Information overload

A report released last week by the US National Academies makes recommendations for tackling the issues surrounding the era of petabyte science.

Geneticists spent more than a decade getting their first complete reading of the 3 billion base pairs of the human genome, which they finally published in 2003. But today's rapid sequencing machines can run through that much DNA in a week, and are busily churning out multiple sequences from an ever-expanding list of species. Meanwhile, astronomers working with the Sloan Digital Sky Survey telescope in New Mexico have mapped some 25% of the sky since 2000, obtaining data on more than 200 million objects. The Large Synoptic Survey Telescope, scheduled for completion atop Chile's Cerro Pachón in 2015, will gather that much data in one night.

Statistics tell a similar story in many scientific fields. This is great news for research: data glut is always better than data famine. But it is also cause for concern, because investigators' ability to amass huge quantities of data has accelerated much faster than have policies and practices for handling those data. Journal editors, in particular, have found themselves grappling with issues such as image manipulation, the preservation of original data, assuring continued access to large data sets, and standards for algorithm and code sharing.

In 2006, these concerns led a number of scientific societies and research journals, including *Nature*, to ask the US National Academy of Sciences to look at the problem. This resulted in the formation of a National Academies study committee, sponsors of which included Nature Publishing Group. The committee was headed by cancer researcher Phillip Sharp and physicist Daniel Kleppner, both of the Massachusetts Institute of Technology in Cambridge, and its report was published on 22 July (see http://tinyurl.com/datasteward).

The report makes 11 recommendations, organized around three major principles: integrity, access and stewardship. The integrity principle affirms that each researcher is ultimately responsible for ensuring the truth and accuracy of the data he or she produces. Individual investigators should adhere to the professional standards in their fields, and institutions should ensure that training is in place to make this possible.

The access principle asserts the value of openness: only if results are shared can other researchers check the data's accuracy, verify analyses and build on previous work. So unless there are very good reasons for researchers to withhold data — reasons that should be publicly posted and available for comment by other researchers — they should make provisions to supply public access in a timely manner, possibly as early as their grant proposals.

Finally, the stewardship principle addresses the need for long-term preservation. Scientific societies and communities need to provide guidelines on which data are worth retaining for future analysis; institutions and funding agencies need to address and support these needs. Journals can play a part in the preservation of the published record, and in the dissemination and enforcement of guidelines. And data professionals should be recognized for their crucial role in stewardship: certainly they deserve more respect and support than researchers sometimes give them.

> "Each researcher is ultimately responsible for ensuring the truth and accuracy of the data he or she produces."

The authors of the report readily admit that they have provided an overview, rather than a resolution, of the complexities that surround digital data. What is needed now is for institutions, consortia and scientific societies to find individual solutions that will work in their fields and physical settings. Funders must take up their responsibilities and increase investment in the upkeep of data, from the individual grant onwards. The scientific enterprise requires that the integrity of its data forms a bond of trust with the public. It is time to strengthen that bond with action. ∎

The shale revolution

The vast reserves of US natural gas must be used judiciously to ease the transition to clean energy.

Several years ago, it looked as though the United States was running short of natural gas. Prices spiked as declining production in old fields collided with increasing industrial demand. Electric utilities shifted from 'clean' gas back to cheap coal, and suppliers began building terminals to import liquefied natural gas from abroad. Yet today, coal-fired power is again on the wane, ports for liquefied natural gas are idling below capacity, and the nation is awash with gas.

So what happened? Clearly, the threat of carbon regulation has curbed industry's appetite for coal, and the sagging economy has depressed energy demand across the board. But just as importantly, natural-gas production is again on the rise. Thanks to advances in drilling technology, including horizontal drilling and more effective rock fracturing, producers have at last unlocked the vast quantities of gas trapped underground in impermeable strata of shale.

The Potential Gas Committee, a volunteer group of industry, government and academic experts headquartered in Golden, Colorado, increased its estimate of recoverable gas reserves by 39% in its biennial report released last month, mostly because of shale gas. The new total, almost 60 trillion cubic metres, is equivalent to about a century's worth of gas at current usage rates.

Policy-makers everywhere should take note. Shale formations similar to those that have upended the US natural-gas market exist all over the world. Early explorations are already under way in Canada and several European countries, many of which are overly reliant on

coal and politically risky Russian gas imports. And there is no reason to think the development will stop there.

The good news is that natural gas is the cleanest fossil fuel available. Compared with coal, burning gas roughly halves carbon dioxide emissions and eliminates the release of toxic chemicals such as mercury and sulphur dioxide. It is often regarded as a bridge fuel to a low-carbon economy, one that can squeeze out coal and supplement wind and solar energy. Indeed, an abundant and relatively cheap supply of natural gas should spell the end of new coal plants and could to a certain extent allow old coal plants to be replaced.

Deployed without forethought, however, natural gas could hamper the transition to clean energy by outcompeting currently more expensive technologies such as wind and solar. Although natural gas seems clean compared with coal, drilling operations scar the landscape, disturb sensitive ecosystems, increase regional air pollution and may, some fear, pollute groundwater. Then there's the carbon dioxide problem. Building new gas-fired plants would lock in emissions for decades to come — unless they have technologies that would allow the carbon dioxide to be captured and either buried or recycled.

Some have recommended that the United States deploy its newly abundant natural-gas resources in the transportation sector, but that would require vast new infrastructure for what is, in the end, a transition fuel. Congress should avoid such single-shot solutions and keep its eye on the target: a solid greenhouse-gas regulatory programme that sets short- and long-term goals while pricing energy according to the damage it inflicts on the environment.

In the short term, regulators and policy-makers should look for ways to encourage the use of natural-gas plants that are currently fired up only when demand is highest. There is a lot of spare capacity; better to use it wherever possible and retire the dirty, inefficient coal plants that are, in any case, unlikely candidates for the carbon-capture retrofit technologies down the road. But the endgame must bring a halt to greenhouse-gas emissions. From this perspective, power plants that run on gas, like coal, will eventually need carbon-capture technology if they are to remain viable.

It is too early to predict how the natural-gas market will play out, and it would be foolhardy to focus on supply to the detriment of energy efficiency, which should be the top priority. Nonetheless, it seems that the world has much more gas at its disposal than was believed only a few years ago. It should be used wisely. ■

Inspiring non-scientists

Those wishing to reveal scientific ideas should learn from the engaging style of TED conference talks.

A conference that charges £4,500 (US$7,440) to attendees, attracts sponsorship from the likes of Nokia and GE, and stuffs 600 participants into a stiflingly hot Oxford theatre (as happened last week) had better deliver. And if what you want is to find yourself in intelligent and engaging company, to be addressed comprehensibly by achievers about their ideas across a diverse range of interests, if you have the money, and if the organizers think you're interesting enough to attend, TED conferences do indeed deliver.

The acronym stands for 'technology, entertainment and design', but in recent years the TED presentations have extended well beyond these topics into culture, management, religion, science, extreme sports and more. Founded in 1984 and long established as annual events in California, TED conferences have recently begun to be held in other countries. They are now run by the not-for-profit Sapling Foundation in New York City, established in 1996 by TED talks curator and one-time publishing entrepreneur Chris Anderson. The visibility of the conferences has expanded hugely since videos of the best talks became available for free on YouTube in 2006 (see http://tinyurl.com/kpmvbo).

TED succeeds in part because participants are encouraged to talk about the unexpected. The title of this year's UK conference was 'The substance of things not seen'. Thus the advertising guru Rory Sutherland's dissection of how Kemal Atatürk, the first president of modern Turkey, sought to prohibit the public use of the veil not by banning it, but by insisting that it be worn by all prostitutes. And thus the activist Evgeny Morozov's discussion of the 'spinternet' — ways in which the Russian and Chinese governments subtly disseminate propaganda using a supposedly open medium. And so on.

But perhaps the most critical key to success is the style of the talks. And here, those scientists wishing to inspire public audiences could take a few tips from the speakers in Oxford who addressed themes as various as biomimicry (Janine Benyus), the neuroscience of other people's rational and moral judgements (Rebecca Saxe) and supermassive black holes (Andrea Ghez). Their videos and many others should become available over the next few weeks.

The talks have a strict time limit of 18 minutes — no interaction with the audience, and no questions except the informal ones asked in the extended conversation breaks. Academics used to talking for 30 to 45 minutes might imagine this to be severely constraining. But TED demonstrates that, for a general audience, 18 minutes is plenty for getting across context and key issues, while still forcing each speaker to focus on a message — whether it be advocacy or the celebration of new knowledge.

There is also a welcome absence of PowerPoint presentations. Instead there are plenty of images — but precious few professional scientific diagrams, which can quickly lose the audience's attention. This forces speakers to craft talks that can engage sophisticated but scientifically untutored listeners at their level. And it also encourages speakers to try for a freely flowing, relaxed presentation style, without notes. This can take hours of practice, and indeed it should — the YouTube postings of these talks offer a potential audience of millions.

After many talks have passed by, a listener may notice another factor at work: TED talks tend to have a strong feel-good aspect, often featuring calls to make the world a better place. Rarely is the audience provoked or seriously challenged. But that's not necessarily bad — the attendees have paid thousands of pounds apiece in this case to have an uplifting time, after all. They are eager to hear about new ideas. And the process does spread those ideas among people who are themselves influential and well connected.

Scientists wishing to inspire non-scientists should look at a few of these talks online and learn a thing or two. ■

RESEARCH HIGHLIGHTS

Chill bill

Science **325,** 468–470 (2009)

Guinness adverts notwithstanding, suggestions for the purpose of a toucan's enormous bill have ranged from sexual ornamentation to peeling fruit. Glenn Tattersall of Brock University in St. Catharines, Canada, and his colleagues have discovered another use: keeping cool.

Toucan bills, which contain a network of blood vessels, are among the most effective heat-loss instrument in any animal yet studied. The authors placed toco toucans (*Ramphastos toco*) in an experiment chamber and subjected them to different temperatures while observing them with infrared imaging (see image, inset). Depending on activity levels and ambient temperature, adult bills accounted for as little as 5% and as much as 100% of total body heat loss.

T. FILADELPHO

G. TATTERSALL

PHYTOLOGY

Tree carbon recalibrated?

New Phytol. doi:10.1111/j.1469-8137.2009.02971.x (2009)

In addition to dispersing into the soil and, ultimately, out into the atmosphere, carbon dioxide produced by tree roots can travel up the stem. The presence of this newfound pathway suggests that scientists have underestimated the amount of energy consumed by underground forest metabolism.

Doug Aubrey and Robert Teskey from the University of Georgia in Athens measured CO_2 levels in the soil as well as in sap and water taken up by four eastern cottonwoods (*Populus deltoides*) at an experimental plantation in South Carolina. They estimated that twice as much CO_2 entered the stem's xylem as diffused into the soil, providing evidence that below-ground respiration might exceed that by leaves and woody tissue.

Although preliminary, the study suggests that scientists need to reconsider the way they calculate forest carbon budgets.

NANOCHEMISTRY

Protein fondue

Angew. Chem. Int. Edn doi:10.1002/anie.200903100 (2009)

Proteins, like many other nanometre-scale objects, cannot normally exist in the liquid state. Instead they sublime or degrade upon heating.

Building on others' work with inorganic nanoparticles, Stephen Mann at the University of Bristol, UK, and his colleagues show that it is possible to generate liquid proteins by substituting a positively charged polymer for their carboxy-terminal groups. They then added another polymer, this time negatively charged. After removing water, the team heated the solid until it formed a liquid crystal, which melted at 50 °C. The protein melt returned to its solid phase at −50 °C.

The solid-to-liquid phase change is possible because the polymer extends the range of intermolecular forces. It is likely that the method could be generalized for a wide range of biological nanostructures.

CANCER BIOLOGY

Cancer's metabolic roots

Science doi:10.1126/science.1175689 (2009)

A mutation in a protein involved in cell respiration makes people susceptible to a rare type of tumour, a new study finds.

Jared Rutter at the University of Utah School of Medicine in Salt Lake City and his colleagues used yeast to uncover the function of SDH5, a previously uncharacterized mitochondrial protein. The protein is required for the assembly of functional succinate dehydrogenase (SDH), a crucial enzyme in both the citric-acid cycle and electron-transport chain. The human version of the protein has the same functions.

Previous work has linked other human mutations that knocked down SDH activity to neuroendocrine tumours called paragangliomas. The authors found that individuals in a family with hereditary paragangliomas had mutations in the gene encoding human SDH5.

MICROFLUIDICS

The sounds of science

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0900043106 (2009)

Microfluidic systems use networks of tiny chambers and channels to mix small volumes of solution. Controlling these systems often requires cumbersome pumps and valves that are much larger than the microfluidic devices themselves.

The problem struck a chord with Mark Burns and his colleagues at the University of Michigan in Ann Arbor. So the team attached a microfluidic device to a series of resonant acoustic cavities that respond to different frequencies (pictured left). When resonant, each cavity applies a specific output pressure to the microfluidic system. The pressure applied — and thus the movement of fluids along channels — can be controlled using a simple melody. The authors hope that their tiny tunes will lead to more compact control systems for microfluidics.
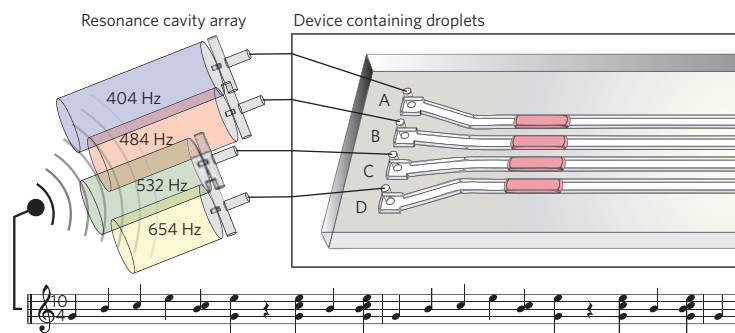
Resonance cavity array

404 Hz

484 Hz

532 Hz

654 Hz

Device containing droplets

A
B
C
D

EARTH MONITORING
Tsunamis from space

Nat. Hazards Earth Syst. Sci. **9**, 1135–1147 (2009)

A team in Colorado has linked surface roughness in the open ocean to a tsunami, offering potential for reliable early warning systems.

Oleg Godin of the Earth System Research Laboratory in Boulder and his colleagues studied radar data from satellite-based altimeters to identify the surface roughness associated with the deadly 2004 Sumatra–Andaman tsunami. Analysis of radar backscattering strength separated surface roughness caused by the tsunami — sometimes referred to as tsunami shadows — from other factors churning the sea. The authors say the report is a starting point for developing algorithms to use satellite-based and airborne radar and microwave radiometers to detect tsunami intensity and direction from surface roughness.

CHEMISTRY
A one-pot shot

Angew. Chem. Int. Edn **121**, 5970–5973 (2009)

Making a complex organic molecule generally involves a long series of reactions and purification steps to build it piece by piece. Much more efficient would be to simply throw all of the components into a flask at the same time, combining them at a stroke.

Romano Orru and his colleagues at the Free University Amsterdam in the Netherlands have now united an unprecedented eight molecular fragments — without intermediate purification steps — to create a complex 'druglike' compound.

Their synthesis has three stages — each designed not to interfere with chemical groups incorporated in previous steps — and lasts for almost three days. But the end product comes out of the flask with an overall yield of 24%, impressive for a synthesis that involves forming nine new chemical bonds.

BEHAVIOUR
Why 'there's never just one'

Behav. Process. **82**, 81–84 (2009)

Even cockroaches develop psychological problems if they are denied a normal social life. Animals reared in solitude are less likely to explore new environments or search for food, are more timid when approaching other cockroaches and are less able to spot the signs of a good mate.

Mathieu Lihoreau and his colleagues at the University of Rennes 1 in France reared nymphs of the gregarious German cockroach (*Blattella germanica*; pictured above) either in isolation or in a group of ten siblings. The effects of solitary confinement parallel those of 'isolation syndrome', the authors say. This behavioural syndrome has been described in a variety of vertebrates, but Lihoreau and colleagues suggest that it may develop when any group-living species is denied company.

PHYSICS
A cold shake

Phys. Rev. Lett. **103**, 045301 (2009)

Scientists have observed turbulence in a Bose–Einstein condensate (BEC), a state of matter that can form when atoms are trapped and cooled almost to absolute zero.

Emanuel Henn at the University of São Paulo in Brazil and his colleagues used an oscillating magnetic field to shake a BEC of rubidium atoms. When the oscillation amplitude passed a certain threshold, vortices distributed in random directions emerged.

A cloud of atoms released from the trap maintained an oval shape while expanding, possibly because vortices kept it from deforming. Further study of the system might help researchers understand turbulence in ordinary fluids, the authors say.

DEVELOPMENT
Starting from scratch

Development **136**, 2695–2703 (2009)

Transplanting a nucleus from a specialized adult cell into an enucleated egg overhauls gene expression in the nucleus, allowing it to direct development into any one of many tissue types. Yet no one is quite sure how.

John Gurdon and his colleagues at the Gurdon Institute in Cambridge, UK, transplanted various nuclei into developing eggs of the frog *Xenopus laevis* and looked at the expression of muscle genes. The transplanted nuclei expressed many muscle genes, but surprisingly their expression did not rely on the activity of proteins such as MyoD that are normally required for muscle-gene expression. This suggests that the egg cell uses special methods to reprogram nuclei.

NIGEL CATTLIN/FLPA

JOURNAL CLUB

Pavel Jungwirth

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic

A chemist realizes that popularity is no measure of strength.

Urea, the water-soluble organic compound found in mammalian urine, has been known for its ability to denature — or unfold — proteins for more than 100 years. To this day, it is among the most widely used protein denaturants. So one could be forgiven for taking it for granted that we know in gory detail what happens when we pour urea into a protein solution. But, alas, nailing down the individual molecular interactions between urea and the chemical groups at a protein's surface is exceedingly difficult.

Experiments and simulations suggest that urea interacts primarily with amide groups in the protein backbone, but every such group in a given protein has its own local environment, leading to fuzzy signals in spectroscopic studies. Paul Cremer's group at Texas A&M University came up with a good means by which to address the problem. They employed a popular protein proxy, poly(*N*-isopropylacrylamide), in which all of the amide groups are chemically equivalent (L. B. Sagle *et al. J. Am. Chem. Soc.* **131**, 9304–9310; 2009). Using infrared spectroscopy combined with measurements of hydrophobic collapse, they showed that urea interacts only weakly with this polymer.

Essentially, Cremer and colleagues' measurements suggest that one needs buckets of urea to see any effect. This is exactly the same situation as that observed for proteins, in which high concentrations of urea are necessary for denaturation. Thus one of the most common denaturants is actually a shockingly weak one. In fact, the strength of its interactions with the protein is little greater than those of harmless water molecules.

In the end, the key to the denaturing mechanism may be the fact that urea is a larger molecule than water — which has subtle entropic consequences — rather than that the two have different hydrogen-binding abilities.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

arguing that Kansas poses a bigger danger of severe tornadoes.

In light of the GAO report, critics are questioning whether the DHS should go back to the drawing board to reconsider a new offshore site to replace the 55-year-old facility on Plum Island. "A careful and comprehensive assessment of risk has not been made," says Richard Ebright, a molecular biologist at Rutgers University in Piscataway, New Jersey. "It's time for those policies to be reconsidered."

But many biosecurity experts point out that disease centres in other countries, such as the Canadian National Centre for Foreign Animal Disease in Winnipeg, operate efficiently and safely in mainland locations. "We have the technology to work with these agents safely in containment," says Corrie Brown, a veterinary pathologist at the University of Georgia in Athens who was interviewed by the GAO for its report.

However, disease outbreaks are not unprecedented. In 2001, a leak at an ageing UK animal-research lab in Pirbright led to an outbreak of foot-and-mouth disease that required the slaughter of 6 million livestock. This week, the British government announced a £100-million (US$165-million) upgrade to the facility. ■
Elie Dolgin

the safety merits of the various applicants. "If a thorough risk assessment was completed that analysed the relative risks of conducting this research at various sites, then ultimately the DHS would conclude that Texas is the safer location," he says,

---

European body told to cut free

The European Commission must make "immediate corrections" to the running of the European Research Council (ERC) or risk the body suffering a "deadly blow", an expert review has found.

On 23 July, a panel led by the former president of Latvia, Vaira Vike-Freiberga, published a review of the ERC — the first pan-European initiative to fund frontier research solely on the basis of excellence.

The ERC was established two years ago and is administered by an executive agency under the commission's control. The panel describes the council's management as a source of "great frustration and low-level conflict".

It recommends recruiting a top scientist with managerial experience from outside the commission to run the executive agency, replacing the commission's current appointee. The review says that existing rules preventing such an appointment should be "urgently remedied".

"Cumbersome" regulations and financial controls imposed on the ERC could prevent the council becoming a world-class institution, the review says. Sources of frustration include funding coming as contracts rather than grants, requiring researchers to document in detail time spent working on ERC-funded projects (see *Nature* **460**, 440–441; 2009). The review calls for the ERC to be allowed to award grants, giving researchers greater flexibility.

Rules requiring reviewers to provide formal identification discourage participation, the review says. It calls on the commission to put in place rules based on "trust and not suspicion and mistrust".

The review says these changes must be made over the coming year. A follow-up review should take place in 2011 to ensure improvements have been made. If they have not, the ERC should become independent from the commission, the panel says.

Janez Potočnik, the research commissioner, said in a statement that the recommendations "coincide with [the] commission's own conclusions" on the ERC. "I believe that we will be able to respond positively to the substance of most recommendations," he says.

The commission's official response to the review is due in October. ■
Natasha Gilbert

---

floor of a University of Texas Medical Branch hospital in Galveston, Avi Markowitz, chief of oncology and haematology, agreed to allow the pharmaceutical industry to provide food for the staff. When hospital officials found out, Markowitz says, they told him to stop, as it contravened university policy. "They had no problem at all letting the students, the trainees and the staff go hungry," he says.

At the meeting, similar frustration was palpable. One attendee complained that he couldn't buy a $12 hamburger for a consultant who had agreed to speak for free. "They're giving us a pro bono service and we're going to ask them to pay for their own lunch?" he lamented.

Stossel and other ACRE leaders argue that stringent policies call into question physicians' integrity. The gift ban, for instance, suggests that physicians "have a corruption problem", he says.

Nissen acknowledges that academics and industry need to work together, but he says firewalls are needed between the two. "When money becomes involved, it no longer becomes a

> **"The notion that a physician is automatically tainted by financial interest in a company is obnoxious."**

scientific collaboration but a commercial one," he says. "So transparency is an absolute minimum."

Markowitz points out that restrictive rules won't stop misconduct. He says the new conflict-of-interest rules are like "cutting off everyone's hands to prevent stealing".

Rather than focusing on "up-front prohibition", Stossel says, institutions should "emphasize vigilance and punishment".

ACRE's membership is small, with only about 100 people. And the group has not yet decided whether to allow industry employees as members.

For now, at least, funding comes from its $200 annual membership fee. Accordingly, the kick-off gathering was far from opulent. At lunchtime, participants gathered at folding tables to eat ham sandwiches and potato crisps.

Weber says the next step is to come up with a proactive code of conduct outlining ethical interactions. "If we can't say what would be appropriate," he says, "then we don't have a position at all." ■
Cassandra Willyard

NEWS

Biodefence lab criticized

US lawmakers investigate site choice of planned facility.

The US government's decision to locate a new national biosecurity lab in Manhattan, Kansas, has been called into question by a congressional watchdog. The Government Accountability Office (GAO) concluded that the Department of Homeland Security (DHS) relied on a flawed risk assessment when deciding the location of the US$650-million National Bio and Agro-Defense Facility.

Last December the DHS chose Kansas State University as the site of a high-containment lab to replace the Plum Island Animal Disease Center in New York state.

But the GAO found that the department used "unrepresentative accident scenarios", "outdated modelling", and "inadequate" information in assessing the chance that highly contagious pathogens might be accidentally released from the facility, according to a draft of the report obtained by *The Washington Post*. The full report was scheduled to be released on 30 July, to coincide with a hearing convened by the House Committee on Energy and Commerce's oversight and investigations subcommittee.

The GAO says that the DHS's environmental impact statement, prepared for all six finalist sites, did not adequately evaluate worst-case scenarios.

Matthew Chandler, a DHS spokesman, defended the agency's site selection process as open, fair and safe. "We stand by our process," he says. A GAO spokeswoman declined to comment on the draft report.

Jerry Jaax, Kansas State University's associate vice-president for research compliance, says that concerns about placing the high-containment lab on the US mainland are "outmoded". He says that no problems have been found at other mainland disease facilities operating at the highest biosecurity level, such as those at the Centers for Disease Control and Prevention in Atlanta, Georgia, and the DHS's biodefence lab in Fort Detrick, Maryland. In addition, worries over a risk of tornados at the site are "disingenuous and stupid", he says. "The safest place in the case of a

> "We have the technology to work with these agents safely."

A planned biosecurity facility in Kansas will conduct animal-disease research.

P. SCHERMEISTER/CORBIS

tornado would be one of these facilities."

Michael Guiffré, a partner at law firm Patton Boggs in Washington DC who represents a consortium seeking to site the biosecurity lab in San Antonio, Texas, applauded the report and called for a full comparison of

Physicians fight back against disclosure rules

BOSTON, MASSACHUSETTS
Even as US legislators work to limit ties between academic physicians and industry, a small group of doctors is calling for greater industry collaboration. On 23 July, they gathered at Brigham and Women's Hospital in Boston to launch a new organization opposed to strict conflict-of-interest rules.

The Association of Clinical Researchers and Educators (ACRE) is the brainchild of Thomas Stossel, a haematologist at Harvard Medical School. Stossel has been a vocal critic of conflict-of-interest rules since the late 1980s, when he was on the board of the Cambridge-based biotech firm Biogen and Harvard adopted its first conflict-of-interest policy.

"Tom, for many years, was pretty much the lone voice," says Laurence Hirsch, an endocrinologist with the medical-technology firm BD in Franklin Lakes, New Jersey.

Now, however, Stossel has company. He and the five other physicians who lead ACRE argue that they are part of a "silent majority" who think the move to curtail physician–industry relations has gone too far. In Massachusetts, a law came into effect on 1 July that requires drug and medical-device manufacturers to inform the state health department about payments of $50 or more made to physicians for marketing activities; gifts are banned. Vermont has adopted a tougher law, and Congress is considering legislation that would set up a national reporting requirement.

ACRE does not yet oppose specific laws, it says, but rather the climate of distrust it claims tough conflict-of-interest rules create. The notion that a physician is automatically tainted by financial interest in a company is "obnoxious", says Michael Weber, a cardiologist at the State University of New York Downstate Medical Center in Brooklyn, and a member of ACRE's

Thomas Stossel: vocal critic of conflict-of-interest rules.

steering committee. "Just because something is good for industry doesn't mean it has to be bad for everyone else."

Still, last year, a congressional investigation revealed that three Harvard psychiatrists had reported to their university only a fraction of the million or more dollars each received from drug companies. And psychiatrist Charles Nemeroff of Emory University in Atlanta, Georgia, stepped down after the same probe uncovered that he failed to report at least $1.2 million he earned from drug companies. "I do not believe that a physician that takes more than a million dollars in money from industry can possibly be objective," says cardiologist Steven Nissen of the Cleveland Clinic in Ohio, a strong supporter of curbing industry–physician ties.

Not all instances are so clear-cut, argued some at the ACRE meeting. After Hurricane Ike destroyed the cafeteria and much of the first

S. G. HENRY

K. CAMPBELL/GETTY

the safety merits of the various applicants. "If a thorough risk assessment was completed that analysed the relative risks of conducting this research at various sites, then ultimately the DHS would conclude that Texas is the safer location," he says,

arguing that Kansas poses a bigger danger of severe tornadoes.

In light of the GAO report, critics are questioning whether the DHS should go back to the drawing board to reconsider a new offshore site to replace the 55-year-old facility on Plum Island. "A careful and comprehensive assessment of risk has not been made," says Richard Ebright, a molecular biologist at Rutgers University in Piscataway, New Jersey. "It's time for those policies to be reconsidered."

But many biosecurity experts point out that disease centres in other countries, such as the Canadian National Centre for Foreign Animal Disease in Winnipeg, operate efficiently and safely in mainland locations. "We have the technology to work with these agents safely in containment," says Corrie Brown, a veterinary pathologist at the University of Georgia in Athens who was interviewed by the GAO for its report.

However, disease outbreaks are not unprecedented. In 2001, a leak at an ageing UK animal-research lab in Pirbright led to an outbreak of foot-and-mouth disease that required the slaughter of 6 million livestock. This week, the British government announced a £100-million (US$165-million) upgrade to the facility. ■
Elie Dolgin

European body told to cut free

The European Commission must make "immediate corrections" to the running of the European Research Council (ERC) or risk the body suffering a "deadly blow", an expert review has found.

On 23 July, a panel led by the former president of Latvia, Vaira Vike-Freiberga, published a review of the ERC — the first pan-European initiative to fund frontier research solely on the basis of excellence.

The ERC was established two years ago and is administered by an executive agency under the commission's control. The panel describes the council's management as a source of "great frustration and low-level conflict".

It recommends recruiting a top scientist with managerial experience from outside the commission to run the executive agency, replacing the commission's current appointee. The review says that existing rules preventing such an appointment should be "urgently remedied".

"Cumbersome" regulations and financial controls imposed on the ERC could prevent the council becoming a world-class institution, the review says. Sources of frustration include funding coming as contracts rather than grants, requiring researchers to document in detail time spent working on ERC-funded projects (see *Nature* 460, 440–441; 2009). The review calls for the ERC to be allowed to award grants, giving researchers greater flexibility.

Rules requiring reviewers to provide formal identification discourage participation, the review says. It calls on the commission to put in place rules based on "trust and not suspicion and mistrust".

The review says these changes must be made over the coming year. A follow-up review should take place in 2011 to ensure improvements have been made. If they have not, the ERC should become independent from the commission, the panel says.

Janez Potočnik, the research commissioner, said in a statement that the recommendations "coincide with [the] commission's own conclusions" on the ERC. "I believe that we will be able to respond positively to the substance of most recommendations," he says.

The commission's official response to the review is due in October. ■
Natasha Gilbert

floor of a University of Texas Medical Branch hospital in Galveston, Avi Markowitz, chief of oncology and haematology, agreed to allow the pharmaceutical industry to provide food for the staff. When hospital officials found out, Markowitz says, they told him to stop, as it contravened university policy. "They had no problem at all letting the students, the trainees and the staff go hungry," he says.

At the meeting, similar frustration was palpable. One attendee complained that he couldn't buy a $12 hamburger for a consultant who had agreed to speak for free. "They're giving us a pro bono service and we're going to ask them to pay for their own lunch?" he lamented.

Stossel and other ACRE leaders argue that stringent policies call into question physicians' integrity. The gift

ban, for instance, suggests that physicians "have a corruption problem", he says.

Nissen acknowledges that academics and industry need to work together, but he says firewalls are needed between the two. "When money becomes involved, it no longer becomes a

"The notion that a physician is automatically tainted by financial interest in a company is obnoxious."

scientific collaboration but a commercial one," he says. "So transparency is an absolute minimum."

Markowitz points out that restrictive rules won't stop misconduct. He says the new conflict-of-interest rules are like "cutting off everyone's hands to prevent stealing".

Rather than focusing on "up-front prohibition", Stossel says, institutions should "emphasize vigilance and punishment".

ACRE's membership is small, with only about 100 people. And the group has not yet decided whether to allow industry employees as members.

For now, at least, funding comes from its $200 annual membership fee. Accordingly, the kick-off gathering was far from opulent. At lunchtime, participants gathered at folding tables to eat ham sandwiches and potato crisps.

Weber says the next step is to come up with a proactive code of conduct outlining ethical interactions. "If we can't say what would be appropriate," he says, "then we don't have a position at all." ■
Cassandra Willyard

LHC students face data drought

Computer simulations are the only option when the world's largest particle accelerator isn't working.

Last November, Sara Bolognesi stood before a committee at the University of Turin in Italy and defended her PhD thesis in experimental high-energy physics. The 180-page document is a treatise on finding the Higgs boson, part of the mechanism believed to endow all other matter with mass. The pages are crammed with dozens of figures and tables, but something is missing: real data.

That's because the Large Hadron Collider (LHC), the world's largest particle accelerator at CERN, outside Geneva in Switzerland, is broken. The 4.6-billion Swiss franc (US$4.3-billion) collider is designed to accelerate protons to near the speed of light and smash them together in four giant detectors spread around its 27-kilometre circumference. Physicists once hoped that the LHC would start its collisions in late 2006, but last September, after a series of delays and soon after the machine was switched on, an electrical short caused extensive damage along a sector of the machine. Repairs have taken longer than expected, and, as of last week, the LHC was not scheduled to restart before mid-November.

The long delays have ended the dreams of a generation of graduate students hoping to use fresh data for their theses. With no machine to deliver results, "people are doing experimental PhDs and effectively doing very little experimenting", says Will Reece, a graduate student at Imperial College London working on a detector known as LHCb. "It's a strange situation."

Strange but not unprecedented, says Rolf-Dieter Heuer, CERN's director-general. During the mid-1980s, physicists were focused on building the Large Electron–Positron collider, the predecessor to the LHC. Over that period, Heuer says, graduate students sometimes wrote theses based on data from detector tests. Today, many of the same physicists work on the LHC project.

But although the electron collider took a few years to build, construction of the LHC took more than a decade, and most testing for the current detectors ended years ago. Aside from a trickle of data created by stray cosmic rays hitting the detectors, there will be no data to be analysed until the collider restarts. "It's a mess," says Burton Richter, a Nobel-prizewinning physicist at the SLAC National Accelerator Laboratory in Menlo Park, California.

European graduate students such as

Sara Bolognesi hopes to move from simulated experiments to real data collection at CERN.

Bolognesi face strict time constraints for completing their PhDs. Most universities require a thesis to be submitted within three to four years, and that means that students cannot wait for their data. Instead, their analyses are being done with data from 'Monte Carlo' simulations — computer programs that replicate what might come out of real collisions.

Not everybody thinks that the simulated data are a problem. "I don't feel that bad about not having data in my thesis," says Carsten Hof, a graduate student at Aachen University in Germany, who is finishing his PhD on software that will automatically analyse real collisions. "All the bugs we found and fixed now will also be fixed for real data." Hof adds that the data drought may actually be an advantage. "You look at everything with no bias," he says.

Heuer says that the situation reflects the growing size and sophistication of high-energy physics experiments. Whereas early experiments could be done in days by a handful of people, machines such as the LHC take thousands of researchers years to complete. The current generation of students may not be familiar with real data, he says, but they have extensive experience in building the huge detectors needed to capture them. Future PhD students will work on software without touching the innards of the detectors, he points out. As long as students get a taste of what's involved with each stage of the project, he says, "I don't think that people are losing anything."

Others are more worried. Although Monte

> **"People are doing experimental PhDs and effectively doing very little experimenting."**

Carlo simulations can reproduce the uncertainties seen in real data, they will never contain a big surprise. That means simulated data can never be as good as the real thing, says Gustaaf Brooijmans, a physicist at Columbia University in New York. "It's like a badly written murder mystery," he says. "In the first chapter you're given enough information that you know who did it, and then you read the rest of the book, and, lo and behold, you get the right answer."

For this reason, Columbia and other US institutions require students to use real data in their PhD theses. That solves the data dilemma, but creates a new problem: US students working on the LHC must move to finish their theses. For students such as Ketino Kaadze of Kansas State University in Manhattan, this meant travelling from Geneva to Batavia, Illinois, the home of the world's other major particle collider, the Tevatron.

Kaadze says that she was initially nervous about the move from one experiment to the other, but she has found it valuable. Although it will take her longer to complete her PhD, she is glad to have made the switch. "I think it's very important to have this experience," she says.

Now at CERN for a postdoctoral fellowship, Bolognesi worries that she will be at a disadvantage compared with students like Kaadze. "Two years from now, I will have to search for work," she says. "I hope they will not discriminate against me." By the time she is looking for a job, the LHC should have completed its first run, and Bolognesi will hopefully have completed a first of her own — an analysis of real collisions. ∎

Geoff Brumfiel

RAY OF HOPE IN HUNT FOR DARK MATTER
Gamma-ray spike in data from Fermi telescope causes excitement.
www.nature.com/news

GENERAL DYNAMICS

Italy introduces performance-related funding

The Italian government has finally given the go-ahead to a national research evaluation agency, ANVUR. But at the same time it published Italy's first university ranking and said performance-related funding would begin immediately.

The 24 July announcement of ANVUR was welcomed by many scientists who say that the country has not adequately rewarded best-performing research institutions. A law creating the agency was passed two years ago, but not put into practice until now. ANVUR should start work in about a year, says Mariastella Gelmini, the minister of universities and research.

Her ranking of universities caused political outcry however. Up to 7% of the approximately €7-billion (US$10-billion) national university budget will be allocated according to the list. Universities in the centre and north of the country generally did well, but most universities in the poorer south ranked at the bottom.

The ranking was based on an internal research evaluation carried out several years ago, combined with other criteria intended to assess teaching quality.

Raffaele Lombardo, president of Sicily, denounced the ranking criteria as discriminatory. Poor infrastructure in many southern universities prevents them from attracting independent research funds, he said in newspaper interviews, and high unemployment rates make it hard for graduates to find jobs, a criterion used to measure effectiveness of teaching. Lombardo and others also criticized the use of old bibliometric data.

All universities are concerned that performance-related funding is being introduced in a year when the government plans a 10% cut in the university budget. Davide Bassi, rector of the top-ranking University of Trento, says that this will be a "disaster for all universities, including our own".

Details of ANVUR's operation have yet to be set, but it is likely to be responsible for fine-tuning the criteria used in ranking exercises. ■
Alison Abbott

Forest growth studies begin to turn up the heat

Across the United States, researchers are firing up experiments to determine how rising temperatures could reshape the nation's forests.

The studies encompass the pines and maples of eastern forests in Massachusetts and North Carolina, the spruce and fir of northern Minnesota, and the alpine tundra ecosystem above the treeline in the Rocky Mountains in Colorado.

Technology to warm soils is more than a decade old, but altering the temperature above ground remains challenging, particularly in a forest. "People criticize just warming the soil, but it's difficult to warm stands of relatively large trees," says ecologist Scott Bridgham, of the University of Oregon in Eugene. "Researchers have tended to shy away from forests." But as the US Department of Energy phases out a number of large-scale forest experiments designed to look at the effects of elevated carbon dioxide levels (see *Nature* **456**, 289; 2008), researchers are clamouring for the next generation of experiments to incorporate temperature control.

For now, the technology available limits such projects to seedlings and young trees on relatively small plots of land. That will still be informative, says Jacqueline Mohan, an ecologist at the University of Georgia in Athens. "The most important determinant of where a species can grow is where the juvenile trees can grow and thrive," she says.

In the next few weeks, Mohan and her colleagues, including Jerry Melillo of the Marine Biological Laboratory in Woods Hole, Massachusetts, will switch on a series of underground heating cables and above-ground, open-topped heating chambers in Harvard Forest, located in Petersham, Massachusetts, and Duke Forest, which spreads across several counties in North Carolina.

Warmed plots show spring greening (right) taking place earlier.

Meanwhile, in Minnesota, Peter Reich, a forest ecologist at the University of Minnesota in St Paul, and his colleagues switched their heaters on in April. Measurements on 1,500 of the 10,000 seedlings his team has planted suggests that a rise of 2–4 °C causes seedlings to put out shoots and leaves earlier in the year. These effects differ across the 11 species that Reich's team is testing, giving an early clue to one mechanism by which increased temperatures can alter the balance of power in a forest.

Later this summer, an experiment in the Rocky Mountains led by Lara Kueppers, an ecologist at the University of California, Merced, will begin heating three sites — one in the alpine tundra above the treeline, one at the edge of the treeline and a lower site within the forest. Her team will also look at how genetic variation and moisture affect which species fare best in higher temperatures.

The teams are using different methods to warm the air. Melillo and Mohan circulate water, heated by propane, in the walls of open-topped chambers. Reich and Kueppers use infrared heating lamps, which have the advantage of not enclosing the sides of test plots, but the disadvantage of heating surfaces directly, rather than simply raising the air temperature.

Many unknowns remain, however. "None of us," says Reich, "really knows whether the way that we're warming will have some effects that might be different from the way that nature will warm forests in the future." ■
Heidi Ledford

A. STEFANSKI

Mice made from induced stem cells

Technical feat shows that the different route to stem cells can indeed make a full mammal body.

Two teams of Chinese researchers have created live mice from induced pluripotent stem (iPS) cells, answering a lingering question about the developmental potential of the cells.

Since Shinya Yamanaka of Kyoto University in Japan created the first iPS cells[1] in 2006, researchers have wondered whether they could generate an entire mammalian body from iPS cells, as they have from true embryonic stem cells. Experiments reported online this week in *Nature*[2] and in *Cell Stem Cell*[3] suggest that, at least for mice, the answer is yes.

For the first study, animal cloners Qi Zhou of the Institute of Zoology in Beijing and Fanyi Zeng of Shanghai Jiao Tong University started by creating iPS cells the same way as Yamanaka, by using viral vectors to introduce four genes into mouse fibroblast cells. The researchers hoped that the introduced factors would 'reprogram' the cells so that they could differentiate into any type of cell in the body.

To check whether the reprogramming had worked, Zhou and Zeng first carried out a standard set of tests, including analysing whether their iPS cells had the same surface markers as embryonic stem cells. Going a step further, they then created a 'tetraploid' embryo by fusing two cells of an early-stage fertilized embryo (see graphic). A tetraploid embryo develops a placenta and other cells necessary for development, but not the embryonic cells that would become the body. It is, in essence, a car without a driver.

When implanted into these embryos, the iPS cells began to steer development. The developing embryo was transferred to a surrogate mother, and 20 days later a mouse was born. It was black, like the mice used to create the iPS cells and unlike the white mice used to create the tetraploid embryo. DNA tests confirmed the mouse, named Xiao Xiao or 'Tiny', had arisen from the iPS cells.

Rudolf Jaenisch, a cloning expert at the Massachusetts Institute of Technology in Cambridge, had tried to do the same experiment in 2007, but didn't succeed in getting beyond late-stage embryos[4]. "There were two possible explanations" for his team's failure, he says. "Either iPS cells aren't pluripotent so it was impossible, or we just hadn't tried hard enough. The first would have been more interesting, but I assumed it was the second explanation."

The Chinese team tried harder, tweaking the culture medium and analysing 250 developing embryos before getting their first mouse.

In the paper, the team reports 27 live births. With their best cell line and optimal recipe, they were able to get 22 live births from 624 injected embryos, a success rate of 3.5%.

Zeng says, however, that the mice seem to have a high death rate, with some dying after just two days, and others displaying physical abnormalities, details of which the team would not reveal. But some of their mice passed one of the most fundamental tests of health: all 12 mice

MAKING AN iPS-CELL MOUSE

Two-cell embryo is fused to generate a tetraploid blastocyst

Two-cell embryo → Electro-fusion → One-cell tetraploid embryo → Tetraploid blastocyst

iPS cells are injected into the tetraploid blastocyst, which then steer development

Developing embryo is implanted in surrogate mother

Surrogate mother → iPS-cell mouse

that were mated produced offspring, and the offspring showed no abnormalities. The team says it now has hundreds of second-generation, and more than 100 third-generation, mice. The team found no tumours in the mice, although they have not systematically looked for them.

The leader of the second team, Shaorong Gao of the National Institute of Biological Sciences in Beijing, also credits persistence for success. His group, which used the same basic technique as Zeng and Zhou, transferred iPS cells to 187 tetraploid complementation embryos to get just two live births (a 1.1% efficiency rate), although one died in infancy. "The chance for generating such a cell line is rare but we tried very hard," he says. Gao's team is now trying to mate its surviving mouse.

Both groups are now trying to understand what differences between iPS cells and embryonic stem cells might explain the abnormalities, high death rates, low efficiency rates and the fact that most iPS cell lines don't seem to work in making mice. Zeng and Zhou found, for one thing, that timing was important: cells that formed iPS cell colonies quickly — after 14 days — were successful, whereas those that formed colonies after 20 or 36 days did not work. Gao suggests that "aberrant reprogramming" might be to blame, at least for the low efficiency rates.

Such mouse studies should help researchers to understand fundamental differences between human embryonic stem cells and iPS cells as well. Earlier this month, researchers at the University of California, Los Angeles, reported that human iPS cells that passed conventional pluripotency tests differed in gene expression from human embryonic stem cells[5]. "iPS cells might do things better or worse than embryonic stem cells," says team member Kathrin Plath. "I don't think we know the answer at this point." Because the tetraploid work cannot be done with human embryos, the Chinese studies can't say much about clinical applications of pluripotent cell lines, adds her colleague William Lowry.

Zhou and Zeng are pursuing several new avenues, including comparing the iPS mice with mice cloned with conventional techniques, and working to prove that the same experiment can be done with adult mice. (The fibroblasts used to create iPS cells in both studies came from late-stage embryos.)
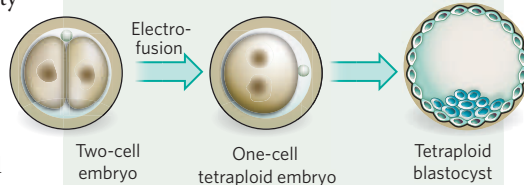
This would essentially be a new way to clone adult mammals — reprogramming DNA from an adult and generating a genetically identical individual. As a potentially easier method that produces fewer abnormalities than conventional cloning, it might evoke interest among mavericks as a tool for human cloning. China recently strengthened its law prohibiting such cloning[6].

Zhou says he hopes that researchers will take advantage of the technology as "an important model for understanding reprogramming". He adds: "It is not intended to be a first step towards using iPS cells to create a human being." ■
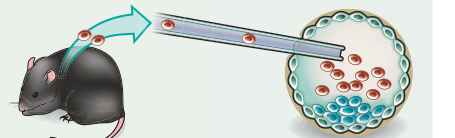
David Cyranoski

1. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
2. Zhao, X.-Y. *et al. Nature* advance online publication doi:10.1038/nature08267 (2009).
3. Kang, L. *et al. Cell Stem Cell* doi:10.1016/j.stem.2009.07.001 (2009).
4. Wernig, M. *et al. Nature* **448,** 318–324 (2007).
5. Chin, M. H. *et al. Cell Stem Cell* **5,** 111–123 (2009).
6. Cyranoski, D. *Nature* **459,** 146–147 (2009).

Legal battle may reshape nanotechnology firm

Oxonica, one of the United Kingdom's leading nanotechnology companies (see *Nature* 446, 963; 2007), has lost its latest legal fight over intellectual-property rights to a fuel additive.

On 9 July, the company lost its appeal against a September 2008 ruling in favour of Neuftec, a company registered in the Commonwealth of Dominica and started by Ronen Hazarika, a former employee of Kidlington-based Oxonica.

Hazarika claims that Oxonica used a form of cerium oxide nanoparticles, invented by him, in its Envirox fuel additive, and then sold large quantities without paying him royalties. The nanoparticles are designed to increase fuel efficiency.

A separate hearing will decide how much of Neuftec's court costs Oxonica will have to pay. Kevin Matthews, Oxonica's chief executive, says the company has set aside £800,000 (US$1.3 million) to pay for costs, and has already paid £394,000 in royalties. In a 9 July statement, Oxonica said it had £1.6 million in the bank.

Oxonica "is obviously in a very precarious financial position", says Michael Holman, research director at Lux Research, a consultancy in New York.

Richard Farleigh, the chairman and biggest shareholder of Oxonica, says the company remains solvent. "We don't have a huge amount of cash," he says. As for paying Neuftec, he says: "We've made provision for it and as far as I'm aware there are adequate finances."

