course as described previously[23]. $Ninj1_{16–45}$ (PGSP GSPD ASPP RW GL RNRP INV NHYA NKK) or $Ninj1_{46–75}$ (SAAE SMLD IALL MANA SQLK AVVE QGNEFA) were synthesized from the peptide service of the Charité as described previously[42]. Anti-CD49d (integrin α_4, anti-VLA4, clone TA-2) antibodies were provided by T. Issekutz. For the motility studies, the peptides or the VLA4 MABs (both $1\,mg\,kg^{-1}$) were infused intravenously during intravital imaging. For the EAE studies, Ninj1 peptides or VLA4 MABs were administered intravenously at $1\,mg\,kg^{-1}$, starting 36 h after T-cell transfer, either three times or once per day, respectively.

Separation of bronchial and alveolar tissue. Lungs of naïve rats were inflated with low-melting agarose as described above. Slices were prepared and stained briefly with methyl blue. Using a Leica-wildM3Z stereomicroscope, bronchiolar or alveolar structures were prepared and transferred into lysis buffer (Qiagen RNeasy mini kit) at 4 °C. Tissue pieces were solubilized using a dounce homogenizer and RNA was processed as described for quantitative PCR.

PTX pre-treatment of T cells. $T_{MBP–GFP}$ blast cells (2 days after re-stimulation) were cultured for 1 h in the presence of $100\,ng\,ml^{-1}$ of PTX (List Biological Laboratories) or its B subunit (Calbiochem). These cells were used for transfer EAE. The chosen dose of PTX was able to block CXCL11- and CCL5-mediated trans-well transmigration of *in-vitro*- and *ex-vivo*-isolated $T_{MBP–GFP}$ cells.

FTY720 treatment. Rats were treated orally with FTY720 (PKF116-812-AA-2, Novartis Pharma) at $1\,mg\,kg^{-1}$ body weight by gavage, as described[43,44]. Treatment was started 1 day before T-cell transfer and was applied daily. EAE was induced as described above by the transfer of T cells.

Overexpression of *Ninj1* in effector T cells. *Rattus norvegicus* ninjurin 1 cDNA (NM_012867, 459 bp) was amplified by PCR using cDNA of effector T cells and the following primers: forward, 5′-ATGGATCCCGGCACCGAGGAGTACG-3′, backward, 5′-CTACTGCCGGGGTGCCACGTCCA-3′ and subsequently cloned in the pGEM-T easy Vector System (Promega). The EcoR1–EcoR1 fragment including *Ninj1* was then cloned upstream of IRES2–eGFP in the multiple cloning site of the murine stem cell virus pMSCVneo (Invitrogen). Retroviral vector construct $pMSCV_{neo}$-Ninj-1-IRES2eGFP was used to transduce $CD4^+$ T lymphocytes. Overexpression of *Ninj1* was confirmed in *in vitro* and *ex vivo* isolated cells by quantitative PCR and immunofluorescence. $T_{MBP–GFP}$ cells overexpressing *Ninj1* ($T_{MBP–GFP-Ninj1}$ cells) did not differ from the control $T_{MBP–GFP}$ cells (established in parallel) in terms of antigen responsiveness, surface antigen phenotype and cytokine production.

31. Eylar, E. H., Kniskern, P. J. & Jackson, J. J. Myelin basic proteins *Methods Enzymol.* **32,** 323–341 (1974).
32. Ben-Nun, A., Wekerle, H. & Cohen, I. R. The rapid isolation of clonable antigen-specific T lymphocyte lines capable of mediating autoimmune encephalomyelitis. *Eur. J. Immunol.* **11,** 195–199 (1981).
33. Kawakami, N. *et al.* Autoimmune CD4+ T cell memory: lifelong persistence of encephalitogenic T cell clones in healthy immune repertoires. *J. Immunol.* **175,** 69–81 (2005).
34. Odoardi, F. *et al.* Instant effect of soluble antigen on effector T cells in peripheral immune organs during immunotherapy of autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **104,** 920–925 (2007).
35. Pintelon, I. *et al.* Selective visualisation of neuroepithelial bodies in vibratome slices of living lung by 4-Di-2-ASP in various animal species. *Cell Tissue Res.* **321,** 21–33 (2005).
36. Huber, W., von Heydebreck, A., Sueltmann, H., Poustka, A. & Vingron, M. Parameter estimation for the calibration and variance stabilization of microarray data. *Stat. Appl. Genet. Mol. Biol.* **2,** article 3 (2003).
37. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
38. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18,** 207–208 (2002).
39. Dahlquist, K. D. *et al.* GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genet.* **31,** 19–20 (2002).
40. Cordiglieri, C. *et al.* Nicotinic acid adenine dinucleotide phosphate-mediated calcium signalling in effector T cells regulates autoimmunity of the central nervous system. *Brain* **133,** 1930–1943 (2010).
41. Reinhardt, R. L. *et al.* Visualizing the generation of memory CD4 T cells in the whole body. *Nature* **410,** 101–105 (2001).
42. Araki, T. *et al.* Mechanism of homophilic binding mediated by ninjurin, a novel widely expressed adhesion molecule. *J. Biol. Chem.* **272,** 21373–21380 (1997).
43. Brinkmann, V. *et al.* The immune modulator FTY720 targets sphingosine 1-phosphate receptors. *J. Biol. Chem.* **277,** 21453–21457 (2002).
44. Kataoka, H. *et al.* FTY720, sphingosine 1-phosphate receptor modulator, ameliorates experimental autoimmune encephalomyelitis by inhibition of T cell infiltration. *Cell. Mol. Immunol.* **2,** 439–448 (2005).

LETTER

doi:10.1038/nature11313

The prokaryote messenger c–di–GMP triggers stalk cell differentiation in *Dictyostelium*

Zhi-hui Chen[1] & Pauline Schaap[1]

Cyclic di-(3′:5′)-guanosine monophosphate (c-di-GMP) is a major prokaryote signalling intermediate that is synthesized by diguanylate cyclases and triggers sessility and biofilm formation[1,2]. We detected the first eukaryote diguanylate cyclases in all major groups of Dictyostelia. On food depletion, *Dictyostelium discoideum* amoebas collect into aggregates, which first transform into migrating slugs and then into sessile fruiting structures[3]. These structures consist of a spherical spore mass that is supported by a column of stalk cells and a basal disk. A polyketide, DIF-1, which induces stalk-like cells *in vitro*, was isolated earlier[4]. However, its role *in vivo* proved recently to be restricted to basal disk formation[5]. Here we show that the *Dictyostelium* diguanylate cyclase, DgcA, produces c-di-GMP as the morphogen responsible for stalk cell differentiation. *Dictyostelium discoideum* DgcA synthesized c-di-GMP in a GTP-dependent manner and was expressed at the slug tip, which is the site of stalk cell differentiation. Disruption of the *DgcA* gene blocked the transition from slug migration to fructification and the expression of stalk genes. Fructification and stalk formation were restored by exposing *DgcA*-null slugs to wild-type secretion products or to c-di-GMP. Moreover, c-di-GMP, but not cyclic di-(3′:5′)-adenosine monophosphate, induced stalk gene expression in dilute cell monolayers. Apart from identifying the long-elusive stalk-inducing morphogen, our work also identifies a role for c-di-GMP in eukaryotes.