Neuftec is also pursuing a separate patent infringement case in English courts involving Envirox, over a deal Oxonica has in Italy and Australia. That case is set to go to court in early 2010; Oxonica has challenged a number of aspects, including the validity of the patent that Neuftec says is being infringed. Hazarika, who has also left Neuftec, is now chief executive of Energenics in Singapore, a company that also develops fuel additives.

OXONICA

This fuel additive is at the heart of Oxonica's court case.

Oxonica was spun out of the University of Oxford in 1999 by materials scientists Peter Dobson and Gareth Wakefield, who had invented a sunscreen using titanium dioxide nanoparticles. In 2005 the company was listed on the UK Alternative Investment Market (AIM), and at its height was valued at around $100 million, says Matthews. "Oxonica has been seen as one of the leading lights for nanotech in the United Kingdom," says Holman.

"I'm profoundly disappointed," comments Dobson, who cut his official ties with Oxonica in 2002 to become academic director of the University of Oxford's Begbroke Science Park. "This was a dream shattered." Dobson also recently became strategic adviser for nanotechnology to the UK Research Councils.

"We are disappointed that we lost the action," Matthews told *Nature*. Matthews resigned on 24 June, and will leave the company at the end of September. On 17 July chief financial officer Richard Clarke also resigned, to leave at the same time as Matthews. It has not been decided who will run the company after that.

"It's not a comfortable position to be in," says Matthews. "But the reality is it's a demanding time for early stage businesses. Obviously losing a legal battle at this point in time is not particularly helpful to us." On 24 July, shareholders approved delisting the company from AIM on 4 August.

Matthews says the company's future is assured, but it will involve "significantly simplifying the business". Oxonica is likely to partner its businesses with other companies, as it did with its Nanoplex diagnostics technology in 2006 by partnering with BD, a medical technology company in Franklin Lakes, New Jersey, in a deal that brought in $7 million.

"It will be a royalties business not an operations business," says Matthews, adding that he is leaving because he isn't needed in a company without operations. ■

Katharine Sanderson

Q&A

US puts flu vaccines on trial

The US National Institute of Allergy and Infectious Diseases (NIAID) announced last week that it will begin five clinical trials for two pandemic H1N1 influenza vaccines in early August. These trials will help inform a likely US mass-vaccination campaign beginning in September. NIAID director **Anthony Fauci** talks about what vaccines were chosen, and why.

To increase the amount of flu vaccine available, the World Health Organization recommends using adjuvants, which boost the body's immune response to the drug. But the five trials you announced last week are for non-adjuvanted vaccines. Why?

We have planned seven priority trials. The five I announced on 22 July are for non-adjuvanted vaccines, but we also plan two more: testing GlaxoSmithKline's AS03 adjuvant with vaccine from Sanofi Pasteur and from CSL Biotherapies. We prioritized non-adjuvanted vaccines as the US government seems likely to recommend using these for vaccinating the first tier of priority groups — expected to include children and groups at higher risk of severe disease, such as those with certain underlying illnesses and pregnant women. We fully intend to proceed with trials of adjuvanted vaccines.

Will you test adjuvanted vaccines in children?

The Europeans have lots of data on the use of adjuvanted flu vaccine in the elderly, but I don't think anybody has really good data on adjuvants in children. The Department of Health and Human Services (DHHS) has therefore decided that we are not going to take the chance, and has made a policy decision that we are not going to give adjuvanted flu vaccines to kids. We don't have the time to collect substantial data.

Might not trials including at-risk groups help inform how vaccines are used?

Yes. But, as in other countries, there are many ethical constraints. We are working with both principal investigators and institutional review boards to draw up protocols for such groups. Even if the Food and Drug Administration (FDA) is considering the initial non-adjuvanted vaccines as simply a strain substitution of seasonal vaccine, it's still a new vaccine, so we want to get data from healthy adults before launching into risk groups.

Dr Fauci (NIH)

What information will the NIAID get that vaccine makers won't from their own trials?

We are asking questions that will inform policy decisions likely to affect how we use vaccines, whereas the focus of vaccine makers is generally directed towards studies needed to get a licence for their vaccine. The sort of information that the FDA, DHHS and the vaccine makers told us they needed most included what levels of antigen per dose are essential to getting an adequate immune response, and whether one or two shots of vaccine will be needed.

So two trials will test both single and double shots of both 15-microgram [the amount in seasonal H1N1 vaccine] or 30-µg doses of antigen, using antigen from Sanofi Pasteur and from CSL Biotherapies. We will give the first doses in the first week or so of August, and the second dose 21 days later. We will learn very quickly after 21 days, when we draw blood, if one dose of 15 µg is enough. And if it isn't, if 30 µg is any better. And if 15 µg is enough, does 30 µg give an even better response? Shortly after 42 days, we will have data on the second doses.

Why are you testing vaccine only from Sanofi Pasteur and CSL Biotherapies, when Novartis accounts for the bulk of vaccine ordered by the United States?

Novartis has quite a sophisticated clinical-trials apparatus. The United States will purchase 45% from Novartis, 26% from Sanofi Pasteur and a little bit less than 19% from CSL Biotherapies. Novartis is able to carry the ball itself, so we made a reasonably well-based decision to fill in the gaps and get information on CSL and Sanofi Pasteur.

You also plan to test co-administration of pandemic 2009 vaccines and seasonal H1N1 vaccines. How will the immune system react to these together?

We don't know. Testing two vaccines against different H1N1s at the same time has never been done. We'll look at three test regimes: giving the pandemic 2009 vaccine before, at the same time as, or after seasonal H1N1 vaccine.

If you give the pandemic H1N1 vaccine first, will subsequently giving the seasonal H1N1 enhance the response to the original dose, or will there be antigenic competition or interference? If you give both at same time, is the body going to have a response that is enhanced against seasonal flu and not do a very good job against the pandemic vaccine, or will it actually amplify the response? And if you give pandemic vaccine after [seasonal vaccine], is that going to have an enhancing or suppressing effect? Immunologically, you can't predict the outcome.

What if we have a vaccine and then the genetics of the virus changes?

You never can predict that, but things look encouraging to me from a molecular-virological standpoint. If you look at the molecular and genetic make-up of the virus from the very first isolates in early April compared to what we are seeing now in late July, it's virtually an identical virus everywhere. So it doesn't look like it is under pressure to mutate to a significant degree. We hope it stays that way for the autumn and winter season.

Interview by Declan Butler

For more H1N1 coverage, see www.nature.com/swineflu.

J. REED/REUTERS

German research bodies draft synthetic-biology plan

Three leading German research organizations have outlined how they think the country could play a major part in synthetic biology.

The DFG (which funds university research), the German Academy of Sciences Leopoldina and the German Academy of Science and Engineering argue in a 27 July report that synthetic biology has great value to society — as long as the field's expansion is balanced with ethical debate.

Germany is particularly sensitive to ethical issues in the life sciences, partly because of previous Nazi abuses of bioscience. The report says that a national centre should be created to host a database of information about newly created stretches of DNA, and to assess their safety.

Ralf Wagner, chief executive of Geneart, a leading manufacturer of synthetic genes based in Regensburg, Germany, says he hopes the report will help create a positive environment for public debate.
For a longer version of this story, see http://tinyurl.com/lfukon

Step-by-step rating system set to improve African labs

An accreditation system that aims to raise the standard of disease diagnosis in African medical laboratories was launched on 27 July in Kigali, Rwanda.

The process, developed by the World Health Organization in collaboration with the US government, will mark African pathology labs on an incremental scale, upping the rating as their quality improves rather than using the 'pass or fail' system of many developed countries. The scheme was launched alongside a training programme for African lab workers.

The US Centers for Disease Control and Prevention in Atlanta, Georgia, which will implement the step-by-step system, estimates that it could see 60 currently

AM. SOC. CLIN. PATHOL.

The number of accredited labs in Africa may rise.

Lucky find of undersea methane bubbles

While testing equipment off the Californian coast last month, a newly refitted research vessel stumbled across plumes of methane gas rising 1,400 metres from the sea floor.

The *Okeanos Explorer*, commissioned last year by the US National Oceanic and Atmospheric Administration (NOAA) after a US$20-million refit, was testing a new multi-beam sonar system in the

NOAA

Mendocino fracture zone (see sonar image, above). On 15 July the ship returned to the site to capture plume material for analysis in the coming weeks.

The plumes, which measure up to 1 kilometre across, typically dissipate about 600 metres below the surface. Cruise scientist Stephen Hammond of the NOAA office in Newport, Oregon, suspects this is because ice with methane gas trapped in its crystal structure melts at the combination of pressure and temperature at that depth. Similar methane plumes have been discovered from the Oregon coast to the Black Sea, but not this large or numerous.
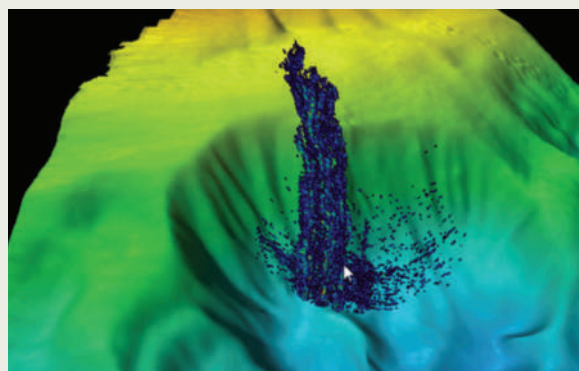
unaccredited African laboratories attain ratings verging on the standard of an average lab in the developed world over the next two years.
For a longer version of this story, see http://tinyurl.com/mlh6x3

UK government urged to disclose evidence base

If the UK government makes policy decisions that contradict recommendations from its science advisers, it should make clear why it has done so, politicians exhorted last week.

In a 23 July report on the use of science in government, the country's Innovation, Universities, Science and Skills Committee said that chief science advisers should challenge the government to publicly acknowledge when policies are not based on evidence.

Committee members also called for prime minister Gordon Brown to strengthen the role of scientific advice in government policies. They recommended that he create a permanent office for science in the cabinet office — where it would be closer to the heart of government policy-making, and have direct access to ministers.

Mauna Kea adds to its family of telescopes

Mauna Kea in Hawaii has beaten off competition from Cerro Armazones, in Chile's Atacama desert, to host the Thirty Meter Telescope. Henry Yang, chancellor of the University of California, Santa Barbara, and chair of the telescope's board

of directors, announced the winning site on 21 July. The decision has been two years in the making, he said.

Mauna Kea, which already hosts many other telescopes, was picked over its Chilean rival for its superior observing climate. It is higher and drier, has less atmospheric turbulence, and its average temperature fluctuates less through the year and over a day, notes board member Richard Ellis of the California Institute of Technology in Pasadena.

Construction of the telescope, which will cost around US$1 billion, is scheduled to begin in 2011 and end in 2018.

Genetic barcode for plants close to agreement

Final consensus on a DNA 'barcode' that could rapidly identify plant species should be reached in the next two months.

Botanists have for years debated which genetic sequences would be most suitable for such a barcode, which could be used to inventory biodiversity or monitor plant shipments (see *Nature* 451, 616; 2008).

David Schindel, executive secretary of the Consortium for the Barcode of Life, says that the consortium should be able to reach a decision quickly now that its plant working group has come to what the group calls "community agreement" on the matter.

The working group's recommendation, published this week (CBOL Plant Working Group *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0905845106; 2009), should also help barcode-related projects obtain funding.
For a longer version of this story, see http://tinyurl.com/kuav82
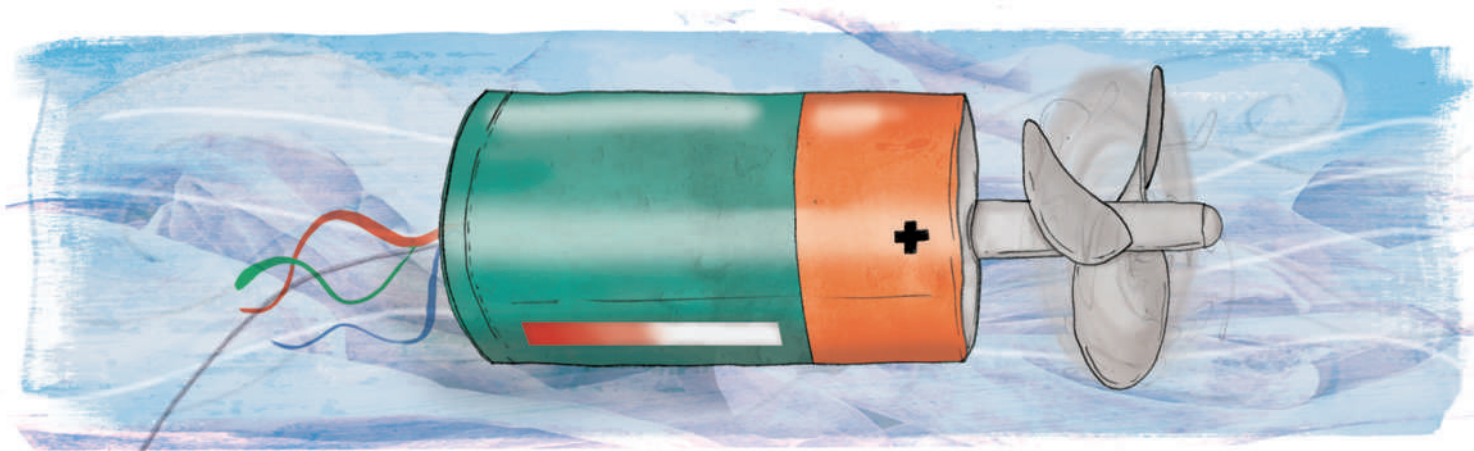
HIGH HOPES

A vast supply of energy is racing around the planet far above the surface. **Erik Vance** meets the engineers trying to bring the power of high-altitude wind down to earth.

A ride on a kite boat might just kill you from fright — if it doesn't crush you first. The vessel is essentially an 8-metre catamaran dragged behind a kite the size of a movie screen — a lot of horsepower for such a small craft. That concern grips me as I skitter across the surface of San Francisco Bay at 40 kilometres per hour. Sudden tugs from the kite jerk the boat from side to side and sometimes nearly out of the water.

But we stay upright, thanks to the skill of the boat's inventor, Don Montague. The 46-year-old pioneer of kite surfing deftly keeps the 200-kilogram craft from flipping up with each gust and crushing its passengers beneath it.

Montague has no intention of selling this floating hazard on the open market — he created it to break speed records and amuse himself. Today he has brought me out to demonstrate the raw power in the skies, just waiting to be tapped. Montague is part of the high-altitude wind movement, promoting an idea on the very fringes of energy development: tethered airborne devices that collect energy from the wind far above the surface. The designs of the aircraft vary widely, as does the height at which they would fly, but all seek to exploit the fact that the farther one goes from the ground, the stronger and steadier the available wind.

For at least a century, engineers have dreamed of pulling electricity from high in the atmosphere. However, only recently have lightweight materials and computer guidance systems emerged that make the idea feasible. In the past five years, what was once seen as a crackpot scheme has entered the early stages of a research-and-development race that is attracting tens of millions of dollars from major private backers such as Google.

Leading the way is a tiny, dedicated community of Californian inventors and kite surfers. As yet, they have generated little electricity and failed to win over mainstream wind experts. But several start-up companies — among them Makani Power, the one Montague has co-founded — plan to put prototypes into the air within the next 18 months.

This moment is overdue for Ken Caldeira, a climate researcher at Stanford University, California, who has long promoted generating power from high-altitude wind. He says wind in some locations contains at least ten times the energy of sunlight, when measured by surface area, but that conventional surface-mounted wind turbines collect only a fraction of that because friction with the ground brakes winds at the surface.

Caldeira's favourite area of study is the powerful high-altitude air currents called jet streams, which taken together contain 100 times as much energy as humans use today. They are "the highest concentration of renewable energy in large quantities," he says. "Sooner or later we will be extracting energy out of high-altitude winds."

There's another advantage to harvesting the wind at altitude. Breezes at the surface blow intermittently, so even in windy sites turbines typically collect only 30–40% of the energy that would be available if they ran continuously. High-altitude devices could push that number towards 80%, thanks to the steadier flow aloft and the possibility of moving higher and lower in the atmosphere to find the best wind.

Recently, Caldeira and Cristina Archer of California State University in Chico published the first comprehensive analysis[1] of global data on wind above 500 metres. They found that the most energetic winds are found in the jet streams, 10,000 metres up, above some of the spots that require large amounts of power, such as Japan and the eastern United States. However, at that height the resource is difficult to predict and even harder to reach.

Most engineers are not planning to go so high. The power of wind grows as the cube of its speed, so even the moderate increase in wind strength a few hundred metres above the surface produces large power gains.

Trade secrets

Leonard Shepard, president of Sky Windpower near Irvine, California, makes that point clear by listing the current wind speeds above Red Bluff, a small town in the northern part of the state that has detailed weather measurements. At ground level, a light breeze is blowing at under 10 kilometres per hour. At 900 metres, the speed reaches 50 kilometres per hour. But conventional wind turbines only stand about 130 metres or so above the ground. So a wind collector flying 900 metres above the town could theoretically gather 125 times as much energy as a turbine on the ground.

Sky Windpower is developing an unmanned helicopter design with four horizontal rotors that would hover at altitude. Shepard says he plans to send a prototype with four 2-metre rotors aloft this year to collect electricity at a few hundred metres. Once it gets up to a

Different ways to tap the winds aloft: JoeBen Bevirt's vision of a turbine array flying 10,000 metres above Earth's surface (right); Makani Power's prototype wing demonstrates its hovering abilities (below); the Sky Windpower concept (below, right); Makani engineers test the design of their carbon-fibre wing (bottom, right).

JOBY ENERGY INC.

A. DUNLAP/MAKANI POWER
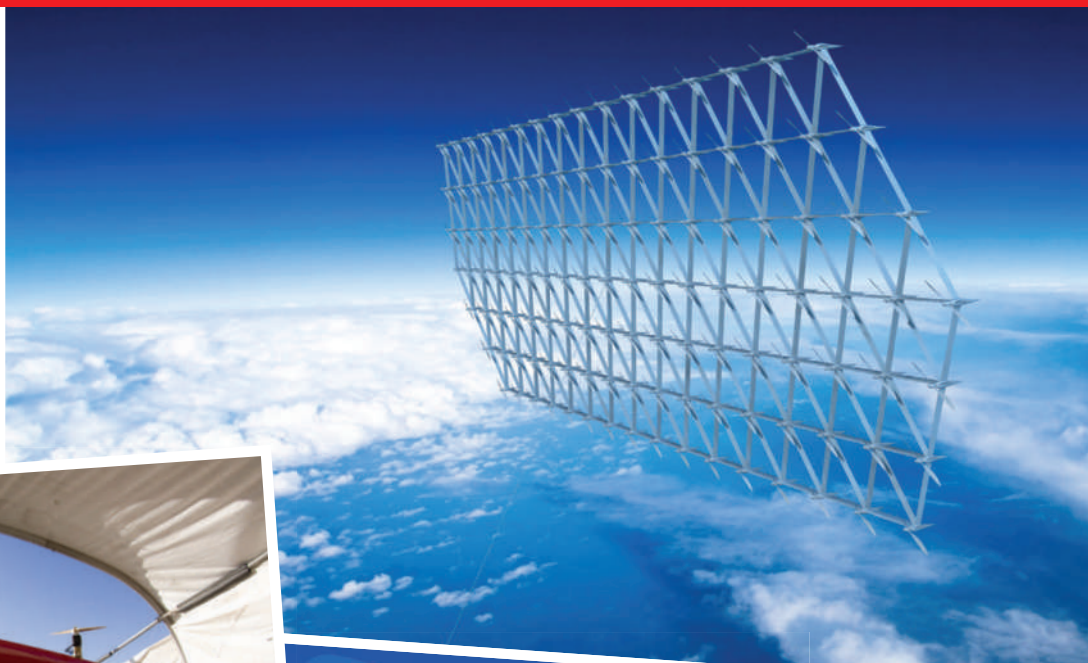
B. SHEPARD

certain height, the wind will spin the rotors and keep it up, but he declines to give many more details than that. A penchant for secrecy is common in the high-altitude wind world: until they have a working prototype, companies are careful not to reveal too much about their designs.

Kite dudes

Corporate secrecy mixes with surf culture at the offices of Makani Power, located on a desolate former naval air base in Alameda, California. In 2006, Montague teamed up with two avid kite surfers, inventor Saul Griffith and Griffith's former student Corwin Hardham, to mix kite know-how with engineering to harness the wind's energy.

The office is in a flight control tower and feels like a 1990s Internet start-up company. Near the entrance hang a dozen high-end bicycles, and windsurfing gear lies about in the workshop. The tower's control room, now the company lounge, has microbrews on tap. And hardly anyone seems to wear shoes.

Despite the Lost-Boys-in-Neverland atmosphere, Makani's design is not fantasy. It starts with a pilotless glider that resembles a giant boomerang made of lightweight carbon fibre. The glider is tethered to the ground by cords at each side, like the kite over Montague's boat. By pulling the cords a controller can send the glider sailing back and forth in a wide loop while four rotors on the leading edge spin to generate energy. Hardham — who is co-chief executive with Montague — says the kite would

ideally fly at 200–1,000 metres. Although wind at this relatively low altitude is slowed by friction with the ground, the back-and-forth movement significantly boosts the speed of the air passing through the rotors, says Hardham. A 30-metre, 1-tonne carbon-fibre glider flying in a moderate 32-kilometre-per-hour wind could produce 750 kilowatts — about as much as a small industrial turbine, but with a fraction of the materials, he says.

"It's surprising how good the opportunity is," Hardham says. "If we can make this work I think it's the best renewable conversion technology there is."

But he quickly tempers this, saying that there is still a lot to be done. For one thing, the power needs to be transferred from the kite to the ground. Both Makani and Sky Windpower are hesitant to go into detail about the tethers they would use, but they would probably be some kind of nylon fibre cord with an electrical cable at its core.

Hardham and others say that the tether is less a concern than automation. The biggest challenge, it seems, is controlling the device as it takes off, lands and flies in unpredictable weather. Their longest test run so far was a 30-hour flight for a 3-metre wing at a breezy site on the island of Maui, Hawaii. As

the device gets larger, controlling it becomes easier, but the consequences of a crash grow more expensive.

Californians do not have a monopoly on high-flying energy efforts. The Dutch astronaut Wubbo Ockels, who once collaborated with Montague, has developed a concept called Laddermill, which would fly multiple kites at various heights, all tethered to one turbine on the ground. As the kites rise, the turbine spins and generates power. The kites are then brought down again and the process repeats. And in Italy, another windsurfing enthusiast and engineer named Massimo Ippolito started a company called Kite Gen in 2007 to produce power by using several kites to turn a giant ring-shaped structure on the ground. Both companies have managed to create prototypes of single kites and have generated up to 10 kilowatts of electricity. Another company, Magenn, based in Kanata, Ontario, Canada, plans in the next year or two to release a spinning dirigible capable of generating 100 kilowatts that may cost around US$500,000.

Even a decade ago, designs such as these

B. PFEIFFER/MAKANI POWER

would have been impossible. However, with lighter materials and advances in pilotless flight, high-altitude wind is starting to attract money. Kite Gen was promised €50 million ($71 million) by the Italian government, although that funding has fallen through because of the financial crisis. Magenn, which is backed by private investors, says it has a capital base in the region of $10 million. Makani is partially funded by the Internet giant Google as a part of its philanthropic wing, Google.org. Google put $15 million into the company in 2007 — one-third of Google's total budget for renewable-energy research. And the company recently injected another dose of funds into Makani. Geoff Sharples of Google's RE<C initiative, which seeks to develop renewable-energy sources that will be cheaper than coal, declines to reveal the exact amount but says it was of the same order as the original investment. It's the only wind project that the company supports.

Sharples knows the wind world well, having previously served as an executive with Clipper Windpower, a major turbine manufacturer based in Carpinteria, California. The conventional earthbound approach is limited, he thinks, by the availability of conveniently located windy land. Although there is a viable alternative — erecting turbines offshore — Google is investing in high-altitude wind because the energy payoff could be so much bigger.

But companies such as Makani have a long way to go. Wind-industry insiders say a new technology needs to be capable of producing at least 1 megawatt — roughly enough to supply 1,000 homes — to get the attention of energy utility companies. To date, no high-altitude wind company has generated anything close to this. In June, the US National Research Council issued a report on renewable energy that briefly mentioned high-altitude wind power as a possibility more than 25 years in the future[2].

For now, most people in the wind-power industry are apparently ignoring high-altitude efforts. "I don't think anyone has given it a serious, objective study who wasn't already committed to the technology," says Robert Thresher, an engineer at the National Renewable Energy Laboratory in Golden, Colorado, which writes certification guidelines for wind technology. Wind experts at the lab are not enthusiastic about high-energy wind and say

Lit by a spotlight, Makani's prototype kite completes a computer-controlled, 30-hour test flight in the skies over Maui, Hawaii.

they wish inventors would focus their talents on conventional turbines. Thresher says he sees much more potential in offshore turbines than flying ones.

Such naysaying has not stopped another Californian inventor from sinking millions of his own dollars into the concept. Deep in the heavily wooded Santa Cruz Mountains, JoeBen Bevirt is chasing the ultimate wind prize — the strong, steady jet streams that scream along at 200 kilometres per hour, 10,000 metres up. Projects that seek to tap this resource draw the most scepticism from conventional wind engineers.

Bevirt has created a number of consumer and industrial products, such as robots for gene sequencing, mobile-phone headsets and a small camera tripod that can grip a branch or pole. In 2007, he sold his robotics company for $50 million.

"I should be retired right now. I should be kite surfing on the beach," Bevirt says. "But I am so passionate about this that I am working 18 or 20 hours a day, driving as hard as I have ever driven in my life."

He has come up with a concept that crosses Sky Windpower's helicopter rotors with Makani's sleek wings, the offspring being a diamond-shaped double-wing frame with rotors at each corner. By itself, it looks somewhat like the Wright Brothers' Flyer with an extra pair of rotors. Because of the strength of the jet streams, he expects each 11-metre rotor to produce a staggering 250–500 kilowatts. However, the design is modular, with many possible configurations: a system may have 4, 32 or even 96 rotors.

The system would lift off like a helicopter, with electricity from the ground used to turn its horizontal rotors to generate lift. Once it reached a desired height, the device would pitch forwards so that the wall of rotors would point into the wind and the frame act as glider wings, keeping the network aloft. Turned

> "I should be retired right now. But I am so passionate about this that I am working 18 or 20 hours a day."
> — JoeBen Bevirt

now by the wind, the rotors would generate electricity.

The design is far from foolproof. As with Makani's wing, take-off and landing require complicated control algorithms. Any tether reaching to 10,000 metres will need to be extremely light. And because a crash could kill people on the ground, the device could fly only over an uninhabited zone at least half the size of London.

Despite these challenges, Bevirt has the confidence of someone who has successfully brought products to market. He has founded a company, Joby Energy in Santa Cruz, to develop his concept. So far, he says he has invested $2 million in the project and plans to keep funding it as long as is necessary.

"I am very selective about the things that I pour my energy into. The fact that I am pouring my energy into this and that I am also willing to pour a lot of money into it lends credibility to the fact that it's not a crackpot scheme," he says.

Coming soon?

Like their energy source, the prospects of success for Bevirt and other high-altitude wind aficionados are up in the air. Makani has generated little more than 10 kilowatts, although it plans to run a bigger prototype this year. Sky Windpower also aims to test an experimental version of its concept this year. The first company to bring a design to market will probably be Canada's Magenn, although the size and drag of its blimp-like units may preclude them from generating enough energy to compete with big industrial turbines.

As a newcomer who started work only in 2008, Bevirt has yet to produce any energy, but he plans to start this summer and get a product to market around 2011. He says he doesn't mind that many engineers write off high-altitude wind power. The proof will come soon enough, he says.

While giving a tour of his workshop, Bevirt demonstrates a 1-metre scale model of his device. During the first test, it buzzes over my head like a wobbly bird. Later, as a mortified employee looks on, the device flips over and snaps in two during take-off.
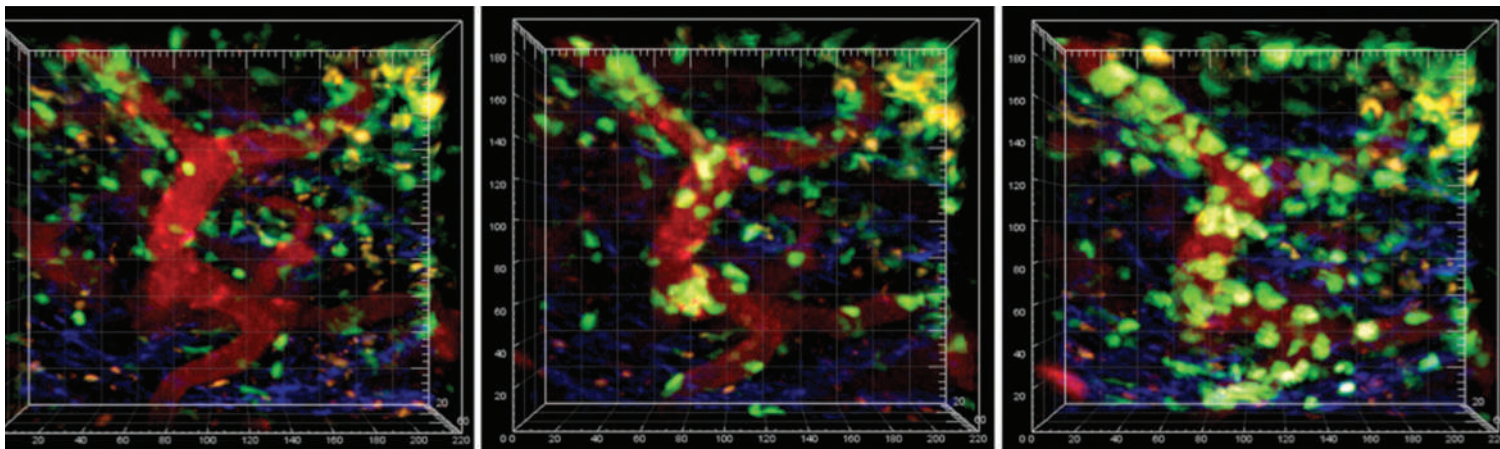
Bevirt just smiles. "Oh well," he says. Sometimes that happens too." ■

Erik Vance is a freelance science writer in Berkeley, California.

1. Archer, C. L. & Caldeira, K. *Energies* **2,** 307–319 (2009).
2. America's Energy Future Panel on Electricity from Renewable Resources *Electricity From Renewable Resources: Status, Prospects, and Impediments* 53 (National Academies Press, 2009).

Lights, camera, infection

Multiphoton microscopy is allowing immunologists to watch infections as they happen. **Jeanne Erdmann** pulls up a seat.

Behind the heavy black curtains of his microscopy room, Mark Miller is shooting an action movie. He gives the settings on the multiphoton microscope a once-over while his senior scientist Vjollca Konjufca checks the sedated mouse on the warmed stage. Then Miller flicks a switch on the scanner and red, blue and green images flicker on the computer monitor. He points to what he's looking for. "Right there," he says. Blue-labelled *Salmonella* bacteria gather near the top of a red villus, a finger-like projection from the wall of the mouse's small intestine. The bacteria look like helicopters buzzing around a mountain. It's early afternoon, and we are settling in to watch these bacteria for the next hour.

Miller's experiments, and others like it, are not just gripping the audience behind the curtain — they are gripping a much broader audience of immunologists. Miller, at Washington University School of Medicine in St Louis, Missouri, runs one of the leading labs using multiphoton microscopy to watch infections in living animals in real time.

Less than a decade ago, Miller and other immunologists mostly studied the process of infection *in vitro*, mixing pathogens and the cells they interact with in a culture dish. These simulations had their limits because the cells, much like animals in a zoo, were removed from the environment that influences their behaviour. The advent of multiphoton microscopy allowed researchers to view deep inside living tissues and watch cell biology live and 'in the wild'. This is particularly valuable for the immune system, in which the component cells roam throughout the body's landscapes interacting with pathogens and surrounding cells.

Miller likens multiphoton imaging to a "naturalist studying a herd of gazelles. You infer their function from the behaviour you observe," he says. "The beauty of multiphoton is that you don't have to know what it is you're looking for ahead of time. It just shows itself. We often find things we didn't expect." One thing that researchers didn't expect was the dramatic changes that take place in the first few minutes or hours of an infection, long before symptoms occur. That drama is what's keeping them glued to the screen. "Reality has been eye-opening," says Ronald Germain, an immunologist at the National Institutes of Health in Bethesda, Maryland, who works with the technique.

Set dressing

Konjufca started her experiment at 11:00 a.m. She took a mouse that had been genetically engineered so that several types of cell in the immune system — including neutrophils, lymphocytes, dendritic cells and macrophages — emit light in the fluorescent microscope. Then she sucked into a syringe some 10,000 *Salmonella*, also labelled with a fluorescent tag, and injected them into the mouse's intestine. She also injected a fluorescent label that will be taken up by epithelial cells lining the villi, painting them red.

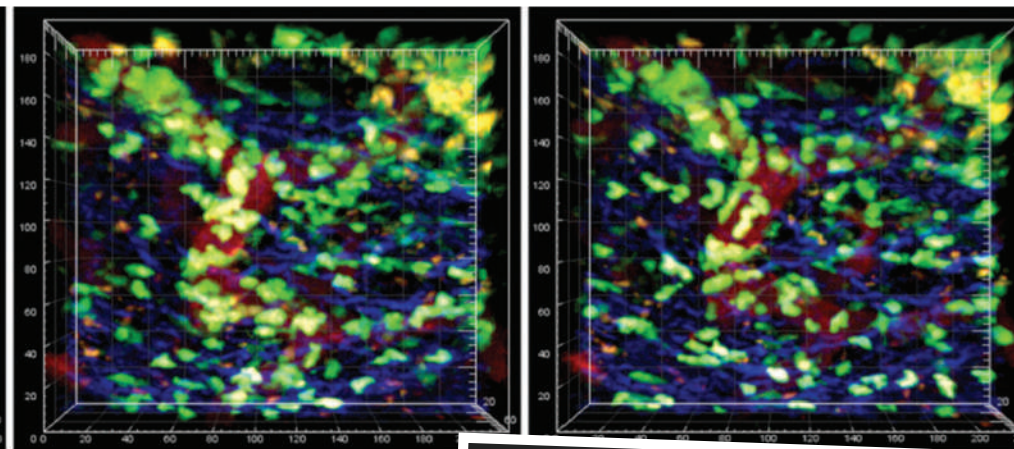> "Multiphoton imaging is like a naturalist studying a herd of gazelles."
> — Mark Miller

Konjufca does about two experiments like this every week as part of a project to study how *Salmonella* invades the body to cause food poisoning, work she hopes might help produce an oral animal vaccine that protects the food supply. She wants to work out the exact sequence of events during an infection: how the bacteria breach the gut wall and various immune cells arrive on the scene; how dendritic cells in the gut villi engulf and process some of the *Salmonella*; and then which cells transport the broken-down *Salmonella*, in the form of antigens, to the lymph nodes and spleen. "We don't necessarily know which cells go where and at which time point," says Konjufca.

For her experiments she uses both wild-type bacterial strains and ones with different virulence genes mutated. Mutations introduced into *Salmonella* to make a harmless vaccine strain often make the bacteria less able to invade the intestine and provoke the immune system, says Konjufca. The trick is to make a harmless *Salmonella* that still induces a strong and lasting immune response. Konjufca is trying to understand this by watching the mouse's response to these differing strains.

None of the work has been published. For now, she is still hammering out the technical details, such as how many bacteria to inject. Too few, and she may not recruit enough of the white blood cells such as neutrophils, the first responders to infection. Too many, and she will

Over about an hour, neutrophils (green) pour out of blood vessels (red) to fight *Salmonella* in a mouse gut. Right: Mark Miller looks on.

no longer be mimicking a physiological response. She also needs to figure out how long to wait after injecting the bacteria. If she waited, say, three to five hours, they may have already exited the small intestine and she would miss what happens there. Today she plans to wait around two hours. Ironing out these technical details is more than half the battle; otherwise there will be nothing to see.

Special effects

What is routine for Konjufca today was ground-breaking in 1990, when multiphoton microscopy made its debut[1]. In standard confocal microscopes, a single photon of light excites an electron in a fluorescent tag called a fluorophore, and light is emitted as the electron drops back to its ground state. But the high energy of light needed to excite the fluorophore quickly breaks it down and is damaging to cells. The innovation of Winfried Denk, now working at the Max Planck Institute for Medical Research in Heidelberg, Germany, was to excite the fluorophore using the simultaneous arrival of two or more photons of longer-wavelength light. These wavelengths are less damaging and, most importantly to immunologists, they typically penetrate at least 200 micrometres into tissue.

Neurobiologists, desperate to see deeper into the living brain, seized on the technique. Immunologists were not far behind. Miller began using the technique in the early 2000s, while working as a postdoc in Michael Cahalan's lab at the University of California, Irvine,

and in collaboration with neurobiologist Ian Parker, also at Irvine. He viewed the compartment of the lymph nodes where naive T cells, a type of lymphocyte, recognize antigen and become activated, a necessary step in readying the cells to respond to future infections. When Miller tried out Parker's multiphoton system on a lymph node snipped out of a mouse, his first reaction was surprise. He thought the T cells would be moving in unison along a gradient of chemicals called cytokines. Instead, he saw them weaving around very fast. "I couldn't believe how quickly and randomly T cells moved," he says. "They looked so purposeful and excited."

Miller used these images to calculate the velocity of the cells, showing that they can reach speeds greater than 25 micrometres a minute. And he suggested that the random movement has a purpose, by helping a T cell range across a broad territory and find the 'antigen-presenting cell' carrying the precise antigen it recognizes. In 2002, the work was published in *Science*[2] as part of a trio of advanced microscopy papers exploring the dynamics of T cells[3,4].

At that time, Miller estimates that only a

handful of immunology labs were using multiphoton microscopy. Researchers were excited to just observe the cells and get a picture of how these dynamics build up the host's defences. "It revolutionized the ideas that immunologists had about how immune responses come about," says Ulrich von Andrian, an immunologist at Harvard Medical School in Boston, Massachusetts.

Live action

These days, Miller says, it seems as though every major immunology department has set up a multiphoton system or wants to. "I get requests for advice all of the time." And much of the work has gone from studying the immune cells on their own to the infection as a whole. This makes for a more complex experiment. The tissue must be kept still, insulated from the pumping of the heart or the contractions of the gut, and the pathogen must be introduced without disturbing the surrounding tissue.

When microbiologist Agneta Richter-Dahlfors, of the Karolinska Institute in Stockholm, mastered the technique to watch the first few hours of a kidney infection, she worked with surgeons to inject bacteria into single rat kidney nephrons the size of a human eyelash. The images, some of the first showing an infection in real time, revealed that just two to three bacteria are enough to attach to the mucosal membrane and set up a colony[5]. Within three hours, the oxygen tension in the nephron drops to zero and blood flow to the area halts. "These analyses cannot be performed in any setting other than a live animal," says Richter-Dahlfors. Researchers have now observed a range of bacterial infections, including *Listeria monocytogenes* in the mouse's footpad, a model system that Miller developed with his postdoc Bernd Zinselmeyer to show how immune cells are recruited in the early stages of infection[6].

Two and a half hours after Konjufca injected the *Salmonella*, she reaches into the cage and gently picks up the sedated mouse. She puts it on the microscope stage, and covers it with a velvety, mouse-sized sheet to keep it as comfortable as possible and maintain a stable body temperature. Miller started working with *Salmonella* in the past two years. He likes studying bacterial infections in the gut because it's a common route of infection. It's also home to

569

V. KONJUFCA/M. NUSSENZWEIG
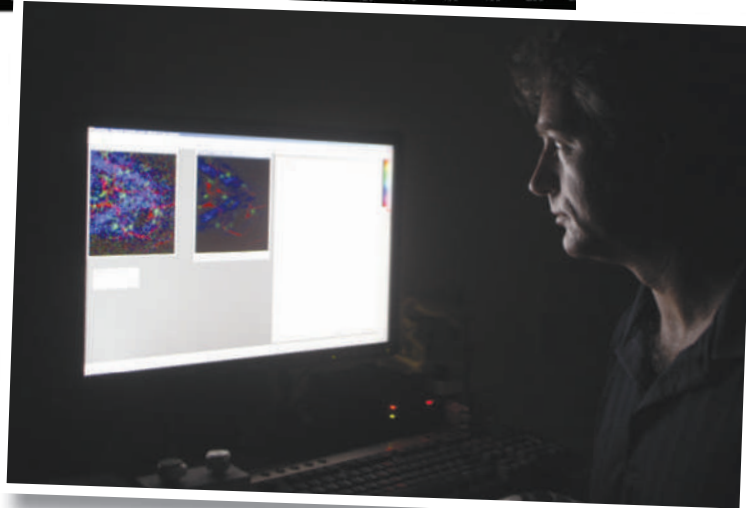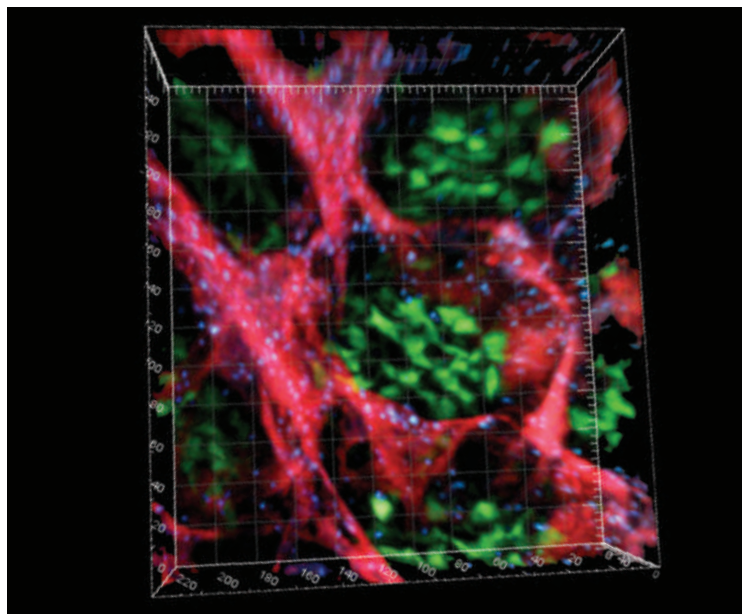
Dendritic cells (green) and *Salmonella* bacteria (blue) face off in the villi of the mouse small intestine (red).

non-pathogenic bacteria, so he can explore how immune cells tell friend from foe.

Miller built his own multiphoton imaging system at a cost of about half-a-million dollars by customizing a standard fluorescent microscope. It can probe hundreds of micrometres deep into tissue and record images at video rate. Today, he moves the microscope stage around until he finds an area of the small intestine largely empty of undigested food. Along the arc of the villus, the square epithelial cells show up like a necklace of perfect teeth. Some of the bacteria have already crossed the epithelial barrier and found their way to the lamina propria, an area at the bottom of the villi that is rich in dendritic cells and other antigen presenting cells.

Towards the talkies

Researchers aren't certain what happens at the epithelium during an infection. One possibility is that the *Salmonella* infect and then kill the epithelial cells, releasing bacteria that dendritic cells engulf. These cells process the bacterial constituents and stick them on their surface as antigens that will stimulate an immune response. Another idea has come from Germain's work suggesting that dendritic cells reach extensions across the lumen and fish bacteria out[7].

Miller collects a stack of images from multiple villi, at intervals of 20–30 seconds, data the computer will compile into three-dimensional time-lapse videos. By 2.30 p.m., the team has watched most of the *Salmonella* reach dendritic cells at the base of the villi. Miller shuts off the microscope, turns on the lights, and pulls back the curtain. The team transfers the data to a computer in an adjoining room.

Microscopy doesn't provide as complete a view of the infection process as researchers might like. "There's a lot of people thinking in molecular terms for their entire careers," says Miller, "and they're resistant to the idea that imaging is providing useful information because they don't feel it's quantitative." But he and other advocates of the technique disagree. They say that the quantification comes from measuring how fast cells move, which direction they move in and whether or how often they stick to other cells. They say that what's missing is a qualitative aspect that gives more information about why cells behave in certain ways. Following cells to an infection site is good, but what happens when they get there? Are the cells signalling, and if so, how?

> "Reality has been eye-opening."
> — Ronald Germain

Immunologist Phillipe Bousso, of the Pasteur Institute in Paris, compares the footage to silent movies: showing how cells move and interact, but not how they communicate biochemically[8]. What would give them 'sound', he says, would be biochemical or genetic assays that could be used in conjunction with microscopy. Jost Enninga, a biochemist also at the Pasteur Institute, is developing and patenting such assays, including one designed to trigger fluorescence in epithelial cells when they encounter pathogens. Another idea is to combine microscopy with real-time monitoring of bacterial gene expression, by dissecting out and analysing tissues at various time points.

Another limitation with microscopy is that the scene is not complete. Any cells not engineered to fluoresce disappear into the black background. "[Researchers] know the cells are there but don't necessarily think it's a disadvantage not to see them," says von Andrian. "If everything revealed itself it would be chaos."

It's a wrap

Later that afternoon, and over the next week, Miller and Konjufca review the videos of the experiment and compare them to earlier ones. They are pleased to see that in the most recent movies the villi are well defined and the bacteria visible on both sides of the epithelium. It shows that they got the set-up right, and that the timing between the injection and the imaging was spot-on. A film of an earlier experiment shows some of the neutrophils in local blood vessels rushing by at several hundred micrometres per second, while other neutrophils respond to inflammatory signals and slow down. The slowed cells then crawl along the inside of the vessel until they reach one spot where they all seem to squeeze through the wall and pour out of the blood vessel like bees swarming through a crack in a fence.

In her next experiments, Konjufca will wait longer before filming, so she can start to study in more detail how the cells cross the epithelium and which cells transport bacteria away from the gut. She'll then follow them to the liver and spleen to study how *Salmonella* interact with T cells and dendritic cells there. She hopes to understand how all this is different in the attenuated vaccine strains. And Miller is working with a computer scientist to develop software to track cells outside the field of view. The microscopic field is small and cells wander in and out of sight. He also wants to analyse their migratory behaviour as it's happening.

Even without these special effects, the surprise and discovery that each movie brings is enough to keep him going, Miller says. "That's the excitement that you get when you're in the lab. That's the motivation," he says. "'What are we going to see today?' We may have some ideas but we may be completely surprised." ■

Jeanne Erdmann is a freelance writer based in Wentzville, Missouri.

1. Denk, W., Strickler, J. H. & Webb, W. W. *Science* **248,** 73–76 (1990).
2. Miller, M. J., Wei, S. H., Parker, I. & Cahalan, M. D. *Science* **296,** 1869–1873 (2002).
3. Stoll, S., Delon, J., Brotz, T. M. & Germain, R. N. *Science* **296,** 1873–1876 (2002).
4. Bousso, P., Bhakta, N. R., Lewis, R. S. & Robey, E. *Science* **296,** 1876–1880 (2002).
5. Månsson, L. E. *et. al. Cell. Microbiol.* **9,** 413–424 (2007).
6. Zinselmeyer, B. H. *et al. Inflamm. Res.* **57,** 93–96 (2008).
7. Chieppa, M., Rescigno, M., Huang, A. Y. & Germain, R. N. *J. Exp. Med.* **203,** 2841–2852 (2006).
8. Bousso, P. *Nature Rev. Immunol.* **8,** 675–684 (2008).

See multiphoton movies at http://tinyurl.com/n35ufp
and http://tinyurl.com/mb6l45

CORRESPONDENCE

Flu: vaccinate to cut risk of chimaeric virus emerging

SIR — The international scientific community and decision-makers on public health are debating how best to manage the anticipated vaccine shortage for the new pandemic strain of influenza A virus, recently emerged from the animal reservoir. Priority distribution of the first product batches must be to individuals at high risk and to crucial employees.

But decisions about priority distribution should also take into account the benefits to the international community of vaccination in developing countries. Viral 'shedding' (the expulsion of virus) is reduced in vaccinated individuals and therefore the risk of reassortment with animal viruses is decreased. This consideration is in line with the 'One Health' vision — a multidisciplinary initiative to improve the health of humans, animals and the environment that is endorsed by the United Nations Food and Agriculture Organization, the World Health Organization and the World Organisation for Animal Health (OIE).

There is a risk of generating novel influenza A viruses through reassortment of the eight genes that result in antigenic shift, which would give rise to strains against which the human population has no immunity. For example, it was reassortment between avian and human influenza viruses that created the human pandemic viruses of 1957 and 1968 (K. Subbarao *et al.* in *Influenza Virology Current Topics* ed. Y. Kawaoka, 229–280, Caister Academic; 2006).

We are at present in a unique situation with the worldwide spread of the latest pandemic H1N1 virus, known as novel animal-origin H1N1 (naoH1N1) virus. Concurrently, and possibly for the first time in history, several developing countries are experiencing widespread infections in poultry by avian influenza viruses of the H5N1 and H9N2 subtypes, which also infect humans. In addition, H5N1 viruses that are widespread in Africa, the Middle East and Asia contain genetic mutations that reflect increased virulence for humans (see, for example, E. De Wit and R. A. M. Fouchier *J. Clin. Virol.* **41,** 1–6; 2008, and G. Cattoli *et al. PLoS ONE* **4,** e4842; 2009).

In countries where animal husbandry practices fall short of accepted biosecurity standards and where immunologically naive animal caretakers infected with naoH1N1 would shed large amounts of infectious virus, there is a significant risk of emergence of a reassortant virus. A reassortant virus containing a combination of genes, including a novel human-adapted influenza virus and H5N1 or H9N2, could result in chimaeric viruses with unknown characteristics.

Fast-tracking vaccination of humans against pandemic influenza in developing countries where flu in poultry is endemic would help prevent reassortment between naoH1N1 or other novel pandemic influenza strains and avian influenza viruses. That would deflect the unpredictable and serious consequences of viral reassortment to humankind worldwide.

Ilaria Capua, Giovanni Cattoli OIE Collaborating Centre for Diseases at the Human–Animal Interface, Istituto Zooprofilattico Sperimentale delle Venezie, 35020 Legnaro, Padova, Italy e-mail: icapua@izsvenezie.it

Flu: weighing up conflicting expert information

SIR — In his Essay 'Pandemics: avoiding the mistakes of 1918' (*Nature* **459,** 324–325; 2009), John M. Barry writes that during an influenza pandemic "telling the public the truth is … paramount". Truth telling is, he notes, the basis for trust and compliance in public-health measures. However, the H1N1 influenza debate in the Netherlands illustrates that things may not be that simple.

In the current 'swine flu' pandemic, the expert sources of information and guidance for the Dutch public are Albert Osterhaus of the National Influenza Centre in Rotterdam and Roel Coutinho at the National Institute for Public Health and the Environment in Bilthoven. These two experts are not strangers to such challenges: both have often been consulted on risks to the public since the 1980s.

Since the early years of the pandemic debate, Osterhaus has kept us informed about the greater picture, warning us of an 'imminent' flu pandemic (see, for example, J. C. de Jong *et al. Nature* **389,** 554; 1997, and the Dutch newspaper *De Volkskrant*, 2 May 2009). Coutinho, by contrast, has stuck to the immediate facts (see, for example, the newspaper *NRC Handelsblad*, 14 March 2006), and cautioned us that "there's no need to panic" (see, for example, the newspaper *Het Parool*, 1 May 2009). In communicating their different perspectives, these virologists were conveying conflicting messages to the general public, even though their aims were the same — to protect the public's health.

Public understanding, therefore, cannot hinge only on experts telling the truth as they see it: also important is how the public interprets mixed messages. It is the public's ability to assess the relative value of expert information that helps to stimulate the trust and compliance needed to follow recommended public-health measures.

That trust will be particularly crucial during the months to come, in implementing the planned vaccination of the entire Dutch population — and the related information campaign — following the announcement made by the World Health Organization on 11 June of a pandemic, which is expected to strike the Netherlands with force this autumn.

Erwin van Rijswoud Institute for Science, Innovation and Society, Radboud University Nijmegen, 6500 GL Nijmegen, the Netherlands e-mail: e.vanrijswoud@science.ru.nl

Where will we find the tritium to fuel hybrid reactors?

SIR — In your News Feature 'The hybrid returns' (*Nature* **460,** 25–28; 2009), you discuss the feasibility of creating a fusion–fission hybrid reactor to generate greenhouse-gas-free and waste-free nuclear energy. However, there is another challenge to add to the factors to be considered: where would the hundreds of grams of tritium needed daily to fuel the deuterium–tritium fusion reaction be produced?

For every neutron produced in this reaction, and in theory destined to burn the radioactive waste from the fission reaction, a tritium atom must be bred in a lithium (or lithium/lead) reactor blanket. The conventional, pure-fusion design demands a breeding ratio that is greater than one, to allow for losses and the decay of tritium; this might be extremely difficult to achieve even with essential neutron multipliers such as beryllium. The significant fraction of neutrons required to treat fissile material means that there could not possibly be enough to breed tritium as well.

At present, gram-scale quantities are available from Canadian sources, but the net production rate is only about 500 grams per year — a trivial amount compared with fusion-reactor requirements. For example, starting up any type of fusion reactor will need several months' supply of tritium just to bring the breeding process on stream. This issue is better addressed now than in 30 to 50 years' time.

J. H. Evans Abingdon, Oxon OX14 2EQ, UK e-mail: jhevans@sky.com

ESSAY

The invention of heroes

The Western public's misapprehension that genius in science is always male and caucasian is partly a legacy of Victorian politics, says **Christine MacLeod**.

In the wake of England's seventeenth-century civil wars, diarist and founder member of the Royal Society John Evelyn praised inventors and scientific discoverers above conquerors.

Such lauding of scientists over leaders was later echoed by many who opposed the costly wars against American colonists and revolutionary France. Indeed in Britain, especially in the eighteenth and nineteenth centuries, it was domestic politics, rather than the desire to promote science or technology per se, that often drove people to bestow heroic status on inventors, engineers and scientists.

Politics influences the public perception of researchers to this day. Yet, in this year of celebration — notably of the works of Charles Darwin, Galileo Galilei and Jean-Baptiste Lamarck — it is salutary to glance back at some early drivers of hero-worship in European science. A brief burst of adulation lavished on 'men of science' in Victorian Britain bolstered public approval of the field and helped improve the country's scientific and technical education. But it has also had a lasting detrimental legacy: it has helped foster the widely held notion that the world's greatest scientists are white, male individualists.

Britain's wave of hero worship began relatively discreetly. Disciples of the eighteenth-century Enlightenment saw in reason a bulwark against tyranny and popery. They expressed admiration of towering intellects, such as Francis Bacon and Isaac Newton, by hanging portraits on their walls and installing monuments in their grounds. After Newton's death in 1727, three marble statues of him were erected in places of note, including Westminster Abbey in London.

Patents and pedestals

The real upsurge of politically inspired celebration came in the early nineteenth century. Initially, engineers and inventors, not scientific intellectuals, took centre stage. In 1824, supported by the Royal Society and leading manufacturers, reform-minded politicians plucked the Scottish inventor and engineer James Watt from posthumous neglect. They began by launching a public subscription for a massive and controversial statue of him, again in Westminster Abbey.

Liberals feared that, in the transition to industrialization, their campaign for parliamentary reform, less taxation and free trade would be frustrated by the wealthy and powerful aristocracy. Lionizing Watt was their attempt to counteract the nation's adulation of Admiral Horatio Nelson and the Duke of Wellington, heroes of Napoleon's defeats at Trafalgar and Waterloo, respectively. The liberal politicians proclaimed Watt to be the war's true hero — his steam engine both history's greatest invention and the foundation of national victory. Moreover, the Industrial Revolution powered by Watt's engine was said to be the real source of the United Kingdom's wealth, international pre-eminence and 'civilizing mission' — an argument that helped clinch the vote for middle-class men in 1832.

Thirty years later, a concerted campaign by free-traders to abolish the patent system provoked a fresh surge of hero worship. The new Patent Act of 1852 was designed to make patents cheaper and more accessible. But during the next 20 years, a faction of economic liberals and engineers, including Isambard Kingdom Brunel, joined a Europe-wide clamour against patents, which they believed were harmful to the economy, unfair and unnecessary. Bennet Woodcroft, a leading figure in patent reform, was tireless in the patent system's defence, collecting portraits and memoirs of inventors, salvaging iconic inventions and launching the Patent Museum (a forerunner of London's Science Museum). Woodcroft even dug up the second Marquess of Worcester in a fruitless attempt to find the model steam engine the seventeenth-century inventor had requested be buried with him.

Proponents of the patent system evidently had all the best tunes. Without a patent, an inventor (typically a working man) was at the mercy of grasping capitalists; with one, he could aspire to fame and fortune. It was a story that chimed with the beliefs of many in the skilled trades that rose on the tide of industrialization. Machine-makers, railwaymen and spinning-machine operators cherished the memory of Watt and others who had succeeded through ingenuity and enterprise. They commemorated them on trade union insignia, ritually toasted their memory and donated generously to their memorials. In Bolton, workers raised a statue to the spinning-machine inventor Samuel Crompton, reminding the 'cotton lords' that a poor man's genius had created their wealth; in Penzance, they strove to ensure that Humphry Davy's invention of the miners' safety lamp and its contribution to coal-owners' profits would never be forgotten.

All this commemoration of lowly men turned successful inventors helped win people's respect for the upper echelons of the working class. Just as the campaign to glorify Watt aided the enfranchisement of middle-class men, so this new wave of commemoration helped convince Parliament to extend the right to vote to skilled male workers in 1867.

Newton apart, Victorian society was slow to extend the accolades it had showered on inventors and engineers to more academically minded scientists; ideas and understanding took a back seat to innovation. It was William Thomson's role in laying the first transatlantic telegraph cable, for example, not his research in electricity and thermodynamics, that brought him celebrity. His many honours included, in 1892, elevation to the House of Lords as Baron Kelvin — a first for a scientist.

A turning tide

Yet wherever possible, scientists took advantage of the nation's appetite for heroes. For instance, after Darwin's death in 1882, leading scientists overturned his wish to rest in his local churchyard, recognizing the propaganda value of a Westminster Abbey funeral.

By 1900, the tide of public enthusiasm for
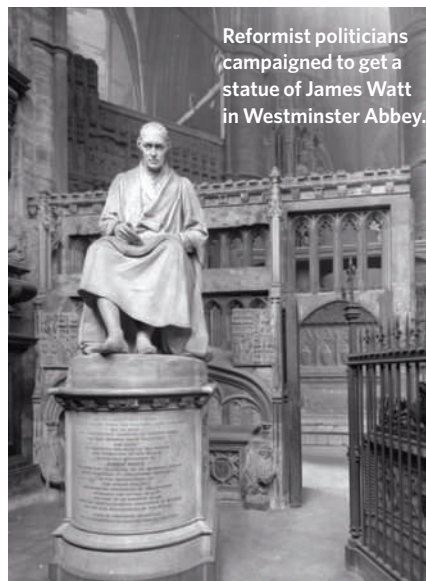
Reformist politicians campaigned to get a statue of James Watt in Westminster Abbey.

COUNTRY LIFE/THE BRIDGEMAN ART LIBRARY
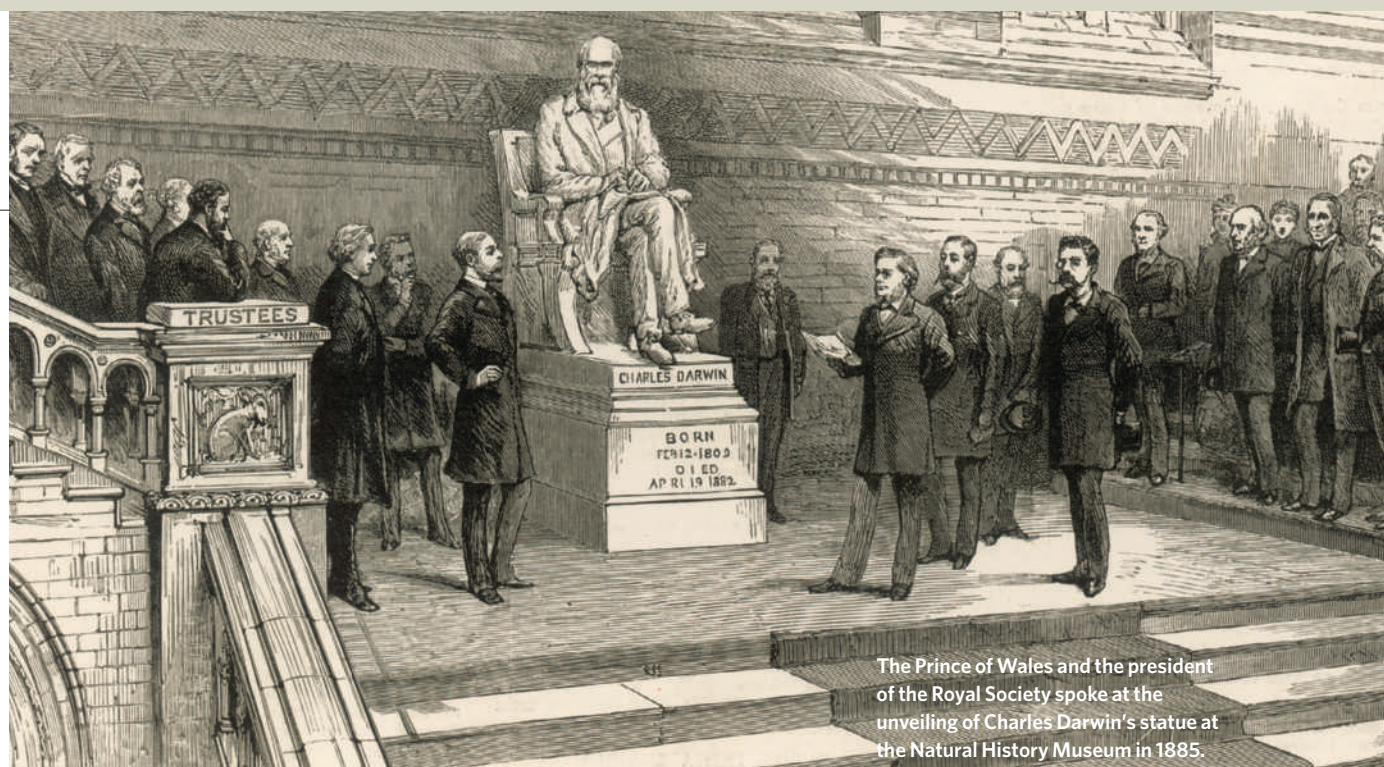
The Prince of Wales and the president of the Royal Society spoke at the unveiling of Charles Darwin's statue at the Natural History Museum in 1885.

both inventors and scientists was turning. Tales of self-help seemed hackneyed, social surveys demonstrated that industrialization had not eradicated poverty and the heroic rhetoric of invention had served its purpose. Half of all working men could now vote, and the patent system had been secured.

Few inventors could match the showmanship of the United States' Thomas Edison, with his dazzling displays of electric light, or of Italy's Guglielmo Marconi, best known for developing wireless telegraphy. Nor could they outdo the bravura of the American Wright brothers in managing to build as well as pilot the world's first successful aeroplane. Meanwhile, imperial exploration and conquest offered more exotic heroes whose widely reported adventures made most British inventors seem lacklustre.

When inventors started to lose popularity, academic scientists tried to present themselves as having a more important role. They started to depict invention either as the automatic spin-off from scientific research or as destined to fail if the inventor lacked a scientific education. Their increasing use of the term 'applied science' neatly expunged the role of the inventor.

An emerging rhetoric of 'pure science' insisted that researchers be funded to pursue their intellectual goals with no concern for utility, but that utility would ultimately reward society's trust. Purveyors of this view fastened, in particular, on Michael Faraday. In 1891, *The Times* newspaper reported how leading physicists celebrating the centenary of his birth had said: "All of us are now enjoying the material results of Faraday's research … but his eye was fixed upon truth itself, and not upon the useful results that

might come from the knowledge of it."

Such views boosted the status of science and scientists above that of inventors and engineers, but they didn't produce a new crop of scientific heroes. After 1914, science, technology and engineering became more anonymous. In corporate research and development labs, discoveries and inventions were claimed by faceless teams and obscured by brand names; in university departments, scientific publications took precedence over patents. The enormous losses of the Great War monopolized public commemorations. Historians played down the role of 'great men', the press grew more cynical, and culture as a whole less deferential.

Death of the science celebrity

Numerous researchers have caught the UK public eye in recent years: Alexander Fleming, James Watson, Francis Crick, Tim Berners-Lee and Stephen Hawking, to name a few. But footballers, musicians and film stars, not scientists, inventors or engineers, now receive the kind of nationwide celebrity status and commemorative fervour once lavished on Newton, Watt, Brunel and Darwin.

The United Kingdom's enthusiasm for scientific heroes may have been short-lived, but the reputations of those celebrated by the Victorians continue to influence people's perception of genius in science. Among the public and media, they reinforce the notion that the country's glory days of science and invention are long gone, and that its engineering has never recovered from the deaths in 1859 of Brunel and Robert Stephenson. Globally, their reputations feed a stereotype of great scientists as male and caucasian.

> "Machine-makers, railwaymen and spinning-machine operators cherished the memory of Watt."

A poll conducted in 2005 to determine who the French public viewed as their country's greatest people in history put Marie Curie fourth — yet she was the only woman in the six scientists and inventors included in the top 100. All 19 scientists and inventors who made it onto the list in an equivalent poll for the United Kingdom were male; as were the nine on a similar list compiled in the United States. Internationally, Europeans such as Darwin, Newton, Albert Einstein and Galileo enjoy iconic status on a scale that far surpasses the acclaim surrounding any non-Western scientist. More than 100 countries have issued Newton and Einstein stamps in recent years, for instance, whereas non-Western scientists rarely appear on stamps outside their own country.

It's unreasonable to heap serious blame on the white, whiskered Victorians — immortalized in stone and marble — for well-known gender and ethnic biases in recruitment. After all, Darwin has hardly deterred women from pursuing biology. But the culture of commemoration, initiated largely in Victorian Britain, carries the subliminal message that genius is male and caucasian. Indeed, it enshrines the very notion of scientific 'genius' and ties it firmly to the past.   ■

Christine MacLeod is in the School of Humanities, University of Bristol, 11 Woodland Road, Bristol, BS8 1TB, UK, and is the author of *Heroes of Invention: Technology, Liberalism and British Identity, 1750–1914* (Cambridge, 2007). e-mail: C.Macleod@bristol.ac.uk

FURTHER READING
Browne, J. *Br. J. Hist. Sci.* **38,** 251–274 (2005).
Fara, P. *Newton: the Making of Genius* (Picador, 2002).
Jordanova, L. *Defining Features: Scientific and Medical Portraits, 1660–2000* (Reaktion, 2000).
Oldenziel, R. *Making Technology Masculine: Men, Women and Modern Machines in America, 1870–1945* (Amsterdam Univ. Press, 1999).
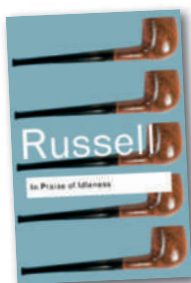
SUMMER BOOKS

ILLUSTRATIONS BY JONATHAN BURTON

A break from the bench

Nature regulars give their recommendations for relaxed, inspiring holiday reading and viewing — from climate-change history to Isaac Newton the detective.

DAVID POEPPEL
In Praise of Idleness: And Other Essays
by Bertrand Russell
Routledge Classics: 2004.

Many academics in the United States will have spent the past few months frenetically applying for support from the sudden infusion of research funds made available by Barack Obama's administration. Now, the over-stimulation of the stimulus package should be balanced with a break from the academy — with philosopher Bertrand Russell.

His essay *In Praise of Idleness*, written in 1932 and published with a group of works on modern life, reflects the delicate balance between frivolity and gravitas that is characteristic of his writing. As well as giving insightful arguments in favour of relaxation, his short pieces are politically edgy. Reading Russell is pure pleasure: just the right mixture of fun, provocation and raw insight. Reading this book on the beach or in a hammock will feel like winning a small moral victory.

David Poeppel is a professor in the Department of Psychology at New York University, New York.

MIKE BROWN
The Control of Nature
by John McPhee
Farrar, Straus and Giroux: 1990.

After I built a small retaining wall last year to prevent my back yard from tumbling into the canyon below, I reread *The Control of Nature*, John McPhee's exploration of human attempts and failures to dominate the natural world. Adding to my delight, these engrossing stories about geological catastrophes in distant places can now be transformed into a visceral confrontation with the real world by looking up the locations on Google Earth. You can track the paths of debris flows from the San Gabriel Mountains in Los Angeles, follow the convolutions of the Mississippi River and view the now-hardened Icelandic basalt — making you realize that these things really happened, right here.

But you don't need a computer to relish passages such as the opening of 'Los Angeles Against the Mountains', in which viscous mud quickly fills a house and presses an apparently doomed family against the ceiling. I know how that story ends, but every time I finish it I find my heart racing and my fingernails missing. And then I walk outside, look at my retaining wall, and add another rock or two.

Mike Brown is a professor of planetary astronomy at the California Institute of Technology, Pasadena.

SUSAN SOLOMON

The Long Thaw: How Humans Are Changing the Next 100,000 Years of Earth's Climate
by David Archer
Princeton University Press: 2008.

A hot summer is the perfect time to learn about climate change and carbon with this enjoyable and fast-paced treatise. Delve into *The Long Thaw* to contemplate how the actions and decisions being made now will affect our climate — not just in our lifetimes but for many millennia. David Archer masterfully sweeps through climate-change history and reviews the intricate cycling of carbon dioxide through the atmosphere, biosphere and ocean.

I have searched long and hard for an easily understandable description of ocean acidification, weathering and the enduring impact of humans on Earth's climate and carbon, and this book delivers on each of these challenges in a compelling way. It is comprehensive, well written and includes numerous useful vignettes from climate history. Archer leads the reader to a simple yet accurate picture of climate changes, ranging from geological time scales to current warming, ice ages and prospects for the future.

Susan Solomon is senior scientist at the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration in Boulder, Colorado.
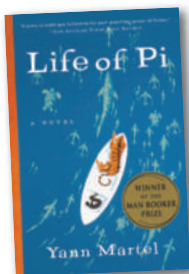
JERRY A. COYNE

Life of Pi
by Yann Martel
Houghton Mifflin Harcourt: 2002.

As I grow older, I find myself consuming much more nonfiction than fiction: there is so much to learn about the world before we leave it. But one book, *Life of Pi* by Yann Martel, stands out. It is the fantasy tale of a zookeeper's son, an Indian boy named Pi Patel. The ship transporting the Patels and their menagerie to Canada founders, leaving Pi in a lifeboat, his only companion a hungry Bengal tiger named Richard Parker.

Their story of survival amid mutual distrust and dependence is an engrossing drama, limned by ruminations about life, faith and the nature of reality. The story is also deeply informed by knowledge of animal behaviour that will appeal to biologists. Pi's defence

of zoos, for example, is a masterpiece of argument. A powerful piece of magical realism, *Life of Pi* is a worthy successor to Salman Rushdie's *Midnight's Children* and Gabriel García Márquez's *One Hundred Years of Solitude*.
Jerry A. Coyne is a professor in the Department of Ecology and Evolution at the University of Chicago, Illinois.
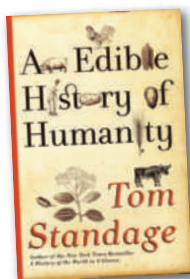
MING-WEI WANG

An Edible History of Humanity
by Tom Standage
Walker & Company: 2009.

"Bread is the staff of life" is an ancient saying. Yet most people never think of the effect that food has had on the history of humankind. Tom Standage's *An Edible History of Humanity* does so comprehensively using historical and analytical approaches (see also J. Hoffman *Nature* **459,** 912–913; 2009). The influence of food production and distribution on social change, political organization, geographical competition, industrial development, military conflict and economic expansion are likened to "an invisible fork that has, at several crucial points in history, prodded humanity and alerted its destiny".

Food is closely linked with the advancement of science and technology: "the adoption of agriculture is the tale of how ancient genetic

engineers developed powerful new tools that made civilization itself possible. In the process, mankind changed plants, and those plants in turn transformed mankind." Standage provides a thorough perspective that includes the genetically modified organisms of today, the relationship between food and poverty, the production of biofuels and the environmental impact of modern agriculture.
Ming-Wei Wang is director of the National Center for Drug Screening, Shanghai, China.

JONATHAN ZITTRAIN

In Search of Jefferson's Moose: Notes on the State of Cyberspace
by David Post
Oxford University Press: 2009.
When Species Meet
by Donna Haraway
University of Minnesota Press: 2007.
The Public Domain: Enclosing the Commons of the Mind
by James Boyle
Yale University Press: 2008.

David Post, Donna Haraway and James Boyle each offers insight into rapidly evolving relationships — among people, ideas and also animals. Post provides a vision of cyberspace through the prism of Thomas Jefferson, the American statesman and polymath who brought a moose to Paris to demonstrate just how different the New World was. To Post, the distributed nature of cyberspace opens up a sense of possibility that is well-mirrored by the frontiers of the late eighteenth century. He laments the idea that borders may eventually carve up this Internetworked space, in ways different from the borders that naturally arise from self-organizing online communities.

'Posthumanities' expert Haraway also concerns herself with unnecessary borders: those that she thinks humans have wrongly established between themselves and other species, particularly domesticated companion species. She explores the blurring of the lines between human and machine, real and virtual, natural and artificial.

Add to this Boyle's definitive work on the public domain, and one can see in common a

deep and spirited paean to 'remix' — a desire for legal, technical, economic and cultural architectures that permit ideas to flow freely from one mind to another, being reshaped in unexpected ways as they go. Boyle's call for digital literacy resonates with Post and Haraway's urging for us to see what is in front of us in a new light. And each draws beautifully on how things used to be to point the way forwards.

Jonathan Zittrain is a professor at Harvard Law School and co-founder of the Berkman Center for Internet & Society, Cambridge, Massachusetts.
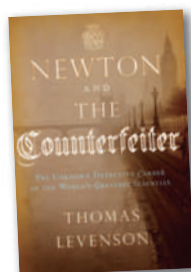
CARL ZIMMER
Newton and the Counterfeiter: The Unknown Detective Career of the World's Greatest Scientist
by Thomas Levenson

Houghton Mifflin Harcourt: 2009.

Imagine Isaac Newton as a policeman, using his brilliant intellect to pursue the most elusive thieves of his day. It sounds like a ridiculous example of historical fiction — except that it actually happened. In *Newton and the Counterfeiter*, Thomas Levenson recounts how Newton became Warden of the Royal Mint and pursued a master forger, William Chaloner. Levenson's account of the struggle between these two masterminds is fascinating on its own, but he also uses this historical episode to show how the modern economy took shape in the late 1600s.

The book's resonance with today's economic upheavals seems almost too good to be true. Would you believe that Isaac Newton lost much of his fortune in a financial crash? The physicists whose models of financial risk lie in ruins today can take some comfort in Levenson's stranger-than-fiction tale.

Carl Zimmer is a writer based in Connecticut.

FELICE FRANKEL
House, M.D. (Season 4 DVD)
Episode 15 ('House's Head') directed by Greg Yaitanes; Episode 16 ('Wilson's Heart') directed by Katie Jacobs

Written by David Shore, Peter Blake, David Foster, Russel Friend, Garrett Lerner and Doris Egan

Put aside your preconception that insightful storytelling never emanates from Hollywood; insert the DVD for season 4 of *House, M.D.* into your computer; and forward it to episode 15. For three-quarters of an hour, you will be engrossed in a mystery with scientifically and medically accurate twists and turns — and will then be compelled to watch episode 16 for the resolution. The interconnected plot of these episodes emphasizes the fragility of our lives and how one simple 'mistake' can change everything. You will witness mesmerizing directing and superb acting.

My concern in recommending only these two episodes is that you will not experience the full pleasure of watching Dr Gregory House, played by Hugh Laurie with deprecating wit and brilliance, nor the intelligent development of the relationships between the characters over the years, nor House's subtly uproarious insights into his colleagues, patients and the human condition — such as "everybody lies". For that, you will have to watch the complete series, which I am not embarrassed to admit I own.

Felice Frankel is a senior research fellow in the Faculty of Arts and Sciences at Harvard University and a research scientist at the Massachusetts Institute of Technology, Cambridge, Massachusetts.
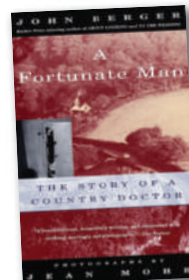
HUGH YOUNG RIENHOFF
A Fortunate Man: The Story of a Country Doctor
by John Berger

Vintage: 1997. First published by Holt, Rinehart and Winston: 1967.

I am reading John Berger's *A Fortunate Man* — again. An anthropologist, Berger manages to inhabit the skin of real-life physician John Sassall and at once take the pulse of his English villagers, society and the human condition. Berger is an invisible presence, documenting the inner and outer life of a man with a conflicted dedication to both science and human need. While tending to births, to the dying and to the suffering in between, Sassall concludes that illness has both a real and imagined component. Sickness is organic and psychic, one leading to the other.

Sassall confronts the plain facts of life without apparent argument. He is a witness to humanity's desperation and terrible loneliness and treats what he can. Sassall's devotion to his patients, his energy and his competence earn him the privilege of trust he uses so carefully. *A Fortunate Man* is not merely the best book written about a general practitioner, it teaches physicians about humanity and humility. Ultimately, it teaches the rest of us that inhumanity is society's greatest sickness.

Hugh Young Rienhoff is director of MyDaughtersDNA.org, based in San Francisco, California.

BRUCE HOOD
Descartes' Baby: How Child Development Explains What Makes Us Human
by Paul Bloom

Arrow Books: 2005. First published by William Heinemann: 2004.

From the opening line, "Having sex with dead animals is disgusting", I knew *Descartes' Baby* was going to be a fun read. Paul Bloom explores the human condition by following the turbulent voyage of a child's mind during development. Arguing that the infant brain is configured to interpret the world in terms of objects and people, and subsequently bodies and minds, he arrives at the mind–body dualism that philosopher René Descartes pondered — a solution that modern neuroscience has rejected.

This is the first popular book to articulate the emerging scientific interest in the idea that dualism of the material and immaterial world leads humans to naturally infer a deeper underlying reality. Such a belief system explains many of our attitudes and preferences to issues as diverse as sex, love, eating and organ transplantation. *Descartes' Baby* is a powerful book that tackles fundamental aspects of the human mind in an accessible style.

Bruce Hood is a professor in the Department of Experimental Psychology at the University of Bristol, Bristol, UK.

NEIL SHUBIN
Microcosm: *E. coli* and the New Science of Life
by Carl Zimmer
William Heinemann/ Pantheon Books: 2008.

What is life? Why do creatures cooperate? Why do living things die? Carl Zimmer, a leading writer on evolution, finds answers to these and other big questions in the most humble of places — the common gut microbe, *Escherichia coli*. In Zimmer's hands, *E. coli* becomes a window on to the basic properties of life and the ways that complex living systems can arise and change.

Zimmer weaves a narrative through the main principles of evolution, genetics and ageing, with stories of the people who made major breakthroughs along the way. His simple way of explaining complex ideas and his fast storytelling pace make for delightful reading. Each chapter contains 'wow' moments about bacteria and the joys and travails of the scientists who study them. The result is a scientific detective story that left me with a new appreciation of the trillions of microbes that live on and inside my body.

Neil Shubin is professor of anatomy at the University of Chicago, Illinois.
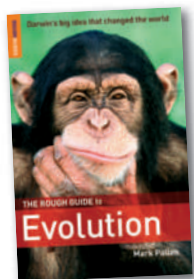
EUGENIE SCOTT
The Rough Guide to Evolution
by Mark Pallen
Rough Guides: 2009.

The 200th anniversary of Charles Darwin's birth, plus the 150th anniversary of the publication of his best-known book, *On the Origin of Species,* make 2009 the year to learn about evolution.

Mark Pallen's *The Rough Guide to Evolution* provides a concise summary of what you need to know: a brief history of the idea that all living things share common ancestry, a complete survey of the mechanisms of evolution and a solid summary of how life originated and then adapted through time to a changing planet. He livens up the story with literary, musical and cultural references so that you never feel you are being told to eat your vegetables. Alas, it is not only non-specialists who don't have a firm grasp of the strength of theory and data supporting the modern understanding of evolution — many scientists outside the field of evolutionary biology struggle too. This entertaining handbook will bring anyone up to date.

Eugenie Scott is executive director of the National Center for Science Education, Oakland, California.

SANDRA KNAPP
On the Origin of Species by Means of Natural Selection
by Charles Darwin
Penguin Classics: 2009.
First published by John Murray: 1859.

On the Origin of Species is the crystallization on paper of what is arguably biology's biggest idea, and it remains one of the clearest explanations of evolution. Although the intricacies of its Victorian prose might put some readers off, the book is as readable as those of Charles Dickens. It is beautifully constructed; Charles Darwin starts with familiar domestic animals, guiding the reader through increasingly complicated ideas such as embryology and hybridization.

The book is one long argument, just like the practice of science. Never mind that Darwin didn't know as much about inheritance as we do today; it is all here in the first articulation of those two great interlinked ideas — descent with modification and natural selection — that allow us to explain how the natural world works. Of the many editions of *Origin*, I prefer the first. It has the spontaneity and verve of thoughts spilling out on to the page, and is untainted by the revisions introduced in later editions in response to critics.

Sandra Knapp is a plant taxonomist in the Department of Botany at the Natural History Museum, London, UK.
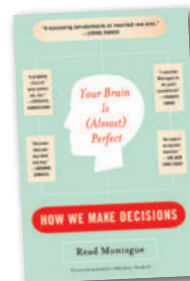
ADAM KEPECS
Your Brain Is (Almost) Perfect: How We Make Decisions
by Read Montague
Plume (Penguin): 2007.

Read Montague brings together neuroscience, psychology and computer science to explain how brains function as efficient computational machines. He first looks at the brain as a physical device, exploring how this energy-efficient information-processing system evolved out of slow and noisy biological components. Montague argues that it is the valuation system that makes brains different from computers, allowing brains to compare disparate options and to 'care' about different goals. Montague elegantly weaves a radical outline of how brains make decisions and ultimately produce the mind.

It is refreshing to read an intelligent book that bucks the trend of many popular expositions of science that simply recount a series of anecdotes, often to the exclusion of deep ideas. This original and clearly written contribution may not be an easy read, but its insights are well worth the effort.

Adam Kepecs is assistant professor of neuroscience at Cold Spring Harbor Laboratory, New York.

NEWS & VIEWS

OPTICS

All smoke and metamaterials

John Pendry

An illusion device, placed near but not enclosing an object of arbitrary shape, manipulates and transforms light scattered off the object so as to give it the appearance of a completely different object.

Seeing is believing — a naive assumption in the case of an illusion device proposed by Lai and colleagues at the Hong Kong University of Science and Technology, and described[1] in *Physical Review Letters*. The new device has the power to 'act at a distance' and therefore covertly alter an object's appearance such that it has no apparent physical connection to the light scattered by the object — although this becomes increasingly difficult to achieve the farther the illusion device is from the object. Lai and colleagues[1] outline a mathematical formalism proving that it is theoretically possible to grab the rays of light emitted by a given object and to reconstruct them so that they seem to come from a completely different object.

But how is this possible? Our brains interpret light reaching our eyes as if it consists of a stream of particles travelling in a straight line from the source. For most practical purposes, this is an excellent approximation. But light has in fact a wave-like character, and the straight-line approximation to light propagation sometimes breaks down. Over the past decade, scientists working in the field of optics have been investigating circumstances in which this breakdown takes an extreme form.

The story begins in 1968, when a Soviet scientist, Victor Veselago, showed[2] that a flat slab of negatively refracting material, one that bends light in the 'wrong' direction, could act as a lens so that objects placed on one side of the slab are seen on the far side as images of exactly the same apparent size as the object. His long-neglected work has attracted much interest over the past decade, partly because the new technology of metamaterials — artificially engineered materials that owe their unique properties to their internal physical structure rather than their chemical composition — enables negative refraction to be experimentally achievable. Other remarkable properties of Veselago's lens have also been revealed. For example, it has been shown that its resolution is not constrained by the normal limits of wavelength, as in a conventional lens (an ordinary lens cannot focus light on an area smaller than the square of the light's wavelength), but only by the perfection with which it can be manufactured[3].
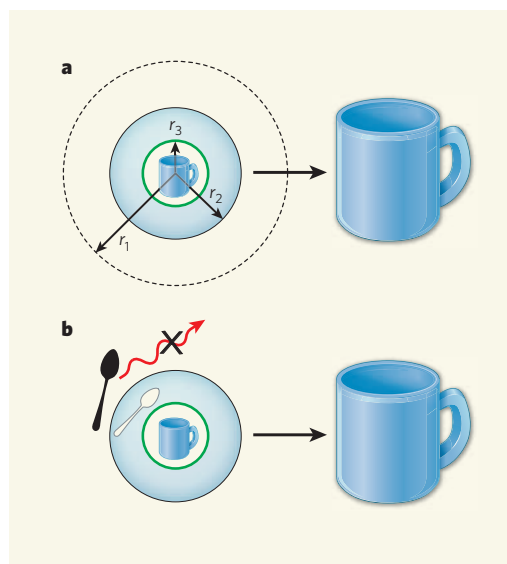

Figure 1 | An illusion device in action. a, A spherical lens system incorporates negatively refracting material — one that bends light in the opposite direction to conventional materials — within the shaded annulus of radius r_2 (blue). The outer annulus of radius r_1 contains only free space, and the innermost sphere of radius r_3 contains a cup. An external observer would see the contents of the inner sphere magnified to fill a sphere of radius r_1. **b,** In Lai and colleagues' illusion device[1], the lens is designed to eliminate any light scattering from a spoon located in free space in the outer annulus: an 'antispoon' in the shaded annulus would effectively cancel any light scattering from the spoon itself. An external observer would see no trace of the spoon, only a magnified image of the cup — the illusion object. A ray of light (red) is shown scattering from the spoon.

Another relevant development is a very powerful design tool for optical systems[4]. Imagine that the optical system in question, such as a lens, is embedded in a rubber medium. We then stretch and pull the rubber, taking with it all the rays of light passing through the lens, until the rays are travelling in the desired directions. Of course, the electric and magnetic fields associated with the light no longer obey Maxwell's equations for the propagation of electromagnetic waves. But this can be remedied by altering two properties of the distorted medium: the electric permittivity and the magnetic permeability. In fact, if we carefully record the changed shape of the rubber medium by measuring the distortion of an embedded system of coordinates, the relationship between the old and new coordinates gives the prescription we need. This tool, known as transformation optics, was originally applied to generalizing Veselago's lens into a magnifying glass, but attained notoriety when it was used to design a 'cloak of invisibility'[5,6].

Figure 1 shows this transformed lens in operation. An object, in this case a cup, is enclosed by a negatively refracting lens of radius r_2, which projects an image of the cup into the volume occupied by an outer sphere of radius r_1. Note that everything outside the shaded annulus, the lens, is free space. It is as if the negatively refracting material 'annihilates' the adjoining space out as far as radius r_1, and the inner sphere of radius r_3 containing the cup expands to fill this space (Fig. 1a). This system is the starting point for Lai and colleagues' illusion device[1].

The authors[1] observe that negatively refracting material, appropriately structured, can annihilate any adjoining positively refracting material[7]. Transformation optics provides the recipe and metamaterials provide the means. If, as in the example suggested by Lai *et al.*, a spoon is placed in the free space beside the illusion device, and an 'antispoon' incorporated in the annulus of radius r_2 is designed to cancel any light scattering from the spoon itself, the inner sphere containing the cup expands to fill this space and effectively replaces the spoon with a cup (Fig. 1b).

But how can the illusion device take control of rays that never strike it? In reality, rays are formed from a collection of waves, and the best approximation to a ray that waves can make is a 'Gaussian beam', which actually has a finite lateral extent. These fuzzy tails on the rays will have a slight overlap with the device which is designed to be peculiarly sensitive to them. The device responds by launching a second ray

(not shown in Fig. 1) calculated to cancel the effect of the first.

Potential applications of Lai and colleagues' illusion device abound, ranging from altering the radar signature of an aircraft to creating large shadows using only small objects. However, it should be noted that this illusion concept is extremely demanding of material properties, and some of the more exotic applications are likely to remain in the realm of theory. ■

John Pendry is in the Condensed Matter Theory

Group, The Blackett Laboratory, Imperial College London, London SW7 2AZ, UK.
e-mail: j.pendry@imperial.ac.uk

1. Lai, Y. et al. Phys. Rev. Lett. 102, 253902 (2009).
2. Veselago, V. G. Sov. Phys. Uspekhi 10, 509–514 (1968).
3. Pendry, J. B. Phys. Rev. Lett. 85, 3966–3969 (2000).
4. Pendry, J. B. in Coherence and Quantum Optics IX (eds Bigelow, N. P. et al.) 42–52 (Opt. Soc. Am., 2009).
5. Pendry, J. B., Schurig, D. & Smith, D. R. Science 312, 1780–1782 (2006).
6. Schurig, D. et al. Science 314, 977–980 (2006).
7. Pendry, J. B. & Ramakrishna, S. A. J. Phys. Condens. Matter 15, 6345–6364 (2003).

STRUCTURAL BIOLOGY

Trimeric ion-channel design

Shai D. Silberberg and Kenton J. Swartz

Cavernous chambers, intricate passages, a gate with a curious lock — the structure of an ATP-activated ion channel reveals its architecture. And this intriguing interior design is found in another type of ion channel too.

Ion channels are membrane-protein complexes that allow the passage of ions into and out of cells. One type of ion channel, the P2X receptor channel[1,2], is activated by extracellular ATP, the nucleotide more commonly known for providing cells with energy. Binding of ATP to the P2X receptor triggers opening of a transmembrane pore, allowing sodium, potassium and calcium ions[3] to flow down their electrochemical gradients, changing membrane voltage or activating intracellular signalling cascades, mediating processes as diverse as pain perception and inflammation[1,2]. Elegant biochemical studies[4] had suggested

that P2X ion channels are trimers, distinguishing them from the better-studied tetrameric and pentameric ion-channel families. On page 592 of this issue, Kawate and colleagues[5] confirm these predictions in their report of the first X-ray crystal structure of a type of P2X receptor, P2X$_4$. In an accompanying paper (page 599), Gonzales and colleagues[6] describe the structure of another trimeric ion channel — an acid-sensing ion channel (ASIC) — which reveals unanticipated similarities with P2X receptors.

For those of us interested in the function of ion channels, an X-ray structure is a game-

changer, offering crucial information about operational mechanisms and providing a firm grounding for further investigations. Obtaining structures of membrane proteins is an arduous task, and P2X was no exception, requiring seven years to come to fruition. The structure shows that P2X's three dolphin-shaped subunits wrap around each other to form the trimer, and that two transmembrane α-helices (TM1 and TM2) from each subunit span the cell membrane (Fig. 1a). The structure was solved in the absence of ATP, and the region of the protein that lies within the membrane lacks a pathway large enough for ions to cross the membrane, suggesting that the structure represents a closed state.

The structure of the P2X receptor[5] is fascinating for many reasons. First, it reveals insight about the region of the protein that forms the transmembrane ion-conducting pore. Previous work had suggested that amino-acid residues in the second transmembrane (TM2) segment influence the type of ion that can pass through the channel[3] and that the three TM2 segments probably line the pore[3,7]. In the P2X receptor structure, the TM2 segments are steeply angled relative to the membrane plane, giving the pore an hourglass appearance consisting of chambers (vestibules) on either side of an extended plug of residues near the middle of the membrane (Fig. 1b). Residues in the region of the plug are inaccessible to ions on the outside of the membrane when the pore is closed, but are highly accessible in the open state[7]. This indicates that the plug functions as a gate that opens when ATP binds to the receptor. The extracellular vestibule above the gate opens to the outside through fenestrations (Fig. 1a; white asterisk), providing a passageway for ions to enter or exit the extracellular side of the pore — a feature that is reminiscent of the fenestrations seen in some potassium channels[8,9] and in the acetylcholine receptor[10].

Another striking feature[5] is the presence of two outer electronegative (acidic) vestibules located within the extracellular domain (Fig. 1b). The proximity of the central vestibule to the pore, and its negative charge, make one wonder whether this central chamber attracts positively charged ions to the channel. Cavities of similar dimensions have been found in a range of proteins, in which they are proposed to facilitate movement between domains by decreasing the contact surface, or to serve as binding sites for cofactors or regulatory compounds[11]. Indeed, Kawate and colleagues[5] found gadolinium (Gd^{3+}), a heavy metal used to solve the structure, in the central vestibule of the P2X receptor, and showed that Gd^{3+} inhibits P2X function. If the outer two vestibules serve as binding sites for regulatory ions, how do such ions get into these chambers, and how do they affect the function of the ion channel? One possibility is that ions access the vestibules through a pathway along the P2X central axis that dilates when ATP binds.

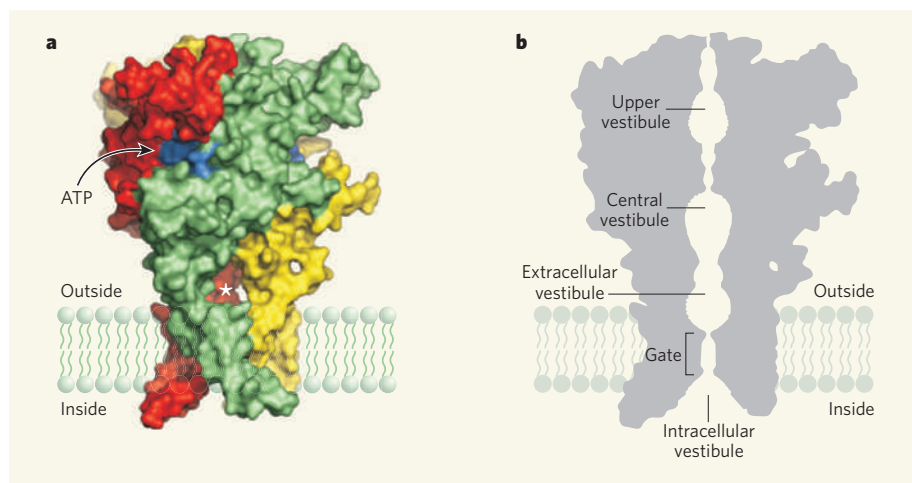So where does ATP bind to open the ion

Figure 1 | Structure of the P2X receptor. Kawate and colleagues[5] report the first X-ray structure of a trimeric P2X$_4$ receptor ion channel. **a**, Surface representation of the P2X$_4$ receptor with the three identical subunits coloured green, red and yellow, and the ATP-binding pockets coloured blue. The white asterisk marks one of three fenestrations between subunits where the extracellular vestibule opens to the exterior, providing a potential passageway for ions to enter or exit from the extracellular side of the pore. **b**, Cartoon representation of the P2X receptor ion channel showing the four vestibules and a gate formed by a plug of hydrophobic amino-acid residues within the pore. Gonzales and colleagues[6] report a new structure of another trimeric ion channel, an acid-sensing ion channel (ASIC), revealing that the transmembrane regions and vestibules in the ASIC are remarkably similar to those in P2X receptors.