Prokaryote diguanylate cyclases (DGCs) contain a signature GGDEF domain[1], which was recognized during annotation of the *D. discoideum* genome[6]. *Dictyostelium discoideum* resides in group four of the four major groups of Dictyostelia[7]. Query of genomes of species that represent the other three groups yielded single genes in *Dictyostelium lacteum* (group 3), *Polysphondylium pallidum* (clade 2B), *Acytostelium subglobosum* (clade 2A) and 13 homologous genes in *Dictyostelium fasciculatum* (group 1). In addition to the GGDEF domain, most proteins also contain a putative transmembrane helix (Fig. 1a). Alignment of the Dictyostelid GGDEF domains with the structurally resolved GGDEF domain of the *Caulobacter crescentus* DGC PleD[8] shows conservation of all residues that are essential for catalysis and binding to the substrate Mg^{2+}-GTP, except for a Gly-to-Ser substitution in the GGDEF domain (Supplementary Fig. 1a). The closest homologue of the Dictyostelid sequences is a DGC from the prokaryote *Gallionella capsiferriformans*. However, phylogenetic inference showed that the *G. capsiferriformans* DGC was still more similar to *C. crescentus* PleD, and that the Dictyostelid proteins formed a separate clade (Fig. 1a). The 13 *D. fasciculatum* GGDEF domain proteins are monophyletic and are most probably derived by gene duplications from a single ancestor (Supplementary Fig. 1c).

To identify a role for a putative DGC in Dictyostelid physiology or development, we disrupted *D. discoideum DgcA* by homologous recombination (Supplementary Fig. 2). The *DgcA*-null (*dgca*$^-$) mutant showed normal growth, aggregation and formation of migrating slugs, but could not form fruiting bodies (Fig. 1b and Supplementary Movie 1). Its slugs continued migration until they were exhausted. Fructification was fully restored by transforming *D. discoideum dgca*$^-$ cells with either the *D. discoideum* or *P. pallidum DgcA* coding sequence (Fig. 1c). This indicates that the fructification-deficient phenotype is due to loss of the *DgcA* gene and that *D. discoideum* and *P. pallidum DgcA* are functionally equivalent. We also tested the ability of a number

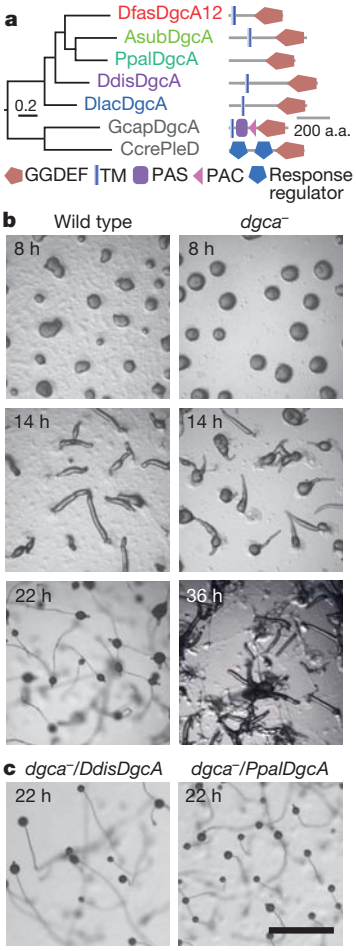


Figure 1 | Identification and disruption of DGCs. a, The *D. discoideum* (*Ddis*) genome[6] contained a single gene with a GGDEF domain. A BLAST search identified orthologues in *D. lacteum* (*Dlac*), *P. pallidum*[27] (*Ppal*) and *A. subglobosum* (*Asub*), and 13 monophyletic genes in *D. fasciculatum*[27] (*Dfas*). The phylogenetic tree shown was constructed from aligned Dictyostelid and prokaryote GGDEF sequences and annotated with the domain architecture of the proteins. a.a., amino acids; PAS, domain found in Per, Arnt and Sim proteins; PAC, motif C-terminal to PAS domain; TM, transmembrane helix. **b**, *Dictyostelium discoideum DgcA* was ablated by homologous recombination. *dgca*$^-$ and wild-type cells were incubated on non-nutrient agar to follow developmental progression. **c**, *Dictyostelium discoideum dgca*$^-$ cells were transformed with fusion constructs of the A15 promoter with either *D. discoideum DgcA* or *P. pallidum DgcA* and developed for 22 h. Scale bar, 1 mm.

[1]College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.

of truncated *DgcA* constructs to restore the *dgca*$^-$ phenotype. Constructs lacking the putative transmembrane helix and low-complexity regions flanking the helix still restored fructification in *dgca*$^-$ cells (Supplementary Fig. 3a). However, constructs lacking the GGDEF domain and/or 60 amino acids amino-terminal of this domain did not, indicating that the GGDEF domain is essential for DgcA function. The full-length gene fused to the yellow fluorescent protein (YFP) reporter was expressed throughout the cytosol (Supplementary Fig. 3b), confirming that the putative transmembrane helix did not function as such.

We next studied the expression of *DgcA* during *D. discoideum* development. Figure 2a shows that *DgcA* messenger RNA (mRNA) was absent from growing cells but increased on starvation to reach a maximum at 10 h when cells are forming slugs. The *dgca*$^-$ cells did not express *DgcA* mRNA, confirming disruption of the gene. To investigate which cells express *DgcA*, we transformed wild-type cells with a gene fusion of the 968-base-pair (bp) *DgcA* 5′ intergenic region and the *lacZ* reporter, and stained developing structures for β-galactosidase activity. Figure 2b shows that *DgcA* is first expressed in cells scattered throughout the aggregate. In slugs, *DgcA* expression is strongest at the anterior tip region (Fig. 2c), whereas in fruiting bodies the tip and stalk show high expression (Fig. 2d). This expression pattern suggests that DgcA functions in tip and stalk cells.

To assess whether the requirement of DgcA for fructification is cell-autonomous, we developed *dgca*$^-$ cells intermixed with increasing numbers of wild-type cells. Wild-type cells in a proportion of 10% were sufficient to restore fully fructification of *dgca*$^-$ cells (Fig. 3a), suggesting that the *dgca*$^-$ phenotype could be due to the lack of production of a secreted stimulatory signal. We tested this directly by incubating *dgca*$^-$ cells on agar that was preconditioned by secretion products from developing *dgca*$^-$, wild-type and *DgcA*-overexpressing cells. Supplementary Fig. 4a shows that agar preconditioned by wild-type cells or cells that overexpress *DgcA*, but not by *dgca*$^-$ cells, restored fructification of the *dgca*$^-$ mutant. The secreted stimulatory signal is likely to be c-di-GMP, the natural product of DGC. Cyclic di-GMP and structurally related compounds were deposited as droplets on top of *dgca*$^-$ slugs at 20 h of development. Cyclic di-GMP fully restored fruiting body formation (Fig. 3a), but solvent, c-di-AMP, cyclic GMP, 5′-GMP, GTP and GDP had no effect (Supplementary Fig. 4b).