(not shown in Fig. 1) calculated to cancel the effect of the first.

Potential applications of Lai and colleagues' illusion device abound, ranging from altering the radar signature of an aircraft to creating large shadows using only small objects. However, it should be noted that this illusion concept is extremely demanding of material properties, and some of the more exotic applications are likely to remain in the realm of theory. ■

John Pendry is in the Condensed Matter Theory

Group, The Blackett Laboratory, Imperial College London, London SW7 2AZ, UK.
e-mail: j.pendry@imperial.ac.uk

1. Lai, Y. et al. Phys. Rev. Lett. 102, 253902 (2009).
2. Veselago, V. G. Sov. Phys. Uspekhi 10, 509–514 (1968).
3. Pendry, J. B. Phys. Rev. Lett. 85, 3966–3969 (2000).
4. Pendry, J. B. in Coherence and Quantum Optics IX (eds Bigelow, N. P. et al.) 42–52 (Opt. Soc. Am., 2009).
5. Pendry, J. B., Schurig, D. & Smith, D. R. Science 312, 1780–1782 (2006).
6. Schurig, D. et al. Science 314, 977–980 (2006).
7. Pendry, J. B. & Ramakrishna, S. A. J. Phys. Condens. Matter 15, 6345–6364 (2003).

STRUCTURAL BIOLOGY

Trimeric ion-channel design

Shai D. Silberberg and Kenton J. Swartz

Cavernous chambers, intricate passages, a gate with a curious lock — the structure of an ATP-activated ion channel reveals its architecture. And this intriguing interior design is found in another type of ion channel too.

Ion channels are membrane-protein complexes that allow the passage of ions into and out of cells. One type of ion channel, the P2X receptor channel[1,2], is activated by extracellular ATP, the nucleotide more commonly known for providing cells with energy. Binding of ATP to the P2X receptor triggers opening of a transmembrane pore, allowing sodium, potassium and calcium ions[3] to flow down their electrochemical gradients, changing membrane voltage or activating intracellular signalling cascades, mediating processes as diverse as pain perception and inflammation[1,2]. Elegant biochemical studies[4] had suggested

that P2X ion channels are trimers, distinguishing them from the better-studied tetrameric and pentameric ion-channel families. On page 592 of this issue, Kawate and colleagues[5] confirm these predictions in their report of the first X-ray crystal structure of a type of P2X receptor, P2X₄. In an accompanying paper (page 599), Gonzales and colleagues[6] describe the structure of another trimeric ion channel — an acid-sensing ion channel (ASIC) — which reveals unanticipated similarities with P2X receptors.

For those of us interested in the function of ion channels, an X-ray structure is a game-

changer, offering crucial information about operational mechanisms and providing a firm grounding for further investigations. Obtaining structures of membrane proteins is an arduous task, and P2X was no exception, requiring seven years to come to fruition. The structure shows that P2X's three dolphin-shaped subunits wrap around each other to form the trimer, and that two transmembrane α-helices (TM1 and TM2) from each subunit span the cell membrane (Fig. 1a). The structure was solved in the absence of ATP, and the region of the protein that lies within the membrane lacks a pathway large enough for ions to cross the membrane, suggesting that the structure represents a closed state.

The structure of the P2X receptor[5] is fascinating for many reasons. First, it reveals insight about the region of the protein that forms the transmembrane ion-conducting pore. Previous work had suggested that amino-acid residues in the second transmembrane (TM2) segment influence the type of ion that can pass through the channel[3] and that the three TM2 segments probably line the pore[3,7]. In the P2X receptor structure, the TM2 segments are steeply angled relative to the membrane plane, giving the pore an hourglass appearance consisting of chambers (vestibules) on either side of an extended plug of residues near the middle of the membrane (Fig. 1b). Residues in the region of the plug are inaccessible to ions on the outside of the membrane when the pore is closed, but are highly accessible in the open state[7]. This indicates that the plug functions as a gate that opens when ATP binds to the receptor. The extracellular vestibule above the gate opens to the outside through fenestrations (Fig. 1a; white asterisk), providing a passageway for ions to enter or exit the extracellular side of the pore — a feature that is reminiscent of the fenestrations seen in some potassium channels[8,9] and in the acetylcholine receptor[10].

Another striking feature[5] is the presence of two outer electronegative (acidic) vestibules located within the extracellular domain (Fig. 1b). The proximity of the central vestibule to the pore, and its negative charge, make one wonder whether this central chamber attracts positively charged ions to the channel. Cavities of similar dimensions have been found in a range of proteins, in which they are proposed to facilitate movement between domains by decreasing the contact surface, or to serve as binding sites for cofactors or regulatory compounds[11]. Indeed, Kawate and colleagues[5] found gadolinium (Gd³⁺), a heavy metal used to solve the structure, in the central vestibule of the P2X receptor, and showed that Gd³⁺ inhibits P2X function. If the outer two vestibules serve as binding sites for regulatory ions, how do such ions get into these chambers, and how do they affect the function of the ion channel? One possibility is that ions access the vestibules through a pathway along the P2X central axis that dilates when ATP binds.
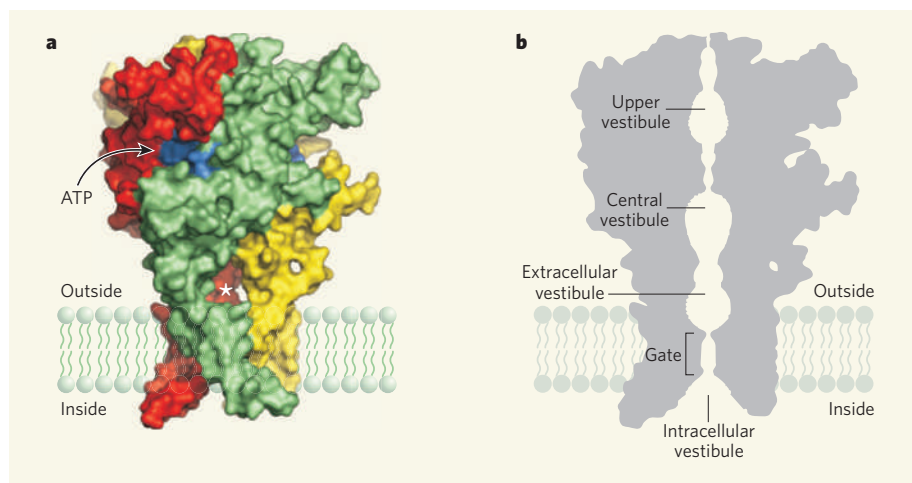
So where does ATP bind to open the ion

Figure 1 | Structure of the P2X receptor. Kawate and colleagues[5] report the first X-ray structure of a trimeric P2X₄ receptor ion channel. **a**, Surface representation of the P2X₄ receptor with the three identical subunits coloured green, red and yellow, and the ATP-binding pockets coloured blue. The white asterisk marks one of three fenestrations between subunits where the extracellular vestibule opens to the exterior, providing a potential passageway for ions to enter or exit from the extracellular side of the pore. **b**, Cartoon representation of the P2X receptor ion channel showing the four vestibules and a gate formed by a plug of hydrophobic amino-acid residues within the pore. Gonzales and colleagues[6] report a new structure of another trimeric ion channel, an acid-sensing ion channel (ASIC), revealing that the transmembrane regions and vestibules in the ASIC are remarkably similar to those in P2X receptors.

channel? Because the P2X structure was solved in the absence of ATP, the binding site isn't visible. But experiments in which amino acids were mutated[12] implicate specific basic and other polar (hydrophilic) residues that line a pocket on the outer surface of the receptor at the interface between each subunit[5] (Fig 1a). Although the structure of this site is unlike that of conventional ATP-binding motifs[13], such motifs are notoriously diverse and the residues lining the P2X pocket are appropriate for ATP binding. It will be interesting to find out whether ATP binds to this region, to determine whether magnesium ions are required for ATP binding, as it is in many other proteins, and to compare the chemistry of ligand binding with that observed in other ATP-binding proteins. Answering these questions will be highly informative for designing new drugs targeting P2X receptors.

The P2X receptor is not the first trimeric ion-channel structure to be solved. A structure had been reported for an acid-sensing ion channel (ASIC)[14], but the protein yielding this structure was non-functional and the transmembrane segments adopted a non-native conformation. The new ASIC structure reported by Gonzales and colleagues[6] was solved from a functional protein, and so provides several novel details. Although P2X receptors and ASICs have unrelated amino-acid sequences and open in response to different ligands (ATP versus protons), their transmembrane regions are remarkably similar. The pore of the ASIC is closed by an extended plug that is similar to that in the P2X receptor. And although the extracellular domains of P2X receptors and ASICs are structurally unrelated, both have the acidic central and upper vestibules. By soaking ASIC crystals in different ions, the authors[6] located an ion-binding site in the extracellular vestibule, providing a first glimpse of how trimeric channels might coordinate (and thereby select for) different permeating ions.

These two papers[5,6] are wonderful examples of how protein structures help to frame a wealth of fascinating questions. How does ligand binding to the extracellular domains trigger opening of the pore, and how does the structure of the pore change during opening? What is the role of the two acidic vestibules in the extracellular domains? Do the similarities in the pore regions of the two channels mean that they gate using similar mechanisms, or are there simply limited ways to construct an ion-conducting pore from so few transmembrane helices? Answering these fundamental questions will undoubtedly keep nerds busy for years to come. ■

Shai D. Silberberg and Kenton J. Swartz are in the Molecular Physiology and Biophysics Section, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mails: shai.silberberg@nih.gov; kenton.swartz@nih.gov

1. Burnstock, G. Cell. Mol. Life Sci. 64, 1471–1483 (2007).
2. Surprenant, A. & North, R. A. Annu. Rev. Physiol. 71, 333–359 (2009).
3. Egan, T. M. & Khakh, B. S. J. Neurosci. 24, 3413–3420 (2004).
4. Nicke, A. et al. EMBO J. 17, 3016–3028 (1998).
5. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Nature 460, 592–598 (2009).
6. Gonzales, E. B., Kawate, T. & Gouaux, E. Nature 460, 599–604 (2009).
7. Li, M., Chang, T.-H., Silberberg, S. D. & Swartz, K. J. Nature Neurosci. 11, 883–887 (2008).
8. Kobertz, W. R., Williams, C. & Miller, C. Biochemistry 39, 10347–10352 (2000).
9. Long, S. B., Campbell, E. B. & MacKinnon, R. Science 309, 897–903 (2005).
10. Unwin, N. J. Mol. Biol. 346, 967–989 (2005).
11. Sonavane, S. & Chakrabarti, P. PLoS Comput. Biol. 4, e1000188 (2008).
12. Evans, R. J. Eur. Biophys. J. 38, 319–327 (2009).
13. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. EMBO J. 1, 945–951 (1982).
14. Jasti, J., Furukawa, H., Gonzales, E. B. & Gouaux, E. Nature 449, 316–323 (2007).

OCEANOGRAPHY

A fishy mix

William K. Dewar

Ocean life is in almost constant motion, and such activity must surely stir things up. Innovative investigations into this concept of 'biogenic mixing' show a role for jellyfish and their brethren.

Well over 100 years ago, scientists speculated that the ocean was driven by the Sun, the Moon, Earth's rotation and the combined motion of all the fishes' tails (M. G. Briscoe, personal communication). Since then, there has been only sporadic interest in the idea that swimming fish and other organisms help to stir the ocean. But, recently, the concept has experienced a resurgence of interest, and the latest example of thinking on the topic appears in the paper by Katija and Dabiri on page 624 of this issue[1]. They offer a new angle: "a viscosity-enhanced mechanism for biogenic ocean mixing".

The broader context here is that the amounts of heat and carbon stored by the ocean dwarf those held by the atmosphere, and that to understand climate change it is essential to understand the processes that affect ocean–atmosphere exchange. The studies involved have led to some counter-intuitive results. First, although ocean properties with length scales of thousands of kilometres matter most to climate, they are sensitive to mixing processes on scales of a few centimetres — think of the way you stir cream into your morning coffee; similarly, tiny whorls mix in the ocean. Second, beneath the surface, starting only about one football field deep, the ocean is a very quiet mixing environment. Roughly speaking, all the energy needed to mix a cubic kilometre of subsurface ocean could be provided by a single hand-held kitchen mixer. Through a remarkable interplay of length and timescales, very weak and small-scale mixing helps to set our climate and affects the burial of atmospheric carbon in the ocean.

It was the venerable Walter Munk[2] who, in 1966, attempted to quantify the effect that activity in the ocean biosphere might have on ocean

Figure 1 | Shoal of *Mastigias* jellyfish. These are the experimental subjects studied by Katija and Dabiri[1], also pictured on the cover of this issue. The study site was a saltwater lake — Jellyfish Lake — on the Pacific island of Palau.

K. KATIJA

channel? Because the P2X structure was solved in the absence of ATP, the binding site isn't visible. But experiments in which amino acids were mutated[12] implicate specific basic and other polar (hydrophilic) residues that line a pocket on the outer surface of the receptor at the interface between each subunit[5] (Fig 1a). Although the structure of this site is unlike that of conventional ATP-binding motifs[13], such motifs are notoriously diverse and the residues lining the P2X pocket are appropriate for ATP binding. It will be interesting to find out whether ATP binds to this region, to determine whether magnesium ions are required for ATP binding, as it is in many other proteins, and to compare the chemistry of ligand binding with that observed in other ATP-binding proteins. Answering these questions will be highly informative for designing new drugs targeting P2X receptors.

The P2X receptor is not the first trimeric ion-channel structure to be solved. A structure had been reported for an acid-sensing ion channel (ASIC)[14], but the protein yielding this structure was non-functional and the transmembrane segments adopted a non-native conformation. The new ASIC structure reported by Gonzales and colleagues[6] was solved from a functional protein, and so provides several novel details. Although P2X receptors and ASICs have unrelated amino-acid sequences and open in response to different ligands (ATP versus protons), their transmembrane regions are remarkably similar. The pore of the ASIC is closed by an extended plug that is similar to that in the P2X receptor. And although the extracellular domains of P2X receptors and ASICs are structurally unrelated, both have the acidic central and upper vestibules. By soaking ASIC crystals in different ions, the authors[6] located an ion-binding site in the extracellular vestibule, providing a first glimpse of how trimeric channels might coordinate (and thereby select for) different permeating ions.

These two papers[5,6] are wonderful examples of how protein structures help to frame a wealth of fascinating questions. How does ligand binding to the extracellular domains trigger opening of the pore, and how does the structure of the pore change during opening? What is the role of the two acidic vestibules in the extracellular domains? Do the similarities in the pore regions of the two channels mean that they gate using similar mechanisms, or are there simply limited ways to construct an ion-conducting pore from so few transmembrane helices? Answering these fundamental questions will undoubtedly keep nerds busy for years to come. ■

Shai D. Silberberg and Kenton J. Swartz are in the Molecular Physiology and Biophysics Section, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mails: shai.silberberg@nih.gov; kenton.swartz@nih.gov

1. Burnstock, G. Cell. Mol. Life Sci. 64, 1471–1483 (2007).
2. Surprenant, A. & North, R. A. Annu. Rev. Physiol. 71, 333–359 (2009).
3. Egan, T. M. & Khakh, B. S. J. Neurosci. 24, 3413–3420 (2004).
4. Nicke, A. et al. EMBO J. 17, 3016–3028 (1998).
5. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Nature 460, 592–598 (2009).
6. Gonzales, E. B., Kawate, T. & Gouaux, E. Nature 460, 599–604 (2009).
7. Li, M., Chang, T.-H., Silberberg, S. D. & Swartz, K. J. Nature Neurosci. 11, 883–887 (2008).
8. Kobertz, W. R., Williams, C. & Miller, C. Biochemistry 39, 10347–10352 (2000).
9. Long, S. B., Campbell, E. B. & MacKinnon, R. Science 309, 897–903 (2005).
10. Unwin, N. J. Mol. Biol. 346, 967–989 (2005).
11. Sonavane, S. & Chakrabarti, P. PLoS Comput. Biol. 4, e1000188 (2008).
12. Evans, R. J. Eur. Biophys. J. 38, 319–327 (2009).
13. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. EMBO J. 1, 945–951 (1982).
14. Jasti, J., Furukawa, H., Gonzales, E. B. & Gouaux, E. Nature 449, 316–323 (2007).

OCEANOGRAPHY

A fishy mix

William K. Dewar

Ocean life is in almost constant motion, and such activity must surely stir things up. Innovative investigations into this concept of 'biogenic mixing' show a role for jellyfish and their brethren.

Well over 100 years ago, scientists speculated that the ocean was driven by the Sun, the Moon, Earth's rotation and the combined motion of all the fishes' tails (M. G. Briscoe, personal communication). Since then, there has been only sporadic interest in the idea that swimming fish and other organisms help to stir the ocean. But, recently, the concept has experienced a resurgence of interest, and the latest example of thinking on the topic appears in the paper by Katija and Dabiri on page 624 of this issue[1]. They offer a new angle: "a viscosity-enhanced mechanism for biogenic ocean mixing".

The broader context here is that the amounts of heat and carbon stored by the ocean dwarf those held by the atmosphere, and that to understand climate change it is essential to understand the processes that affect ocean–atmosphere exchange. The studies involved have led to some counter-intuitive results. First, although ocean properties with length scales of thousands of kilometres matter most to climate, they are sensitive to mixing processes on scales of a few centimetres — think of the way you stir cream into your morning coffee; similarly, tiny whorls mix in the ocean. Second, beneath the surface, starting only about one football field deep, the ocean is a very quiet mixing environment. Roughly speaking, all the energy needed to mix a cubic kilometre of subsurface ocean could be provided by a single hand-held kitchen mixer. Through a remarkable interplay of length and timescales, very weak and small-scale mixing helps to set our climate and affects the burial of atmospheric carbon in the ocean.

It was the venerable Walter Munk[2] who, in 1966, attempted to quantify the effect that activity in the ocean biosphere might have on ocean

Figure 1 | Shoal of _Mastigias_ jellyfish. These are the experimental subjects studied by Katija and Dabiri[1], also pictured on the cover of this issue. The study site was a saltwater lake — Jellyfish Lake — on the Pacific island of Palau.

K. KATIJA

mixing. His result neither strongly dismissed nor supported the idea, but for decades afterwards conventional wisdom held that fish could be ignored in ocean mixing. It often happens in science that a flurry of unconnected activities on a common topic emerge almost simultaneously, however, and such has been the case for biogenic ocean mixing.

In 2004, Huntley and Zhou[3] pointed out that the expected levels of turbulence in schools of fish are comparable to those associated with storms. In a subsequent paper, my colleagues and I argued[4] that the kinetic energy expended by the biosphere is sizeable compared with global mixing requirements; we further suggested that the true swimmers (fish), when all lumped together, provide about half of the biosphere input, with the balance coming from zooplankton. Exciting, direct confirmation of hugely elevated turbulence levels in vertically migrating shrimp-like animals followed from Kunze and colleagues[5].

A major question is how efficiently biogenic turbulence actually mixes the ocean. The answer hinges on length scales. Very small whorls introduced into a fluid will be quickly damped by friction, and thus will not mix the fluid. To illustrate, consider a tall coffee cup with a slight gradient in creaminess from top to bottom; small whorls at the bottom would have little effect on the cream at the top before dying a frictional death. Guidance on the size at which turbulence changes from unimportant to important in mixing is provided by the Ozmidov scale, which takes into account how stratified a fluid is and how strong the turbulence is. Given that many zooplankton are comparable to or smaller than oceanic Ozmidov scales, one view is that biogenic mixing is negligible[6]. The story is not yet complete, however. It could be that zooplankton schooling introduces larger scales and increases mixing efficiency[7]. Although an attempt[8] to observe such an effect failed to do so, the search continues.

Into this mix comes the paper by Katija and Dabiri[1]. The authors emphasize that the mere act of swimming implies that some water travels with the swimmer. Whereas viscosity lessens the effect of turbulent mixing, here it is found to increase the total transport. In remarkable videos obtained by scuba divers in shoals of jellyfish (Fig. 1), dye releases clearly show the process (see Supplementary Information[1]). One wonders what the jellyfish made of all this, but that would be another story.

The relevance to mixing, however, can be simply described. Suppose a jellyfish is in cold water, and swims vertically to warmer zones. Some amount of cold water will follow (the videos suggest a surprisingly large amount). Once there, mixing of the local fluid properties ensues. From energetics estimates based on the dye's behaviour, the effect seems to be sizeable. This mechanism is implicit in previous energetics estimates, but it has escaped explicit notice until now and lessens doubts,

based on Ozmidov scales, about the possible strength of biogenic mixing.

Translation of Katija and Dabiri's results from anecdotes to assessments of possible global impacts remains to be carried out. Should the overall idea of significant biogenic mixing survive detailed scrutiny, climate science will have experienced a paradigm shift. To quote Carl Wunsch[9], modellers will "need to start thinking about the fluid dynamics of biology", to which he added, "that's a tough one" — as, indeed, it is. ∎

William K. Dewar is in the Department of

Oceanography, Florida State University, Tallahassee, Florida 32306-4320, USA. e-mail: dewar@ocean.fsu.edu

1. Katija, K. & Dabiri, J. O. *Nature* **460,** 624–626 (2009).
2. Munk, W. H. *Deep-Sea Res.* **13,** 707–730 (1966).
3. Huntley, M. E. & Zhou, M. *Mar. Ecol. Prog. Ser.* **273,** 65–79 (2004).
4. Dewar, W. K. *et al. J. Mar. Res.* **64,** 541–561 (2006).
5. Kunze, E. *et al. Science* **313,** 1768–1770 (2006).
6. Visser, A. W. *Science* **316,** 838–839 (2007).
7. Catton, K. B., Webster, D. R. & Yen, J. 'Can krill mix the ocean?' *2008 Ocean Sci. Mtg,* Orlando, Florida, abstr. (ASLO, 2008).
8. Gregg, M. C. & Horne, J. K. *J. Phys. Oceanogr.* (in the press).
9. Schiermeier, Q. *Nature* **447,** 522–524 (2007).

PLANETARY SCIENCE

Windy clues to Saturn's spin

Adam P. Showman

Saturn's rotation period has been a mystery. An estimate based on its meteorology comes with implications for our understanding of the planet's atmospheric jet streams and interior structure.

The rate at which a planet rotates is a fundamental property that informs our understanding of its formation, evolution, internal dynamics and meteorology. For planets with solid surfaces, the spin rate can simply be determined by tracking the motion of landforms as they rotate across the surface. But for the gas giants Jupiter, Saturn, Uranus and Neptune, which lack any solid surfaces, determining the rotation rates of their interiors is more difficult. Saturn has proved the most enigmatic, and in recent years our imprecise understanding of its rotation rate has become obvious[1]. On page 608 of this issue, Read and colleagues[2] use clues from Saturn's dynamic meteorology to derive a new estimate for its rotation rate.

Tracking cloud motions over time shows that Saturn's atmosphere, like all atmospheres, does not rotate as a solid body but contains several east–west jet streams. Air at the equator circles the planet once every 10 hours 12 minutes, whereas air at higher latitudes can take up to 30 minutes longer to do so[3]. These cloud-tracked wind measurements imply that Saturn's atmosphere contains a broad equatorial jet — extending from 30° N to 30° S latitude — that flows eastward at speeds that are up to 450 m s^{-1} faster than air at higher latitudes. High-latitude atmospheric regions (outside the equatorial jet) are further subdivided into differentially rotating latitude bands whose relative speeds typically differ by 100 m s^{-1}.

But what is the rotation rate of Saturn's interior? Despite the planet's fluid nature, electromagnetic forces in the electrically conducting interior should keep the interior rotation at nearly a single value. But is this interior rotation rate faster, slower or intermediate between the wide range of atmospheric
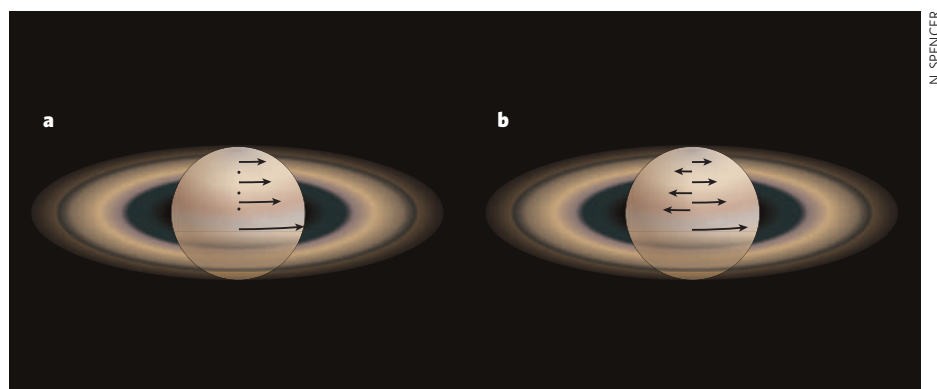

Figure 1 | Saturn's swinging winds. a, The rotation period of Saturn is traditionally deduced from periodicities in the planet's radio emission. These measurements have long suggested that the planet's atmospheric winds move solely in an eastward direction (right-pointing arrows), varying in strength with latitude and interspersed with nulls in speed (dots). **b,** Read and colleagues' new estimate[2] of Saturn's rotation period, which is based instead on the planet's dynamical meteorology, implies that the winds alternate between eastward and westward (left-pointing arrows) with latitude.

mixing. His result neither strongly dismissed nor supported the idea, but for decades afterwards conventional wisdom held that fish could be ignored in ocean mixing. It often happens in science that a flurry of unconnected activities on a common topic emerge almost simultaneously, however, and such has been the case for biogenic ocean mixing.

In 2004, Huntley and Zhou[3] pointed out that the expected levels of turbulence in schools of fish are comparable to those associated with storms. In a subsequent paper, my colleagues and I argued[4] that the kinetic energy expended by the biosphere is sizeable compared with global mixing requirements; we further suggested that the true swimmers (fish), when all lumped together, provide about half of the biosphere input, with the balance coming from zooplankton. Exciting, direct confirmation of hugely elevated turbulence levels in vertically migrating shrimp-like animals followed from Kunze and colleagues[5].

A major question is how efficiently biogenic turbulence actually mixes the ocean. The answer hinges on length scales. Very small whorls introduced into a fluid will be quickly damped by friction, and thus will not mix the fluid. To illustrate, consider a tall coffee cup with a slight gradient in creaminess from top to bottom; small whorls at the bottom would have little effect on the cream at the top before dying a frictional death. Guidance on the size at which turbulence changes from unimportant to important in mixing is provided by the Ozmidov scale, which takes into account how stratified a fluid is and how strong the turbulence is. Given that many zooplankton are comparable to or smaller than oceanic Ozmidov scales, one view is that biogenic mixing is negligible[6]. The story is not yet complete, however. It could be that zooplankton schooling introduces larger scales and increases mixing efficiency[7]. Although an attempt[8] to observe such an effect failed to do so, the search continues.

Into this mix comes the paper by Katija and Dabiri[1]. The authors emphasize that the mere act of swimming implies that some water travels with the swimmer. Whereas viscosity lessens the effect of turbulent mixing, here it is found to increase the total transport. In remarkable videos obtained by scuba divers in shoals of jellyfish (Fig. 1), dye releases clearly show the process (see Supplementary Information[1]). One wonders what the jellyfish made of all this, but that would be another story.

The relevance to mixing, however, can be simply described. Suppose a jellyfish is in cold water, and swims vertically to warmer zones. Some amount of cold water will follow (the videos suggest a surprisingly large amount). Once there, mixing of the local fluid properties ensues. From energetics estimates based on the dye's behaviour, the effect seems to be sizeable. This mechanism is implicit in previous energetics estimates, but it has escaped explicit notice until now and lessens doubts,

based on Ozmidov scales, about the possible strength of biogenic mixing.

Translation of Katija and Dabiri's results from anecdotes to assessments of possible global impacts remains to be carried out. Should the overall idea of significant biogenic mixing survive detailed scrutiny, climate science will have experienced a paradigm shift. To quote Carl Wunsch[9], modellers will "need to start thinking about the fluid dynamics of biology", to which he added, "that's a tough one" — as, indeed, it is. ∎

William K. Dewar is in the Department of Oceanography, Florida State University, Tallahassee, Florida 32306-4320, USA.
e-mail: dewar@ocean.fsu.edu

1. Katija, K. & Dabiri, J. O. *Nature* **460,** 624–626 (2009).
2. Munk, W. H. *Deep-Sea Res.* **13,** 707–730 (1966).
3. Huntley, M. E. & Zhou, M. *Mar. Ecol. Prog. Ser.* **273,** 65–79 (2004).
4. Dewar, W. K. *et al. J. Mar. Res.* **64,** 541–561 (2006).
5. Kunze, E. *et al. Science* **313,** 1768–1770 (2006).
6. Visser, A. W. *Science* **316,** 838–839 (2007).
7. Catton, K. B., Webster, D. R. & Yen, J. 'Can krill mix the ocean?' *2008 Ocean Sci. Mtg,* Orlando, Florida, abstr. (ASLO, 2008).
8. Gregg, M. C. & Horne, J. K. *J. Phys. Oceanogr.* (in the press).
9. Schiermeier, Q. *Nature* **447,** 522–524 (2007).

PLANETARY SCIENCE

Windy clues to Saturn's spin

Adam P. Showman

Saturn's rotation period has been a mystery. An estimate based on its meteorology comes with implications for our understanding of the planet's atmospheric jet streams and interior structure.

The rate at which a planet rotates is a fundamental property that informs our understanding of its formation, evolution, internal dynamics and meteorology. For planets with solid surfaces, the spin rate can simply be determined by tracking the motion of landforms as they rotate across the surface. But for the gas giants Jupiter, Saturn, Uranus and Neptune, which lack any solid surfaces, determining the rotation rates of their interiors is more difficult. Saturn has proved the most enigmatic, and in recent years our imprecise understanding of its rotation rate has become obvious[1]. On page 608 of this issue, Read and colleagues[2] use clues from Saturn's dynamic meteorology to derive a new estimate for its rotation rate.

Tracking cloud motions over time shows that Saturn's atmosphere, like all atmospheres, does not rotate as a solid body but contains several east–west jet streams. Air at the equator circles the planet once every 10 hours 12 minutes, whereas air at higher latitudes can take up to 30 minutes longer to do so[3]. These cloud-tracked wind measurements imply that Saturn's atmosphere contains a broad equatorial jet — extending from 30° N to 30° S latitude — that flows eastward at speeds that are up to 450 m s^{-1} faster than air at higher latitudes. High-latitude atmospheric regions (outside the equatorial jet) are further subdivided into differentially rotating latitude bands whose relative speeds typically differ by 100 m s^{-1}.

But what is the rotation rate of Saturn's interior? Despite the planet's fluid nature, electromagnetic forces in the electrically conducting interior should keep the interior rotation at nearly a single value. But is this interior rotation rate faster, slower or intermediate between the wide range of atmospheric
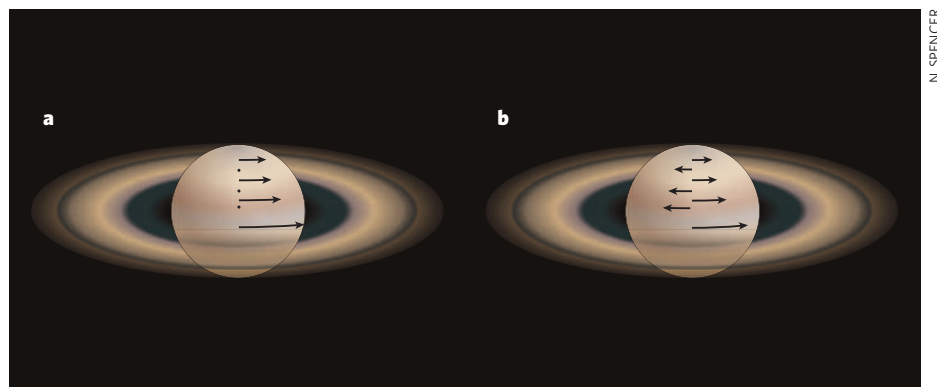

N. SPENCER

Figure 1 | Saturn's swinging winds. a, The rotation period of Saturn is traditionally deduced from periodicities in the planet's radio emission. These measurements have long suggested that the planet's atmospheric winds move solely in an eastward direction (right-pointing arrows), varying in strength with latitude and interspersed with nulls in speed (dots). **b,** Read and colleagues' new estimate[2] of Saturn's rotation period, which is based instead on the planet's dynamical meteorology, implies that the winds alternate between eastward and westward (left-pointing arrows) with latitude.

rotation rates determined by cloud tracking? Answering this question has broad implications not only for the planet's interior structure[4], but also for our understanding of whether the jet streams are primarily eastward or westward relative to the interior — and hence for the degree of angular-momentum exchange between the atmosphere and the interior, for the thermal structure below the clouds, and for the formation mechanisms of the jet streams.

For Jupiter, Uranus and Neptune, the key to unlocking this puzzle lies in the radio emission from sources in the planets' magnetospheres (the region of space near the planet where the planetary magnetic field dominates over that of the solar wind). Because the magnetic dipoles of these planets are tilted relative to their axes of rotation, magnetospheric emissions exhibit a periodicity that allows the rotation period of the magnetic field — and therefore of the planetary interior where the field is generated — to be determined. Saturn's magnetic dipole does not seem to exhibit such a tilt, however. Although Saturn does emit radio waves whose periodic modulations were long assumed to define the rotation rate[5], recent measurements[1] show that this period varies by about 1% over intervals of months and so cannot represent the interior rotation.

Read et al.[2] adopt a radically different approach. They propose that Saturn's rotation rate can be determined by considering the dynamical stability of the planet's jet streams. Using observational estimates of winds and temperatures at and above the clouds, they expand on previous work[6] in which they showed that, at many latitudes, the pattern of the jet streams is almost neutrally stable — lying very near the boundary between stability and instability — according to a stability theorem developed by Vladimir Arnol'd. In the neutral configuration, this theorem relates a flow's east–west wind speed to the latitudinal gradient of a quantity called the potential vorticity — essentially, the rate at which individual air columns spin, divided by a measure of the vertical thickness of the columns. Because the potential vorticity is independent of the reference frame, but the east–west wind speed is not, knowledge of the potential vorticity can be used to determine the reference frame in which the east–west wind speed must be evaluated for the neutrality condition to be valid. This allows an estimate of the interior rotation rate.

Read and colleagues' analysis[2] builds on previous theoretical and observational work suggesting that such a near-neutral configuration is plausible for both Jupiter and Saturn[7–10]. But why would a flow adopt a state that is neutrally stable? Under appropriate conditions, Saturn's loss of heat to space, and the turbulent transports of momentum that help to pump the jet streams, may force the jets to become unstable. But because Saturn's radiated heat flux is meagre, the timescales for radiative cooling and jet pumping — and hence for the jets to gradually become unstable — are probably years to

decades. By contrast, the natural timescale for a strongly unstable jet to naturally develop eddies that rob the jet of energy — thereby making it less unstable — is typically days to weeks. Because of this mismatch in timescales, these competing processes could drive the flow into a configuration that is almost neutrally stable. Analogous arguments have been put forward to explain the configuration of the large-scale air flow in Earth's mid-latitudes, but in that case the instability timescales are not well separated from the radiative timescales, weakening the argument for such a neutral configuration.

Interestingly, the planet's interior rotation rate of 10 hours 34 minutes proposed by Read et al.[2] — like that suggested in another recent attempt to estimate the rotation rate[11] — is intermediate between the fastest and slowest atmospheric rotation rates determined from cloud tracking. Such a value suggests that Saturn's winds exhibit an alternating pattern, with eastward-flowing jets at some latitudes and westward-flowing jets at others (Fig. 1). This is significantly different from the original estimate from Saturn's radio emission[5], which implied a slower interior rotation period of 10 hours 39 minutes for which all the observed winds would be eastward. Because of the dynamical linkage between winds and temperatures, the new rotation rate has additional implications for the latitudinal gradient of temperatures below the clouds, as well as for the mass of Saturn's putative rocky core.

Because Jupiter's jet streams also alternate

between eastward and westward, the revised rotation period gives Saturn a more Jupiter-like countenance than previously appreciated. Nevertheless, Saturn's winds are stronger than Jupiter's, its banded cloud patterns and populations of hurricane-like vortices differ considerably, and its magnetic field, which is almost symmetrical about its axis — a puzzle in its own right — contrasts with Jupiter's tilted dipole. These contrasts indicate that the planets are cousins rather than twins, whose intriguing mix of similarities as well as differences will keep planetary scientists engaged for years to come. ■

Adam P. Showman is in the Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: showman@lpl.arizona.edu

1. Gurnett, D. A. et al. Science 316, 442–445 (2007).
2. Read, P. L., Dowling, T. E. & Schubert, G. Nature 460, 608–610 (2009).
3. Sánchez-Lavega, A. Icarus 49, 1–16 (1982).
4. Helled, R., Schubert, G. & Anderson, J. D. Icarus 199, 368–377 (2009).
5. Desch, M. D. & Kaiser, M. L. Geophys. Res. Lett. 8, 253–256 (1981).
6. Read, P. L. et al. Planet. Space Sci. doi:10.1016/j.pss.2009.03.004 (2009).
7. Dowling, T. E. J. Atmos. Sci. 50, 14–22 (1993).
8. Stamp, A. P. & Dowling, T. E. J. Geophys. Res. 98, 18847–18855 (1993).
9. Read, P. L. et al. Q. J. R. Meteorol. Soc. 132, 1577–1603 (2006).
10. Lian, Y. & Showman, A. P. Icarus 194, 597–615 (2008).
11. Anderson, J. D. & Schubert, G. Science 317, 1384–1387 (2007).

EARTH SCIENCE

Trickle-down geodynamics

Nicholas Arndt

Analysis of the platinum-group elements in a particular type of ancient volcanic rock provides clues about Earth's early history as well as a fresh approach to understanding mantle dynamics.

In a paper on page 620 of this issue[1], Maier et al. provide a provocative hypothesis for those engaged in the study of Earth's evolution. Their evidence comes from measurements of the concentrations of platinum-group elements (PGE) in ancient volcanic rocks known as komatiites, which originated from deep within the mantle.

Maier et al. find that the contents of these elements in 3.5-billion- to 3.2-billion-year-old komatiites from Barberton in South Africa and Pilbara in Western Australia are lower than those in younger komatiites. To explain the difference, they propose that the deep-mantle source of the older komatiites was deficient in PGE. According to a widely accepted hypothesis, the mantle acquired these elements as a surface layer of meteoritic material, the 'late veneer', whose deposition terminated with the

'late heavy bombardment' 3.8 billion years ago. Maier et al. propose that alloys of iron and the PGE trickled slowly down through the mantle and — so their thinking goes — older komatiites did not receive their full dose whereas younger ones did.

For Earth scientists, the value of the PGE is that they provide unique information about the composition and evolution of the mantle. Three of these elements, osmium (Os), iridium (Ir) and ruthenium (Ru), normally behave compatibly — that is, they tend to be retained in the solid phase during partial melting of the mantle or crystallization of rock. The other three, platinum (Pt), palladium (Pd) and rhodium (Rh), as well as the geochemically similar element rhenium (Re), are incompatible — they partition preferentially into the melt. This contrasting behaviour leads to changes in the

rotation rates determined by cloud tracking? Answering this question has broad implications not only for the planet's interior structure[4], but also for our understanding of whether the jet streams are primarily eastward or westward relative to the interior — and hence for the degree of angular-momentum exchange between the atmosphere and the interior, for the thermal structure below the clouds, and for the formation mechanisms of the jet streams.

For Jupiter, Uranus and Neptune, the key to unlocking this puzzle lies in the radio emission from sources in the planets' magnetospheres (the region of space near the planet where the planetary magnetic field dominates over that of the solar wind). Because the magnetic dipoles of these planets are tilted relative to their axes of rotation, magnetospheric emissions exhibit a periodicity that allows the rotation period of the magnetic field — and therefore of the planetary interior where the field is generated — to be determined. Saturn's magnetic dipole does not seem to exhibit such a tilt, however. Although Saturn does emit radio waves whose periodic modulations were long assumed to define the rotation rate[5], recent measurements[1] show that this period varies by about 1% over intervals of months and so cannot represent the interior rotation.

Read *et al.*[2] adopt a radically different approach. They propose that Saturn's rotation rate can be determined by considering the dynamical stability of the planet's jet streams. Using observational estimates of winds and temperatures at and above the clouds, they expand on previous work[6] in which they showed that, at many latitudes, the pattern of the jet streams is almost neutrally stable — lying very near the boundary between stability and instability — according to a stability theorem developed by Vladimir Arnol'd. In the neutral configuration, this theorem relates a flow's east–west wind speed to the latitudinal gradient of a quantity called the potential vorticity — essentially, the rate at which individual air columns spin, divided by a measure of the vertical thickness of the columns. Because the potential vorticity is independent of the reference frame, but the east–west wind speed is not, knowledge of the potential vorticity can be used to determine the reference frame in which the east–west wind speed must be evaluated for the neutrality condition to be valid. This allows an estimate of the interior rotation rate.

Read and colleagues' analysis[2] builds on previous theoretical and observational work suggesting that such a near-neutral configuration is plausible for both Jupiter and Saturn[7–10]. But why would a flow adopt a state that is neutrally stable? Under appropriate conditions, Saturn's loss of heat to space, and the turbulent transports of momentum that help to pump the jet streams, may force the jets to become unstable. But because Saturn's radiated heat flux is meagre, the timescales for radiative cooling and jet pumping — and hence for the jets to gradually become unstable — are probably years to

decades. By contrast, the natural timescale for a strongly unstable jet to naturally develop eddies that rob the jet of energy — thereby making it less unstable — is typically days to weeks. Because of this mismatch in timescales, these competing processes could drive the flow into a configuration that is almost neutrally stable. Analogous arguments have been put forward to explain the configuration of the large-scale air flow in Earth's mid-latitudes, but in that case the instability timescales are not well separated from the radiative timescales, weakening the argument for such a neutral configuration.

Interestingly, the planet's interior rotation rate of 10 hours 34 minutes proposed by Read *et al.*[2] — like that suggested in another recent attempt to estimate the rotation rate[11] — is intermediate between the fastest and slowest atmospheric rotation rates determined from cloud tracking. Such a value suggests that Saturn's winds exhibit an alternating pattern, with eastward-flowing jets at some latitudes and westward-flowing jets at others (Fig. 1). This is significantly different from the original estimate from Saturn's radio emission[5], which implied a slower interior rotation period of 10 hours 39 minutes for which all the observed winds would be eastward. Because of the dynamical linkage between winds and temperatures, the new rotation rate has additional implications for the latitudinal gradient of temperatures below the clouds, as well as for the mass of Saturn's putative rocky core.

Because Jupiter's jet streams also alternate

between eastward and westward, the revised rotation period gives Saturn a more Jupiter-like countenance than previously appreciated. Nevertheless, Saturn's winds are stronger than Jupiter's, its banded cloud patterns and populations of hurricane-like vortices differ considerably, and its magnetic field, which is almost symmetrical about its axis — a puzzle in its own right — contrasts with Jupiter's tilted dipole. These contrasts indicate that the planets are cousins rather than twins, whose intriguing mix of similarities as well as differences will keep planetary scientists engaged for years to come. ∎

Adam P. Showman is in the Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: showman@lpl.arizona.edu

1. Gurnett, D. A. *et al. Science* **316,** 442–445 (2007).
2. Read, P. L., Dowling, T. E. & Schubert, G. *Nature* **460,** 608–610 (2009).
3. Sánchez-Lavega, A. *Icarus* **49,** 1–16 (1982).
4. Helled, R., Schubert, G. & Anderson, J. D. *Icarus* **199,** 368–377 (2009).
5. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8,** 253–256 (1981).
6. Read, P. L. *et al. Planet. Space Sci.* doi:10.1016/j.pss.2009.03.004 (2009).
7. Dowling, T. E. *J. Atmos. Sci.* **50,** 14–22 (1993).
8. Stamp, A. P. & Dowling, T. E. *J. Geophys. Res.* **98,** 18847–18855 (1993).
9. Read, P. L. *et al. Q. J. R. Meteorol. Soc.* **132,** 1577–1603 (2006).
10. Lian, Y. & Showman, A. P. *Icarus* **194,** 597–615 (2008).
11. Anderson, J. D. & Schubert, G. *Science* **317,** 1384–1387 (2007).

EARTH SCIENCE

Trickle-down geodynamics

Nicholas Arndt

Analysis of the platinum-group elements in a particular type of ancient volcanic rock provides clues about Earth's early history as well as a fresh approach to understanding mantle dynamics.

In a paper on page 620 of this issue[1], Maier *et al.* provide a provocative hypothesis for those engaged in the study of Earth's evolution. Their evidence comes from measurements of the concentrations of platinum-group elements (PGE) in ancient volcanic rocks known as komatiites, which originated from deep within the mantle.

Maier *et al.* find that the contents of these elements in 3.5-billion- to 3.2-billion-year-old komatiites from Barberton in South Africa and Pilbara in Western Australia are lower than those in younger komatiites. To explain the difference, they propose that the deep-mantle source of the older komatiites was deficient in PGE. According to a widely accepted hypothesis, the mantle acquired these elements as a surface layer of meteoritic material, the 'late veneer', whose deposition terminated with the

'late heavy bombardment' 3.8 billion years ago. Maier *et al.* propose that alloys of iron and the PGE trickled slowly down through the mantle and — so their thinking goes — older komatiites did not receive their full dose whereas younger ones did.

For Earth scientists, the value of the PGE is that they provide unique information about the composition and evolution of the mantle. Three of these elements, osmium (Os), iridium (Ir) and ruthenium (Ru), normally behave compatibly — that is, they tend to be retained in the solid phase during partial melting of the mantle or crystallization of rock. The other three, platinum (Pt), palladium (Pd) and rhodium (Rh), as well as the geochemically similar element rhenium (Re), are incompatible — they partition preferentially into the melt. This contrasting behaviour leads to changes in the

proportions of the two types of element, which can be used to monitor how the mantle melts. The compatible isotope ^{187}Os is produced by radioactive decay of incompatible ^{187}Re, and the resultant changes in the ratio of radiogenic ^{187}Os to stable ^{188}Os can be used to establish the timing of mantle melting.

Basalt, the most common mantle-derived magma, is the product of 'low-degree melting'. Under these conditions, small amounts of sulphides and alloys remain in the residue of melting, as do most of the PGE, which are all highly compatible with these minor phases. Komatiitic magmas form by high levels of mantle melting, when even the refractory sulphides and alloys enter the melt, and the komatiites thereby acquire high concentrations of all the PGE. Analyses of PGE have served to date komatiites and trace the composition of their mantle source[2–4], but until the work by Maier *et al.*[1] temporal changes in their PGE contents had received little attention. Although there was some evidence that the oldest komatiites contained a lower content of PGE than younger komatiites[5,6], a systematic analysis had not been done.

Maier *et al.* chose Pt and Ru to represent the overall behaviour of all the PGE. They show that the concentrations of these elements are lowest in 3.5-billion-year-old komatiites; are slightly higher in samples aged about 3.2 billion years; and are higher still and relatively constant in komatiites between 2.7 billion and 90 million years old. Having considered and rejected alternative explanations for the increase in PGE levels, the authors opt for an explanation that invokes the late veneer and a protracted timescale for mixing of PGE into the mantle.