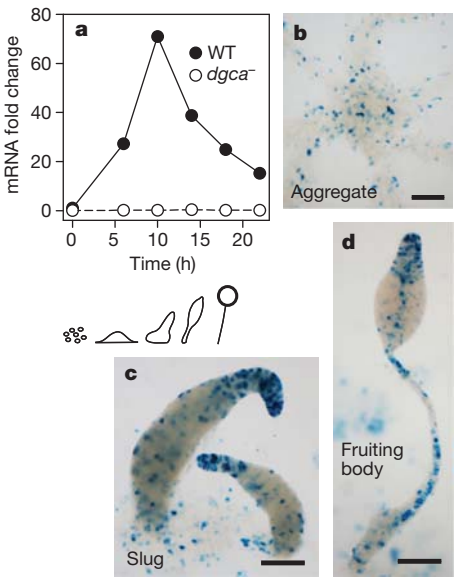


Figure 2 | *DgcA* expression pattern. **a**, Total RNA was isolated during development of *D. discoideum* wild-type (WT) and *dgca*$^-$ cells on non-nutrient agar. *DgcA* RNA levels were measured by real-time quantitative PCR (qRT–PCR) with *DgcA*-specific primers DgcAf and DgcAr (Supplementary Table 1). **b**–**d**, *Dictyostelium discoideum* cells were transformed with a fusion of 968-bp *DgcA* 5′ intergenic sequence and the *lacZ* reporter gene. β-galactosidase activity was visualized with X-gal in fixed aggregates (**b**), slugs (**c**) and fruiting bodies (**d**). Scale bar, 100 μm.

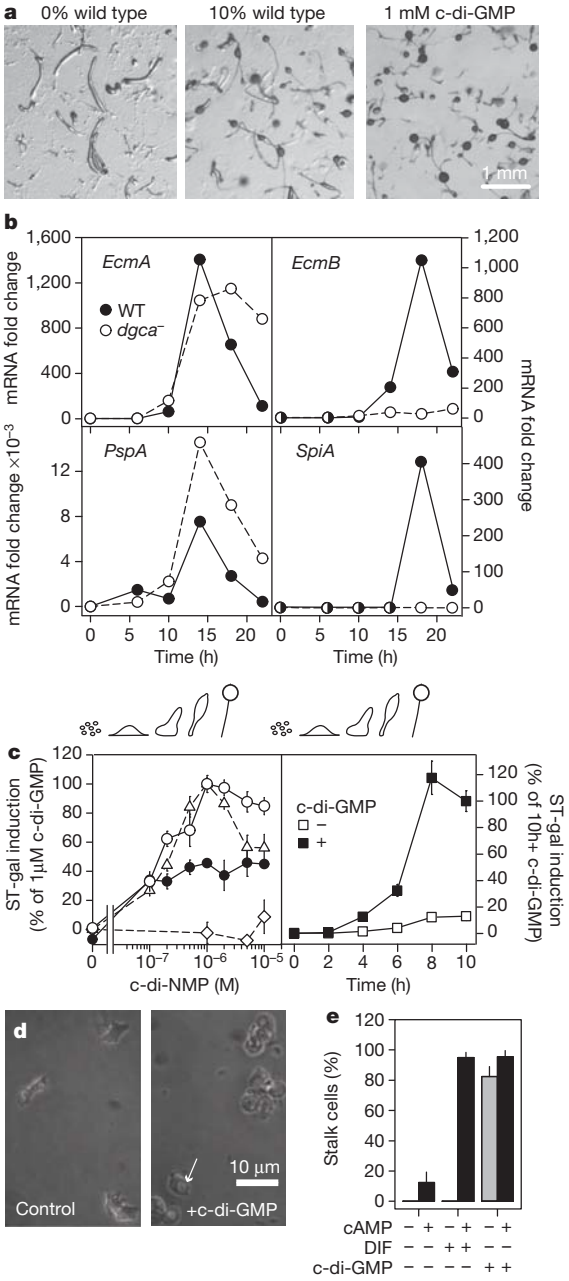


Figure 3 | Biological role of c-di-GMP. **a**, *dgca*$^-$ cells were mixed with 0 or 10% wild-type cells and developed for 24 h, or *dgca*$^-$ slugs were exposed to 1 mM c-di-GMP and observed after 10 h. **b**, The wild-type and *dgca*$^-$ RNA time series (Fig. 2a) were used to amplify *EcmA*, *EcmB*, *PspA* and *SpiA* by qRT–PCR using gene-specific primers (Supplementary Table 1). **c**, Dissociated aggregates of *dgca*$^-$ (open circles, diamonds, filled circles) or *dmta*$^-$ (triangles), transformed with ST-gal, were incubated at 10^6 cells ml^{-1} for 8 h with increasing concentrations of cyclic di-(3′:5′)-adenosine monophosphate (c-di-AMP; diamonds), c-di-GMP (open circles, triangles) or c-di-GMP and 1 mM cAMP (filled circles), followed by β-galactosidase assay. *dgca*$^-$/ST-gal was additionally incubated for variable time periods with (filled squares) or without (open squares) 1 μM c-di-GMP before β-galactosidase assay. Means and s.e.m. ($n = 9$). **d**, **e**, *Dictyostelium discoideum* V12M2 cells were incubated in monolayers[18] with 10 μM c-di-GMP, 100 nM DIF and/or 1 mM cAMP. After 30 h, cells were photographed (**d**) and the proportion of stalk cells to total cells was determined (**e**). Means and s.d. ($n = 2$).

Fructification initiates when the slug projects its tip and lays down a cellulose tube at its centre. The anterior prestalk cells then move into the tube and differentiate into stalk cells[9]. The *dgca*$^-$ slugs do project up occasionally, but do not follow this with stalk formation, suggesting that they fail either to form the stalk tube or to differentiate into stalk cells. Coordination of cellulose synthesis was the first established role of bacterial DGCs[10], and although the *D. discoideum* cellulose synthase, DcsA, is more similar to bacterial than eukaryote enzymes[11], it does not have the PilZ domain that mediates regulation by c-di-GMP, negating a role for c-di-GMP in cellulose deposition. We therefore focused on expression of marker genes for stalk and spore differentiation in the *dgca*$^-$ mutant. The prespore gene *PspA*[12]. was somewhat over-expressed in the *dgca*$^-$ mutant, whereas the prestalk gene *EcmA*[13]. was initially normally expressed but failed to be downregulated after 20 h, as occurred in wild-type cells. The spatial expression pattern of both genes in the slug stage was the same in *dgca*$^-$ and wild-type cells (Supplementary Fig. 5). In contrast, the spore gene *SpiA*[14]. was not expressed at all (Fig. 3b), and the stalk/basal disk gene *EcmB*[13]. was expressed very poorly. These data suggests that c-di-GMP could have a role in induction of stalk cell differentiation. Spores only mature when stalk formation is almost complete. The lack of *SpiA* expression could therefore be a derived effect.

To test acute effects of c-di-GMP on stalk gene expression, we used cells transformed with a gene fusion of the β-galactosidase reporter gene with the proximal stalk-specific region of the *ecmB* promoter (ST-gal), which was defined previously[15,16]. Aggregates were dissociated into single cells, which were incubated with increasing concentrations of c-di-GMP. Figure 3c shows that c-di-GMP activates ST-gal expression within 8 h, with half-maximal induction occurring at a c-di-GMP concentration of ~200 nM. Cyclic di-AMP had little effect up to 10 μM, and cyclic AMP (cAMP) partly inhibited ST-gal induction by c-di-GMP. Cyclic di-GMP also activated ST-gal expression in *dmta*$^-$ cells that cannot synthesize DIF-1 (ref. 17), indicating that c-di-GMP does not exert its effects by activating DIF-1 synthesis.

We also tested whether c-di-GMP could induce fully differentiated stalk cells in monolayers of *D. discoideum* V12M2 cells[18]. Figure 3d shows that after 30 h of c-di-GMP treatment, V12M2 cells had formed the central large vacuole (arrow) that characterizes stalk cells. In the V12M2 test system, DIF-1 also induces stalk cell differentiation, but additionally requires the presence of 1–5 mM cAMP. However, this is not the case for c-di-GMP (Fig. 3e), on which cAMP, if anything, has an inhibitory effect, as was also evident in the ST-gal induction assay (Fig. 3c). The fact that DIF-1 requires cAMP as a co-factor is probably because it actually induces basal disk cells[5], which are morphologically similar to stalk cells. In normal development, basal disk cells are derived from the prespore cell population, which themselves require cAMP for differentiation[19]. Because the spore gene *SpiA* was also not expressed in the *dgca*$^-$ mutant, we tested whether c-di-GMP was required for or could induce spore formation in monolayers. However, neither proved to be the case (Supplementary Fig. 6), highlighting the specific role for c-di-GMP in stalk formation.

Induction of ST-gal gene expression provided us with a bioassay to ascertain that *D. discoideum* DgcA has diguanylate cyclase activity. A fusion protein of YFP and the essential carboxy-terminal region of DgcA (ΔTL-DgcA–YFP; Supplementary Fig. 3) was purified from *D. discoideum* cell lysates by immunoprecipitation with YFP antibody. The immunoprecipitate was incubated with the DGC substrate Mg^{2+}-GTP for 60 min and then tested for ST-gal induction activity. Figure 4 shows that the ΔTL-DgcA–YFP immunoprecipitate synthesized a stalk-gene-inducing factor in a time-, GTP- and Mg^{2+}-dependent manner. The factor was confirmed to be c-di-GMP by mass spectrometry (Supplementary Fig. 7).

We showed that species representing all major groups of Dictyostelia contain one or more conserved diguanylate cyclases that were previously only found in eubacteria. *Dictyostelium discoideum* DgcA is essential for stalk formation and this requirement can be replaced by secretion products of wild-type cells, by c-di-GMP and by DgcA from a distantly related Dictyostelid. *DgcA* is expressed in the organizing tip of structures, where stalk cell differentiation is initiated. Combined, these data indicate that c-di-GMP is an apically secreted conserved signal for the induction of stalk cell differentiation in Dictyostelia. The polyketide DIF-1, which was identified after painstaking isolation from *D. discoideum* cells, has been for 30 years the primary candidate for a stalk-inducing factor[4]. However, identification and abrogation of its synthetic enzymes revealed that it was required for formation of the basal disk but not for formation of the stalk[5,17].

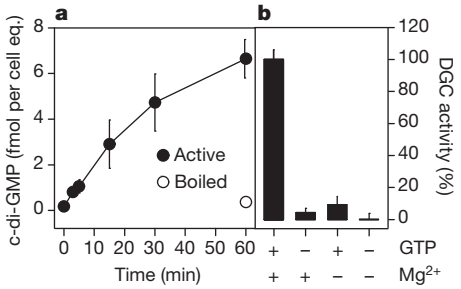


Figure 4 | Bioassay of DGC activity. **a**, Preboiled or active immunoprecipitates of the ΔTL-DgcA–YFP fusion protein were incubated with 0.1 mM GTP and 10 mM $MgCl_2$ at 22 °C. After boiling, reaction mixtures were tested for ST-gal induction activity. Cyclic di-GMP concentrations in the mixtures were estimated by comparison with a dose–response curve of ST-gal induction by c-di-GMP, and standardized on the amount of cells from which the immunoprecipitate was derived. Cell eq., cell equivalent. **b**, DGC activity was measured for 60 min with combinations of 0.1 mM GTP and 10 mM $MgCl_2$, as indicated. Data represent means and s.e.m. of two experiments assayed in triplicate.

Cyclic di-GMP and its synthetic enzyme DGC appeared more recently at the forefront of bacterial signal transduction. Cyclic di-GMP has a principal role in the loss of motility and the synthesis of adhesins and exopolysaccharide matrix components by bacteria, which marks their transition from a swarming phase to a sessile biofilm associated life style. In *C. crescentus*, this also involves formation of a stalk[1,2]. In Dictyostelia, c-di-GMP also triggers the transition from motile slugs into sessile fruiting bodies, and the question arises of whether this represents either convergent evolution or deep ancestral connections between Dictyostelid stalk formation and bacterial sessility. In the second case, DGCs should also be present in other Amoebozoa, such as the polyphyletic Protosteloid amoebas, which differentiate into a single spore sitting on its own stalk[20]. No protostelid genome sequences are yet available to address this issue. Of immediate interest are the questions of how Dictyostelid DGCs are themselves regulated and how the c-di-GMP signal is processed to activate stalk gene expression. These questions will be the topic of our future research.

METHODS SUMMARY

Expression constructs. The full length *D. discoideum* and *P. pallidum* DgcA coding regions were amplified by PCR from *D. discoideum* and *P. pallidum* genomic DNA, using primer pairs DgcF/DgcR and PpDgcF/PpDgcR (Supplementary Table 1), respectively. PCR products were digested with BamHI and XhoI and inserted into vector pB17S-EYFP[21], which was transformed into *D. discoideum* *dgca*$^-$ cells.

Fusion of *DgcA* promoter and *lacZ*. The 968-bp *DgcA* 5′ intergenic region was amplified from genomic DNA with primers DgcApr5′ and DgcApr3′ (Supplementary Table 1) and inserted into pDdGal-17 (ref. 22). After transformation in *D. discoideum*, β-galactosidase activity was visualized with X-gal in developing structures[23].

RNA analysis. RNA was extracted using RNeasy kits (Qiagen) with on-column DNA digestion, and transcribed into complementary DNA using ImPromII reverse transcriptase (Promega). Real-time quantitative PCR was performed using iQ SYBR Green Supermix (Bio-Rad). Gene expression levels were normalized to expression of the constitutively expressed *IG7* gene[24].

Gene induction assay. *dgca*$^-$ cells, transformed with ST-gal[15,16], were incubated with variables for up to 10 h. Cells were lysed and spectrophotometrically assayed for β-galactosidase activity[25].