It is generally thought[7] that the PGE from the late veneer of meteoritic material mixed rapidly through the mantle, but Maier *et al.*[1] suggest that this material took several hundred million years to percolate down to the lower mantle. Working on the principle that komatiites formed through melting of mantle plumes that rose from the base of the mantle, they propose that, 3.5 billion to 3.2 billion years ago, this part of the mantle had not received its full quota of PGE. By 2.7 billion years ago, trickle-down was complete, and from then on komatiites were derived from a homogenized PGE-rich source.

But how firm are their data? To measure PGE content, Maier *et al.*[1] used a technique called the Ni-sulphide collection method, which may underestimate Pt contents by 15% or more. But they present convincing evidence that their analyses are sufficiently accurate for their purposes. Data obtained using a more precise method — the Carius-tube approach — yield comparably low PGE contents in some Barberton komatiites[5]. Moreover, any systematic bias should equally affect komatiites that are both older and younger than 3 billion years.

The bigger picture of komatiite geochemistry has to take into account studies of neodymium (Nd) isotopes[8]. These analyses tell us that the mantle source of komatiites had become depleted in incompatible elements very early in Earth's history, perhaps during extraction of the melt that formed the first continental crust[9]. The extraction of this material should also have depleted Re (which is incompatible) more than Os (which is compatible), leading to an Os isotope ratio that is distinctly lower than that of the original, undifferentiated mantle[7]. Yet the Os isotopic compositions of most komatiites are almost identical to that of the meteoritic material that we accept as representative of the mantle. Maier and colleagues' results provide an explanation for the contrasting nature of the two isotope systems — the Nd isotopes in komatiites retain a record of differentiation during early melt extraction, but their Os isotopic composition was swamped by later input of meteoritic material from the late veneer.

Maier and colleagues' work sets the scene for analysis of the trace-element and Nd and Os isotopic compositions of komatiites, allied with modelling of mixing of meteoritic material into the lower mantle. Such a combination has the potential to set important constraints on the evolution and geodynamics of the early Earth. ∎

Nicholas Arndt is at LGCA, UMR5025 CNRS, Université de Grenoble, St Martin d'Heres 38401, France.
e-mail: arndt@ujf-grenoble.fr

1. Maier, W. D. *et al. Nature* **460,** 620–623 (2009).
2. Luck, J. M. & Arndt, N. T. *Terra Cognita* **5,** 323 (1986).
3. Keays, R. L., Ross, J. R. & Woolrich, P. W. *Econ. Geol.* **76,** 1645–1674 (1981).
4. Puchtel, I. S., Brügmann, G. E., Hoffmann, A. W., Kulikov, V. S. & Kulikova, V. V. *Contr. Mineral. Petrol.* **140,** 588–599 (2001).
5. Puchtel, I. S., Walker, R. J., Anhaeusser, C. R. & Gruau, G. *Chem. Geol.* **262,** 355–369 (2009).
6. Maier, W. D., Roelofse, F. & Barnes, S.-J. *J. Petrol.* **44,** 1787–1804 (2003).
7. Walker, R. J. *Chemie Erde* **69,** 101–125 (2009).
8. Arndt, N. T., Barnes, S. J. & Lesher, C. M. *Komatiite* (Cambridge Univ. Press, 2008).
9. Harrison, T. M. *et al. Science* **310,** 1947–1950 (2005).

DEVELOPMENTAL BIOLOGY

Skeletal muscle comes of age

Terry Partridge

A regulatory protein thought to be crucial for maintaining the muscle stem-cell pool throughout life is shown to be dispensable in the adult. Muscle biologists are left wondering what fundamental things apply as time goes by.

When, as we grow up, we abandon our beloved teddy bear, we don't throw it away — we merely ignore it, a desertion that has parallels in skeletal-muscle development. Lepper *et al.*[1] (page 627 of this issue) show that the Pax7 transcription factor, which is required for early juvenile muscle growth in mice, becomes inessential in the adult, although it is still expressed.

The sequence of cellular events that gives rise to skeletal muscle in vertebrates is similar throughout prenatal development and postnatal growth, and during muscle regeneration in the adult. The proliferation of precursor cells is followed by the differentiation and fusion of most cells into muscle fibres containing multiple nuclei. In all instances, some undifferentiated cells remain as reserves that can give rise to muscle tissue. In mice, just before birth, most of these undifferentiated cells become enclosed in the basement membrane that develops around each muscle fibre, forming a reservoir of muscle stem cells known as satellite cells. This reservoir is the main source of cells for postnatal muscle growth and repair.

The similarities in cellular events in the different phases of muscle development are paralleled by strong similarities in molecular mechanisms. The transcription factors Pax3 and Pax7, which activate genes that are essential for muscle development, perform an intriguing *pas de deux* throughout embryonic and fetal muscle development[2]. Their functions seem largely to overlap, although Pax3 seems to be essential for some processes, such as migration of precursor muscle cells from their site of origin in the somites (the tissue that develops on either side of the embryonic central axis) to sites of muscle formation in the limbs and body wall[2] (Fig. 1a). More recently, Pax7 was shown to be required for the maintenance of the satellite-cell population after birth. In mice that lack the *Pax7* gene (*Pax7* germline-null mice), muscle develops normally up to birth, but fails to grow or regenerate after this[3]. It was initially proposed[3] that Pax7 is necessary to specify satellite cells, but it is now believed that it is required for satellite-cell survival[4,5]. How Pax7 performs this function is unknown, but it is generally thought that, in limb skeletal muscle, Pax7 takes over from Pax3 to support satellite-cell survival in postnatal life (Fig. 1b). Indeed, Pax7 expression is used as a definitive marker of satellite cells.

The work by Lepper *et al.*[1] profoundly upsets this status quo. The authors use an elegant approach to knock out *Pax7* after birth in mice. In these animals, only one of the two *Pax7* genes is functional, and the functional *Pax7* gene is engineered to be ablated, solely

proportions of the two types of element, which can be used to monitor how the mantle melts. The compatible isotope [187]Os is produced by radioactive decay of incompatible [187]Re, and the resultant changes in the ratio of radiogenic [187]Os to stable [188]Os can be used to establish the timing of mantle melting.

Basalt, the most common mantle-derived magma, is the product of 'low-degree melting'. Under these conditions, small amounts of sulphides and alloys remain in the residue of melting, as do most of the PGE, which are all highly compatible with these minor phases. Komatiitic magmas form by high levels of mantle melting, when even the refractory sulphides and alloys enter the melt, and the komatiites thereby acquire high concentrations of all the PGE. Analyses of PGE have served to date komatiites and trace the composition of their mantle source[2-4], but until the work by Maier et al.[1] temporal changes in their PGE contents had received little attention. Although there was some evidence that the oldest komatiites contained a lower content of PGE than younger komatiites[5,6], a systematic analysis had not been done.

Maier et al. chose Pt and Ru to represent the overall behaviour of all the PGE. They show that the concentrations of these elements are lowest in 3.5-billion-year-old komatiites; are slightly higher in samples aged about 3.2 billion years; and are higher still and relatively constant in komatiites between 2.7 billion and 90 million years old. Having considered and rejected alternative explanations for the increase in PGE levels, the authors opt for an explanation that invokes the late veneer and a protracted timescale for mixing of PGE into the mantle.

It is generally thought[7] that the PGE from the late veneer of meteoritic material mixed rapidly through the mantle, but Maier et al.[1] suggest that this material took several hundred million years to percolate down to the lower mantle. Working on the principle that komatiites formed through melting of mantle plumes that rose from the base of the mantle, they propose that, 3.5 billion to 3.2 billion years ago, this part of the mantle had not received its full quota of PGE. By 2.7 billion years ago, trickle-down was complete, and from then on komatiites were derived from a homogenized PGE-rich source.

But how firm are their data? To measure PGE content, Maier et al.[1] used a technique called the Ni-sulphide collection method, which may underestimate Pt contents by 15% or more. But they present convincing evidence that their analyses are sufficiently accurate for their purposes. Data obtained using a more precise method — the Carius-tube approach — yield comparably low PGE contents in some Barberton komatiites[5]. Moreover, any systematic bias should equally affect komatiites that are both older and younger than 3 billion years.

The bigger picture of komatiite geochemistry has to take into account studies of neodymium (Nd) isotopes[8]. These analyses tell us that the mantle source of komatiites had become depleted in incompatible elements very early in Earth's history, perhaps during extraction of the melt that formed the first continental crust[9]. The extraction of this material should also have depleted Re (which is incompatible) more than Os (which is compatible), leading to an Os isotope ratio that is distinctly lower than that of the original, undifferentiated mantle[7]. Yet the Os isotopic compositions of most komatiites are almost identical to that of the meteoritic material that we accept as representative of the mantle. Maier and colleagues' results provide an explanation for the contrasting nature of the two isotope systems — the Nd isotopes in komatiites retain a record of differentiation during early melt extraction, but their Os isotopic composition was swamped by later input of meteoritic material from the late veneer.

Maier and colleagues' work sets the scene for analysis of the trace-element and Nd and Os isotopic compositions of komatiites, allied with modelling of mixing of meteoritic material into the lower mantle. Such a combination has the potential to set important constraints on the evolution and geodynamics of the early Earth.						■

Nicholas Arndt is at LGCA, UMR5025 CNRS, Université de Grenoble, St Martin d'Heres 38401, France.
e-mail: arndt@ujf-grenoble.fr

1. Maier, W. D. et al. Nature **460,** 620–623 (2009).
2. Luck, J. M. & Arndt, N. T. Terra Cognita **5,** 323 (1986).
3. Keays, R. L., Ross, J. R. & Woolrich, P. W. Econ. Geol. **76,** 1645–1674 (1981).
4. Puchtel, I. S., Brügmann, G. E., Hoffmann, A. W., Kulikov, V. S. & Kulikova, V. V. Contr. Mineral. Petrol. **140,** 588–599 (2001).
5. Puchtel, I. S., Walker, R. J., Anhaeusser, C. R. & Gruau, G. Chem. Geol. **262,** 355–369 (2009).
6. Maier, W. D., Roelofse, F. & Barnes, S.-J. J. Petrol. **44,** 1787–1804 (2003).
7. Walker, R. J. Chemie Erde **69,** 101–125 (2009).
8. Arndt, N. T., Barnes, S. J. & Lesher, C. M. Komatiite (Cambridge Univ. Press, 2008).
9. Harrison, T. M. et al. Science **310,** 1947–1950 (2005).

DEVELOPMENTAL BIOLOGY

Skeletal muscle comes of age

Terry Partridge

A regulatory protein thought to be crucial for maintaining the muscle stem-cell pool throughout life is shown to be dispensable in the adult. Muscle biologists are left wondering what fundamental things apply as time goes by.

When, as we grow up, we abandon our beloved teddy bear, we don't throw it away — we merely ignore it, a desertion that has parallels in skeletal-muscle development. Lepper et al.[1] (page 627 of this issue) show that the Pax7 transcription factor, which is required for early juvenile muscle growth in mice, becomes inessential in the adult, although it is still expressed.

The sequence of cellular events that gives rise to skeletal muscle in vertebrates is similar throughout prenatal development and postnatal growth, and during muscle regeneration in the adult. The proliferation of precursor cells is followed by the differentiation and fusion of most cells into muscle fibres containing multiple nuclei. In all instances, some undifferentiated cells remain as reserves that can give rise to muscle tissue. In mice, just before birth, most of these undifferentiated cells become enclosed in the basement membrane that develops around each muscle fibre, forming a reservoir of muscle stem cells known as satellite cells. This reservoir is the main source of cells for postnatal muscle growth and repair.

The similarities in cellular events in the different phases of muscle development are paralleled by strong similarities in molecular mechanisms. The transcription factors Pax3 and Pax7, which activate genes that are essential for muscle development, perform an intriguing *pas de deux* throughout embryonic and fetal muscle development[2]. Their functions seem largely to overlap, although Pax3 seems to be essential for some processes, such as migration of precursor muscle cells from their site of origin in the somites (the tissue that develops on either side of the embryonic central axis) to sites of muscle formation in the limbs and body wall[2] (Fig. 1a). More recently, Pax7 was shown to be required for the maintenance of the satellite-cell population after birth. In mice that lack the *Pax7* gene (*Pax7* germline-null mice), muscle develops normally up to birth, but fails to grow or regenerate after this[3]. It was initially proposed[3] that Pax7 is necessary to specify satellite cells, but it is now believed that it is required for satellite-cell survival[4,5]. How Pax7 performs this function is unknown, but it is generally thought that, in limb skeletal muscle, Pax7 takes over from Pax3 to support satellite-cell survival in postnatal life (Fig. 1b). Indeed, Pax7 expression is used as a definitive marker of satellite cells.

The work by Lepper et al.[1] profoundly upsets this status quo. The authors use an elegant approach to knock out *Pax7* after birth in mice. In these animals, only one of the two *Pax7* genes is functional, and the functional *Pax7* gene is engineered to be ablated, solely
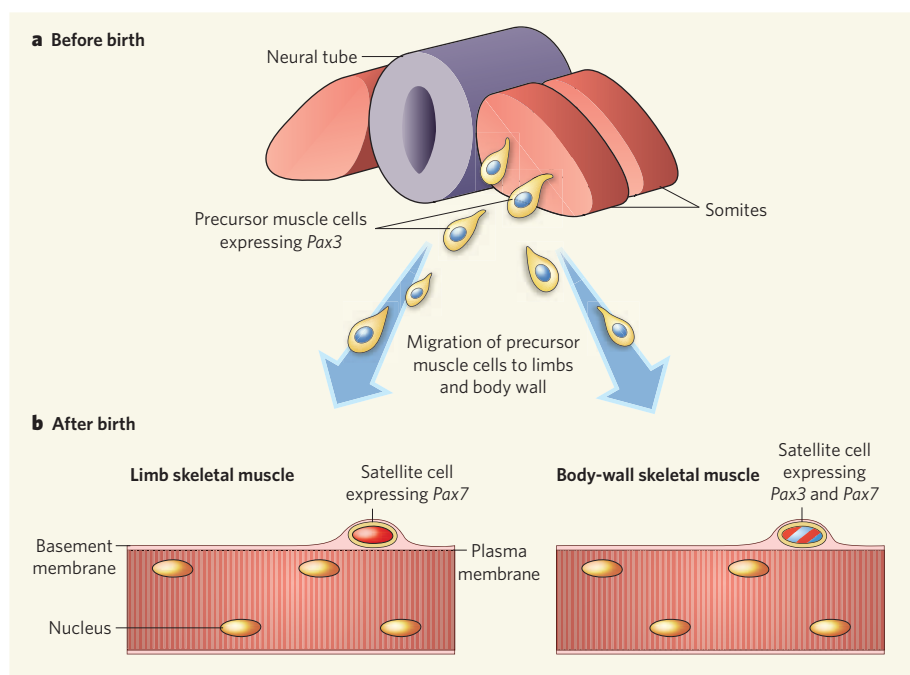
Figure 1 | Skeletal muscle and the *Pax* genes. **a**, During embryonic development in mice, precursor muscle cells arise in the somites on either side of the developing neural tube and migrate to sites of muscle formation in the limbs and body wall, where they give rise to muscle fibres. *Pax3* expression (blue) is required for the migration of cells to sites of muscle formation. **b**, After birth, satellite cells in limb skeletal muscle express only *Pax7* (red), whereas satellite cells in body-wall skeletal muscle express both *Pax3* and *Pax7*. Lepper *et al.*[1] show that, in limb skeletal muscle, *Pax7* is necessary for maintenance of the satellite-cell population only in the immediate postnatal period and not beyond.

in cells that express it (satellite cells), on giving the drug tamoxifen. The same mechanism that ablates *Pax7* simultaneously induces expression of a reporter molecule, allowing this specific satellite-cell population to be tracked.

When *Pax7* was ablated at birth, limb skeletal muscle failed to grow or regenerate after injury, similar to mice that lack *Pax7* from conception (the germline-null mutants). However, surprisingly, when *Pax7* was ablated in mice at 2 to 3 weeks of age and beyond, limb skeletal-muscle growth and regeneration was normal. And by detecting the reporter molecule, the authors showed[1] that the cells undertaking muscle regeneration were satellite cells that had expressed Pax7 at the time of the tamoxifen-induced knockout. This last point is significant, because several studies have shown that muscle cells can arise from non-satellite cells associated with blood vessels or interstitial tissue, prompting suggestions[6] that such cells are important for regeneration and are a precursor population of the satellite cell. Both notions are challenged, at least in the short term, by the finding that virtually all satellite cells in regenerated muscle were marked as being derived from *Pax7*-ablated progenitors.

An alternative explanation for the retention of the regenerative ability of Pax7-deficient satellite cells is that these cells might re-express Pax3, which may compensate for the absence of Pax7. But Lepper and colleagues[1] disproved this by showing that ablating both *Pax3* and *Pax7* in adult mice did not impair limb skeletal-muscle regeneration.

Thus, it seems that the cells that give rise to limb skeletal muscle, having been dependent on either Pax3 or Pax7 throughout most of prenatal life, abruptly cast off these developmental shackles towards the end of the postnatal growth period. Moreover, this liberation from Pax dependence seems to be intrinsic to the satellite cells, as ablation of *Pax7* in tissue culture leads to impaired muscle formation in precursor muscle cells isolated from young but not older muscle.

Intriguingly, the transition of satellite cells to Pax7 independence occurs during the period when there was shown to be a sharp decrease in fusion of Pax7-expressing cells into the growing muscle fibres[1]. It perhaps marks the time when muscle fibres begin to respond predominantly to environmental signals, and to undertake regeneration in response to injury, or enlargement in response to the demands of physiological work, as opposed to processes driven predominantly by developmental mechanisms.

The development of adult muscle cells in the absence of Pax7 and Pax3 has left researchers with a lot of explaining to do. Given the increasing appreciation of the diversity of different types of skeletal muscle, it would be worthwhile investigating the effects of *Pax3* and *Pax7* ablation in some muscles of the head, which are not Pax3 dependent[7], and in muscles of the diaphragm and body wall, where satellite cells continue to express Pax3 in the adult[5] (Fig. 1b). It is also important to examine the longer-term effects of ablation of the *Pax* genes and to try to relate the phases of muscle development in the mouse to the quite different time frame of muscle development and use in humans. ■

Terry Partridge is at the Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC 20010, USA.
e-mail: tpartridge@cnmcresearch.org

1. Lepper, C., Conway, S. J. & Fan, C.-M. *Nature* **460,** 627–631 (2009).
2. Montarras, D. & Buckingham, M. in *Skeletal Muscle Repair and Regeneration* (eds Schiaffano, S. & Partridge, T.) 19–44 (Springer, 2008).
3. Seale, P. *et al. Cell* **102,** 777–786 (2000).
4. Oustanina, S., Hause, G. & Braun, T. *EMBO J.* **23,** 3430–3439 (2004).
5. Relaix, F. *et al. J. Cell Biol.* **172,** 91–102 (2006).
6. Péault, B. *et al. Mol. Ther.* **15,** 867–877 (2007).
7. Sambasivan, R. *et al. Dev. Cell* **16,** 810–821 (2009).

Phosphorus caged

Kenneth N. Raymond

Violent criminals are imprisoned to keep them under control. Similarly, incarceration in a molecular jail stops white phosphorus from bursting into flames — but on release, it regains its fiery character.

White phosphorus — P_4, one of two forms of the element — has earned a notorious reputation for being highly reactive. When it reacts slowly with oxygen, it merely glows; indeed, the words 'phosphorescence' and 'phosphorus' both originate from the Greek word for 'light-carrier'. But on exposure to high concentrations of oxygen (such as in air), phosphorus spontaneously ignites, burning fiercely to produce a great heat. Reporting in *Science*, Mal *et al.*[1] describe a way to tame this fiery element.

When its molecules are trapped in nanoscale 'flasks', white phosphorus becomes inert, and will no longer react even slowly with oxygen.

Mal and colleagues' nanoflasks are actually molecular complexes that self-assemble from smaller subunits (molecules and ions). The cavities of similar 'host' complexes are known to isolate, protect and sometimes stabilize other 'guest' molecules. In nature, for example, viruses use protein shells known as capsids to protect their nucleic acid. Similarly,
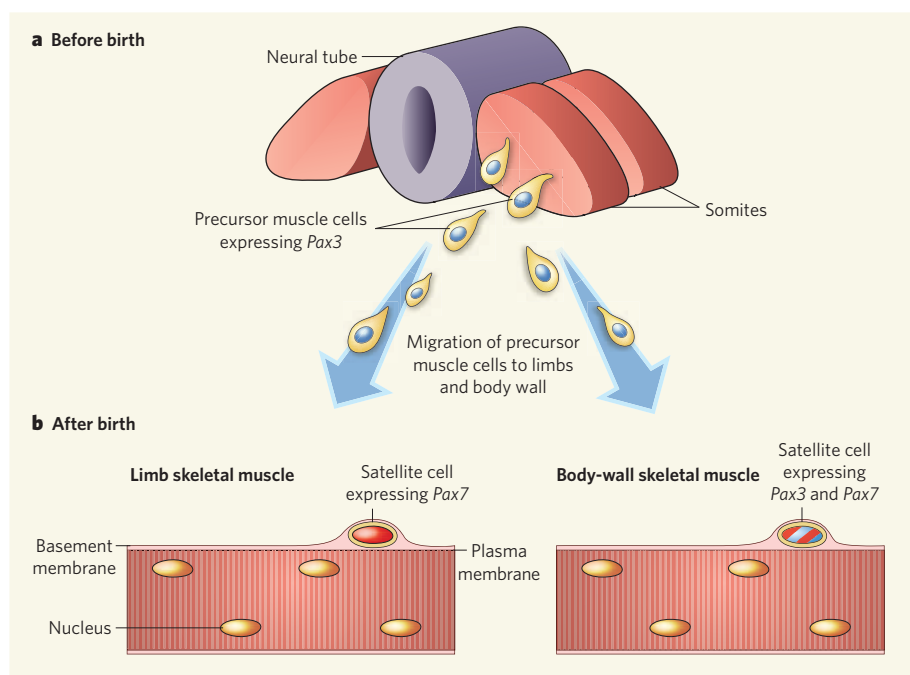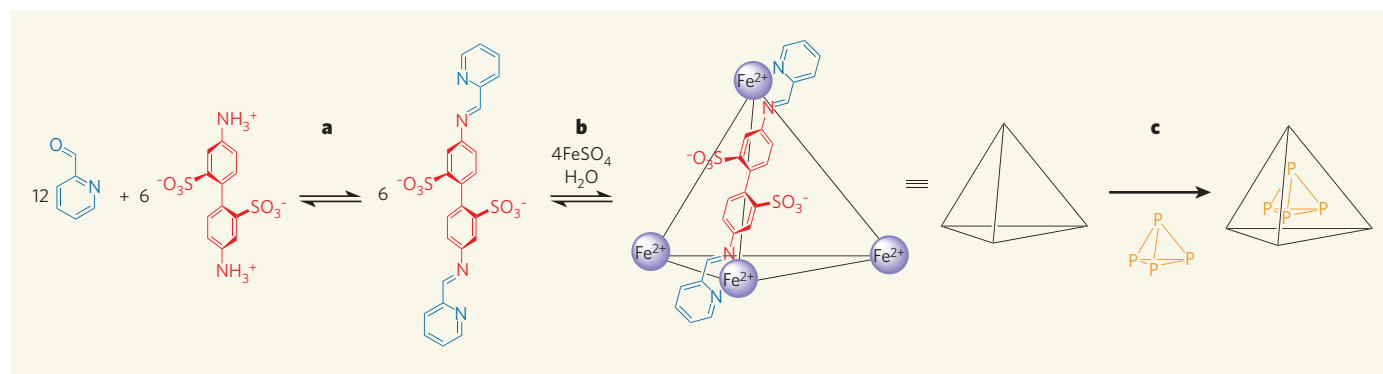
Figure 1 | Skeletal muscle and the *Pax* genes. **a**, During embryonic development in mice, precursor muscle cells arise in the somites on either side of the developing neural tube and migrate to sites of muscle formation in the limbs and body wall, where they give rise to muscle fibres. *Pax3* expression (blue) is required for the migration of cells to sites of muscle formation. **b**, After birth, satellite cells in limb skeletal muscle express only *Pax7* (red), whereas satellite cells in body-wall skeletal muscle express both *Pax3* and *Pax7*. Lepper *et al.*[1] show that, in limb skeletal muscle, *Pax7* is necessary for maintenance of the satellite-cell population only in the immediate postnatal period and not beyond.

in cells that express it (satellite cells), on giving the drug tamoxifen. The same mechanism that ablates *Pax7* simultaneously induces expression of a reporter molecule, allowing this specific satellite-cell population to be tracked.

When *Pax7* was ablated at birth, limb skeletal muscle failed to grow or regenerate after injury, similar to mice that lack *Pax7* from conception (the germline-null mutants). However, surprisingly, when *Pax7* was ablated in mice at 2 to 3 weeks of age and beyond, limb skeletal-muscle growth and regeneration was normal. And by detecting the reporter molecule, the authors showed[1] that the cells undertaking muscle regeneration were satellite cells that had expressed Pax7 at the time of the tamoxifen-induced knockout. This last point is significant, because several studies have shown that muscle cells can arise from non-satellite cells associated with blood vessels or interstitial tissue, prompting suggestions[6] that such cells are important for regeneration and are a precursor population of the satellite cell. Both notions are challenged, at least in the short term, by the finding that virtually all satellite cells in regenerated muscle were marked as being derived from *Pax7*-ablated progenitors.

An alternative explanation for the retention of the regenerative ability of Pax7-deficient satellite cells is that these cells might re-express Pax3, which may compensate for the absence of Pax7. But Lepper and colleagues[1] disproved this by showing that ablating both *Pax3* and *Pax7* in adult mice did not impair limb skeletal-muscle regeneration.

Thus, it seems that the cells that give rise to limb skeletal muscle, having been dependent on either Pax3 or Pax7 throughout most of prenatal life, abruptly cast off these developmental shackles towards the end of the postnatal growth period. Moreover, this liberation from Pax dependence seems to be intrinsic to the satellite cells, as ablation of *Pax7* in tissue culture leads to impaired muscle formation in precursor muscle cells isolated

from young but not older muscle.

Intriguingly, the transition of satellite cells to Pax7 independence occurs during the period when there was shown to be a sharp decrease in fusion of Pax7-expressing cells into the growing muscle fibres[1]. It perhaps marks the time when muscle fibres begin to respond predominantly to environmental signals, and to undertake regeneration in response to injury, or enlargement in response to the demands of physiological work, as opposed to processes driven predominantly by developmental mechanisms.

The development of adult muscle cells in the absence of Pax7 and Pax3 has left researchers with a lot of explaining to do. Given the increasing appreciation of the diversity of different types of skeletal muscle, it would be worthwhile investigating the effects of *Pax3* and *Pax7* ablation in some muscles of the head, which are not Pax3 dependent[7], and in muscles of the diaphragm and body wall, where satellite cells continue to express Pax3 in the adult[5] (Fig. 1b). It is also important to examine the longer-term effects of ablation of the *Pax* genes and to try to relate the phases of muscle development in the mouse to the quite different time frame of muscle development and use in humans. ∎

Terry Partridge is at the Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC 20010, USA.
e-mail: tpartridge@cnmcresearch.org

1. Lepper, C., Conway, S. J. & Fan, C.-M. *Nature* **460,** 627–631 (2009).
2. Montarras, D. & Buckingham, M. in *Skeletal Muscle Repair and Regeneration* (eds Schiaffano, S. & Partridge, T.) 19–44 (Springer, 2008).
3. Seale, P. *et al. Cell* **102,** 777–786 (2000).
4. Oustanina, S., Hause, G. & Braun, T. *EMBO J.* **23,** 3430–3439 (2004).
5. Relaix, F. *et al. J. Cell Biol.* **172,** 91–102 (2006).
6. Péault, B. *et al. Mol. Ther.* **15,** 867–877 (2007).
7. Sambasivan, R. *et al. Dev. Cell* **16,** 810–821 (2009).

SUPRAMOLECULAR CHEMISTRY

Phosphorus caged

Kenneth N. Raymond

Violent criminals are imprisoned to keep them under control. Similarly, incarceration in a molecular jail stops white phosphorus from bursting into flames — but on release, it regains its fiery character.

White phosphorus — P_4, one of two forms of the element — has earned a notorious reputation for being highly reactive. When it reacts slowly with oxygen, it merely glows; indeed, the words 'phosphorescence' and 'phosphorus' both originate from the Greek word for 'light-carrier'. But on exposure to high concentrations of oxygen (such as in air), phosphorus spontaneously ignites, burning fiercely to produce a great heat. Reporting in *Science*, Mal *et al.*[1] describe a way to tame this fiery element.

When its molecules are trapped in nanoscale 'flasks', white phosphorus becomes inert, and will no longer react even slowly with oxygen.

Mal and colleagues' nanoflasks are actually molecular complexes that self-assemble from smaller subunits (molecules and ions). The cavities of similar 'host' complexes are known to isolate, protect and sometimes stabilize other 'guest' molecules. In nature, for example, viruses use protein shells known as capsids to protect their nucleic acid. Similarly,

Figure 1 | Molecular prisons. Mal *et al.*[1] have made tetrahedral molecular complexes that encapsulate molecules of white phosphorus, P_4. The complexes self-assemble from various subunits. **a**, Six bridging subunits (red) combine with 12 terminal subunits (blue) to make six larger components. **b**, These larger components bind to four iron(II) ions (Fe^{2+}, obtained from iron(II) sulphate, $FeSO_4$) to form the tetrahedral host complex. Only one of the six components is shown for simplicity. **c**, When a solution of the host in water is left in contact with white phosphorus, P_4 guest molecules become trapped inside it.
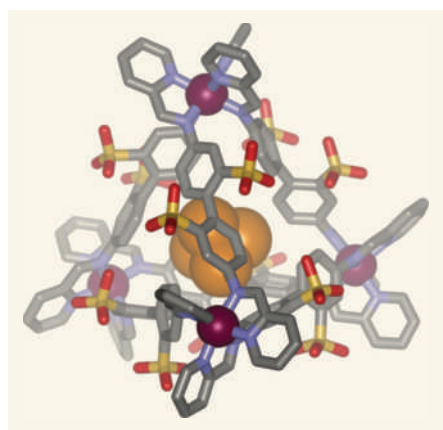
Figure 2 | X-ray crystal structure of the host–guest complex, as obtained by Mal and colleagues[1]**.** Iron atoms are purple, carbon atoms grey, nitrogen atoms blue and phosphorus atoms orange. Charged sulphonate groups (yellow and red) help the cluster to dissolve in water.

human cells wrap iron in a protein shell, forming a complex called ferritin; without this protection, the iron would be toxic. Chemists are increasingly using this strategy in their work, by synthesizing host assemblies and using them to facilitate new reactions or to isolate unstable chemical species.

The nanoscale cages made by Mal *et al.* are tetrahedra that assemble spontaneously in water from four iron(II) ions (Fe^{2+}, which become the apexes of the tetrahedra) and six organic molecules (Fig. 1). These six molecules assemble in turn from the reactions of smaller molecular subunits: six 'bridging' components (which form the edges of the tetrahedra) and 12 'terminal' components (which bind to the Fe^{2+} ions at the apexes of the tetrahedra). Although the reactions of the bridging and terminal subunits are easily reversible, the products are stabilized by their subsequent coordination to the Fe^{2+} ions. The entire assembly is therefore stable, but in dynamic equilibrium with the starting components.

The tetrahedral complex has a negative charge that makes it water-soluble, but its inner space is relatively hydrophobic, and just big enough to encapsulate a P_4 guest. When Mal *et al.* left an oxygen-free solution of the complex in water in contact with solid white phosphorus, a host–guest complex formed — as shown both by spectroscopic experiments on the resulting solution, and by X-ray crystallography on isolated crystals of the product (Fig. 2). The caged phosphorus is both water-soluble and air-stable. Indeed, the authors found that solutions of the host–guest complex remained unchanged after four months' exposure to air; compare this with the instant conflagration of unencapsulated white phosphorus. So what is going on?

White phosphorus reacts with oxygen to produce an oxide (P_2O_5). This oxide then reacts with any water that is around to form phosphoric acid. The phosphorus–phosphorus bonds of P_4 are weak compared with the stronger phosphorus–oxygen bonds of P_2O_5; in other words, the oxide is thermodynamically much more stable than white phosphorus, and this drives the reaction to such an extent that white phosphorus spontaneously combusts in air.

One might therefore assume that Mal and colleagues' nanoflasks simply isolate P_4 molecules from oxygen. But this isn't the case: oxygen molecules are smaller than P_4 molecules, and must therefore also be able to gain access to the flasks' interiors. Instead, the tight confinement of P_4 molecules prevents the formation of phosphorus–oxygen bonds during the first steps of phosphorus oxidation — there simply isn't room for the reaction intermediates to form. This is the first time that a reactive species has been stabilized by such an effect, and represents a fundamental advance for the field.

Mal *et al.*[1] found that encapsulated P_4 molecules could be easily released from their miniature prisons. When the authors floated an equal volume of benzene on top of an aqueous solution of the host–guest complex, benzene molecules replaced P_4 as guests. The ejected phosphorus is insoluble in water, but soluble in common organic solvents, and so the authors could extract it into the benzene layer, where it reacted fully with air within 24 hours. By contrast, the organic solvent *n*-heptane does not replace P_4 in the host–guest complex, presumably because it doesn't fit well into the host's chamber.

Other tetrahedral clusters have been reported to self-assemble from metal ions and organic molecules[2,3], and are increasingly being used as molecular flasks for new types of reaction[4]. They have also been used to make 'ship-in-a-bottle' molecules, by assembling around guests that are too large to leave the resulting cavity. Some complexes even mimic enzymes, catalysing reactions by stabilizing the transient intermediates for those reactions as guests. Still other host complexes assemble from organic subunits alone, using hydrogen bonding or other weak interactions, and catalyse reactions because their cavities force guest molecules into the optimal alignment for those reactions[5].

Mal *et al.*[1] are currently trying to prepare bigger host complexes that they hope will isolate and stabilize larger, more complicated guests. Ultimately, they hope that their encapsulation strategy will provide a general way to control the release of reactive or sensitive molecules, or even as a method for removing harmful substances from the environment. Such applications are a long way off, but in the meantime we have every reason to expect more chemical surprises and capabilities as we continue to explore this rich field of endeavour. ■

Kenneth N. Raymond is in the Department of Chemistry, University of California, Berkeley, and in the Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
e-mail: raymond@socrates.berkeley.edu

1. Mal, P., Breiner, B., Rissanen, K. & Nitschke, J. R. *Science* **324,** 1697–1699 (2009).
2. Saalfrank, R. W., Hörner, B., Stalke, D. & Salbeck, J. *Angew. Chem. Int. Edn* **32,** 1179–1182 (1993).
3. Caulder, D. L., Powers, R. E., Parac, T. N. & Raymond, K. N. *Angew. Chem. Int. Edn* **37,** 1840–1843 (1998).
4. Yoshizawa, M., Klosterman, J. K. & Fujita, M. *Angew. Chem. Int. Edn* **48,** 3418–3438 (2009).
5. Hou, J.-L., Ajami, D. & Rebek, J. Jr *J. Am. Chem. Soc.* **130,** 7810–7811 (2008).

Recent progress in the biology and physiology of sirtuins

Toren Finkel[1], Chu-Xia Deng[2] & Raul Mostoslavsky[3]

The sirtuins are a highly conserved family of NAD$^+$-dependent enzymes that regulate lifespan in lower organisms. Recently, the mammalian sirtuins have been connected to an ever widening circle of activities that encompass cellular stress resistance, genomic stability, tumorigenesis and energy metabolism. Here we review the recent progress in sirtuin biology, the role these proteins have in various age-related diseases and the tantalizing notion that the activity of this family of enzymes somehow regulates how long we live.

"Chekhov's gun" refers to the seemingly irrelevant prop or observation that is introduced early on in a drama (such as the gun in Act I of *Uncle Vanya*) but whose true significance is only revealed at the play's conclusion. Often, science has its own version of the Russian playwright's famous maxim. In our case, the drama begins in the autumn of 1914, with the seemingly irrelevant observations of a young professor describing the beneficial effects of rats placed on a calorie-restricted diet. Some suspect that these remarkable observations were largely ignored and forgotten because of concerns about the scientist himself. Indeed the young professor, F. Peyton Rous, had only a few years earlier shocked the scientific community by postulating that in some situations, cancer could be transmitted by viruses—a notion that was roundly ridiculed and derided by his older colleagues. For whatever reason, Rous's curious observations remained dormant for nearly two decades. Over time, however, other scientists gradually extended these ideas and demonstrated the benefits of caloric restriction in species ranging from humans to yeast. Indeed, it was in the latter species that dissection of the molecular basis of caloric restriction first implicated a family of NAD$^+$-dependent enzymes, now collectively termed the sirtuins. In the following sections, we review the recent progress in our understanding of sirtuin biology with particular emphasis on their role in mammalian species. Although much remains unanswered and controversies remain unresolved, as with Chekhov's gun it is only with the passage of time that the true significance of Rous's observations can finally be understood.

A new family of deacetylases

The founding member of the sirtuin family, yeast Sir2 (silent information regulator 2), was originally isolated in a screen for silencing factors[1]. Ultimately, four proteins (named Sir1–Sir4) were established as important regulators of silencing at the mating type locus as well as telomeric DNA. The Sir family was rescued from relative obscurity when Guarente and colleagues, using *Saccharomyces cerevisiae* as a model system, independently identified these proteins as key regulators of lifespan[2,3]. Subsequent analysis revealed that Sir2 functioned biochemically as a histone deacetylase, in a unique reaction requiring the energetic intermediate NAD$^+$ as a co-factor[4]. Remarkably, analysis in other model organisms (including *Caenorhabditis elegans* and *Drosophila*) also implicated Sir2 homologues as a determinant of lifespan[5]. In both yeast and flies, Sir2 levels increased following caloric-restriction treatment. Perhaps more

importantly, in sirtuin-deficient yeast and mice, extension of lifespan by caloric restriction is abolished[5,6]. Subsequent studies have qualified some of these observations and have demonstrated that in yeast, the effect of Sir2 in caloric restriction and lifespan extension might be strain-, species- and context-dependent[7,8].

To date, seven mammalian homologues have been identified, with mammalian SIRT1 closest evolutionarily to yeast Sir2. Specific mouse models in which one or more sirtuin genes has been knocked out have been recently created (Fig. 1). Cell biological studies have further demonstrated different subcellular compartments for each family member, with SIRT6 and SIRT7 being nuclear proteins, SIRT3, SIRT4 and SIRT5 mitochondrial proteins, and SIRT1 and SIRT2 being found both in the nucleus and the cytoplasm, in a cell- and tissue-dependent context[9]. The observation that these proteins
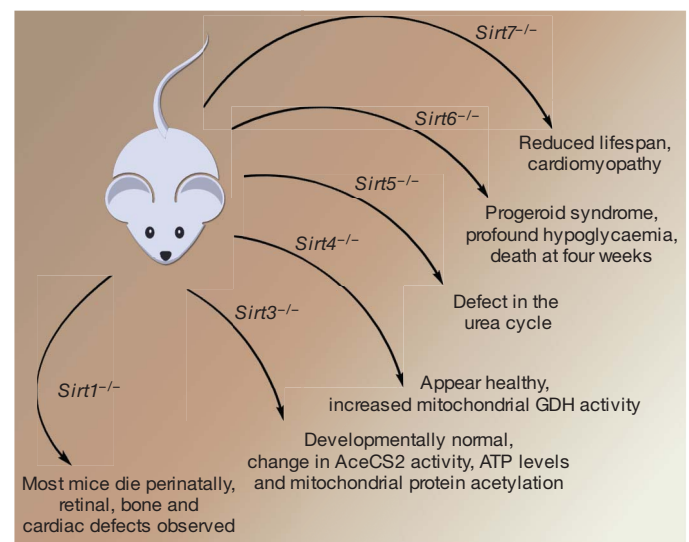

Figure 1 | Mouse knockout models as tools for exploring sirtuin function. Gene targeting of SIRT1, SIRT3, SIRT4, SIRT5, SIRT6 and SIRT7 has been reported. The phenotypes include a reduction in median lifespan, ranging from a usual survival of days (SIRT1) to weeks (SIRT6) or months (SIRT7). In contrast, although biochemical phenotypes have been reported, $Sirt3^{-/-}$ and $Sirt4^{-/-}$ mice appear outwardly normal. Initial reports suggest that $Sirt5^{-/-}$ mice also exhibit no obvious phenotype[49].

---

[1]Translational Medicine Branch, National Heart Lung and Blood Institute, [2]Genetics of Development and Disease Branch, National Institute of Diabetes, Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. [3]The Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02114, USA.

require NAD$^+$ as part of their enzymatic action immediately suggested a mechanistic link between sirtuin activity and intracellular energetics. However, it is important to stress that currently there is no definitive evidence that these proteins play any direct role in mammalian lifespan regulation, as neither pharmacological sirtuin activators nor overexpression of SIRT1 has been demonstrated to extend lifespan in mice. Nonetheless, the explosion of interest in the sirtuins centres on the nagging notion that in some fashion, these proteins provide the connection between what we eat and how long we live.

Sirtuins in DNA repair

A role for sirtuins in maintaining genomic integrity was suggested by early work in model organisms. For instance, yeast Sir2 was shown to inhibit recombination of ribosomal DNA as well as to relocalize to sites of DNA breaks[5,10]. Furthermore, yeast sirtuin family members are capable of deacetylating histone proteins and the absence of this activity results in silencing defects, increased genomic instability and sensitivity to DNA damage[5,10]. These properties appear to also extend to the mammalian sirtuins. The first example came with analysis of SIRT6-deficient cells that demonstrated increased sensitivity to genotoxic damage and accumulation of chromosomal abnormalities[11]. Recent studies have shown that SIRT6 also acts as a histone deacetylase that can influence the telomeres of human cells[12].

Mammalian SIRT1 also appears to influence genomic stability. First, SIRT1 can deacetylate various factors linked to the repair of DNA damage, including the Werner helicase and NBS1 (ref. 13). A role for SIRT1 in genomic integrity was further substantiated by the recent demonstration of increased chromosomal aberrations and impaired DNA repair in $Sirt1^{-/-}$ embryos[14]. In addition, following oxidative damage, SIRT1 has been shown to be recruited to sites of DNA breaks. Although such recruitment appears to be important in protecting against genomic instability, this response is accompanied by the derepression of previously silenced genes[15]. These observations suggest that like yeast Sir2, SIRT1 appears to regulate epigenetic silencing and chromatin modification, and recent evidence suggests this is achieved, at least in part, through direct regulation of modifying enzymes, such as the histone methyltransferase SUV39H1 (refs 16, 17).

Sirtuins and cell fate

Mammalian sirtuins also appear to have an important role in regulating cellular stress resistance and modulating the threshold for cell death. In part, this increased stress resistance comes from the interaction with the Forkhead box class O (FOXO) family of transcription factors. These mammalian transcription factors regulate both energy status and stress resistance, two properties intimately connected to lifespan extension. SIRT1 can bind and deacetylate FOXO3a, leading to a selective augmentation of FOXO-regulated stress resistance genes[18,19]. Subsequent experiments have extended these observations to demonstrate that various different sirtuin and FOXO family members can interact. In addition, a recent report suggested that SIRT1 can also protect cells against stress by regulating the heat shock response[20]. The interaction of SIRT1 and p53 can also modulate the threshold for cell death in the setting of exogenous stress[21,22]. Indeed, FOXO proteins and p53 proteins can directly interact under stress conditions[18,23], and p53 can regulate SIRT1 expression through various means that include p53-dependent microRNAs[23–25]. Besides p53, SIRT1 can regulate other targets linked to cell death, including Ku70, E2F1 and TGF-β signalling[26]. In addition, the coordinated action of SIRT3 and SIRT4 appear to inhibit cell death by maintaining mitochondrial NAD$^+$ levels following stress[27]. Although in most examples sirtuin activity appears to antagonize stress-induced cell death pathways, SIRT1 can also deacetylate components of the NF-κB complex, leading to increased apoptosis[28].

Given the potential link between sirtuins and ageing, it is appealing to envisage a role for these proteins in mediating processes such as cellular senescence and stem cell function. Consistent with such a role, SIRT1 expression can antagonize the premature senescence seen in mouse embryonic fibroblasts following forced oncogene expression[29]. Similarly, through its regulation of NF-κB activity, knockdown of SIRT6 appears to augment senescence in primary keratinocytes[30]. However satisfying it may be to conclude that increased sirtuin activity can forestall senescence, it is important to note that an analysis of studies of $Sirt1^{-/-}$ mouse embryonic fibroblasts has come to the opposite conclusion[31]. With regards to differentiation, sirtuins have been shown to inhibit adipogenesis by modulating PPAR-γ (peroxisome proliferative activated receptor γ) and to effect muscle and neuronal differentiation[32,33]. Recent studies in mouse embryonic stem cells also suggest a role for SIRT1 in regulating stem cell reactive oxygen species homeostasis and differentiation[34]. The observation that SIRT1 can also interact with members of the polycomb family of repressors, a family of proteins intricately connected to stem cell self-renewal, suggests that additional links between sirtuins, differentiation and stem cell function are likely to emerge[35].

Sirtuins and metabolic regulation

In mammals, blood glucose concentration is maintained within a narrow range under a variety of physiological conditions. During starvation, maintenance of serum glucose is achieved in part by implementing a program of hepatic gluconeogenesis. Increasing evidence suggests an important role for sirtuins in this physiological adaptation. The peroxisome proliferator-activated receptor gamma-coactivator-1α (PGC-1α) is a known target of SIRT1-dependent deacetylation[36,37], and this coactivator also plays a fundamental part in regulating gluconeogenesis and fatty acid oxidation pathways within the liver. The ability of PGC-1α to modulate these latter two pathways appears to require SIRT1 (refs 37, 38). Recently, distinct roles for protein acetylation and SIRT1-dependent deacetylation have been shown to regulate the hepatic response to both short term (<6 h) and long term (>18 h) fasting[39]. In this case, the opposing actions of SIRT1 and the p300/CBP acetyltransferase choreograph hepatic glucose production in the setting of nutrient stress. Finally, the observation that SIRT6-deficient mice demonstrate severe hypoglycaemia suggests a potential role for other sirtuins in glucose production and homeostasis[11].

In addition to regulating hepatic gluconeogenesis, sirtuins also modulate serum glucose levels by regulating pancreatic insulin secretion. A transgenic mouse with β cell-specific SIRT1-overexpression was noted to have increased glucose-stimulated insulin secretion and improved glucose tolerance compared to control mice[40], while $Sirt1^{-/-}$ mice have impaired glucose-stimulated insulin secretion[41]. Finally, SIRT4 through its ADP-ribosylation of glutamate dehydrogenase (GDH)[42] as well as its interactions with insulin degrading enzyme[43] also appears to be involved in regulating pancreatic insulin secretion.

The sirtuin family has a much broader role in metabolism than the regulation of glucose homeostasis (Fig. 2). As was previously discussed, SIRT1 through its regulation of PPAR-γ and PGC-1α activity has a significant regulatory role in fat mobilization and fatty acid oxidation[32,44]. The regulation of PGC-1α activity also suggests a role for sirtuins in the generation of new mitochondria, as PGC-1α is a key regulator of mitochondrial biogenesis. These observations, coupled with recent links between SIRT1 and autophagy[45], suggest that sirtuins might regulate the flux of mitochondria within cells by balancing PGC-1α mediated generation with autophagy-dependent clearance. In addition to the connection between SIRT1 and the mitochondria, SIRT3-dependent deacetylation regulates the activity of the mitochondrial enzyme acetyl coenzyme A synthetase 2 (AceCS2)[46,47] as well as Complex I of the electron transport chain[48]. Mitochondrial lysates from $Sirt3^{-/-}$ mice reveal >30 proteins whose acetylation is markedly increased, suggesting that other important targets undoubtedly exist[48,49]. Finally, a very recent study has implicated SIRT5, another mitochondrial sirtuin, as an important regulator of the urea cycle[50].