Immunoprecipitation. Cells transformed with ΔTL-DgcA–YFP (Supplementary Fig. 3) were filter lysed in DGC assay buffer[26]. Cleared supernatant (500 μl) was incubated with 20 μl GFP-Trap_A beads (Chromotek) for 60 min. After three washes with 150–500 mM NaCl, beads were resuspended in 20 μl assay buffer.

DGC assay. Immunoprecipitate (5 μl) was incubated at 22 °C with 95 μl of 0.1 mM GTP and 10 $MgCl_2$ (ref. 26) in DGC assay buffer. Reactions were terminated by boiling. The reaction mixture (10 μl) was incubated for 8 h with 90 μl of *dgca*$^-$/ST-gal cells and assayed for β-galactosidase activity.

Received 17 February; accepted 11 June 2012.
Published online 5 August 2012.

1. Jenal, U., &. Malone, J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.* **40,** 385–407 (2006).
2. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nature Rev. Microbiol.* **7,** 263–273 (2009).
3. Schaap, P. The evolution of size and pattern in the social amoebas. *Bioessays* **29,** 635–644 (2007).
4. Morris, H. R., Taylor, G. W., Masento, M. S., Jermyn, K. A. & Kay, R. R. Chemical structure of the morphogen differentiation inducing factor from *Dictyostelium discoideum. Nature* **328,** 811–814 (1987).
5. Saito, T., Kato, A. & Kay, R. R. DIF-1 induces the basal disc of the *Dictyostelium* fruiting body. *Dev. Biol.* **317,** 444–453 (2008).
6. Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum. Nature* **435,** 43–57 (2005).
7. Schaap, P. *et al.* Molecular phylogeny and evolution of morphology in the social amoebas. *Science* **314,** 661–663 (2006).
8. Chan, C. *et al.* Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl Acad. Sci. USA* **101,** 17084–17089 (2004).
9. Raper, K. B. & Fennell, D. I. Stalk formation in *Dictyostelium. Bull. Torrey Botan. Club* **79,** 25–51 (1952).
10. Ross, P. *et al.* Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325,** 279–281 (1987).
11. Blanton, R. L., Fuller, D., Iranfar, N., Grimson, M. J. & Loomis, W. F. The cellulose synthase gene of *Dictyostelium. Proc. Natl Acad. Sci. USA* **97,** 2391–2396 (2000).
12. Early, A. E. *et al.* Structural characterization of *Dictyostelium discoideum* prespore-specific gene D19 and of its product, cell surface glycoprotein PsA. *Mol. Cell. Biol.* **8,** 3458–3466 (1988).
13. Williams, J. G. Transcriptional regulation of *Dictyostelium* pattern formation. *EMBO Rep.* **7,** 694–698 (2006).
14. Richardson, D. L. & Loomis, W. F. Disruption of the sporulation-specific gene spiA in *Dictyostelium discoideum* leads to spore instability. *Genes Dev.* **6,** 1058–1070 (1992).
15. Ceccarelli, A., Zhukovskaya, N., Kawata, T., Bozzaro, S. & Williams, J. Characterisation of a DNA sequence element that directs *Dictyostelium* stalk cell-specific gene expression. *Differentiation* **66,** 189–196 (2000).
16. Ceccarelli, A., Mahbubani, H. & Williams, J. G. Positively and negatively acting signals regulating stalk cell and anterior-like cell differentiation in *Dictyostelium. Cell* **65,** 983–989 (1991).
17. Thompson, C. R. & Kay, R. R. The role of DIF-1 signaling in *Dictyostelium* development. *Mol. Cell* **6,** 1509–1514 (2000).
18. Brookman, J. J., Town, C. D., Jermyn, K. A. & Kay, R. R. Developmental regulation of stalk cell differentiation-inducing factor in *Dictyostelium discoideum. Dev. Biol.* **91,** 191–196 (1982).
19. Wang, M., Van Driel, R. & Schaap, P. Cyclic AMP-phosphodiesterase induces dedifferentiation of prespore cells in *Dictyostelium discoideum* slugs: evidence that cyclic AMP is the morphogenetic signal for prespore differentiation. *Development* **103,** 611–618 (1988).
20. Shadwick, L. L., Spiegel, F. W., Shadwick, J. D., Brown, M. W. & Silberman, J. D. *Eumycetozoa = Amoebozoa*?: SSUrDNA phylogeny of protosteloid slime molds and its significance for the amoebozoan supergroup. *PLoS ONE* **4,** e6754 (2009).
21. Meima, M. E., Biondi, R. M. & Schaap, P. Identification of a novel type of cGMP phosphodiesterase that is defective in the chemotactic *stmF* mutants. *Mol. Biol. Cell* **13,** 3870–3877 (2002).
22. Harwood, A. J. & Drury, L. New vectors for expression of the *E.coli lacZ* gene in *Dictyostelium. Nucleic Acids Res.* **18,** 4292 (1990).
23. Dingermann, T. *et al.* Optimization and in situ detection of *Escherichia coli* beta-galactosidase gene expression in *Dictyostelium discoideum. Gene* **85,** 353–362 (1989).
24. Williams, J. G. *et al.* Direct induction of Dictyostelium prestalk gene expression by DIF provides evidence that DIF is a morphogen. *Cell* **49,** 185–192 (1987).
25. Schaap, P. *et al.* Cell-permeable non-hydrolyzable cAMP derivatives as tools for analysis of signaling pathways controlling gene regulation in *Dictyostelium. J. Biol. Chem.* **268,** 6323–6331 (1993).
26. Paul, R. *et al.* Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* **18,** 715–727 (2004).
27. Heidel, A. *et al.* Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res.* **21,** 1882–1891 (2011).

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

Acknowledgements We thank R. R. Kay for alerting us to the presence of a putative diguanylate cyclase in the *D. discoideum* genome. U. Jenal is gratefully acknowledged for advice and an EAL-PDE construct in an early phase of the project. We are grateful to the late H. MacWilliams for plasmids PspA-ile-gal, EcmA-ile-gal and PstO-ile-gal, and to C.Thompson for *dmta*$^-$ cells. We thank W. Chen and D. Lamont for mass spectrometry and C. Sugden for guidance with qRT–PCR. We are grateful to H. Urushihara and the *A. subglobosum* genome project (http://acytodb.biol.tsukuba.ac.jp/cgi-bin/index.cgi?org=as) for the *A. subglobosum DgcA* sequence. This research was supported by Wellcome Trust Project grant 090276 and BBSRC grant BB/G020426.

Author Contributions Z.C. and P.S. designed the experiments and wrote the manuscript. Z.C. performed the experiments.

Author Information DNA sequences for *D. lacteum* and *A. subglobosum DgcA* have been submitted to Genbank under accession numbers JQ676836 and JQ676837, respectively. 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 P.S. (p.schaap@dundee.ac.uk).

M. WIGGINS/IKON/CORBIS

JOB SATISFACTION

Turbulent times

Nature's 2012 Salary and Satisfaction Survey suggests that many scientists are content with their work, but uneasy about finances.

BY GENE RUSSO

Stacy Gelhaus has given a lot of thought to the career prospects of postdocs. Two years ago, she was chairwoman of the board of directors for the US National Postdoctoral Association (NPA) in Washington DC — so she knows that positions in academia come up relatively rarely.

And yet despite the risks, she chose to pursue academia after her own postdoctoral fellowship. Now an assistant research professor in pharmacology at the University of Pittsburgh in Pennsylvania, she is living through some of the issues that she explored as an NPA officer, including the challenges of publishing, acquiring funding and balancing the demands of work and personal life. Her job is not yet on the tenure track — perhaps, she says, securing a good publication record and a sought-after grant will give her a chance to change that. For now, Gelhaus relishes having relatively few administrative duties. But, she says, she is already finding herself "competing for grants in a situation where the overall funding dollar is decreasing as well as the number of tenure-track positions".

Gelhaus enjoys her work, but uncertainty about the future has generated some anxiety. As an NPA officer, she learned that all career tracks have their downsides. "It really comes down to what you can live with, what you are willing to accept," she says. "I'm sure that everyone has different deal-breakers."

Gelhaus's experience is common. The ▶

NATIONAL DIFFERENCES

The geography of happiness

Which nation has the most contented scientists? Among the countries with considerable numbers of respondents to *Nature*'s survey, Sweden, Switzerland and Denmark earned top honours. Globally, 67% of respondents were satisfied to some degree with their jobs, in comparison with 86% in Switzerland, 82% in Denmark and 81% in Sweden.

Although salary is not the primary driver of satisfaction, it is a contributing factor. Some 78% of Swiss respondents were somewhat or very satisfied with their salaries, whereas UK and US respondents were closer to the global average of 36%. We found that Swiss earnings are indeed among the highest, with an average of approximately US$104,000, although this did not take into account differences in the cost of living. The mean for industry positions in Europe was about $86,000, in comparison with $64,000 in academia.

Denmark, Sweden and Switzerland also had the fewest respondents agreeing that the recession has had a negative effect on satisfaction; they were joined in that by Germany, Australia and Brazil. Not surprisingly, given the state of their economies, Spain and Portugal had the most.

Ralph Eichler, president of the Swiss Federal Institute of Technology in Zurich, attributes Swiss satisfaction to reliable government funding and a good quality of life. (One measure of quality of life is a country's 'happiness index', tracked in the World Database of Happiness (http://www1.eur.nl/fsw/happiness/); our results suggest that this is often highest in the countries with the most contented scientists.) But Eichler does not take his country's position for granted. "Switzerland is in a very good economic state," he says. But because exports are important to the nation, it "depends on the world economic status". Swiss scientists also reported high satisfaction with salary and independence in *Nature*'s 2010 salary survey, as did those in Denmark, the Netherlands, Sweden and Germany.

Several countries reported that the amount of guidance at work was a problem. About 59% of respondents from Japan said that it had a negative impact on satisfaction, as did half of UK and German scientists. In general, guidance was a bigger concern for postdocs and assistant and associate professors than for those in more senior academic positions.

As in 2010, satisfaction with 'degree of independence' was high across the board, with more than 70% of respondents somewhat or very satisfied in most countries. India, China and Japan scored the lowest, with just 48% satisfied in Japan. Although scientists relish the freedom to pursue projects independently, they also seek mentors, as many made clear (a finding also familiar from *Nature*'s 2011 survey of graduate students[3]).

China and India both rated worse than the global average for scientists who were 'not at all satisfied' with salary, guidance, total hours worked and independence. India also rated much worse than the global average for benefits such as retirement plan and health care. But scientists perceive improvements: India and China were the only countries in which satisfaction had increased for more than 50% of respondents in the past year.

When the results are broken down by continent and career stage, satisfaction looks similar in 2010 and 2012. However, the proportion of associate professors in North America reporting satisfaction has dipped from 75% to 68%, perhaps because of concerns about budgets and funding. Satisfaction for full professors also fell slightly. By contrast, the proportion of satisfied postdocs in Asia has risen, from 44% to 51%.

Martin Hynes, chief executive of the European Science Foundation in Strasbourg, France, notes that in the 2012 results, as in 2010, satisfaction for assistant professors and lecturers in Europe is lower than for those in North America, possibly because of lack of empowerment or tenure-track positions. But he is surprised that in both years, the satisfaction levels of full professors converged at a relatively high rating across continents. "It is remarkable that they all cluster so closely," he says. "Where is the image of the grumpy, eminent professor?" G.R.

▶ results of *Nature*'s 2012 Salary and Satisfaction Survey show that most scientists feel engaged and pleased with their research, while tolerating modest salaries and benefits. But even those lucky enough to secure an academic post often remain uneasy about their future (see 'Income, independence and instability').

ECONOMY EXACERBATES ANXIETIES

The survey paints a sobering picture of concerns for many scientists as the global recession squeezes research budgets around the world. It also reinforces the overall conclusion of *Nature*'s previous survey[1], two years ago: when it comes to job satisfaction for scientists of all stripes, factors such as the independent pursuit of research and the prospects for learning and advancement are generally as important or more important than salary concerns.

More than 11,500 scientists from various career stages and academic fields replied to this year's survey. We had respondents from more than 100 nations, including considerable sample sizes from more than a dozen (see 'The geography of happiness'). The results illustrate challenges and illuminate trends in the global scientific workforce.

The financial and other hurdles that researchers face are not new — especially in academia. But the precarious state of the global economy has intensified them. The 2012 survey included a suite of questions about the effects of the recession on job satisfaction.

66% of respondents were somewhat or very satisfied with their jobs

Some respondents expressed despair. One said: "The shortfall in grant funding and the decline in state revenues both mean substantially more work writing grants that have a lower chance of funding. The increasing administrative burden also adds to the list of tasks that must be completed in the same amount of time. Very demoralizing and exhausting." Another wrote: "Universities are under more 'performance' pressure, leading to more interference from management."

But several respondents said that although the recession and reduced funding levels were worrying, they had little effect on job satisfaction. One wrote: "My job satisfaction is based on my intellectual freedom. There is no connection between the markets and the pursuit of my scientific interests." And location matters. Another respondent, mourning his country's economic contraction, said: "I am a Spaniard living in Spain, bearing the various penances this implies."