Although the role of sirtuins in regulating metabolism has centred on key metabolic organs, such as liver and pancreas, early studies in
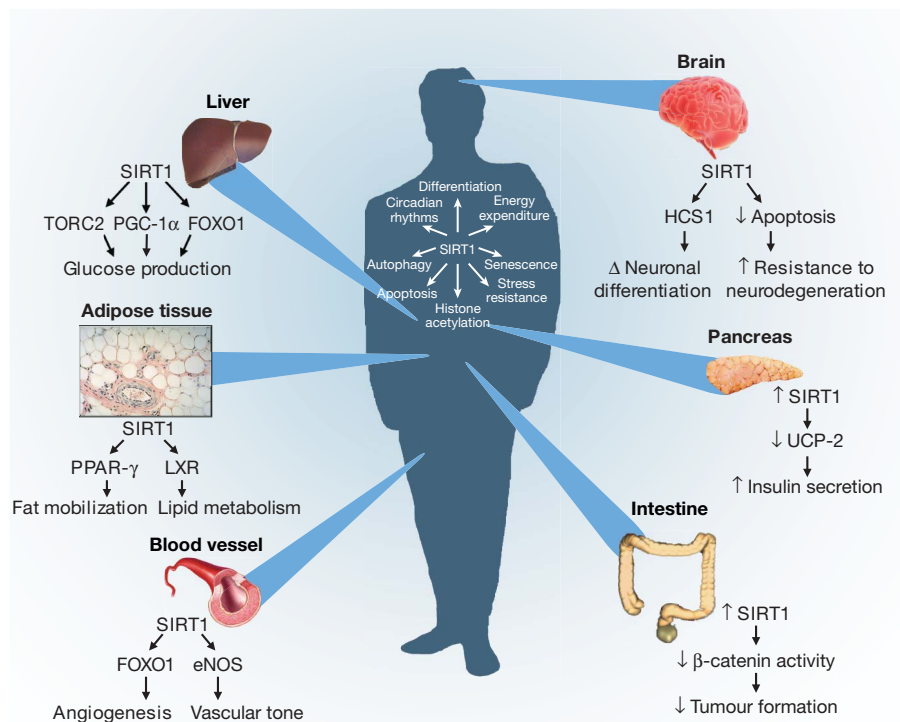
Figure 2 | The diverse physiological roles of the sirtuins. Shown are examples of the organ-specific physiology of SIRT1, along with some of the direct or indirect targets of sirtuin regulation (see text for details). In addition, examples of some of the SIRT1-regulated intracellular parameters are presented, ranging from modulation of progenitor differentiation to altering the threshold for apoptosis. At the beginning of a compound name or process (for example, 'Tumour formation' at bottom right), down-arrow indicates 'decreasing', up-arrow indicates 'increasing' and 'Δ' indicates 'change in'.

human subjects undergoing voluntary caloric restriction suggest that levels of SIRT1 rise in tissues as diverse as skeletal muscle and circulating mononuclear cells[51,52]. The role of sirtuins in the metabolic adaptation of these cell types is largely unexplored. The intriguing connection between SIRT1 and circadian rhythms provides a glimpse of the wide spectrum of metabolic effects sirtuins may eventually be shown to regulate[53,54]. The observation that SIRT1 can directly deacetylate core components of the circadian clock machinery is particularly fascinating, as the ultimate goal of such rhythms is to coordinate the sleep-wake cycle of an organism with environmental cues, including coordinating and matching intracellular metabolism to external food availability.

Sirtuins and age-related diseases

Given the wealth of data connecting sirtuins to glucose homeostasis and insulin secretion, it seemed reasonable to suspect that these proteins might also regulate the susceptibility to developing insulin resistance and diabetes. Two recent studies using moderate transgenic overexpression of SIRT1 demonstrate that such engineered animals do indeed exhibit improved glucose tolerance when challenged with high-fat diets[55,56]. Interestingly, these two studies came to two different conclusions regarding the effects of moderately increased SIRT1 expression on basal energy homeostasis. In addition to these gain of function studies, inhibition of SIRT1 by genetic or pharmacological means can induce insulin resistance[57]. These data support the notion that manipulating sirtuin activity might result in protection from a host of metabolic derangements. Such notions have also been aided by the association of SIRT1 genetic variants with human energy expenditure and obesity[58,59]. In animal models, treatment with the sirtuin activator resveratrol appears to protect against diet-induced obesity and glucose intolerance[60,61]. Some concerns have been raised regarding whether resveratrol in mammalian systems works exclusively or even predominantly through regulation of sirtuin activity. For instance, resveratrol also activates the AMPK pathway, although recent evidence suggests that AMPK activity is intimately connected to sirtuin function[62]. Some of the concerns raised regarding the specificity of resveratrol have also been addressed by the use of newer and seemingly more specific sirtuin activators, which again appear to provide some protection in animal models of diet-induced obesity and insulin resistance[63,64].

Although the link between sirtuins and metabolic disease has been intensely studied, there is a growing body of evidence that similar relationships may hold for a wide swath of age-related maladies. For instance, the ability of SIRT1 to regulate p53 activity potentially implicates the sirtuins in tumorigenesis. Indeed, depending on the context, arguments have been made that SIRT1 might either increase or decrease cancer risks[65]. Recently, the notion that SIRT1 might act as a non-traditional tumour suppressor was highlighted by observations that in the context of a *p53*[+/−] mouse, haploinsufficiency of SIRT1 resulted in increased tumour formation[14] while overexpressed SIRT1 reduced tumours[15]. In addition, a direct connection between the tumour suppressor Brca1 and SIRT1 expression has recently been established[66]. Finally, overexpression of SIRT1 can inhibit tumorigenesis in a mouse model of colon cancer triggered by constitutive Wnt signalling[67]. With observations that SIRT6 regulates DNA repair, and that SIRT2 activity is necessary for proper cytokinesis, a strong possibility exists that other sirtuin family members will also affect cancer predisposition.

The sirtuins also appear to have a prominent role in vascular biology, and may regulate aspects of age-dependent atherosclerosis. Part of these effects may come through regulation of lipid and cholesterol metabolism, including the ability of SIRT1 to modulate the activity of the nuclear receptor LXR, a critical factor in reverse cholesterol transport[68]. In addition, a conditional deletion of SIRT1 in endothelial cells has been demonstrated to impair the angiogenic response following an ischaemic insult[69]. SIRT1 can also deacetylate and regulate endothelial nitric oxide synthase (eNOS) activity, a key regulator of vascular tone[70]. These results are particularly interesting, because in mouse models, caloric restriction has been shown to induce mitochondrial biogenesis through an eNOS-dependent pathway[71]. Other sirtuins can also affect cardiovascular physiology, as SIRT7-deficient mice exhibit cardiac abnormalities[72] and the important vasoconstrictor angiotensin II regulates SIRT3 expression[73].

A unified picture of the role of sirtuins in neurological diseases has yet to emerge. Initial reports had implicated SIRT1 in a specific model of axonal degeneration[74], although subsequent studies have raised the possibility that the neuronal pathology observed in this model may occur through a sirtuin-independent pathway[75]. Similarly, although (as discussed above) SIRT1 is in general thought to be protective against oxidative stress, there is evidence both *in vitro*
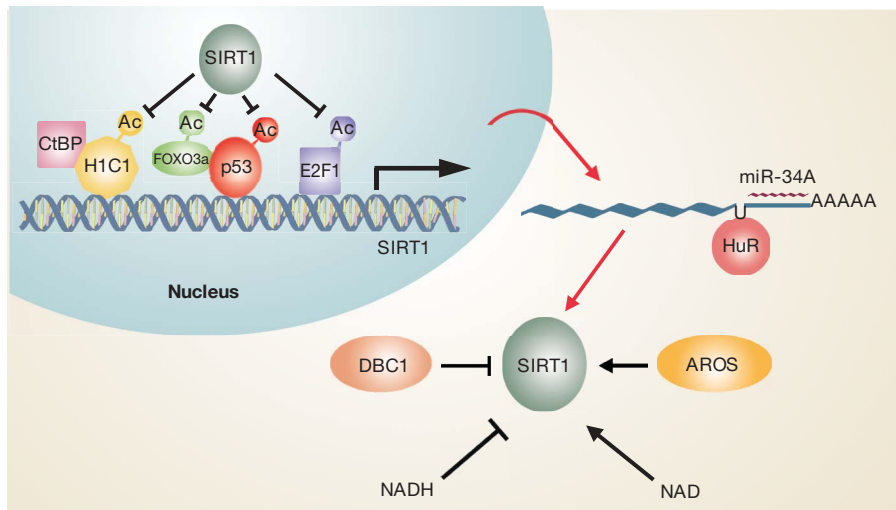
Figure 3 | Complex regulation of SIRT1 activity.
The promoter of SIRT1 is positively and negatively regulated by the binding of various transcription factors and repressors, including H1C1, CtBP, p53, FOXO3A and E2F1. The acetylation and hence activity of many of these factors are in turn controlled by SIRT1. The SIRT1 message is also regulated by the RNA binding protein HuR and the p53-regulated microRNA miR-34A. Finally, SIRT protein activity is regulated positively and negatively by interacting proteins such as AROS and DBC1, as well as by the overall metabolic state, as reflected in the NAD/NADH ratio.

and *in vivo* that this may not hold true in the brain[76]. Despite these observations, in a mouse model of Alzheimer's disease, brain specific overexpression of SIRT1 appears to reduce neurodegeneration[77,78]. Surprisingly, pharmacological or genetic inhibition of SIRT2 was also protective in both a cell-based and a *Drosophila* model of Parkinson's disease[79]. These results suggest that increased SIRT1 or decreased SIRT2 both reduce neurodegeneration, potentially raising cautionary flags for the development of broadly acting sirtuin-based pharmaceuticals for these types of conditions.

Concluding remarks

From their humble origins as silencing factors in yeast, members of the sirtuin family have emerged as broad regulators of cellular fate and mammalian physiology. As impressive as the recent progress has been, much of our understanding has come from studies involving only SIRT1, whereas the other six mammalian sirtuins—presumably equally important—have received significantly less attention. Even for SIRT1, many questions remained unanswered, including a full understanding of how the activity of the enzyme is regulated under both normal and stress conditions, as well as during ageing (Fig. 3). The recent discovery of interacting proteins that can both positively and negatively regulate SIRT1 deacetylase activity[80–82] suggests that similar interacting proteins may exist for the other sirtuin family members. Pharmacological manipulation of sirtuin activity that began with the use of resveratrol has now extended to a variety of newer agents that appear to have greater specificity. Some of these agents have already begun human clinical trials[83]. Whereas the primary benefit of these drugs may lie in preventing diet-induced metabolic disorders, direct effects on vascular health, tumorigenesis and even overall lifespan are at least within the realm of possibility, given the wide spectrum of sirtuin biology.

From an evolutionary standpoint, it remains unclear why the sirtuin family has emerged as a key regulator of so many seemingly varied processes. Perhaps some of this prominence can be attributed to the enzyme's dependence on NAD^+, forever linking sirtuin activity to the underlying metabolic state of the cell. In the 1940s and 50s, it became clear that ingested sugar, fat and protein was eventually reduced to a single, simple and versatile intermediate. Fifty years later, it was realized that this same intermediate, acetyl-CoA, could also be used to modify histones and hence regulate gene expression. As such, acetylation and NAD^+-dependent deacetylation have emerged as perhaps the most immediate and versatile connection between intracellular energetics and intracellular fate. Indeed, the evolving and intricate connection between acetylation, energetics and gene expression is now just beginning to be revealed[84]. In this regard, although numerous individual studies have linked sirtuins to changes in gene expression, regulation of genomic stability or

alterations in intracellular metabolism, the real excitement lies in the understanding that these disparate functions are perhaps all interconnected—that sirtuins serve as the bridge between what we eat and what we are. More twists and turns await us in this drama, as we are indeed far from the final curtain. But for now at least, many of the curious observations of the past are beginning to make sense. After waiting for nearly a century, the significance of Chekhov's gun may have finally been realized.

1. Rine, J., Strathern, J. N., Hicks, J. B. & Herskowitz, I. A suppressor of mating-type locus mutations in *Saccharomyces cerevisiae*: evidence for and identification of cryptic mating-type loci. *Genetics* **93**, 877–901 (1979).
2. Kennedy, B. K., Austriaco, N. R. Jr, Zhang, J. & Guarente, L. Mutation in the silencing gene SIR4 can delay aging in *S. cerevisiae*. *Cell* **80**, 485–496 (1995).
3. Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
4. Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
5. Guarente, L. & Picard, F. Calorie restriction—the SIR2 connection. *Cell* **120**, 473–482 (2005).
6. Boily, G. *et al.* SirT1 regulates energy metabolism and response to caloric restriction in mice. *PLoS One* **3**, e1759 (2008).
7. Bass, T. M., Weinkove, D., Houthoofd, K., Gems, D. & Partridge, L. Effects of resveratrol on lifespan in *Drosophila melanogaster* and *Caenorhabditis elegans*. *Mech. Ageing Dev.* **128**, 546–552 (2007).
8. Kaeberlein, M., Kirkland, K. T., Fields, S. & Kennedy, B. K. Sir2-independent life span extension by calorie restriction in yeast. *PLoS Biol.* **2**, E296 (2004).
9. Haigis, M. C. & Guarente, L. P. Mammalian sirtuins—emerging roles in physiology, aging, and calorie restriction. *Genes Dev.* **20**, 2913–2921 (2006).
10. Denu, J. M. Linking chromatin function with metabolic networks: Sir2 family of NAD^+-dependent deacetylases. *Trends Biochem. Sci.* **28**, 41–48 (2003).
11. Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
12. Michishita, E. *et al.* SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* **452**, 492–496 (2008).
13. Yuan, Z., Zhang, X., Sengupta, N., Lane, W. S. & Seto, E. SIRT1 regulates the function of the Nijmegen breakage syndrome protein. *Mol. Cell* **27**, 149–162 (2007).
14. Wang, R. H. *et al.* Impaired DNA damage response, genome instability, and tumorigenesis in SIRT1 mutant mice. *Cancer Cell* **14**, 312–323 (2008).
15. Oberdoerffer, P. *et al.* SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. *Cell* **135**, 907–918 (2008).
    Refs 14 and 15 describe a role for SIRT1 in genomic stability and how disruption of this activity might contribute to cancer and ageing.
16. Murayama, A. *et al.* Epigenetic control of rDNA loci in response to intracellular energy status. *Cell* **133**, 627–639 (2008).
17. Vaquero, A. *et al.* SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation. *Nature* **450**, 440–444 (2007).
18. Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
19. Motta, M. C. *et al.* Mammalian SIRT1 represses forkhead transcription factors. *Cell* **116**, 551–563 (2004).
20. Westerheide, S. D., Anckar, J., Stevens, S. M. Jr, Sistonen, L. & Morimoto, R. I. Stress-inducible regulation of heat shock factor 1 by the deacetylase SIRT1. *Science* **323**, 1063–1066 (2009).

21. Luo, J. et al. Negative control of p53 by Sir2α promotes cell survival under stress. Cell 107, 137–148 (2001).

22. Vaziri, H. et al. hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. Cell 107, 149–159 (2001).

23. Nemoto, S., Fergusson, M. M. & Finkel, T. Nutrient availability regulates SIRT1 through a forkhead-dependent pathway. Science 306, 2105–2108 (2004).

24. Chen, W. Y. et al. Tumor suppressor HIC1 directly regulates SIRT1 to modulate p53-dependent DNA-damage responses. Cell 123, 437–448 (2005).

25. Yamakuchi, M., Ferlito, M. & Lowenstein, C. J. miR-34a repression of SIRT1 regulates apoptosis. Proc. Natl Acad. Sci. USA 105, 13421–13426 (2008).

26. Michan, S. & Sinclair, D. Sirtuins in mammals: insights into their biological function. Biochem. J. 404, 1–13 (2007).

27. Yang, H. et al. Nutrient-sensitive mitochondrial NAD⁺ levels dictate cell survival. Cell 130, 1095–1107 (2007).

28. Yeung, F. et al. Modulation of NF-κB-dependent transcription and cell survival by the SIRT1 deacetylase. EMBO J. 23, 2369–2380 (2004).

29. Langley, E. et al. Human SIR2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence. EMBO J. 21, 2383–2396 (2002).

30. Kawahara, T. L. et al. SIRT6 links histone H3 lysine 9 deacetylation to NF-κB-dependent gene expression and organismal life span. Cell 136, 62–74 (2009).

31. Chua, K. F. et al. Mammalian SIRT1 limits replicative life span in response to chronic genotoxic stress. Cell Metab. 2, 67–76 (2005).

32. Feige, J. N. & Auwerx, J. Transcriptional targets of sirtuins in the coordination of mammalian physiology. Curr. Opin. Cell Biol. 20, 303–309 (2008).

33. Prozorovski, T. et al. Sirt1 contributes critically to the redox-dependent fate of neural progenitors. Nature Cell Biol. 10, 385–394 (2008).

34. Han, M. K. et al. SIRT1 regulates apoptosis and Nanog expression in mouse embryonic stem cells by controlling p53 subcellular localization. Cell Stem Cell 2, 241–251 (2008).

35. Kuzmichev, A. et al. Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc. Natl Acad. Sci. USA 102, 1859–1864 (2005).

36. Nemoto, S., Fergusson, M. M. & Finkel, T. SIRT1 functionally interacts with the metabolic regulator and transcriptional coactivator PGC-1α. J. Biol. Chem. 280, 16456–16460 (2005).

37. Rodgers, J. T. et al. Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. Nature 434, 113–118 (2005).

38. Rodgers, J. T. & Puigserver, P. Fasting-dependent glucose and lipid metabolic response through hepatic sirtuin 1. Proc. Natl Acad. Sci. USA 104, 12861–12866 (2007).

39. Liu, Y. et al. A fasting inducible switch modulates gluconeogenesis via activator/coactivator exchange. Nature 456, 269–273 (2008).
A detailed look at the role of protein acetylation and SIRT1-dependent deacetylation in the hepatic response to starvation.

40. Moynihan, K. A. et al. Increased dosage of mammalian Sir2 in pancreatic beta cells enhances glucose-stimulated insulin secretion in mice. Cell Metab. 2, 105–117 (2005).

41. Bordone, L. et al. Sirt1 regulates insulin secretion by repressing UCP2 in pancreatic beta cells. PLoS Biol. 4, e31 (2006).

42. Haigis, M. C. et al. SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. Cell 126, 941–954 (2006).

43. Ahuja, N. et al. Regulation of insulin secretion by SIRT4, a mitochondrial ADP-ribosyltransferase. J. Biol. Chem. 282, 33583–33592 (2007).

44. Rodgers, J. T., Lerin, C., Gerhart-Hines, Z. & Puigserver, P. Metabolic adaptations through the PGC-1 alpha and SIRT1 pathways. FEBS Lett. 582, 46–53 (2008).

45. Lee, I. H. et al. A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. Proc. Natl Acad. Sci. USA 105, 3374–3379 (2008).

46. Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S. & Verdin, E. Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2. Proc. Natl Acad. Sci. USA 103, 10224–10229 (2006).

47. Hallows, W. C., Lee, S. & Denu, J. M. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. Proc. Natl Acad. Sci. USA 103, 10230–10235 (2006).

48. Ahn, B. H. et al. A role for the mitochondrial deacetylase Sirt3 in regulating energy homeostasis. Proc. Natl Acad. Sci. USA 105, 14447–14452 (2008).

49. Lombard, D. B. et al. Mammalian Sir2 homolog SIRT3 regulates global mitochondrial lysine acetylation. Mol. Cell. Biol. 27, 8807–8814 (2007).

50. Nakagawa, T., Lomb, D. J., Haigis, M. C. & Guarente, L. SIRT5 deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle. Cell 137, 560–570 (2009).
A recent demonstration of the expanding connection between sirtuin family members and metabolic regulation.

51. Crujeiras, A. B., Parra, D., Goyenechea, E. & Martinez, J. A. Sirtuin gene expression in human mononuclear cells is modulated by caloric restriction. Eur. J. Clin. Invest. 38, 672–678 (2008).

52. Civitarese, A. E. et al. Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. PLoS Med. 4, e76 (2007).

53. Asher, G. et al. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. Cell 134, 317–328 (2008).

54. Nakahata, Y. et al. The NAD⁺-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. Cell 134, 329–340 (2008).
Refs 53 and 54 provide the first link between sirtuins and circadian rhythms.

55. Banks, A. S. et al. SirT1 gain of function increases energy efficiency and prevents diabetes in mice. Cell Metab. 8, 333–341 (2008).

56. Pfluger, P. T., Herranz, D., Velasco-Miguel, S., Serrano, M. & Tschop, M. H. Sirt1 protects against high-fat diet-induced metabolic damage. Proc. Natl Acad. Sci. USA 105, 9793–9798 (2008).

57. Sun, C. et al. SIRT1 improves insulin sensitivity under insulin-resistant conditions by repressing PTP1B. Cell Metab. 6, 307–319 (2007).

58. Weyrich, P. et al. SIRT1 genetic variants associate with the metabolic response of Caucasians to a controlled lifestyle intervention — the TULIP Study. BMC Med. Genet. 9, doi:10.1186/1471-2350-9-100 (2008).

59. Peeters, A. V. et al. Association of SIRT1 gene variation with visceral obesity. Hum. Genet. 124, 431–436 (2008).

60. Lagouge, M. et al. Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1α. Cell 127, 1109–1122 (2006).

61. Baur, J. A. et al. Resveratrol improves health and survival of mice on a high-calorie diet. Nature 444, 337–342 (2006).

62. Canto, C. et al. AMPK regulates energy expenditure by modulating NAD⁺ metabolism and SIRT1 activity. Nature 458, 1056–1060 (2009).
Provides a link between sirtuins and other energy sensing pathways in the cell.

63. Feige, J. N. et al. Specific SIRT1 activation mimics low energy levels and protects against diet-induced metabolic disorders by enhancing fat oxidation. Cell Metab. 8, 347–358 (2008).

64. Milne, J. C. et al. Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. Nature 450, 712–716 (2007).

65. Deng, C. X. SIRT1, is it a tumor promoter or tumor suppressor? Int. J. Biol. Sci. 5, 147–152 (2009).

66. Wang, R. H. et al. Interplay among BRCA1, SIRT1, and Survivin during BRCA1-associated tumorigenesis. Mol. Cell 32, 11–20 (2008).

67. Firestein, R. et al. The SIRT1 deacetylase suppresses intestinal tumorigenesis and colon cancer growth. PLoS One 3, e2020 (2008).

68. Li, X. et al. SIRT1 deacetylates and positively regulates the nuclear receptor LXR. Mol. Cell 28, 91–106 (2007).

69. Potente, M. et al. SIRT1 controls endothelial angiogenic functions during vascular growth. Genes Dev. 21, 2644–2658 (2007).

70. Mattagajasingh, I. et al. SIRT1 promotes endothelium-dependent vascular relaxation by activating endothelial nitric oxide synthase. Proc. Natl Acad. Sci. USA 104, 14855–14860 (2007).

71. Nisoli, E. et al. Calorie restriction promotes mitochondrial biogenesis by inducing the expression of eNOS. Science 310, 314–317 (2005).

72. Vakhrusheva, O. et al. Sirt7 increases stress resistance of cardiomyocytes and prevents apoptosis and inflammatory cardiomyopathy in mice. Circ. Res. 102, 703–710 (2008).

73. Benigni, A. et al. Disruption of the Ang II type 1 receptor promotes longevity in mice. J. Clin. Invest. 119, 524–530 (2009).

74. Araki, T., Sasaki, Y. & Milbrandt, J. Increased nuclear NAD biosynthesis and SIRT1 activation prevent axonal degeneration. Science 305, 1010–1013 (2004).

75. Fainzilber, M. & Twiss, J. L. Tracking in the Wlds—the hunting of the SIRT and the luring of the Draper. Neuron 50, 819–821 (2006).

76. Li, Y., Xu, W., McBurney, M. W. & Longo, V. D. SirT1 inhibition reduces IGF-I/IRS-2/Ras/ERK1/2 signaling and protects neurons. Cell Metab. 8, 38–48 (2008).

77. van Ham, T. J. et al. C. elegans model identifies genetic modifiers of α-synuclein inclusion formation during aging. PLoS Genet. 4, e1000027 (2008).

78. Kim, D. et al. SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis. EMBO J. 26, 3169–3179 (2007).

79. Outeiro, T. F. et al. Sirtuin 2 inhibitors rescue α-synuclein-mediated toxicity in models of Parkinson's disease. Science 317, 516–519 (2007).

80. Zhao, W. et al. Negative regulation of the deacetylase SIRT1 by DBC1. Nature 451, 587–590 (2008).

81. Kim, J. E., Chen, J. & Lou, Z. DBC1 is a negative regulator of SIRT1. Nature 451, 583–586 (2008).
Refs 80 and 81 demonstrate the role of protein–protein interactions in modulating SIRT1 function and suggest that this mode of regulation probably exists for other sirtuins.

82. Kim, E. J., Kho, J. H., Kang, M. R. & Um, S. J. Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity. Mol. Cell 28, 277–290 (2007).

83. Elliott, P. J. & Jirousek, M. Sirtuins: novel targets for metabolic disease. Curr. Opin. Investig. Drugs 9, 371–378 (2008).

84. Wellen, K. E. et al. ATP-citrate lyase links cellular metabolism to histone acetylation. Science 324, 1076–1080 (2009).

ARTICLES

Crystal structure of the ATP-gated P2X$_4$ ion channel in the closed state

Toshimitsu Kawate[1], Jennifer Carlisle Michel[1], William T. Birdsong[1] & Eric Gouaux[1,2]

P2X receptors are cation-selective ion channels gated by extracellular ATP, and are implicated in diverse physiological processes, from synaptic transmission to inflammation to the sensing of taste and pain. Because P2X receptors are not related to other ion channel proteins of known structure, there is at present no molecular foundation for mechanisms of ligand-gating, allosteric modulation and ion permeation. Here we present crystal structures of the zebrafish P2X$_4$ receptor in its closed, resting state. The chalice-shaped, trimeric receptor is knit together by subunit–subunit contacts implicated in ion channel gating and receptor assembly. Extracellular domains, rich in β-strands, have large acidic patches that may attract cations, through fenestrations, to vestibules near the ion channel. In the transmembrane pore, the 'gate' is defined by an ∼8 Å slab of protein. We define the location of three non-canonical, intersubunit ATP-binding sites, and suggest that ATP binding promotes subunit rearrangement and ion channel opening.

ATP is most commonly known as the vital carrier of free energy, having multifaceted roles in energy metabolism, biosynthesis and intracellular signal transduction. A non-canonical role for ATP in extracellular signal transduction emerged from studies showing that ATP is released from sensory nerves and promotes vasodilatation[1]. Subsequently, the concept of ATP-mediated signalling, termed purinergic signalling, was provided by Burnstock as a ubiquitous mechanism for extracellular communication[2]. Interest in this field redoubled after molecular cloning and characterization of two different ATP receptors: ionotropic P2X, and G-protein-coupled P2Y receptors[3–6]. Although the physiological importance of purinergic signalling is now generally accepted[7], the determination of the molecular mechanisms of ATP binding and the subsequent signal transduction has been hindered by the absence of high-resolution structures for any ATP receptors.

Ionotropic P2X receptors are widely distributed throughout the human body and participate in diverse physiological processes, from the nervous system to the immune system[8]. In the central nervous system, presynaptic neurons expressing P2X receptors enhance the release of neurotransmitters such as glutamate[9,10] and GABA (γ-aminobutyric acid)[11,12], whereas expression in postsynaptic neurons is required to evoke an ATP-induced postsynaptic current[13,14]. In the peripheral nervous system, afferent neurons carrying P2X receptors sense a variety of stimuli, such as taste[15], pain[16,17] and distention of the bladder[18]. Furthermore, P2X-receptor-deficient mice demonstrate the involvement of these receptors in blood pressure regulation and vascular remodelling, autoregulation of blood flow in the retina, and interleukin-1β production from macrophages[19–22]. Because P2X receptors are integral to many signal transduction pathways, it is perhaps not surprising that the dysfunction of P2X-receptor-mediated signalling is implicated in cancer[23] and inflammatory[24], cardiovascular, and neuronal diseases. P2X receptors are therefore promising targets for new therapeutic agents.

P2X receptors are cation permeable, ATP-gated ion channels derived from seven different subtypes (P2X$_{1–7}$) found only in eukaryotes[25]. Intact receptors are composed of three subunits assembled as either homomeric or heteromeric complexes contingent on the specific subunits and the cellular context[26–29]. Gating kinetics and pharmacology vary widely between different homomeric and heteromeric receptor assemblages. Whereas homomeric P2X$_1$ receptors exhibit rapid, nearly complete desensitization and high sensitivity to suramin and pyridoxalphosphate-6-azophenyl-2',4'-disulphonic acid (PPADS), homomeric P2X$_4$ receptors show slow, incomplete desensitization and insensitivity to common P2X receptor antagonists[30]. Secondary structure prediction and hydropathy plots indicate that each subunit has two transmembrane segments arranged such that the intracellular domain is formed by the amino and the carboxy termini. Although the transmembrane topologies of P2X receptors are similar to acid-sensing ion channels (ASICs, also known as ACCNs), epithelial sodium channels (ENaCs), and degenerin channels (DEGs)[31], there is little, if any, relationship between their primary amino acid sequences.

Ascertaining the structure of a P2X receptor will not only elaborate on the architecture of this important class of ligand-gated ion channels and, thus, form the basis for molecular mechanisms of function, but it will also provide new insights into the molecular principles of agonist and antagonist binding, in turn spurring the design of new therapeutic agents. Here we show the crystal structure of a zebrafish P2X$_4$ receptor at 3.1 Å resolution, verifying that these receptors are trimers with previously unseen subunit folds and non-canonical ATP-binding sites. The closed transmembrane pore, consistent with crystallization of the receptor in the absence of ATP, defines the ion channel gate in a closed, resting state.

Crystallization and structure determination

P2X receptors tend to aggregate or dissociate in the presence of detergents commonly used for crystallization (Supplementary Fig. 1). We therefore used fluorescence-detection size-exclusion chromatography (FSEC) to rapidly and efficiently evaluate the stability and monodispersity of 35 P2X orthologues expressed in transiently transfected HEK293 cells[32]. The zebrafish P2X$_{4.1}$ receptor (encoded by *p2rx4a*; P2X$_4$) emerged as a promising candidate for crystallization trials because it has a sharp and symmetrical elution profile (Fig. 1d, grey trace). The full-length zebrafish P2X$_4$ receptor is activated by ATP with an effector concentration for half-maximum response (EC$_{50}$) of ∼800 μM (Fig. 1a and Supplementary Fig. 2a)[33]. To improve crystallization behaviour, however, we analysed a series of N and C termini

[1]Vollum Institute, [2]Howard Hughes Medical Institute, Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Oregon 97239, USA.
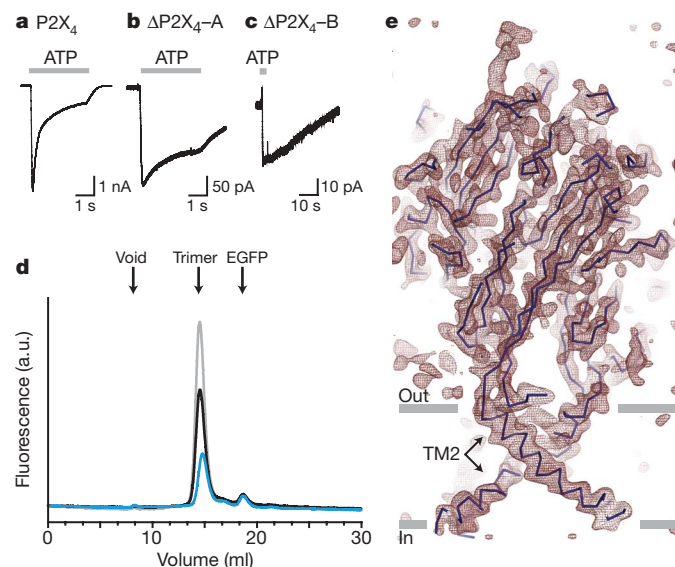
Figure 1 | A functional P2X$_4$ receptor for structural studies. a–c, Whole cell recordings of ATP-evoked current (1 mM, 3 s, grey bars) from the full-length P2X$_4$-EGFP construct (**a**), the ΔP2X$_4$-EGFP-A construct (**b**), and the ΔP2X$_4$-EGFP-B construct (**c**). **d**, FSEC profiles for P2X$_{4.1}$-EGFP (grey), ΔP2X$_4$-EGFP-A (black), and ΔP2X$_4$-EGFP-B (blue) expressed in tsA201 cells. The arrows indicate the estimated elution position of the void volume, the P2X$_4$-EGFP receptor (trimer) and free EGFP. a.u., arbitrary units. **e**, $2F_o - F_c$ electron density map contoured at 1.2σ. The blue line represents the Cα trace and the grey bars suggest the boundaries of the outer (out) and inner (in) leaflets of the membrane bilayer. The featured slice depicts TM2 helices but not TM1.

deletion mutants, settling on a minimal yet functional construct (ΔP2X$_4$-A; Fig. 1d, black trace). Further optimization to avoid non-native disulphide bond formation and to reduce heterogeneity resulting from glycosylation yielded a derivative of ΔP2X$_4$-A containing three point mutations (Cys51Phe/Asn78Lys/Asn187Arg; termed ΔP2X$_4$-B; Fig. 1d, blue trace). Electrophysiological experiments showed that both ΔP2X$_4$-A and -B are activated by 1 mM ATP (Fig. 1b, c), although the peak current amplitudes are smaller than those recorded from the full-length receptor (Fig. 1a)—an observation consistent with the lower expression levels of the mutants (Fig. 1d and Supplementary Fig. 2b–d). The ΔP2X$_4$-A structure was solved by single-wavelength anomalous diffraction (SAD) using a gadolinium derivative, and the ΔP2X$_4$-B structure was solved by molecular replacement (Supplementary Tables 1 and 2). Because the ΔP2X$_4$-B crystals diffract to higher resolution than those formed from the ΔP2X$_4$-A construct, we used the ΔP2X$_4$-B structure in most of the analysis reported here.

Architecture

The homotrimeric zebrafish P2X$_4$ receptor has a chalice-like shape, with the large extracellular domain protruding ~70 Å above the membrane plane, and the comparatively smaller transmembrane stem extending ~28 Å through the membrane (Fig. 2a). In the ΔP2X$_4$-A receptor complex each of the three crystallographically independent subunits adopts a similar conformation (Supplementary Fig. 7), and in the ΔP2X$_4$-B structure the subunits are related by the crystallographic three-fold axis of symmetry passing through the receptor centre, perpendicular to the putative membrane plane (Supplementary Fig. 8).

The shape of the transmembrane region is reminiscent of an hourglass and is formed by six transmembrane helices, two from each of the three subunits. Within a subunit, the transmembrane helices are oriented approximately antiparallel to one another and are angled nearly 45° from the membrane normal. The inner TM2 helices cross each other about halfway across their membrane-spanning lengths, constricting the transmembrane pore and defining the closed, resting

state of the channel (Fig. 1e). At the cytoplasmic termini of TM1 and TM2 the electron density is weak and we were not able to fit all of the residues to electron density features.

In contrast to the left-handed twist of the transmembrane helices, as seen from the cytoplasmic termini, the extracellular region of each subunit wraps around its neighbour with a right-handed twist, gripping adjacent subunits with extensive contact interfaces (Fig. 2a). The large extracellular domain, when viewed perpendicular to the crystallographic three-fold axis of symmetry, has a corrugated profile, replete with protruding N-linked glycosylation moieties. Seen parallel to the three-fold axis, the extracellular domain is shaped like an equilateral triangle (Fig. 2b). Although the transmembrane topology of P2X receptors is similar to that of ASICs and other members of the ENaC/DEG superfamily, the fold of the extracellular domains and the corresponding trimeric quaternary architecture is entirely different from ASICs, tetrameric ionotropic glutamate receptors and pentameric Cys-loop receptors.

Subunit fold and interfaces

The P2X$_4$ subunit resembles the shape of a dolphin, with the transmembrane helices and the extracellular region akin to the flukes and the upper body, respectively (Fig. 3a). The central architecture of the extracellular body domain is characterized by a transthyretin-like β-sandwich motif[34]. This domain appears structurally rigid and perhaps even resistant to conformational changes because the two β-sheets in the β-sandwich are knitted together by extensive contacts. Interestingly, the upper regions of the core β-sheets in the body domain contact neighbouring subunits, whereas there are no contacts between adjacent subunits at the base of the extracellular domain, proximal to the transmembrane domain. This conformation may allow the transmembrane helices, which are connected directly to the lower region of the body domain, the latitude to move to an open conformation after ligand-induced rearrangement of the upper regions. Attached to the body domain are the head domain and three structurally different elements: the dorsal fin, the right flipper and the left flipper. The head domain adopts a fold similar to an oligosaccharide-binding protein[35], and is defined by three antiparallel β-strands and one α-helix. Electron density is weak between Lys 136 and Asp 141, and we have introduced a corresponding break in the polypeptide chain. We find that all ten conserved cysteine residues in the extracellular region form pairings previously predicted by mutagenesis and electrophysiological studies[36,37] (Supplementary Fig. 9).

Subunit–subunit interactions are largely mediated by the extracellular domains, and a single subunit buries ~3,750 Å^2 of surface area upon trimer formation. The three major subunit–subunit interfaces are: body-to-body, head-to-body, and left-flipper-to-dorsal-fin (Fig. 3b). On the one hand, the residues forming the core β-sheets in the body domain are highly conserved, suggesting that the body-to-body interactions represent contacts common to all P2X receptors. On the other hand, the residues in the head, the left flipper, and the dorsal fin are less conserved (Supplementary Figs 3 and 10). We speculate that the head-to-body and the left-flipper-to-dorsal-fin interactions may encode some of the chemical and structural information that guides assembly of homomeric or heteromeric receptors. Subunit–subunit contacts are also probably involved in receptor function and, consistent with this hypothesis, experiments have shown that a single mutation in the left flipper of the P2X$_3$ receptor, Asp266Ala (Asp 283 in zebrafish P2X$_4$), considerably slows the rate of receptor desensitization[38]. A plausible explanation for the lack of function in homotrimeric P2X$_6$ receptors is that owing to ~9 missing residues in the left flipper, subunit–subunit contacts are compromised, decoupling agonist binding from ion channel gating.

Closed, resting state

The ion channel domain consists of three TM2 helices arranged around the three-fold axis of molecular symmetry, positioned to define most of the ion conducting pathway and surrounded by
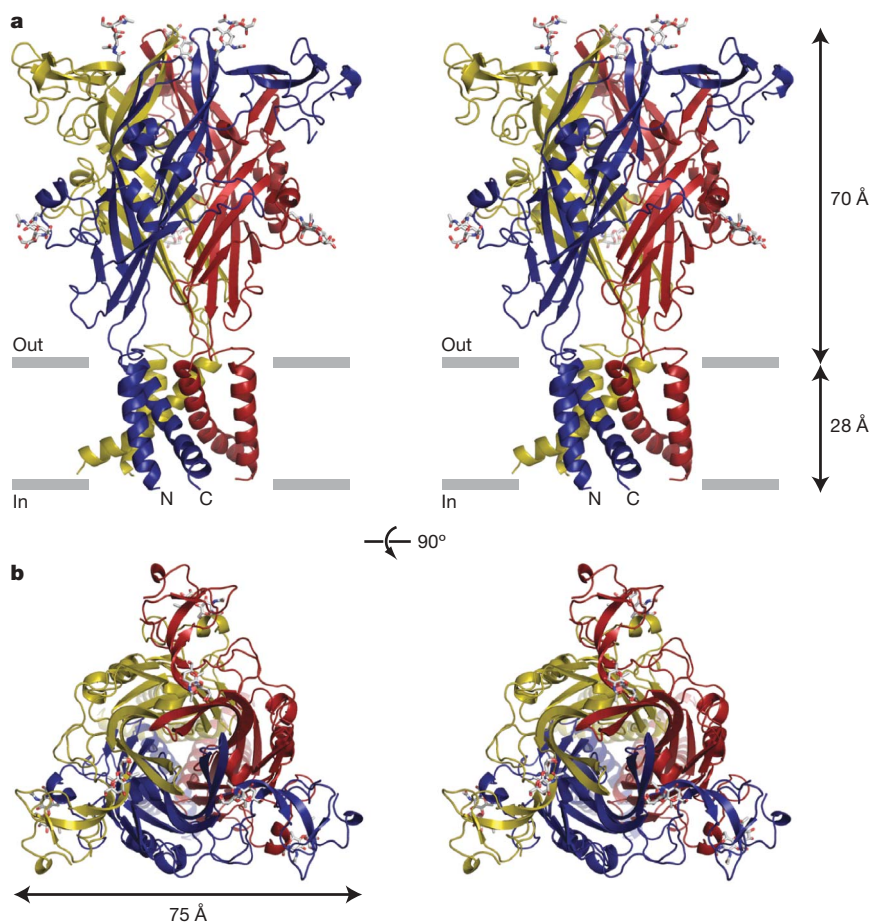
Figure 2 | The architecture of P2X receptors. a, Stereoview of the homotrimeric ΔP2X$_4$ structure viewed parallel to the membrane. Each subunit is depicted in a different colour. NAG and glycosylated asparagine residues are shown in stick representation. The grey bars suggest the boundaries of the outer (out) and inner (in) leaflets of the membrane bilayer. **b**, Stereoview of the homotrimeric ΔP2X$_4$ structure parallel to the molecular three-fold axis from the extracellular side of the membrane.

three peripheral TM1 helices (Fig. 4). A solvent-accessible surface representation clearly shows that the extracellular vestibule extends only a fraction of the distance across the membrane bilayer, to residues Leu 340 and Asn 341. On the cytoplasmic side of this occlusion, probably 5–10 Å from bulk intracellular solution, there is a

solvent-accessible intracellular vestibule. Because the receptor was crystallized in an agonist-free, apo state, and because the putative ion permeation pathway is unambiguously occluded, the present structure provides an atomic model for the closed, resting state of P2X receptors.
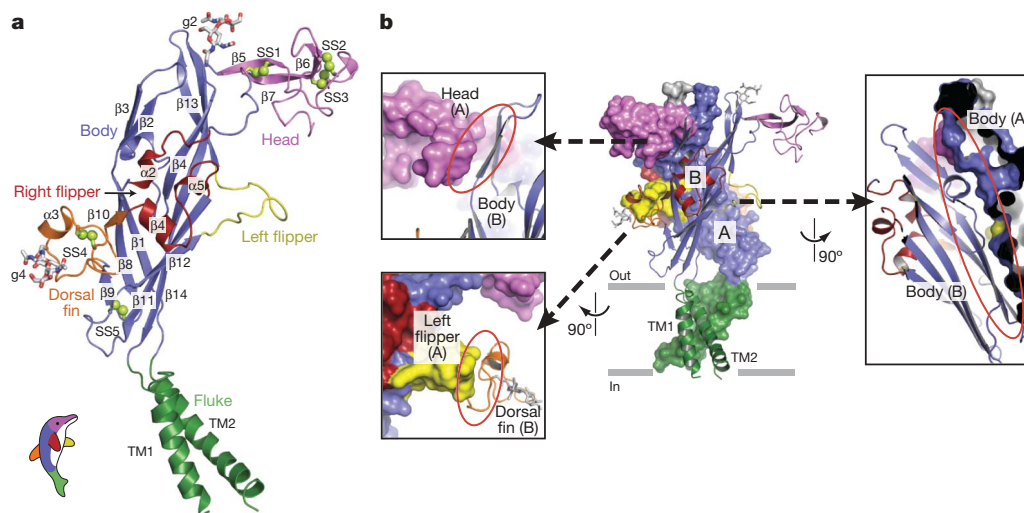
Figure 3 | Subunit fold and intersubunit contacts. a, The ΔP2X$_4$ subunit has a dolphin-like shape. α-helices (TM1–2 and α2–5), β-strands (β1–14), disulphide bonds (SS1–5), and attached glycans (g2 and g4) are indicated. **b**, Interface of two adjacent subunits. Subunits A and B are shown in a solvent-accessible surface model and a cartoon representation, respectively. The three major subunit–subunit interfaces are emphasized in different panels in which the red ellipses highlight the contacts between the two subunits. Models are coloured according to domains as in panel **a**.
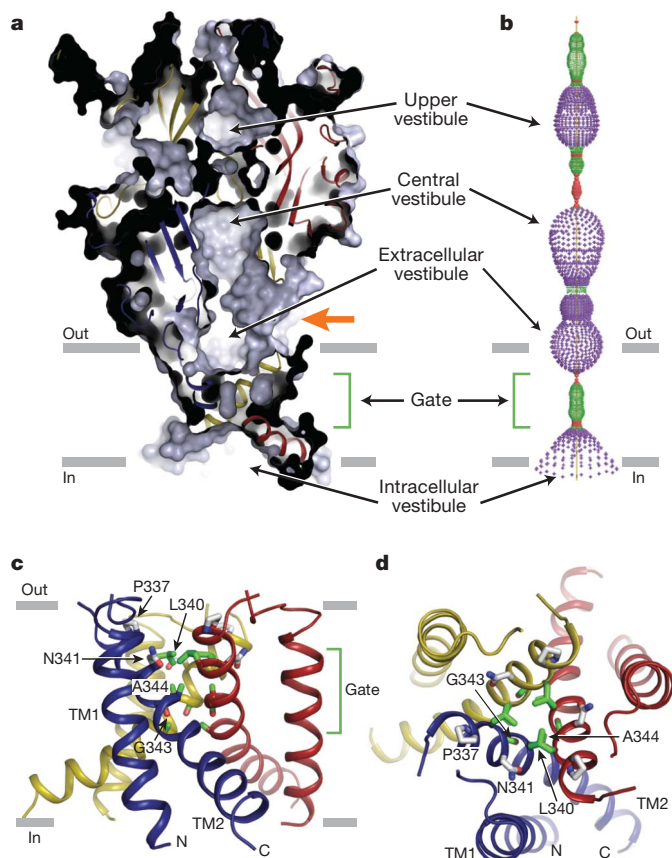
Figure 4 | Closed, resting conformation. a, A sagittal section reveals a closed conformation of the pore and shows that the gate is located about halfway across the membrane bilayer. Four vestibules (upper, central, extracellular and intracellular vestibules) are located on the molecular three-fold axis, with the extracellular vestibule connected to the bulk solution through a fenestration (orange arrow). **b,** Pore-lining surface calculated by the HOLE[49] program. Each colour represents a different radius range measured from the receptor centre (red: <1.15 Å, green: 1.15–2.3 Å, and purple: >2.3 Å). **c,** Cartoon representations of the transmembrane domain viewed parallel to the membrane plane. Phe 337 and Asn 341 are shown in grey, and potential gate residues (Leu 340, Gly 343 and Ala 344) are shown in green. **d,** Transmembrane domain viewed perpendicular to the membrane plane.