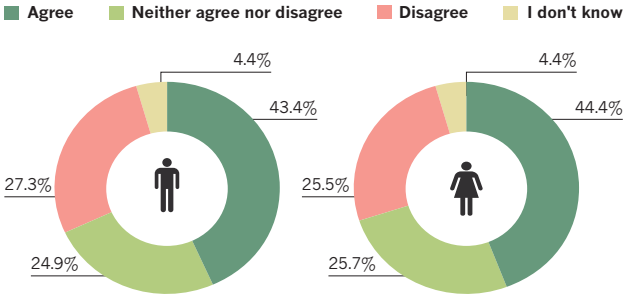
The more senior the scientists, the more likely they were to report that the recession had negatively affected their job satisfaction. Thirty-six per cent of respondents who had earned their highest degree in the past one to two years agreed strongly or to some extent that "the global recession has had a negative impact on my overall job satisfaction". By 11–15 years after earning the degree, agreement was 46%. Among those who had received their ▶

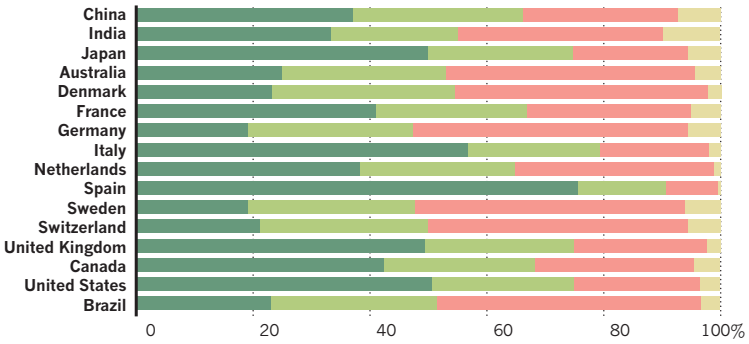
Income, independence and instability

Selected statistics from *Nature*'s international Satisfaction and Salary Survey illustrate the factors that affect the outlooks of scientists at various stages of their careers. The survey had more than 11,500 responses from researchers.

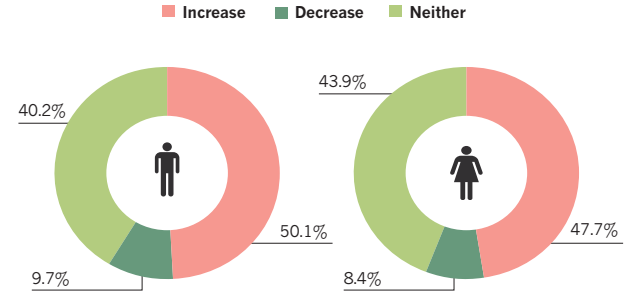
Q To what extent do you agree or disagree with the following statement?

"The global recession has had a negative impact on my overall job satisfaction."

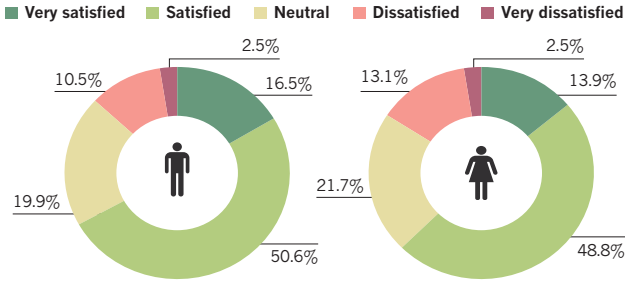




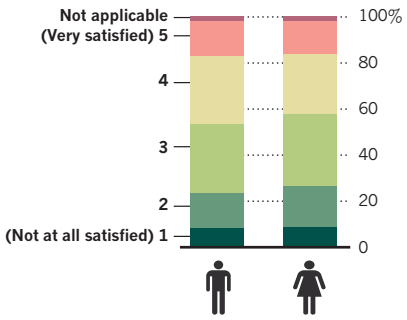
Q Did you receive an increase or decrease to your base salary in 2011?



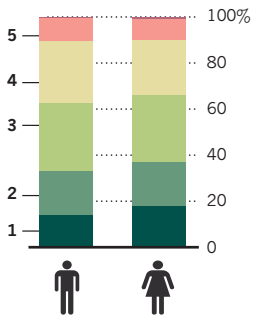
Q Overall, how satisfied are you with your current job?



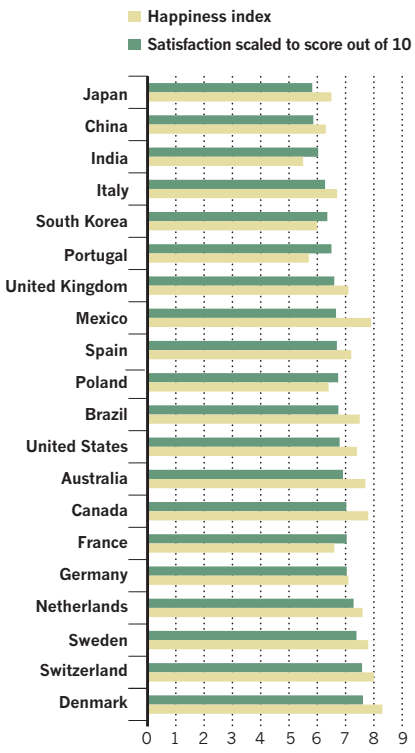
Q To what extent are you satisfied with the total hours worked per week in your primary job?



Q To what extent are you satisfied with the salary in your primary job?

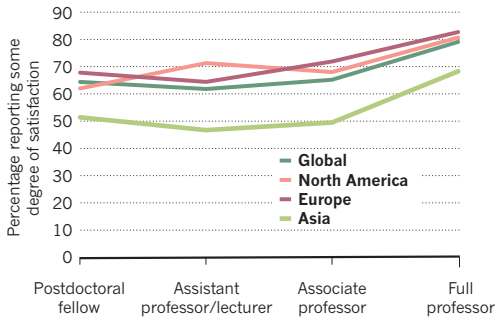


OVERALL JOB SATISFACTION BY COUNTRY

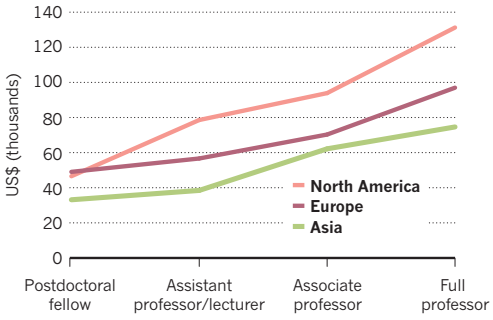


Salaries not adjusted for costs of living. Source: (happiness index) World Database of Happiness, Erasmus Univ. Rotterdam.

LEVELS OF SATISFACTION BY CONTINENT AND ACADEMIC CAREER STAGE



MEAN SALARIES BY BY CONTINENT AND ACADEMIC CAREER STAGE



GENDER

Unequal pay, similar satisfaction

Large salary disparities persist between male and female researchers, the results of *Nature*'s survey suggest. But the gulf did not seem to affect most satisfaction ratings.

In some countries, the mean salaries reported for women were 75% of those reported for men, or even less. But several factors could contribute to that. From some nations, for example, we received many more responses from male senior scientists than from female senior scientists, and the apparent salary gulf could reflect this sampling bias. But that bias in itself could reflect the trend of more men advancing to higher positions, rather than suggesting that men are paid more for a given post — although both trends are troubling.

Other numbers suggest more similarities than divides. Some 50% of men reported a pay rise in 2011, in comparison with 48% of women. Yet despite apparent salary discrepancies, women do not report much less satisfaction with their earnings. About 37% of men said they were somewhat or very satisfied, as did 33% of women.

Other attributes also showed comparable satisfaction. Men were slightly more content with the total number of hours worked per week (44% versus 40% for women), but women were more likely to be pleased with guidance from superiors or co-workers (35% of men were satisfied to some degree, in comparison with 40% of women).