Ion channel access

Inspection of the zebrafish $P2X_4$ structure suggests two pathways by which ions in extracellular solution might access the transmembrane ion channel. The first pathway is through any of three fenestrations located directly above the transmembrane domains, proximal to the extracellular leaflet of the membrane bilayer (Fig. 4a, orange arrow). With openings as large as ~8 Å in diameter, these fenestrations should readily allow Na^+, K^+ and Ca^{2+} ions to access the channel. A second possible pathway runs the length of the extracellular domain, along the three-fold axis of symmetry and through two conspicuous vestibules rich in acidic residues (Fig. 4a, b). In this apo, closed-state structure the constrictions flanking the upper vestibule are too narrow for hydrated ions to pass (~2.3 Å). However, agonist binding may induce conformational changes between subunits, expanding these constrictions and, thus, enabling ions to access the transmembrane ion channel. On the cytoplasmic side of the ion channel, we propose that ions exit or enter the pore by the intracellular vestibule, an inverted cone-like structure that includes the conserved aspartic acid residue, Asp 357, a residue important to receptor assembly[39].

Ion channel gate

What are the solvent-accessible boundaries of the ion channel gate, and what residues or elements of protein structure define the gate? Viewed from the extracellular surface, residues Leu 340 and Asn 341 define the extracellular boundary of the ion channel gate, with the hydrophobic side chain of Leu 340 occluding the pore. On the opposite side of the membrane, the cytoplasmic gate is defined by Ala 347 and the side chain of Leu 346 (Fig. 4c, d and Supplementary Fig. 11). The 'centre' of the gate is Ala 344 and it defines the closest association of the TM2 helices. Therefore, the P2X receptor ion channel gate is primarily flanked by hydrophobic residues, includes about two turns of the TM2 α-helix, is composed of a slab of packed protein that is ~8 Å thick, and is consistent with recent cysteine-accessibility studies[40].

Ion selectivity

On the basis of the analysis of the current $P2X_4$ structure, together with the recently solved structure of a minimal functional chicken ASIC1 construct (ASIC1mfc)[41], we speculate on the molecular basis of cation selectivity in P2X receptors and suggest two distinct yet complementary mechanisms. First, the presence of several acidic residues in the central vestibule, immediately above the ion channel, together with Asp 59 and Asp 61 near the extracellular fenestrations, may not only enable the direct binding of cations, but may also create a long-range negative electrostatic potential that serves to concentrate cations near the extracellular entrance of the ion channel (Fig. 5g). Second, we suggest that permeant ions interact directly and specifically with the main-chain and side-chain oxygen atoms in the transmembrane ion channel. Although we have not yet determined the conducting, open channel structure, we speculate that the side-chain oxygen atom of Asn 341 may interact directly with permeant ions, perhaps in a similar manner to the interactions between Asp 433 and Cs^+ in the ASIC1mfc structure[41]. As ions progress towards the cytoplasm, main-chain carbonyl oxygen atoms from carbonyl groups slightly tipped off of the TM2 helix axis may participate in further protein–ion interactions.

Modulation by Gd^{3+}

To solve the structure of the $\Delta P2X_4$-A construct, we used a Gd^{3+} derivative and found four highly occupied sites. One site is located in the central vestibule, on the three-fold axis of non-crystallographic symmetry, and is coordinated by carboxylate groups of Glu 98 residues from each of the three subunits (Fig. 5a, c, d, g). The other three Gd^{3+} sites are located at the periphery of the receptor (one site for each subunit) and are coordinated by the carboxylate group of Asp 184 and the hydroxyl group of an N-acetyl-D-glucosamine (NAG) residue attached to Asn 187 (Fig. 5a, b, d). Because P2X receptors are commonly modulated by divalent and trivalent cations, and because the Gd^{3+} ions were bound to sites on the $P2X_4$ receptor that might have a role in ion channel function, we asked whether Gd^{3+} altered ATP-dependent receptor gating.

Whole-cell patch-clamp recordings of mammalian tsA201 cells transfected with the $\Delta P2X_4$-A construct demonstrated that in the presence of 100 μM Gd^{3+}, and at a holding potential of −60 mV, coapplication of 30 μM ATP failed to elicit inward current, suggesting that Gd^{3+} is an antagonist (Fig. 5e, f). At positive holding potentials, Gd^{3+} continued to antagonize ATP-dependent receptor activation, thus demonstrating that Gd^{3+} was not acting solely as a pore blocker (Supplementary Fig. 12a). Although increasing ATP concentrations concomitantly extinguished Gd^{3+} antagonism (Supplementary Fig. 12c), raising the possibility that Gd^{3+} might simply sequester ATP, a direct action of Gd^{3+} on the receptor is supported by the fact that preapplication of Gd^{3+} greatly attenuated channel activation when ATP was subsequently applied in a Gd^{3+}-free solution (Supplementary Fig. 12a, b). Furthermore, Gd^{3+} speeds the rate of ion channel deactivation (Supplementary Fig. 12c). Finally, the direct action of Gd^{3+} on the receptor is further bolstered by the presence of four highly occupied Gd^{3+}-binding sites, one of which is at the 'top' of the profoundly acidic central vestibule, a cavity that also serves to attract and concentrate cations (Fig. 5a, g).
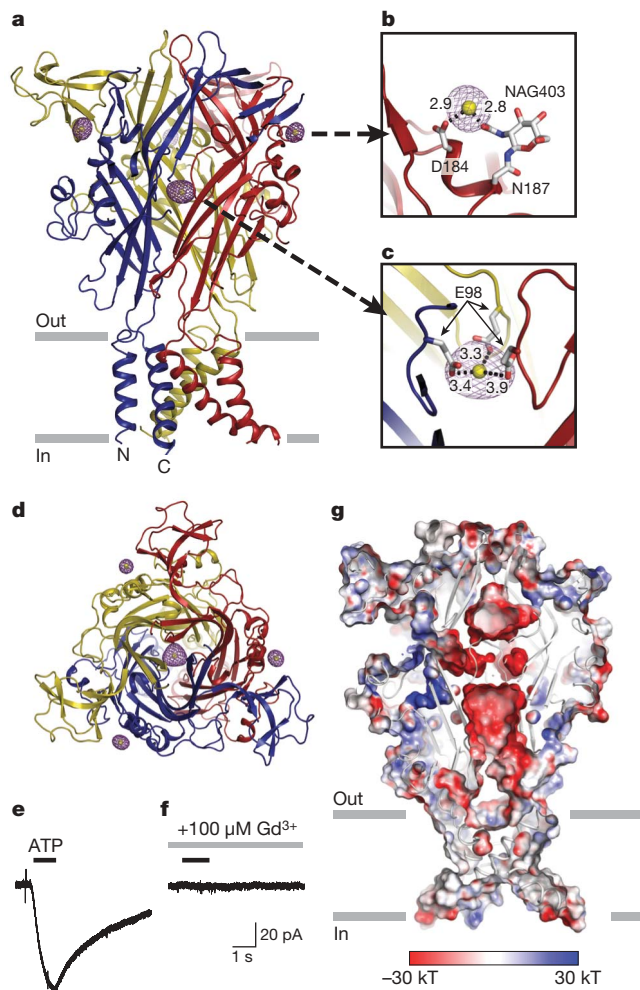
Figure 5 | Gadolinium (Gd³⁺)-binding sites. a, Anomalous difference Fourier map contoured at 8.0σ (purple) and the modelled Gd³⁺ ions (yellow). **b**, A peripheral Gd³⁺-binding site in chain B. Distances between Gd³⁺ and the coordinating carboxyl or hydroxyl groups are shown in angstrom. **c**, The central Gd³⁺-binding site is coordinated by three Glu 98 residues. **d**, View of the Gd³⁺-binding sites parallel to the molecular three-fold axis from the extracellular side of the membrane. Panels **a–d** are derived from diffraction data measured from crystals of the ΔP2X₄-A construct. **e, f**, Gd³⁺ antagonizes ΔP2X₄-EGFP-A whole-cell currents measured by patch-clamp electrophysiology. ATP (30 μM, 1 s, black bar) evokes an inward current in tsA201 cells expressing ΔP2X₄-EGFP-A (**e**). Pre-application of Gd³⁺ (100 μM, 5 s, grey bar) inhibits the current evoked by ATP (30 μM, 1 s, black bar) (**f**). **g**, Acidic surface on the central vestibule of ΔP2X₄. Electrostatic potential surface and cartoon representations of ΔP2X₄-B sliced as in Fig. 4a show an acidic patch located in the middle of the three subunits. The surface is coloured on the basis of the electrostatic potential contoured from −30 kT (red) to +30 kT (blue). White denotes 0 kT. Surface potential was calculated using APBS tools[50] for a ΔP2X₄-B model in which side-chain atoms were added to residues without side-chain atoms in the crystal structure. The following residues were excluded from the calculation: Tyr 53, Asn 78 and Asn 187.

ATP-binding site

The location of the ATP-binding site remains unknown. We suggest that deep grooves on the outside of the trimer, 45 Å from the ion channel domain and spanning neighbouring subunits, are the binding sites for ATP (Fig. 6a, b). These intersubunit grooves are populated by eight conserved residues implicated in ATP-dependent P2X receptor gating[42–46] (Fig. 6c), and whose amino acid composition is compatible with an ATP-binding motif. This putative ATP site, shaped like an open jaw, is one of three in the receptor. It is surrounded by the head domain, the body domain, the right flipper and the dorsal fin, and
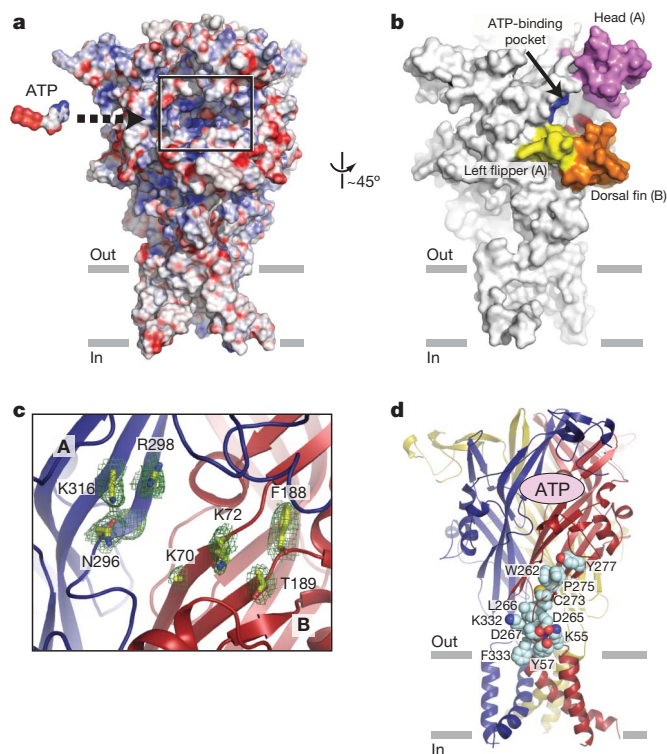
Figure 6 | ATP-binding site. a, A plausible ATP-binding pocket located between two neighbouring subunits is highlighted in the black rectangle on an electrostatic potential surface representation of the trimeric ΔP2X₄-B receptor. The surface is coloured on the basis of the electrostatic potential contoured from −30 kT (red) to +30 kT (blue). White denotes 0 kT. An ATP molecule, scaled appropriately, is also shown. **b**, A surface representation viewed ~45° from panel **a**. The head, dorsal fin and left flipper domains forming the jaw-shaped ATP-binding pocket are coloured as in Fig. 3. The putative ATP-binding residues are in blue for subunit A (Asn 296, Arg 298 and Lys 316) and in red for subunit B (Lys 70, Lys 72 and Thr 189). **c**, Close-up view of the highlighted region in **a** illustrating subunits A (blue) and B (red). Conserved residues implicated in ATP binding[42–46] are labelled, and side chains are in stick representation. Contours from a 2F₀ − F_c electron density map drawn around the side chains are in green. The electron density for the side chain of Lys 70 is weak and it has been built as an alanine. **d**, Conserved residues, shown in space-filling representation, are located between the ATP-binding site and the transmembrane domain–extracellular domain interface. Only residues for a single subunit are shown.

includes residues Lys 70, Lys 72, Phe 188 and Thr 189 from one subunit, and residues Asn 296, Phe 297, Arg 298 and Lys 316 from the neighbouring subunit. Among those residues, Lys 70, Lys 72, Thr 189, Asn 296, Arg 298 and Lys 316 are oriented towards the groove of the pocket, indicating that they may bind to ATP directly. In contrast, both Phe 188 and Phe 297 are oriented away from the groove, suggesting that they may participate in transducing conformational changes from the binding pocket to the ion channel (Supplementary Fig. 13). We speculate that ATP binding induces movement of the head, right flipper and dorsal fin domains, effectively closing these 'jaws' around the agonist, and thus resulting in conformational changes within and between subunits.

Antagonist-binding site

A recent study has shown that the Phe95Leu mutation in human P2X₇ (Ile 94 in zebrafish P2X₄) drastically reduces the sensitivity to allosteric antagonists such as N2-(3,4-difluorophenyl)-N1-(2-methyl-5-(1-piperazinylmethyl)phenyl)glycinamide dihydrochloride (GW791343) and 4-(4-fluorophenyl)-2-(4-methylsulphinylphenyl)-5-(4-pyridyl)1H-imidazole[47] (SB203580). Likewise, the Arg126Gly mutation (Ala 126 in zebrafish P2X₄) reduces the potency of PPADS. In a different study, Lys 138 in human P2X₁ (Asp 141 in

zebrafish P2X₄) was shown to be important for the inhibitory effect of suramin[48]. Notably, all residues are located in the vicinity of the predicted ATP-binding pocket. Although speculative, we suggest that these antagonists block conformational rearrangements by occupying part or all of the ATP site and precluding closure of the head, right flipper and dorsal fin domain jaws.

Conclusion

Here we present the first, to our knowledge, crystal structure of an ATP-gated P2X ion channel in a closed, resting state at 3.1 Å resolution, providing atomic-resolution evidence that these receptors are trimeric in subunit stoichiometry, with each subunit being composed of two continuous, transmembrane α-helices, intracellular termini and a large disulphide-bond-rich extracellular domain. We propose that ATP binds to a non-canonical site ∼45 Å from the ion channel domain, in a deep cleft, inducing conformational changes within and between subunits. We speculate that these changes, in turn, are propagated to the ion channel by conserved residues located at the interface between the transmembrane domain and the extracellular domain (Fig. 6d), opening the ion channel pore.

METHODS SUMMARY

Thirty-five P2X receptor genes fused to EGFP were separately and rapidly screened by transient transfection in human embryonic kidney (HEK293) cells, followed by FSEC. On the basis of a sharp, symmetric elution profile, the zebrafish P2X₄.₁ receptor was identified as a highly promising construct for X-ray crystallographic studies. The shortest well-behaved constructs of P2X₄.₁ (ΔP2X₄-A or -B) were expressed in Sf9 cells using a baculovirus infection system, the membranes were solubilized in n-dodecyl-β-D-maltoside (DDM), and the receptor was purified by metal-ion-affinity and size-exclusion chromatography. ΔP2X₄-A crystals were grown in 10–12% PEG 4000, 100 mM sodium acetate (pH 4.2–4.6), 100 mM ammonium sulphate, and 1 mM GdCl₃ in D₂O, whereas ΔP2X₄-B crystals were obtained with 20% PEG 2000, 100 mM Tris, pH 8.4, 300 mM MgNO₃ and 1 mM GdCl₃. The ΔP2X₄-A structure was solved by SAD using data measured at the gadolinium (Gd) L_III edge. SOLVE was used to determine Gd ion positions and to calculate SAD phases. These initial phases were subsequently improved by density modification using programs in the CCP4 package. Iterative model building and refinement were performed using the crystallography software COOT and PHENIX. The ΔP2X₄-B structure was determined by molecular replacement. Whole-cell patch-clamp recordings were performed on tsA201 cells transfected with plasmid DNA encoding P2X₄.₁-EGFP, ΔP2X₄-A-EGFP or ΔP2X₄-B-EGFP constructs.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Holton, F. A. & Holton, P. The capillary dilator substances in dry powders of spinal roots; a possible role of adenosine triphosphate in chemical transmission from nerve endings. *J. Physiol. (Lond.)* **126**, 124–140 (1954).
2. Burnstock, G. Purinergic nerves. *Pharmacol. Rev.* **24**, 509–581 (1972).
3. Valera, S. et al. A new class of ligand-gated ion channel defined by P₂ₓ receptor for extracellular ATP. *Nature* **371**, 516–519 (1994).
4. Lustig, K. D., Shiau, A. K., Brake, A. J. & Julius, D. Expression cloning of an ATP receptor from mouse neuroblastoma cells. *Proc. Natl Acad. Sci. USA* **90**, 5113–5117 (1993).
5. Webb, T. E. et al. Cloning and functional expression of a brain G-protein-coupled ATP receptor. *FEBS Lett.* **324**, 219–225 (1993).
6. Brake, A. J., Wagenbach, M. J. & Julius, D. New structural motif for ligand-gated ion channels defined by an ionotropic ATP receptor. *Nature* **371**, 519–523 (1994).
7. Schwiebert, E. M. & Zsembery, A. Extracellular ATP as a signaling molecule for epithelial cells. *Biochim. Biophys. Acta* **1615**, 7–32 (2003).
8. Surprenant, A. & North, R. A. Signaling at purinergic P2X receptors. *Annu. Rev. Physiol.* **71**, 333–359 (2008).
9. Khakh, B. S. & Henderson, G. ATP receptor-mediated enhancement of fast excitatory neurotransmitter release in the brain. *Mol. Pharmacol.* **54**, 372–378 (1998).
10. Gu, J. G. & MacDermott, A. B. Activation of ATP P2X receptors elicits glutamate release from sensory neuron synapses. *Nature* **389**, 749–753 (1997).
11. Hugel, S. & Schlichter, R. Presynaptic P2X receptors facilitate inhibitory GABAergic transmission between cultured rat spinal cord dorsal horn neurons. *J. Neurosci.* **20**, 2121–2130 (2000).
12. Donato, R. et al. GABA release by basket cells onto Purkinje cells, in rat cerebellar slices, is directly controlled by presynaptic purinergic receptors, modulating Ca²⁺ influx. *Cell Calcium* **44**, 521–532 (2008).
13. Edwards, F. A., Gibb, A. J. & Colquhoun, D. ATP receptor-mediated synaptic currents in the central nervous system. *Nature* **359**, 144–147 (1992).
14. Sim, J. A. et al. Altered hippocampal synaptic potentiation in P2X₄ knock-out mice. *J. Neurosci.* **26**, 9006–9009 (2006).
15. Finger, T. E. et al. ATP signaling is crucial for communication from taste buds to gustatory nerves. *Science* **310**, 1495–1499 (2005).
16. Cook, S. P., Vulchanova, L., Hargreaves, K. M., Elde, R. & McCleskey, E. W. Distinct ATP receptors on pain-sensing and stretch-sensing neurons. *Nature* **387**, 505–508 (1997).
17. Souslova, V. et al. Warm-coding deficits and aberrant inflammatory pain in mice lacking P2X₃ receptors. *Nature* **407**, 1015–1017 (2000).
18. Cockayne, D. A. et al. Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X₃-deficient mice. *Nature* **407**, 1011–1015 (2000).
19. Chessell, I. P. et al. Disruption of the P2X₇ purinoceptor gene abolishes chronic inflammatory and neuropathic pain. *Pain* **114**, 386–396 (2005).
20. Yamamoto, K. et al. Impaired flow-dependent control of vascular tone and remodeling in P2X4-deficient mice. *Nature Med.* **12**, 133–137 (2006).
21. Inscho, E. W., Cook, A. K., Imig, J. D., Vial, C. & Evans, R. J. Renal autoregulation in P2X₁ knockout mice. *Acta Physiol. Scand.* **181**, 445–453 (2004).
22. Solle, M. et al. Altered cytokine production in mice lacking P2X₇ receptors. *J. Biol. Chem.* **276**, 125–132 (2001).
23. White, N. & Burnstock, G. P2 receptors and cancer. *Trends Pharmacol. Sci.* **27**, 211–217 (2006).
24. Di Virgilio, F. Liaisons dangereuses: P2X₇ and the inflammasome. *Trends Pharmacol. Sci.* **28**, 465–472 (2007).
25. North, R. A. Molecular physiology of P2X receptors. *Physiol. Rev.* **82**, 1013–1067 (2002).
26. Aschrafi, A., Sadtler, S., Niculescu, C., Rettinger, J. & Schmalzing, G. Trimeric architecture of homomeric P2X₂ and heteromeric P2X₁₊₂ receptor subtypes. *J. Mol. Biol.* **342**, 333–343 (2004).
27. Barrera, N. P., Ormond, S. J., Henderson, R. M., Murrell-Lagnado, R. D. & Edwardson, J. M. Atomic force microscopy imaging demonstrates that P2X₂ receptors are trimers but that P2X₆ receptor subunits do not oligomerize. *J. Biol. Chem.* **280**, 10759–10765 (2005).
28. Nicke, A. et al. P2X₁ and P2X₃ receptors form stable trimers: a novel structural motif of ligand-gated ion channels. *EMBO J.* **17**, 3016–3028 (1998).
29. Nicke, A., Rettinger, J. & Schmalzing, G. Monomeric and dimeric byproducts are the principal functional elements of higher order P2X₁ concatamers. *Mol. Pharmacol.* **63**, 243–252 (2003).
30. North, R. A. & Surprenant, A. Pharmacology of cloned P2X receptors. *Annu. Rev. Pharmacol. Toxicol.* **40**, 563–580 (2000).
31. Kellenberger, S. & Schild, L. Epithelial sodium channel/degenerin family of ion channels: a variety of functions for a shared structure. *Physiol. Rev.* **82**, 735–767 (2002).
32. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
33. Diaz-Hernandez, M. et al. Cloning and characterization of two novel zebrafish P2X receptor subunits. *Biochem. Biophys. Res. Commun.* **295**, 849–853 (2002).
34. Blake, C. C., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* **121**, 339–356 (1978).
35. Williams, D. C. Jr, Lee, J. Y., Cai, M., Bewley, C. A. & Clore, G. M. Crystal structures of the HIV-1 inhibitory cyanobacterial protein MVL free and bound to Man3GlcNAc2: structural basis for specificity and high-affinity binding to the core pentasaccharide from n-linked oligomannoside. *J. Biol. Chem.* **280**, 29269–29276 (2005).
36. Ennion, S. J. & Evans, R. J. Conserved cysteine residues in the extracellular loop of the human P2X₁ receptor form disulfide bonds and are involved in receptor trafficking to the cell surface. *Mol. Pharmacol.* **61**, 303–311 (2002).
37. Clyne, J. D., Wang, L. F. & Hume, R. I. Mutational analysis of the conserved cysteines of the rat P2X₂ purinoceptor. *J. Neurosci.* **22**, 3873–3880 (2002).
38. Fabbretti, E. et al. Identification of negative residues in the P2X₃ ATP receptor ectodomain as structural determinants for desensitization and the Ca²⁺-sensing modulatory sites. *J. Biol. Chem.* **279**, 53109–53115 (2004).
39. Duckwitz, W., Hausmann, R., Aschrafi, A. & Schmalzing, G. P2X₅ subunit assembly requires scaffolding by the second transmembrane domain and a conserved aspartate. *J. Biol. Chem.* **281**, 39561–39572 (2006).
40. Li, M., Chang, T. H., Silberberg, S. D. & Swartz, K. J. Gating the pore of P2X receptor channels. *Nature Neurosci.* **11**, 883–887 (2008).
41. Gonzales, E. B., Kawate, T. & Gouaux, E. Pore architecture and ion sites in acid-sensing ion channels and P2X receptors. *Nature* doi:10.1038/nature08218 (this issue).
42. Ennion, S., Hagan, S. & Evans, R. J. The role of positively charged amino acids in ATP recognition by human P2X₁ receptors. *J. Biol. Chem.* **275**, 29361–29367 (2000).
43. Roberts, J. A. & Evans, R. J. ATP binding at human P2X₁ receptors. Contribution of aromatic and basic amino acids revealed using mutagenesis and partial agonists. *J. Biol. Chem.* **279**, 9043–9055 (2004).
44. Jiang, L. H., Rassendren, F., Surprenant, A. & North, R. A. Identification of amino acid residues contributing to the ATP-binding site of a purinergic P2X receptor. *J. Biol. Chem.* **275**, 34190–34196 (2000).

45. Roberts, J. A. & Evans, R. J. Contribution of conserved polar glutamine, asparagine and threonine residues and glycosylation to agonist action at human P2X$_1$ receptors for ATP. *J. Neurochem.* **96**, 843–852 (2006).
46. Marquez-Klaka, B., Rettinger, J., Bhargava, Y., Eisele, T. & Nicke, A. Identification of an intersubunit cross-link between substituted cysteine residues located in the putative ATP binding site of the P2X1 receptor. *J. Neurosci.* **27**, 1456–1466 (2007).
47. Michel, A. D. *et al.* Identification of regions of the P2X$_7$ receptor that contribute to human and rat species differences in antagonist effects. *Br. J. Pharmacol.* **155**, 738–751 (2008).
48. Sim, J. A., Broomhead, H. E. & North, R. A. Ectodomain lysines and suramin block of P2X$_1$ receptors. *J. Biol. Chem.* **283**, 29841–29846 (2008).
49. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
50. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).

METHODS

Expression and purification. The shortest well-behaved construct of zebrafish P2X$_{4.1}$ (ΔP2X$_4$-A)[51,52] was determined by examining twelve different combinations of N and C termini deletions in Sf9 cells by rapid FSEC analysis. Similarly, the FSEC screening strategy was exploited to identify a well-behaved derivative of ΔP2X$_4$-A that carries mutations at two of four glycosylation sites (Asn78Lys/Asn187Arg) and a point mutation (Cys51Phe) in TM1 (P2X$_4$-B). The ΔP2X$_4$-A protein was expressed as an N-terminal EGFP fusion with an octa-histidine affinity tag (EGFP–8$\times$His) in baculovirus-infected Sf9 cells. Infected Sf9 cells were cultured in serum-free medium (Invitrogen) at 27 °C for 24 h after infection, after which time the temperature was reduced to 20 °C. Cells were collected 72 h after infection by centrifugation at 6,200g and broken by sonication in TBS (50 mM Tris, pH 8.0, 150 mM NaCl) supplemented with 1 mM phenylmethanesulphonylfluoride, 5.2 µg ml^{-1} aprotinin, 2 µg ml^{-1} leupeptin, and 1.4 µg ml^{-1} pepstatin A (all from Sigma Aldrich). Cell debris was cleared by a low-speed spin (10,000g). Membranes were collected by a high-speed spin (19,000g) and solubilized in TBS containing 40 mM DDM (Anatrace). The detergent-soluble fraction was incubated with cobalt-charged metal ion affinity resin (Clontech), and ΔP2X$_4$ was eluted with 250 mM imidazole (Fluka) and 1 mM DDM in TBS. After thrombin digestion to remove the EGFP–8$\times$His tag, ΔP2X$_4$ was isolated by size-exclusion chromatography in 20 mM HEPES, pH 7.0, 80 mM NaCl, 20 mM KCl and 0.5 mM DDM. Peak fractions were pooled, concentrated to 2 mg ml^{-1}, and used for crystallization. All steps after Sf9 cell culture were carried out on ice or at 4 °C. For the purification of ΔP2X$_4$-B, all steps were identical with the exception that 15% glycerol was included in the solubilization, metal-ion affinity and size-exclusion chromatography buffers. For production of selenomethionine (SeMet)-labelled receptor, baculovirus-infected Sf9 cells were cultured for 1 day at 27 °C, collected by centrifugation at 1,000g for 5 min, re-cultured in serum-free medium without methionine for 4 h at 27 °C, and then supplemented with 50 mg l^{-1} SeMet (Anatrace). After 10 h incubation at 27 °C, the temperature was shifted to 20 °C, and the cells were cultured for another 2 days before collection. SeMet proteins were purified as described earlier.

Crystallization. For ΔP2X$_4$-A, crystals were obtained at 4 °C in 3–4 weeks by vapour diffusion by mixing 1:1 or 2:1 (v/v) ratios of protein and a reservoir solution containing 10–12% PEG 4000, 100 mM sodium acetate (pH 4.2–4.6), 100 mM ammonium sulphate and 1 mM GdCl$_3$ in D$_2$O. Crystals were dehydrated and cryoprotected by adding glycerol in 2.5% steps (final 12.5%) followed by increasing the PEG 4000 concentration by 2.5% steps (final 25%). For native and SeMet crystals, GdCl$_3$ was excluded from the final cryoprotection solution. For ΔP2X$_4$-B, crystals were obtained at 4 °C in 9 months by vapour diffusion by mixing 1:1 or 2:1 ratios of protein and a reservoir solution containing 20% PEG 2000, 300 mM Mg(NO$_3$)$_2$, 100 mM Tris, pH 8.4, and 1 mM GdCl$_3$. Crystals were cryoprotected by adding glycerol in 2.0% steps (final 18%). Crystals were flash-frozen in liquid nitrogen and used for X-ray diffraction data collection.

Structure determination. X-ray data sets were collected at the Advanced Light Source (beam lines 5.0.2, 8.2.1 and 8.2.2) and at the Advanced Photon Source (beamline 24-ID-E). The diffraction frames were indexed, integrated and scaled using HKL2000. The structure of ΔP2X$_4$-A was solved using data from a SAD experiment (Advanced Light Source beamline 5.0.2). The program SOLVE was used to find heavy-atom positions and to calculate phases. The phases were improved by density modification that included three-fold non-crystallographic symmetry averaging, as carried out by the computer program DM. This R3 space group crystal form has one receptor trimer in the asymmetric unit (Supplementary Fig. 4). Initially, several poly-alanine chains were built into the electron density map using COOT. Subsequently, specific protein sequences, together with correct side-chain atoms, were fitted to the electron density map on the basis of Met and Cys locations derived from anomalous Fourier difference maps calculated from SeMet data derived from wild-type, Cys51Met or Leu349Met constructs (Supplementary Fig. 5) and from native data collected at a low energy ($\lambda = 1.6$ Å), respectively. Finally, iterative model building using COOT and CCP4 led to a continuous protein model that nevertheless contained several Ala residues at positions where the native, longer amino acid side chains were disordered. The resulting structure was manually rebuilt and refined using programs in the CCP4, COOT and PHENIX packages with the following NCS restraints: B-factor weight, 10; coordinate sigma, 0.1 for residues 36–64, 119–169 and 326–352, and 0.04 for residues 65–118 and 170–235. The structure of ΔP2X$_4$-B was obtained by molecular replacement with the refined model of ΔP2X$_4$-A using the program Phaser. There is one subunit in the asymmetric unit of the R32 ΔP2X$_4$-B crystal form. The resulting model was manually rebuilt and refined using programs in the CCP4, COOT and PHENIX packages. Electron densities for the transmembrane helices 1 and 2 are shown in Supplementary Fig. 6. The structures were validated by the computer program PROCHECK and MolProbity.

Electrophysiology. Whole-cell patch-clamp experiments were performed on a mammalian cell line (tsA201) transiently expressing EGFP, P2X$_{4.1}$-EGFP, ΔP2X$_4$-A-EGFP or ΔP2X$_4$-B-EGFP constructs using methods previously described. The holding potential was -70 mV unless noted. The pipette solution contained (in mM): 115 K-methanesulphonate, 20 NaCl, 1.5 MgCl$_2$, 10 HEPES and 10 BAPTA; pH was adjusted to 7.4 with KOH. Standard extracellular solution contained (mM): 140 NaCl, 5 KCl, 2 CaCl$_2$, 1 MgCl$_2$, 10 HEPES and 10 MES; pH was adjusted to 7.4 using N-methyl-D-glucamine. ATP test solutions were made from serial dilutions of standard extracellular solution supplemented with 10 mM Na$_2$ATP, pH 7.4, with N-methyl-D-glucamine. Dilutions were made in standard extracellular solution supplemented with 20 mM NaCl to maintain an equivalent sodium concentration among all test solutions. External solutions were exchanged on cells within 20 ms using computer actuated solenoid valves controlling flow through an array of 10 µl pipettes positioned within several hundred micrometres of the cell. All recordings were at room temperature ($\sim$23 °C). Data were collected using pClamp (Molecular Devices), analysed with Clampfit (Molecular Devices) and Origin Lab software, and organized using Excel (Microsoft). To account for run-down of current responses, the peak current amplitude for each ATP test was measured and scaled relative to the current evoked by a preceding test of 100 µM ATP (a concentration which did not cause run-down). Data were plotted relative to the current evoked by 100 µM ATP and fit to the Hill equation to determine the ATP concentration required to evoke a half-maximal current. For presentation, data were normalized to the maximum-evoked current (I_{max}) of the best fit to the Hill equation.

51. Diaz-Hernandez, M. *et al.* Cloning and characterization of two novel zebrafish P2X receptor subunits. *Biochem. Biophys. Res. Commun.* **295**, 849–853 (2002).
52. Kucenas, S. *et al.* Molecular characterization of the zebrafish P2X receptor subunit gene family. *Neuroscience* **121**, 935–945 (2003).

Pore architecture and ion sites in acid-sensing ion channels and P2X receptors

Eric B. Gonzales[1], Toshimitsu Kawate[1] & Eric Gouaux[1,2]

Acid-sensing ion channels are proton-activated, sodium-selective channels composed of three subunits, and are members of the superfamily of epithelial sodium channels, mechanosensitive and FMRF-amide peptide-gated ion channels. These ubiquitous eukaryotic ion channels have essential roles in biological activities as diverse as sodium homeostasis, taste and pain. Despite their crucial roles in biology and their unusual trimeric subunit stoichiometry, there is little knowledge of the structural and chemical principles underlying their ion channel architecture and ion-binding sites. Here we present the structure of a functional acid-sensing ion channel in a desensitized state at 3 Å resolution, the location and composition of the ~8 Å 'thick' desensitization gate, and the trigonal antiprism coordination of caesium ions bound in the extracellular vestibule. Comparison of the acid-sensing ion channel structure with the ATP-gated P2X$_4$ receptor reveals similarity in pore architecture and aqueous vestibules, suggesting that there are unanticipated yet common structural and mechanistic principles.

Acid-sensing ion channels (ASICs, also known as ACCNs) are voltage-independent, sodium-selective, ligand-gated ion channels expressed in diverse organisms[1–3], and are members of the epithelial sodium channel/degenerin (ENaC/DEG) family[4]. These proton-gated channels form functional homomeric and heteromeric channels with a spectrum of pH sensitivities and desensitization profiles[5]. The crystal structure of chicken ASIC1 (after amino- and carboxy-terminal deletions, termed ΔASIC1)[6], together with subsequent microscopy[7] and biochemical[8] studies, proved that ASICs are trimers, and by extension, that ENaCs and DEG channels are also, thus ending the stoichiometry controversy[9,10]. The structure of ΔASIC1 defined the architecture of the extracellular domain, and the locations of a chloride ion and residues implicated in proton binding. Because the ΔASIC1 construct does not exhibit proton-dependent gating[6], however, it left unresolved the structure of the ion channel, the location and chemical composition of the ion channel gate, and the stereochemistry and location of ion-binding sites.

ASICs and ENaCs are sodium-selective channels, favouring Na$^+$ over K$^+$ by tenfold[11] and 100-fold[12,13], respectively. The trimeric sub-unit arrangement of ASICs and ENaCs, together with their selectivity for sodium, raises questions related to the architecture of the ion channel, the shape and chemical composition of the pore, the location of ion channel gates, and the position(s) of residue(s) implicated in ion permeation and block. In ENaCs, the steric volume of Ser 589 (α subunit) in TM2 has a profound effect on selectivity, with permeation of larger ions correlated with residues of larger steric volume, suggesting that residues at this position modulate the diameter of the selectivity filter[10,14,15]. To illuminate fundamental principles of ion channel architecture and gating in ASICs and related ion channels, and to define the location of residues such as Ser 589, we have studied the atomic structure of a functional ASIC. Here we describe the crystal structure of a minimal functional chicken ASIC1 construct (ASIC1mfc), show the location and composition of the desensitization gate of the ion channel, define the location and stereochemistry of caesium-binding sites in the extracellular vestibule, and assign the position of residues implicated in ion selectivity.

The ASIC/ENaC/DEG superfamily is not the only group of trimeric ion channels. P2X receptors are also trimeric ligand-gated ion channels activated by ATP[16–20]. Although ASICs and P2X receptors share a common transmembrane topology, with intracellular termini and two predicted transmembrane segments, together with large, cysteine-rich extracellular domains, there are no significant amino acid sequence relationships between ASICs and P2X receptors. In a companion paper, we describe the crystal structure of the zebrafish P2X$_4$ receptor[21]. Comparison of the functional chicken ASIC structure and the P2X$_4$ structure provides unanticipated insights into the structural similarities of ASIC and P2X receptor ion channel domains, the locations and chemical compositions of the ion channel gates of the closed and desensitized states, and the presence of negatively charged aqueous vestibules in the extracellular domains.

Crystallization of a functional ASIC

We defined a functional receptor amenable to crystallization by examining chicken ASIC1 deletion mutants by patch-clamp electro-physiology and fluorescence-detection size-exclusion chromatography (FSEC)[22]. A construct that included the N terminus—a region of ASICs and ENaCs crucial to activation gating[23,24]—yet which stopped at residue 466, shortly after the second transmem-brane segment, preserved the pH-dependent gating and sodium selectivity of the full-length channel and yielded a sharp and sym-metrical peak by FSEC (Fig. 1a–c). The ASIC1mfc produced crystals at pH 6.5 that diffract to 3 Å resolution (Supplementary Table 1 and Supplementary Fig. 1).

The ASIC1mfc crystal structure was solved by molecular replace-ment using the extracellular domain of a ΔASIC1 protomer[6] as a search probe. Initial ASIC1mfc electron density maps indicated that the conformation of the transmembrane domains differed from the ΔASIC1 structure, and thus we manually built this portion of the structure (Supplementary Figs 2 and 3). The model was subjected to cycles of refinement and rebuilding, ultimately yielding a structure with R_{work} and R_{free} values of 22.5% and 27.4%, respectively, and good stereochemistry (Supplementary Tables 1 and 2). The present struc-ture includes residues 46–451; N (2–45) and C (452–466) terminal residues were not seen in electron density maps. Unlike in the ΔASIC1 crystal form, there are no visible lattice contacts between the

[1]Vollum Institute, [2]Howard Hughes Medical Institute, Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA.
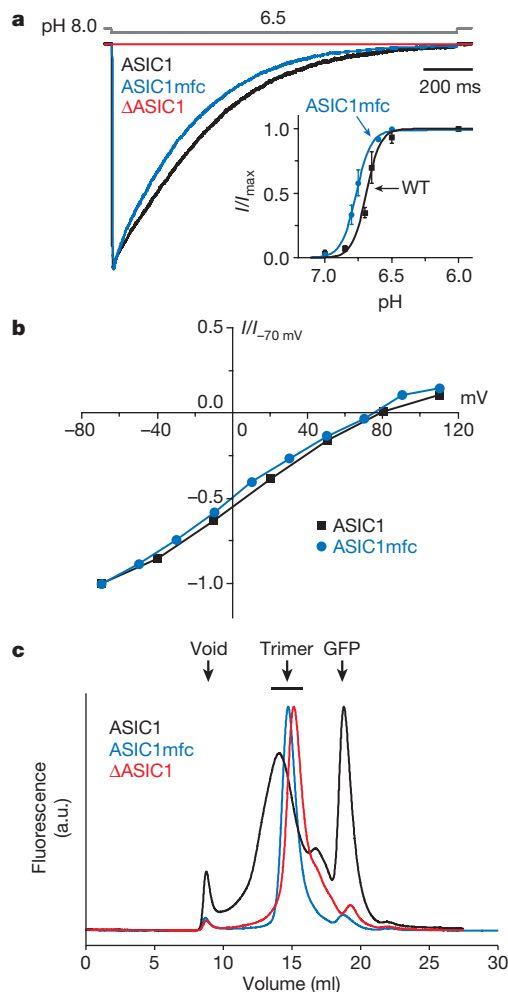
Figure 1 | Identification of a minimally functional chicken ASIC1 construct.
a, Typical current generated by chicken ASIC1 constructs in outside-out
patch recordings (holding potential −60 mV). Channels were activated by
stepping from pH 8.0 to 6.5 for the wild-type (black), ΔASIC1 (red) and
ASIC1mfc (blue) constructs (open-tip junction potential is grey). Typical
pH concentration-response relationships are shown in the inset. The wild-
type channel had a half-maximal effective pH (pH$_{50}$) of 6.68 ± 0.01 and n
value of 8.35 ± 1.91. The ASIC1mfc had a pH$_{50}$ and n value of 6.77 ± 0.01
and 7.86 ± 1.22, respectively. Points on the curve are the mean $\pm$ s.e.m. of at
least four cells. **b**, Normalized current–voltage relationship. Current
observed at low pH and various holding potentials was normalized to the
current observed at −70 mV. **c**, FSEC of chicken ASIC1 constructs. The
arrows indicate the estimated elution position of the void volume, the
ASIC1mfc-GFP receptor (trimer) and free GFP. a.u., arbitrary units.

transmembrane domains and neighbouring protein molecules
(Supplementary Fig. 1).

Ion channel architecture

In this desensitized state, the ASIC1mfc ion channel has the shape of an
hourglass with wide vestibules near the extracellular and cytoplasmic
surfaces of the membrane bilayer and a narrow constriction at the
middle (Fig. 2a–c). The channel domain is defined by six continuous
transmembrane helices in which two helical segments (TM1 and TM2)
are contributed by each of three subunits related by the three-fold axis
of crystallographic symmetry. The TM2 helices are positioned closest
to the three-fold axis, are tilted by ~50° from the membrane normal,
and cross each other about halfway across the putative membrane
bilayer (Supplementary Fig. 2a, b). The TM1 helical segments form
extensive interactions with the TM2 helix of the same subunit, but also
make contacts with TM1 and TM2 helices on adjacent subunits. The
TM1 helices therefore make most contacts with the lipid bilayer,
whereas the TM2 helices line the putative ion channel pore.
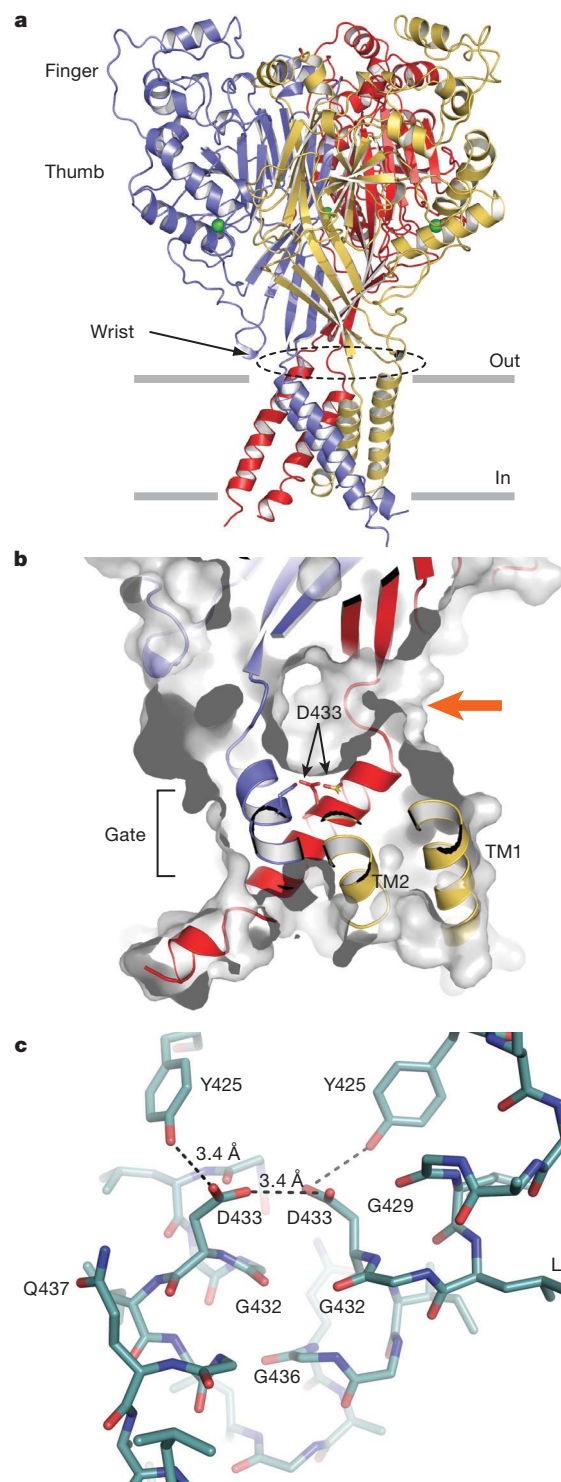
Figure 2 | Structure of ASIC1mfc. a, View of the functional ASIC1mfc
trimer. Chloride ions are green spheres. The 'thumb', 'finger' and wrist
regions are labelled. Grey bars suggest the boundaries of the outer (out) and
inner (in) leaflets of the membrane bilayer. **b**, A vertical slice through a
solvent-accessible surface representation of the transmembrane domain, the
extracellular vestibule and fenestrations. One of the three equivalent
fenestrations is indicated by an arrow. Asp 433 defines the bottom of the
extracellular vestibule. **c**, Key interactions between symmetry-related
Asp 433 carboxyl and Tyr 425 hydroxyl groups.

Access to the pore

Ions and small molecules access the ASIC1mfc pore through three
oval-shaped fenestrations that lead directly to the extracellular ves-
tibule (Supplementary Fig. 4). These portals are located at the 'wrist'[6],

the juncture between the transmembrane domains and the extracellular domains, near the boundary of the membrane bilayer and the extracellular solution (Fig. 2a). Framing the ~4 Å × ~10 Å fenestrations are the C terminus of TM1 and the beginning of β1, together with the end of β12 and the N terminus of TM2. We suggest that ions might also reach the pore by a path along the three-fold axis of symmetry. Although this pathway is occluded in the present low pH structure, modest conformational changes that might accompany opening of the ion channel may expand the constrictions along the three-fold axis in the extracellular domain (Fig. 3 and Supplementary Fig. 5).

Once through the fenestrations, ions move into the broad extracellular vestibule, which is located above residues Gly 432 and Asp 433 of the TM2 domain and is ~8 Å wide and 12 Å high (Fig. 2b, c). If we use Phe 70 of TM1 to define the boundary between the lipid head groups and aqueous solution, the 'lower' half of the extracellular vestibule resides within the membrane plane. Bounding the vestibule on the 'top' are the methyl groups of Val 75, and on the 'bottom' are three symmetry-related carboxyl moieties of Asp 433, positioned to occlude the ion channel pore (Fig. 2c). On the cytoplasmic side of the closed pore there is another vestibule, shaped like an inverted cone, with a diameter at its base of ~15 Å and a height of ~10 Å, forming the bottom portion of the transmembrane 'hourglass' (Fig. 3).

Vestibules in the extracellular domain have profound negative electrostatic potentials, arising from the concentration of 12 acidic residues in each of the central (Glu 80, Glu 374, Glu 412 and Glu 417) and extracellular vestibules (Asp 79, Glu 420, Glu 426 and Asp 433). We suggest that these vestibules act as a 'cation reservoir', concentrating cations near the ion channel pore leading to robust channel conductance[25]. Attraction of divalent cations within the vestibule may also reduce both the ambient Na^+ concentration and the resulting Na^+ current, as seen in diminished ASIC3 (also known as ACCN3) sodium current at high calcium concentration[26].

The desensitization gate

There is not a continuous solvent-accessible path from one vestibule to the other in the low pH desensitized ASIC1mfc structure, and the obvious ion conduction pathway along the molecular three-fold axis is occluded by a constriction formed by the crossing of the TM2 domains (Fig. 3). This constriction describes the ASIC desensitization gate, an ~8 Å thick occlusion bounded on the extracellular side by the carboxyl groups of Asp 433, on the cytoplasmic side by Gly 436, and in the middle by one turn of the three TM2 α-helices. The conserved Gly 436 residue allows symmetry-related TM2 helices to pack close together, with Gly 436 Cα–Cα distances of ~3.8 Å, thus unambiguously demonstrating that the desensitization gate is a physical block of the transmembrane pore. Beyond Gly 436, towards the cytoplasm, the transmembrane helices diverge, opening up the ion channel to bulk solution.