Nor did gender make a difference to the perceived impact of the global recession: 43% of men agreed that the recession had reduced overall job satisfaction, as did 44% of women. G.R.

degrees 16 years ago or more, it was 50%.

Responses to the statement "I feel that my salary would have increased in 2011 were it not for the global recession" showed a similar pattern, although this is probably partly because the salaries of postdocs, from whom we received many responses, vary little regardless of the state of the economy.

Scientists at all career stages seemed to be satisfied with most other aspects of their jobs — and, in many respects, men and women had similar satisfaction levels (see 'Unequal pay, similar satisfaction'). As in the 2010 survey, scientists were generally very happy with their degree of independence. The proportion increased as career stage advanced: 70% of postdocs are somewhat or very satisfied with their degree of independence, in comparison with 87% of professors.

NOW AND THEN

The preliminary results of a separate, ongoing survey by Euroscience, a non-profit organization in Strasbourg, France, that represents scientists and promotes science in Europe, underscore the divide between current job satisfaction and anxieties about the future among early-career scientists. On the basis of responses from young European academic researchers, it found that nearly 80% expected still to be in academia in one year's time. But only 40% saw themselves still working at a research institute or university in five years. "I was satisfied as a postdoc, but definitely had anxiety looking ahead," says Gelhaus. At the time, she adds, "there was a postdoc office and benefits were good, so there wasn't really too much to be unhappy about."

"There are a lot of existential fears," agrees Ruth Müller, a research fellow focusing on science and technology studies and innovation at the Austrian Institute of International Affairs in Vienna. "Will this translate into an occupation that will feed me, allow me to have a family?" Müller and her research team have completed more than 50 in-depth interviews with life scientists at different career stages to investigate how professional pressures affect their work practices and working relationships, especially during the postdoc period. One of her findings[2] was that the compulsion to pursue first-author publications often leads postdocs to eschew collaborations.

But, as with *Nature*'s survey, Müller and her colleagues found that early-career researchers were passionate about their work. And it's clear from the *Nature* survey results on satisfaction with various job attributes that money is not the most important factor at any career stage, and especially among early-career researchers.

Still, economic conditions have heightened anxieties. "I do not think the recession dramatically impacts me as a postdoc currently," noted one respondent to *Nature*'s survey. "But the recession does add to my anxiety about finding a faculty position in an increasingly competitive job market with no clear indication that funding will increase in the near future." ■

Gene Russo *is* Nature*'s Careers editor.* **Fiona Watt** *and* **Laura Harper** *contributed to the study design and data analysis.*

1. Russo, G. *Nature* **465,** 1104–1107 (2010).
2. Müller, R. *Higher Edu. Policy* **25,** 289–311 (2012).
3. Russo, G. *Nature* **475,** 533–535 (2011).

GROWTH

The bitter taste of success.

BY WILLIAM MEIKLE

Climate change became moot on the 24 June 2026. The Sun went dim early that morning, and never recovered. One month into the *dimming*, it was all too obvious that the problem was very real. Crops were going to fail all across the Northern Hemisphere and NASA told us that it wasn't going to get better any time soon.

That's where I came in.

I'd spent the past 20 years in dingy laboratories trying to perfect a cheap foodstuff. My focus was on something that could provide for long, deep-space missions, and I had been quite successful with a genetically modified fungus I had developed that grew in the dark and was spectacularly prolific. Someone in NASA took note, put two and two together, and suddenly I had money, resources and lab techs thrown at me; it became my job to save the world.

JACEY

I don't need to remind you of that first winter. I was shielded from the worst of it by my new-found security, which meant I was warm, fed and safe; three things that became a distant memory for many that season. As matters in the wider world worsened, I took to ignoring the news broadcasts and lost myself in the tedium of lab work. Men in suits tried, in increasingly shrill tones, to hurry me up, and attempts were made first to bribe, then to bully me into cutting corners; to get *the product* into production as soon as possible.

It wasn't any coercion that finally swayed me into releasing the samples to the wider community. What did it was the pictures that I could no longer avoid; of rioting in the streets of the cities, of the forests of the Amazon and New Guinea rotting in the darkness, of vast hordes of people on the move, like migrating wildebeest. It was my job to save the day.

So I tried.

It almost worked, for a while. I released *Boletus edulis Watsonii* to select labs around the world. They in turn were quick to start production. Soon there were factories all over the planet dedicated solely to the growth and distribution of my new mushroom. In the first summer after the dimming my creation managed to keep a proportion of the population alive.

Others were not so fortunate. Wars raged across much of Africa and the Middle East, a new plague hit South America, and it was estimated that more than a billion people died in the year after the Sun went dark.

But my *Boletus* had given the survivors hope. It might not taste of much, but it was plentiful, and filled empty stomachs well enough. For a while the governments of the industrialized nations even started to think that they had matters under some kind of control.

Until the escapes.

No one will ever discover where or how it happened. Given our propensity for catastrophe, I suspect that there was not any single source of the outbreak. But however it happened, the *Boletus*, in spore form at first, escaped into the wider environment. It quickly discovered that it liked what it found there. The new dark, wet environment under the dimmed Sol proved to be perfect conditions for its growth, and there was no shortage of rotting vegetation for the mycelia to feed on. By the time the second winter came round, my *Boletus* was on its way to becoming the dominant form of vegetation on the planet.

Even then the powers that be were not unduly worried; well, not about the spread of the fungus. They had other concerns by then, as populations tried to shift to the equator to take advantage of the slightly warmer weather there. More wars ensued, as they usually do. All we managed to do was put more ash, smoke and particulates into the atmosphere, ensuring that the cooling accelerated.

By halfway through the second winter, anything that wasn't covered in ice was being eaten by the *Boletus*. It was only then that the full implications of rushing its release became apparent. My creation discovered that it not only liked rotting vegetation, but that the mycelium could grow just as well on, and through, any organic material.

I saw the first pictures to come in. A mound of bodies lay in an empty town square. The camera zoomed in to show mycelia spreading in white spider's web profusion over all areas of exposed flesh. A time-lapse segment of footage showed the fruiting bodies burst wetly from arms, legs and faces to spread their parasols high. A breeze came up, and the view was filled with a fine powder, quickly dispersed as the spores went looking for fresh feeding grounds.

I've been in prison since then, awaiting trial, charged with crimes against humanity. Any news I've had has been at third-hand from my jailers, but they speak of a planet going to rot, of a rampant mycelium infection that resists all antibiotics, is immune to all known treatments.

They say I will die for what I have done. But I know that already, for today I found a small, but perfectly formed, fruiting body between my toes. It's only a matter of time.

They have allowed me one last look out of the window. My prison is high above what used to be Manhattan. It is now a beautiful field of *Boletus* parasols swaying in the wind, stretching as far as the eye can see.

I can die happy.

I may not have saved mankind.

But it looks like I have saved the planet. ■

➲ NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mtoodm

William Meikle *is a Scottish writer resident in Canada with 10 novels published in the genre press and more than 200 short-story credits in 13 countries.*