Residue Asp 433 defines the bottom of the extracellular vestibule, at the boundary of the densensitization gate, and is implicated in low affinity, voltage-dependent Ca^{2+} block[27] (Fig. 2b). The close proximity of the three Asp 433 carboxyl groups (~3.4 Å) suggests that their pK_a values are perturbed, with either one or more protonated (that is, neutral), or that there are monovalent or divalent (Ca^{2+} or Mg^{2+}) cations bound, especially at low proton concentration. Gly 432 is 'below' the side chain of Asp 433 and is the site of the degenerin mutation[28,29]. The substitution of larger amino acid residues at this position in MEC-4 channels results in neuronal death[30] due to constitutive channel activation[31]. In ENaCs and ASICs, similar mutations increase open probability[32] and yield constitutively active channels[33], respectively. Using simple model building, side-chain atoms of non-glycine residues at position 432 sterically clash with symmetry-related TM2 domains, thus providing a mechanism by which mutations at the degenerin site perturb gating.

Ion-binding sites

To probe for monovalent cation-binding sites, we soaked ASIC1mfc crystals in solutions containing Cs^+, a weakly permeant ion (Supplementary Fig. 6) with a substantial anomalous signal (Fig. 4 and Supplementary Fig. 7). The resulting anomalous difference map showed a cylindrical electron density feature located on the crystallographic three-fold axis stretching from the Asp 433 carboxyl groups to the carbonyl oxygens of Gly 432 (Fig. 4a). The two Cs^+ ions manually fit to this elongated peak are separated by ~3.4 Å, suggesting that both sites are probably not simultaneously occupied. The Cs^+ ion at position 1, immediately 'above' the base of the extracellular vestibule, is coordinated by three carboxyl oxygen atoms from symmetry-related Asp 433 residues (Fig. 4b).

The Cs^+ ion at site 2 is coordinated by three main-chain carbonyl oxygen atoms from symmetry-related Gly 432 residues, and three oxygens from the Asp 433 side chains (Fig. 4b). The main-chain carbonyl groups of Gly 432 can coordinate the Cs^+ ion at site 2 because the carbonyl group is turned away from the helix axis, a conformation most accessible to a glycine residue. We suggest that coordination of permeant ions in ASICs, and related ion channels, involves main-chain carbonyl oxygen atoms, derived primarily from glycine residues, adopting conformations in which the carbonyl group is tipped off of the helix axis. Thus, main-chain carbonyl oxygen atoms at glycine residues 436, 439 and 443 may coordinate permeant ions in the open-channel state.

The two Cs^+-binding sites in the extracellular domain are within the thumb and finger domains—regions implicated in proton-dependent ASIC activation[6] (Fig. 4 and Supplementary Fig. 7). One Cs^+ site is located near putative proton-binding sites (Asp 238–Asp 350 and Glu 239–Asp 346) in a solvent-filled crevice. The other Cs^+-binding site is enveloped by backbone carbonyl oxygens beneath the thumb α4-helix. The functional consequence of Cs^+ binding to these sites is unknown.

Insights into ion permeation

What can the Cs^+ sites tell us about the principles of ion binding in a trimeric ion channel? Even though the ionic radius of Cs^+ (1.69 Å) is substantially larger than Na^+ (0.95 Å)[34], analysis of these sites is instructive. The Cs^+ ion at site 2 is bound by six ligands arranged in a trigonal antiprism geometry that arises because the three ligands on the 'upper' triangular plane (Asp 433) adopt a staggered arrangement
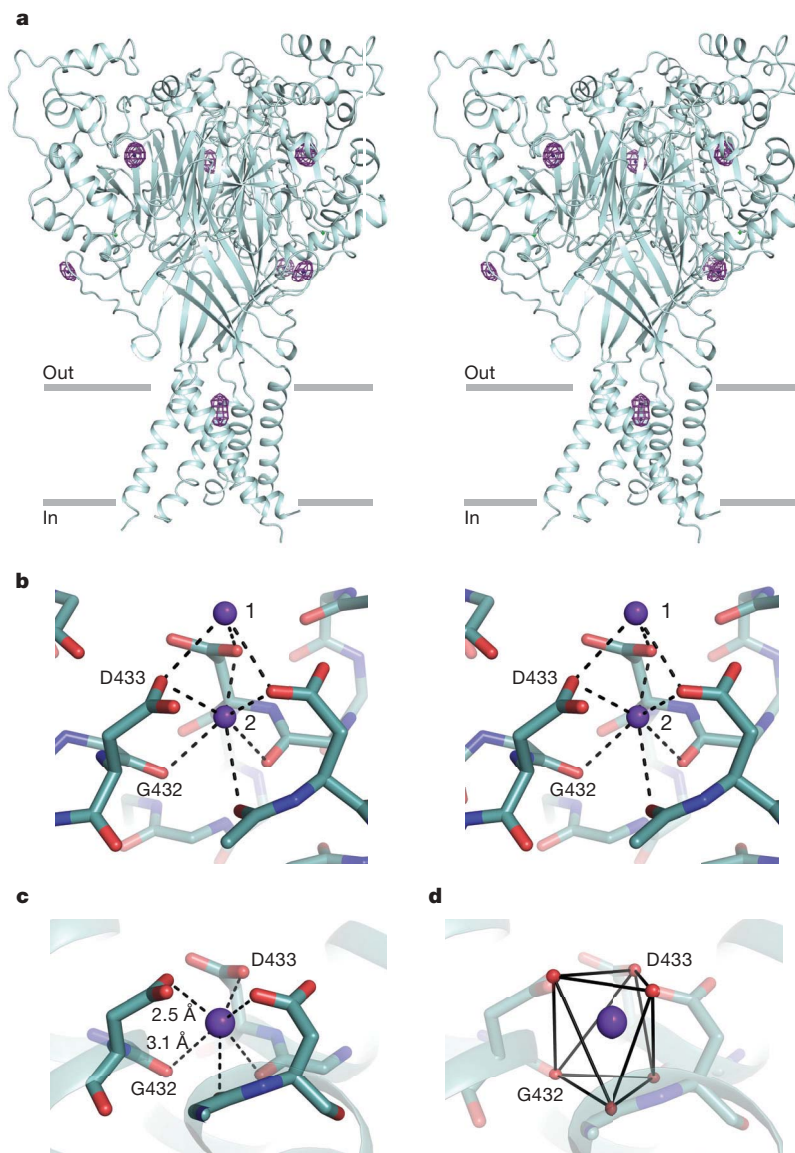

Figure 3 | Vestibules and possible ion permeation pathways. a, An electrostatic potential surface and cartoon representation of ASIC1mfc sliced along the molecular three-fold axis of symmetry. The surface is coloured based on electrostatic potential, contoured from −50 kT (red) to +30 kT (blue). White is 0 kT. **b**, Illustration of the radius of possible pathways along the three-fold axis, generated using the HOLE program (red < 1.4 Å < green < 2.3 Å < purple).

a

b

c

d

Figure 4 | **Cs$^+$-binding sites. a**, Stereo image of the electron density peaks (3.5σ) from anomalous difference Fourier maps calculated using diffraction data measured from crystals soaked in CsCl. **b**, Key interactions between Cs$^+$ ions at sites 1 and 2 with the main-chain and side-chain oxygen atoms of Gly 432 and Asp 433, respectively. **c**, Stick representation of ASIC1mfc interaction with Cs$^+$ at site 2. **d**, Trigonal antiprism coordination of the Cs$^+$ ion by the Gly 432 carbonyl and Asp 433 carboxyl oxygens. Oxygen atoms (red spheres) form the vertices, whereas solid lines represent the sides of each of the two staggered triangles of the antiprism.

in comparison to the three ligands on the lower triangular plane (Gly 432; Fig. 4c). This coordination geometry is notable because it provides for an appropriate number of ligands for coordination of a Na$^+$ ion (six partial negative charges)[35] while perfectly accommodating the underlying molecular symmetry of the protein. We predict that if permeant ions maintain trigonal antiprism coordination through the open pore, then they would shed most of their hydration shell. Indeed, mutations in the ENaC selectivity-filter reduced the Na$^+$ conductance, but became permeable to K$^+$, possibly because of a perturbation of precise channel-ion hydration geometry[10,14,36].

The trigonal antiprism coordination of the Cs$^+$ ion at site 2, which is topologically equivalent to octahedral coordination, is reminiscent of potassium coordination by valinomycin[37] (Supplementary Fig. 8). We suggest that sodium-selective ion channels such as ASICs and ENaCs possess ion-binding sites with trigonal antiprism coordination geometry, providing permeant ions with six ligands arranged with an ideal ion-to-ligand distance (Fig. 4d). Indeed, the trigonal antiprism coordination of the site 2 ion is stereochemically analogous to the eight-fold square antiprism coordination of larger K$^+$ ions in potassium channels[38]—in both cases the symmetry and subunit number of the protein matrix is matched to the optimal coordination requirements of the permeant ion[35].

By simply translating sites 1 and 2 along the three-fold axis, towards the cytoplasm, we predict a third ion-binding site in the open, conducting state of the channel. This site is defined by the

carbonyl oxygens of Gly 432 and Gly 436, all oriented towards the three-fold axis, and arranged with trigonal antiprism coordination geometry. Other residues probably contribute, directly and indirectly to the permeation pathway, including Ser 445, which in ENaCs changes ion selectivity on mutation[10,14,15]. In ASIC1mfc, however, the side chain of Ser 445 is oriented away from the channel pore, and thus it either indirectly affects selectivity or is brought into direct interaction with ions after a considerable conformational change[39].

ASICs and P2X receptors

The low pH ASIC1mfc crystal structure describes the conformation of a trimeric, proton-gated ion channel in a desensitized state. In a companion paper, the crystal structure of the trimeric ATP-gated P2X$_4$ receptor in an agonist-free, closed state is described[21]. Both ASICs and P2X receptors have a similar transmembrane topology[19,40] and share common functional properties: both are ligand-gated, cation-selective ion channels. However, there is no significant amino acid sequence identity between ASICs and P2X receptors.

Upon solving the structures of ASIC1mfc and the zebrafish ΔP2X$_4$ receptor, we observed that the transmembrane domains of the two proteins were remarkably similar. Superposition of the TM2 helices using residues 427–445 of ASIC1mfc and residues 335–353 of ΔP2X$_4$ yielded a root mean squared deviation of 2.3 Å, demonstrating that the organization of the TM2 helices are roughly equivalent (Fig. 5a). Thus, the fundamental conformations of the ion channel domains
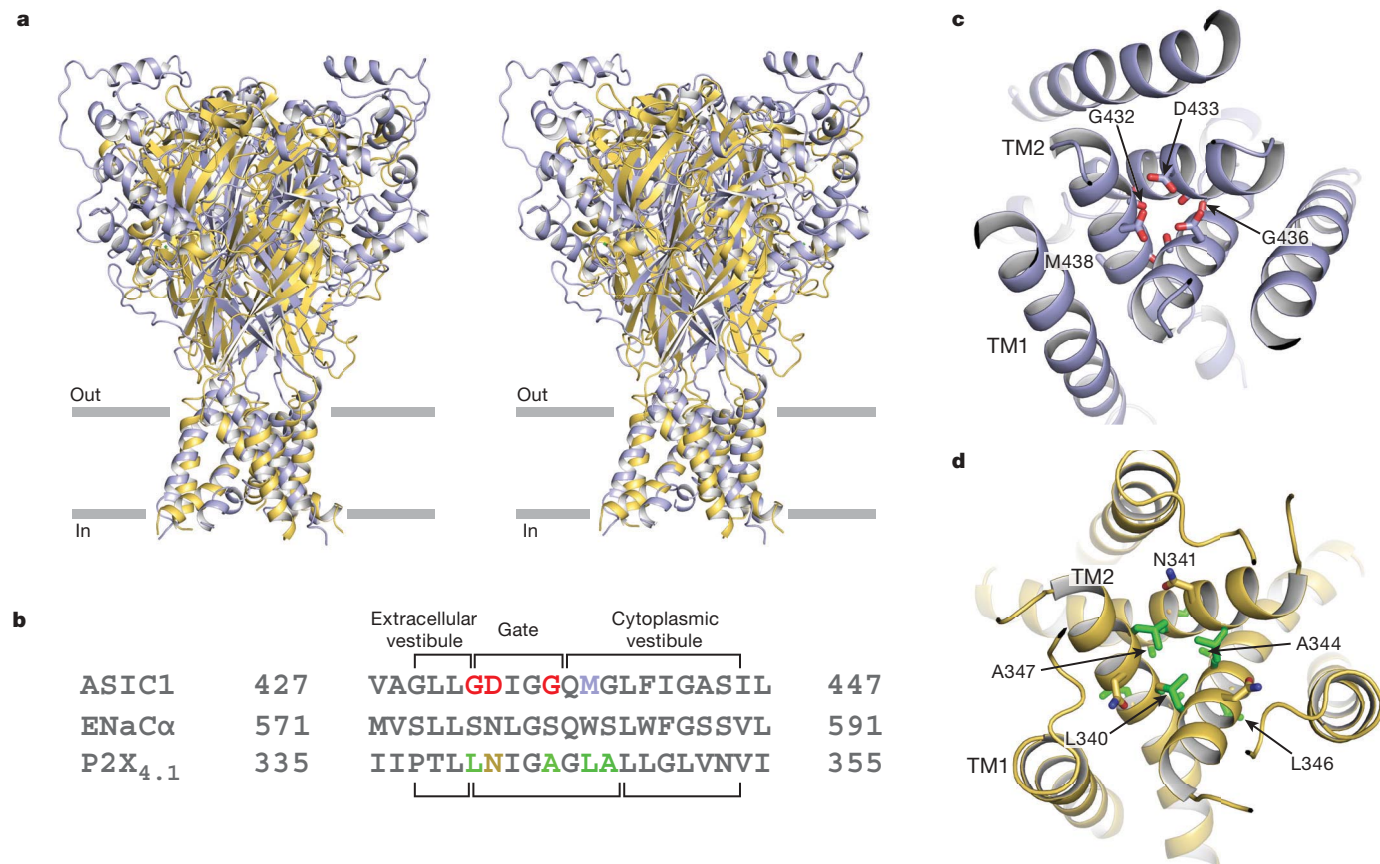
Figure 5 | ASIC and P2X receptors share a common pore architecture.
a, Stereoview of ASIC1mfc (purple) and ΔP2X$_4$ (gold) crystal structure overlay based on superposition of the respective TM2 domains. **b**, A structure-based sequence alignment of chicken ASIC1 and zebrafish P2X$_4$ TM2 domains is shown. Corresponding residues in ASIC1mfc and P2X$_4$ are highlighted by side-chain colour. A rat ENaCα sequence is shown for comparison. **c, d**, Views of the ASIC1mfc (**c**) and ΔP2X$_4$ (**d**) ion channel pores from the extracellular side of the membrane. Selected side-chain residues of ΔP2X$_4$ are in green. The ASIC1mfc Met 438 is modelled as an alanine.

are similar in the resting state of the ΔP2X$_4$ receptor and in the desensitized state of ASIC1mfc (Fig. 5a).

The structural superposition of the two receptors allows us to align the amino acid sequences of the TM2 domains, revealing unanticipated relationships (Fig. 5b). For example, we see that there is a common signature sequence associated with the base of the extracellular vestibule, 'Asp-Ile-Gly' for ASIC1 and 'Asn-Ile-Gly' for P2X$_4$, with the Asp (ref. 27) and Asn (ref. 41) residues both crucial for ion channel function. In the zebrafish P2X$_4$ receptor, however, there is a leucine residue (Leu 340) at the equivalent ASIC degenerin position, yet the receptor is not constitutively active. If the ion channel domains are truly related, one must ask how the P2X$_4$ receptor accommodates this bulky amino acid. Leucine is probably tolerated in ΔP2X$_4$ owing to the presence of a proline (Pro 337) one turn of the helix upstream from Leu 340, which bends the helix, making space for the bulky side chain. In both cases the respective gates of the channels are defined by the close crossing of the TM2 helices, located at ASIC1mfc residue Gly 436 and ΔP2X$_4$ residue Ala 344, and are composed of a continuous stretch of five to eight amino acid residues (Fig. 5b, c).

The presence of similar cavities and vestibules with deep negative electrostatic potentials in the extracellular domains of the ASIC1mfc (Fig. 3) and ΔP2X$_4$ receptors represent further common features. The upper and 'central' vestibules are located along the molecular three-fold vertical axis, are similar in size, and the central vestibule is especially rich in acidic residues. In zebrafish ΔP2X$_4$ and chicken ASIC1, experimental[21] and computational studies[42] demonstrate and predict, respectively, the binding of cations in these central vestibules. The central vestibule is accessible to ions in the resting state ΔP2X$_4$ structure, but not in the ASIC1mfc desensitized state structure. It has been shown previously that residues in the central

vestibule of rat ASIC3 are accessible in the resting, but not desensitized state[43]. We speculate that accessibility to the central vestibule is modulated by conformational changes in the wrist/palm region, near Thr 76, during the transition from the closed to the desensitized state.

The ASIC1mfc structure may tell us how permeant cations are coordinated in the P2X receptor channel pore. We speculate that the presence of glycine residues in the ΔP2X$_4$ receptor may also allow main-chain carbonyl groups to tip off of the helix axis so as to allow the oxygen atoms to coordinate cations in the pore. Because the cation selectivity of P2X receptors is not as stringent as ASICs, however, there must be features of their respective ion channel pores that are distinct.

Mechanism

On the basis of the ASIC1mfc and ΔP2X$_4$ crystal structures, we propose that proton binding to sites throughout the extracellular domain of acid-sensing ion channels results in movements of the thumb and wrist regions that in turn open the activation gate by a rotation of the transmembrane helices. The opening of the pore brings main-chain carbonyl oxygens, as well as oxygen atoms from side-chain residues such as Asp 433, into an ion conduction pathway along the molecular three-fold axis of symmetry. Extended sojourns to low pH cause the wrist and adjacent portions of β1 and β12 to draw together, occluding the middle vestibule and triggering closure of the ion channel through the formation of the desensitization gate, as observed in the current ASIC1mfc crystal structure.

METHODS SUMMARY

The ASIC1mfc construct was identified by systematic deletion mutagenesis, patch-clamp electrophysiology, and FSEC studies. The ASIC1mfc protein was purified

from baculovirus-infected Sf9 cells using metal-ion-affinity and size-exclusion chromatography. Crystals were obtained in 150–350 mM NaCl, 100 mM HEPES, pH 6.5, 23–28% (v/v) PEG 400 and 10 mM taurine. PHASER[44] was used in molecular replacement, using the extracellular domain of a ΔASIC1 subunit. Density modification using programs in the CCP4 suite[45] improved the initial phases, and the structure was refined using REFMAC[46] and COOT[47]. Patch-clamp recording was performed in tsA201 and CHO-K1 cells 24–48 h after transfection with an N-terminal green fluorescent protein (GFP)-tagged ASIC1mfc construct. The endogenous proton-mediated current observed in tsA201 cells was negligible in outside-out patch recordings of chicken ASIC1 currents.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 February; accepted 16 June 2009.

1. Waldmann, R., Champigny, G., Bassilana, F., Heurteaux, C. & Lazdunski, M. A proton-gated cation channel involved in acid-sensing. *Nature* **386**, 173–177 (1997).
2. Coric, T., Zheng, D., Gerstein, M. & Canessa, C. M. Proton sensitivity of ASIC1 appeared with the rise of fishes by changes of residues in the region that follows TM1 in the ectodomain of the channel. *J. Physiol. (Lond.)* **568**, 725–735 (2005).
3. Coric, T., Passamaneck, Y. J., Zhang, P., Di Gregorio, A. & Canessa, C. M. Simple chordates exhibit a proton-independent function of acid-sensing ion channels. *FASEB J.* **22**, 1914–1923 (2008).
4. Kellenberger, S. & Schild, L. Epithelial sodium channel/degenerin family of ion channels: a variety of functions for a shared structure. *Physiol. Rev.* **82**, 735–767 (2002).
5. Hesselager, M., Timmermann, D. B. & Ahring, P. K. pH dependency and desensitization kinetics of heterologously expressed combinations of acid-sensing ion channel subunits. *J. Biol. Chem.* **279**, 11006–11015 (2004).
6. Jasti, J., Furukawa, H., Gonzales, E. B. & Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* **449**, 316–323 (2007).
7. Carnally, S. M. *et al.* Direct visualization of the trimeric structure of the ASIC1a channel, using AFM imaging. *Biochem. Biophys. Res. Commun.* **372**, 752–755 (2008).
8. Zha, X. M. *et al.* Oxidant regulated inter-subunit disulfide bond formation between ASIC1a subunits. *Proc. Natl Acad. Sci. USA* **106**, 3573–3578 (2009).
9. Coscoy, S., Lingueglia, E., Lazdunski, M. & Barbry, P. The Phe-Met-Arg-Phe-amide-activated sodium channel is a tetramer. *J. Biol. Chem.* **273**, 8317–8322 (1998).
10. Kellenberger, S., Hoffmann-Pochon, N., Gautschi, I., Schneeberger, E. & Schild, L. On the molecular basis of ion permeation in the epithelial Na$^+$ channel. *J. Gen. Physiol.* **114**, 13–30 (1999).
11. Bassler, E. L., Ngo-Anh, T. J., Geisler, H. S., Ruppersberg, J. P. & Gründer, S. Molecular and functional characterization of acid-sensing ion channel (ASIC) 1b. *J. Biol. Chem.* **276**, 33782–33787 (2001).
12. Benos, D. J., Mandel, L. J. & Simon, S. A. Cationic selectivity and competition at the sodium entry site in frog skin. *J. Gen. Physiol.* **76**, 233–247 (1980).
13. Palmer, L. G. Ion selectivity of the apical membrane Na channel in the toad urinary bladder. *J. Membr. Biol.* **67**, 91–98 (1982).
14. Kellenberger, S., Gautschi, I. & Schild, L. A single point mutation in the pore region of the epithelial Na$^+$ channel changes ion selectivity by modifying molecular sieving. *Proc. Natl Acad. Sci. USA* **96**, 4170–4175 (1999).
15. Snyder, P. M., Olson, D. R. & Bucher, D. B. A pore segment in DEG/ENaC Na$^+$ channels. *J. Biol. Chem.* **274**, 28484–28490 (1999).
16. Brake, A. J., Wagenbach, M. J. & Julius, D. New structural motif for ligand-gated ion channels defined by an ionotropic ATP receptor. *Nature* **371**, 519–523 (1994).
17. Valera, S. *et al.* A new class of ligand-gated ion channel defined by P$_{2x}$ receptor for extracellular ATP. *Nature* **371**, 516–519 (1994).
18. Nicke, A. *et al.* P2X$_1$ and P2X$_3$ receptors form stable trimers: a novel structural motif of ligand-gated ion channels. *EMBO J.* **17**, 3016–3028 (1998).
19. Aschrafi, A., Sadtler, S., Niculescu, C., Rettinger, J. & Schmalzing, G. Trimeric architecture of homomeric P2X$_2$ and heteromeric P2X$_{1+2}$ receptor subtypes. *J. Mol. Biol.* **342**, 333–343 (2004).
20. Surprenant, A. & North, R. A. Signaling at purinergic P2X receptors. *Annu. Rev. Physiol.* **71**, 333–359 (2008).
21. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Crystal structure of the ATP-gated P2X$_4$ ion channel in the closed state. *Nature* doi:10.1038/nature08198 (this issue).
22. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
23. Gründer, S. *et al.* A mutation causing pseudohypoaldosteronism type 1 identifies a conserved glycine that is involved in the gating of the epithelial sodium channel. *EMBO J.* **16**, 899–907 (1997).
24. Pfister, Y. *et al.* A gating mutation in the internal pore of ASIC1a. *J. Biol. Chem.* **281**, 11787–11791 (2006).
25. Dani, J. A. Ion-channel entrances influence permeation. Net charge, size, shape, and binding considerations. *Biophys. J.* **49**, 607–618 (1986).
26. Immke, D. C. & McCleskey, E. W. Protons open acid-sensing ion channels by catalyzing relief of Ca^{2+} blockade. *Neuron* **37**, 75–84 (2003).
27. Paukert, M., Babini, E., Pusch, M. & Gründer, S. Identification of the Ca^{2+} blocking site of acid-sensing ion channel (ASIC) 1: implications for channel gating. *J. Gen. Physiol.* **124**, 383–394 (2004).
28. Chalfie, M. & Wolinsky, E. The identification and suppression of inherited neurodegeneration in *Caenorhabditis elegans*. *Nature* **345**, 410–416 (1990).
29. Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
30. Hong, K. & Driscoll, M. A transmembrane domain of the putative channel subunit MEC-4 influences mechanotransduction and neurodegeneration in *C. elegans*. *Nature* **367**, 470–473 (1994).
31. Brown, A. L., Fernandez-Illescas, S. M., Liao, Z. & Goodman, M. B. Gain-of-function mutations in the MEC-4 DEG/ENaC sensory mechanotransduction channel alter gating and drug blockade. *J. Gen. Physiol.* **129**, 161–173 (2007).
32. Snyder, P. M., Bucher, D. B. & Olson, D. R. Gating induces a conformational change in the outer vestibule of ENaC. *J. Gen. Physiol.* **116**, 781–790 (2000).
33. Champigny, G., Voilley, N., Waldmann, R. & Lazdunski, M. Mutations causing neurodegeneration in *Caenorhabditis elegans* drastically alter the pH sensitivity and inactivation of the mammalian H$^+$-gated Na$^+$ channel MDEG1. *J. Biol. Chem.* **273**, 15418–15422 (1998).
34. Hille, B. *Ion Channels of Excitable Membranes* 3rd edn, Ch. 10 (Sinauer Associates, 2001).
35. Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
36. Kellenberger, S., Auberson, M., Gautschi, I., Schneeberger, E. & Schild, L. Permeability properties of ENaC selectivity filter mutants. *J. Gen. Physiol.* **118**, 679–692 (2001).
37. Neupert-Laves, K. & Dobler, M. The crystal structure of a K$^+$ complex of valinomycin. *Helv. Chim. Acta* **58**, 432–442 (1975).
38. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K$^+$ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
39. Askwith, C. C., Benson, C. J., Welsh, M. J. & Snyder, P. M. DEG/ENaC ion channels involved in sensory transduction are modulated by cold temperature. *Proc. Natl Acad. Sci. USA* **98**, 6459–6463 (2001).
40. Jiang, L. H. *et al.* Subunit arrangement in P2X receptors. *J. Neurosci.* **23**, 8903–8910 (2003).
41. Silberberg, S. D., Chang, T. H. & Swartz, K. J. Secondary structure and gating rearrangements of transmembrane segments in rat P2X$_4$ receptor channels. *J. Gen. Physiol.* **125**, 347–359 (2005).
42. Shaikh, S. A. & Tajkhorshid, E. Potential cation and H$^+$ binding sites in acid sensing ion channel-1. *Biophys. J.* **95**, 5153–5164 (2008).
43. Cushman, K. A., Marsh-Haffner, J., Adelman, J. P. & McCleskey, E. W. A conformation change in the extracellular domain that accompanies desensitization of acid-sensing ion channel (ASIC) 3. *J. Gen. Physiol.* **129**, 345–350 (2007).
44. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
45. Collaborative Computational Project, Number 4. The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D.* **50**, 760–763 (1994).
46. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D.* **53**, 240–255 (1997).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
49. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–383 (2007).

nature

METHODS

Expression and purification. The ASIC1mfc construct, consisting of residues Met 2–Arg 466, was used for crystallographic trials. This minimal functional construct was expressed as an N-terminal fusion with GFP from baculovirus-infected Sf9 cells. Infected cells were incubated at 27 °C in serum-free media for 48 h before collecting. Cells were collected by centrifugation and sonicated in the presence of 150 mM NaCl, 20 mM Tris, pH 8.0, 4 mM phenylmethylsuphonyl fluoride and protease inhibitor cocktail (Roche cOmplete tablets). Crude membranes were solubilized in 15 ml of 150 mM NaCl, 20 mM Tris, pH 8.0, and 40 mM n-dodecyl-β-D-maltoside (DDM) per gram of membrane for 1 h at 4 °C. The resulting solution was clarified by centrifugation (19,000g) at 4 °C for 1 h (Ti45 rotor). Protein was bound to TALON resin and eluted with 150 mM NaCl, 20 mM Tris, pH 8.0, 1 mM DDM and 250 mM imidazole. The octa-histidine (8×His) GFP-tag was cleaved with a 1-h thrombin digestion at room temperature, and the receptor was further purified by size-exclusion chromatography using a mobile phase composed of 150 mM NaCl, 20 mM Tris, pH 8.0, 1 mM DDM, 1 mM dithiothreitol and 1 mM EDTA. Peak fractions were pooled and concentrated to 1.8–2.2 mg ml^{-1} (calculated molar extinction of 53,570 at 280 nm) before crystallization.

Crystallization. Crystal trials were carried out using the hanging-drop vapour diffusion method by mixing a 1:1 ratio of protein solution and reservoir solution consisting of: 150–350 mM NaCl, 100 mM HEPES, pH 6.5, 23–28% PEG 400 and 10 mM taurine. Crystals were grown and collected at 4 °C. Cryoprotection was achieved by soaking crystals in reservoir solutions supplemented with DDM and with increasing concentrations of PEG 400 (final concentration 38–40%). The crystals were then flash-cooled in liquid nitrogen. For Cs$^+$-soaked crystals, 250 mM CsCl was substituted for NaCl in the cryoprotection solution.

Structure determination. X-ray diffraction data sets were collected at the Advanced Light Source (beam line 5.0.2). Cs$^+$-soaked crystal diffraction data were measured using low energy X-rays (9,000 eV). Collected data were processed using HKL2000 (ref. 48). Molecular replacement using the extracellular domain of the ΔASIC1 structure was used to obtain the initial phases. The phases were improved by density modification using DM. The resulting electron density maps demonstrated that whereas the ΔASIC1 extracellular domain was a faithful structure for the ASIC1mfc extracellular domain, the transmembrane domains associated with the ASIC1mfc maps adopted a significantly different conformation that the transmembrane domains of the ΔASIC1 structure. Therefore, the ASIC1mfc transmembrane domains were manually rebuilt. Iterative model building and refinement was performed using COOT[47] and REFMAC5 (ref. 46), respectively. The stereochemical quality of the final structure validity was assessed using PROCHECK and MolProbity[49].

Electrophysiology. Whole-cell and outside-out patch recording were carried out with tsA201 and CHO-K1 cells 24–48 h after transfection using plasmids encoding the fluorescently tagged ASIC1mfc construct. We did not observe endogenously expressed ASIC current in outside-out patches from tsA201 cells. Patch pipettes were pulled and polished to 2–5 MΩ resistance and filled with the following internal solution (in mM): 100 KCl, 5 NaCl, 40 HEPES, 10 EGTA and 5 MgCl$_2$, adjusted to pH 7.4 with KOH. External solutions for whole-cell recording consisted of (in mM): 150 NaCl, 5 KCl, 5 HEPES, 5 MES and 1.8 CaCl$_2$. In outside-out patch recording, the external solution contained 168 mM NaCl and 5 mM sucrose to enhance current and improve visibility between the two solutions, respectively. For CsCl substitution, the external solution contained 168 mM CsCl instead of NaCl. The pH values of all external solutions were adjusted with N-methyl-D-glucamine. Rapid solution exchange was achieved by placing outside-out patches containing ASIC1mfc receptors at the interface between high pH (8.0) and test solutions streaming from a theta-tube attached to a piezo-electric stack. Test solution was applied for 1 s at a holding potential of −60 mV. The open-tip junction potential was obtained after each experiment. Concentration-response profiles were recorded, normalized to maximal current (I_{max}), and fitted to the Hill equation.

LETTERS

Resonant stripping as the origin of dwarf spheroidal galaxies

Elena D'Onghia[1,2], Gurtina Besla[2], Thomas J. Cox[2] & Lars Hernquist[2]

Dwarf spheroidal galaxies are the most dark-matter-dominated systems in the nearby Universe[1–3] and their origin is one of the outstanding puzzles of how galaxies form. Dwarf spheroidals are poor in gas and stars, making them unusually faint[4–6], and those known as ultra-faint dwarfs[7,8] have by far the lowest measured stellar content of any galaxy[9,10]. Previous theories[11] require that dwarf spheroidals orbit near giant galaxies like the Milky Way, but some dwarfs have been observed in the outskirts of the Local Group[12]. Here we report simulations of encounters between dwarf disk galaxies and somewhat larger objects. We find that the encounters excite a process, which we term 'resonant stripping', that transforms them into dwarf spheroidals. This effect is distinct from other mechanisms proposed to form dwarf spheroidals, including mergers[13], galaxy–galaxy harassment[14], or tidal and ram pressure stripping, because it is driven by gravitational resonances. It may account for some of the observed properties of dwarf spheroidals in the Local Group. Within this framework, dwarf spheroidals should form and interact in pairs, leaving detectable long stellar streams and tails.

In the widely accepted cold dark matter theory, bound systems form hierarchically through gravitational collapse. Because gravity acts in the same manner on all forms of mass, it is expected that galaxies should contain a ratio of dark to ordinary luminous material roughly similar to the cosmic mean. However, dwarf spheroidals have much larger dark to luminous mass ratios[1–3], requiring some mechanism to efficiently separate gas and stars from dark matter in order to explain their origin. Two classes of models have been proposed to account for the properties of dwarf spheroidals. In one class, gas is either prevented from collecting in shallow potential wells of dark matter by photoheating during cosmic reionization or is blown out subsequently by feedback from star formation before a significant mass in stars can accumulate[15,16]. However, there are no clear signatures of reionization in the star formation histories of these galaxies[12], which in many cases continued to form stars beyond the reionization epoch, and for most of them the dark halo masses are probably too large (10^8–$10^9 M_\odot$, where $M_\odot$ is the solar mass; refs 3, 9, 10) for supernova-driven winds to remove gas[17].

The other class of models relies on impulsive heating by gravitational tidal forces to strip stars and reshape the stellar component of a disk-dominated dwarf galaxy. Variously referred to as ''tidal stirring''[11] or ''tidal shocking''[18], this process operates when the gravitational potential acting on a small galaxy varies rapidly as it orbits in a larger system and can, in principle, convert a rotationally supported disk of stars into a spheroid. However, gas can radiatively dissipate energy; thus, scenarios based on gravitational heating require an additional process, that is, ram pressure stripping, to remove the gas[11]. Because gravitational heating and ram pressure stripping occur with varying efficiencies, models coupling them may rely on particular conditions to produce dwarf spheroidal galaxies.

This second class of models also requires that dwarfs orbit close to giant galaxies. Whereas several of the brightest dwarf spheroidals in the Local Group cluster around the large galaxies, some, like Cetus and Tucana, reside at great distances from both the Milky Way and Andromeda[12,19]. Furthermore, many nearby dwarf spheroidals lie in the same plane as the orbits of the Magellanic Clouds and the Magellanic Stream[20]. A thin planar distribution may be difficult to explain if dwarf galaxies fall in individually[21,22] but can arise naturally if dwarfs fall into larger systems as members of groups of dwarfs[23]. Examples of such groups have been found nearby[24]. Moreover, interactions between dwarfs should be common in these environments, especially in the past when the Universe was younger and more dense.

Here we describe numerical experiments to investigate the consequences of encounters between dwarf galaxies. Figure 1 shows the time evolution of an interaction between a pair of dwarf galaxies. This could represent an interaction between dwarfs in a small group or between a dwarf and the forming Milky Way at high redshift. After two billion years (2 Gyr), nearly ~80% of the stars are stripped away from the smaller dwarf but its surrounding dark matter halo is less strongly affected, leading to a change in the ratio of dark to luminous matter. This outcome is surprising, because the stars sit at the bottom of the potential well and therefore comprise the most tightly bound material in the galaxy.

This unexpected result is caused by a gravitational process, which we term 'resonant stripping'. The efficient removal of stars from the smaller dwarf over the course of the interaction shown in Fig. 1 is mediated by a resonance between the spin angular frequency (Ω_s) of its disk and the angular frequency of its orbit (Ω_0) around the larger dwarf. When these frequencies are comparable in magnitude and the spin and orbital angular momenta are somewhat aligned (that is, a prograde encounter), the gravitational perturbation from the larger dwarf on a patch of the smaller dwarf's disk is always directed outwards from its centre, leading to significant stripping. However, for a perfectly retrograde collision, stars in the disk are alternately pulled inward and outward with little net result. The minimal response from the dark matter is expected because these particles move on random orbits, and the net perturbation on the halo mostly averages out.

Although resonances have previously been invoked to produce bridges and tails in collisions between galaxies of comparable mass[25], our work shows that resonant stripping can alter the mass to light ratio of dwarfs by removing luminous material more efficiently than dark matter. In Fig. 2 the luminous fraction of the smaller dwarf is plotted as a function of time during its interaction with a galaxy 100 times its mass (filled circles), compared to another simulation designed to follow a similar dwarf orbiting about the Milky Way today (filled squares). Resonant stripping is most effective for galaxies interacting on prograde orbits, as in the example shown in the top panels in Fig. 1. However, our calculations further indicate that gravitational torques

[1]Institute for Theoretical Physics, University of Zurich, CH-8057 Zurich, Switzerland. [2]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS 51, Cambridge, Massachusetts 02138, USA.
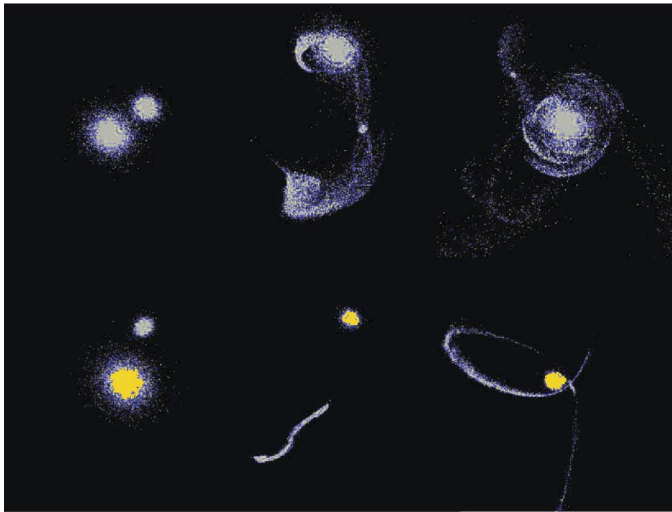
Figure 1 | Encounters between galaxies. Top row, interaction between a dwarf galaxy with a mass of $1.7 \times 10^8 \, M_\odot$ orbiting around a larger dwarf with 100 times its mass. Only the stellar components are plotted. Top left, the initial set-up where the two dwarfs approach one another on a somewhat prograde orbit (the disks are seen face-on). Top middle, the state of the system after 2 Gyr, following the first pericentric passage. Top right, the appearance of the galaxies after 7 Gyr. Each panel displays a region 100 kpc on a side. An outcome like the one illustrated in the top row occurs preferentially when one of the interacting galaxies is between 10 and 100 times more massive than the other one. If the galaxies have nearly the same mass they will merge quickly, masking the effects of resonant stripping because nearly all the luminous matter will remain bound to the remnant. Bottom row, the orbit of the same small galaxy (in white) around the Milky Way today (in yellow), which has 10,000 times its mass. Although the encounter is mostly prograde, the spin and orbital frequencies are no longer well-matched and the resonant response is suppressed. Bottom left, the initial set-up, displaying a region 150 kpc on a side. Bottom middle and bottom right panels show an expanded view 300 kpc on a side, and give the state of the system after 2 Gyr and 7 Gyr, respectively.

can alter the internal structure of the galaxies and the relative alignments of spin and orbital angular momenta, allowing the resonance to eventually act even for orbits that are initially somewhat retrograde (open circles in Fig. 2).

Resonant stripping is distinct from other processes proposed to drive galaxy evolution, such as mergers[13], galaxy–galaxy harassment[14] or more general heating processes, and tidal or ram pressure stripping. In particular, mechanisms that can be treated using the impulse approximation do not account for resonances because the particles in the perturbed system are assumed to remain roughly stationary over the course of the encounter. Because resonant stripping will affect gas and stars in a similar manner in a rotationally supported disk, it is simpler than models that require separate effects to strip the gas versus the stars.

Resonant stripping can drive the morphological evolution of dwarfs. When operating in low mass groups, this mechanism can pre-process dwarfs by transforming disk galaxies into spheroids before they are accreted by larger galaxies like the Milky Way. We find that dwarf spheroidal galaxies formed in this manner have properties similar to those of dwarfs observed in the Local Group. This is demonstrated in Figs 3 and 4, which show, respectively, that the final radial surface mass profiles and kinematic properties of a dwarf disk galaxy undergoing resonant stripping in our simulations are similar to those of observed dwarf spheroidals in the Local Group[26].

Our model makes definite predictions that can be tested in the future. In particular, resonant stripping should be visible *in situ* in associations of dwarfs. Unlike in previous theories, dwarf spheroidals are thus expected to be found with detectable stellar tails and shells, marking their formation. If this is indeed their dominant production mechanism, our model predicts that dwarf spheroidals should have similar properties in different environments, which is supported by the
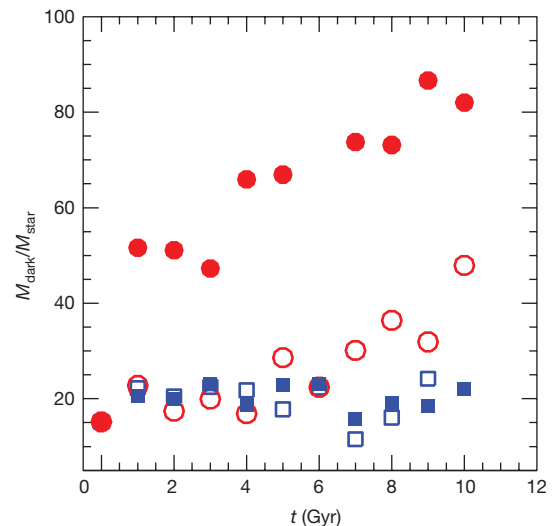

Figure 2 | The time evolution of the dark-mass to luminous-mass ratio. The dark-mass to luminous-mass ratio, M_{dark}/M_{star}, of the smaller dwarf being resonantly stripped is computed at the tidal radius and marked by the filled red circles. The same case but for a mostly retrograde encounter is illustrated by the open red circles. A resonance occurs if $\Omega_s = \Omega_0 \rightarrow \frac{v}{r} \approx \frac{V_0}{R_{peri}} \sqrt{(1+e)}$, where v is the rotation velocity, r is the size of the smaller dwarf, V_0 is the orbital velocity, R_{peri} the pericentric distance, and e the eccentricity of the orbit. Note that this resonance condition is not dependent on the specific choice of the pericentric distance alone but rather on the combination of the internal structure (for example, the rotation curve) of the small dwarf and the orbital parameters. In other words, if the pericentric distance changes, the resonance condition could still be satisfied provided that the disk rotation speed were modified accordingly. If the orbit is more retrograde, stars are not preferentially removed immediately. So, after 2 Gyr, the net change in M_{dark}/M_{star} is a factor of 4 larger for the prograde versus retrograde cases illustrated. However, after 4 Gyr the internal structure of the smaller dwarf and the orbit are affected by gravitational torques, allowing resonant stripping to occur. The ratio M_{dark}/M_{star} of the small galaxy orbiting about the Milky Way today is plotted for prograde (filled blue squares) and retrograde (open blue squares) orbits. In both these cases, the spin and orbital frequencies of the galaxies are no longer comparable and the resonant interaction is suppressed, even in the prograde case.
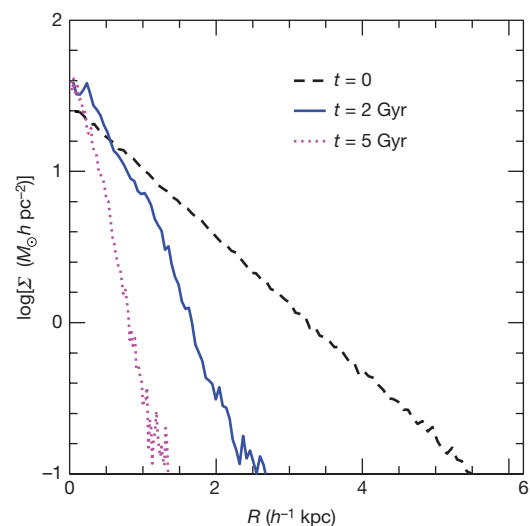

Figure 3 | Radial stellar surface mass density profile of the smaller dwarf. The profile is plotted for a prograde encounter at the initial time of the simulation (dashed black line), after the first pericentric passage (2 Gyr; solid blue line), and after 5 Gyr (magenta dotted line). The profile is initially an exponential distribution with effective radius, R_e, appropriate for dwarf disk galaxies. However, it evolves immediately into a more concentrated profile with $R_e \approx 0.5 h^{-1}$ kpc (blue line), and after 5 Gyr the disk is converted into a compact spheroid with a smaller effective radius.
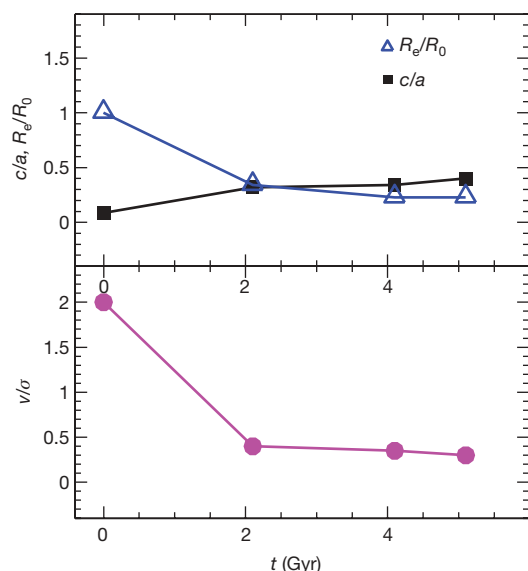
Figure 4 | The time evolution of the structural properties of the small dwarf galaxy. The morphological evolution of the small dwarf galaxy over the course of the prograde encounter is quantified at characteristic times (0, 2, 4 and 5 Gyr). Top panel, Evolution of the effective radius, R_e, normalized to its value at the beginning of the simulation (R_0), and marked by blue open triangles. The evolution of the minor- to major-axis ratio of the bound stellar component (c/a) measured at R_e is plotted as filled squares. The squares indicate that the initial stellar disk ($c = 0$) changes its shape immediately after 2 Gyr into a more spheroidal object ($c \approx 0.5$). This evolution is caused by the complete removal of the outer stellar disk via resonant stripping. Bottom panel, Evolution of the kinematic properties of the dwarf, measured in terms of the ratio of the centrifugal rotational velocity of the small dwarf, v, to its random motions, σ. The ratio v/σ (filled circles) evolves after the first pericentric passage (2 Gyr), indicating a change from a state where the rotational support at R_e is twice that from random motions to a value for v/σ of less than 0.5: such a low value is similar to that of dwarf spheroidals not only in the Local Group but also in galaxy clusters[26].

observed similarities between dwarf spheroidals in the Perseus cluster[27] and those in the Local Group.

The efficiency of resonant stripping depends on various factors, including the internal structure of the smaller dwarf (Supplementary Information). In the example shown in Fig. 1, the interacting dwarfs were chosen to have characteristics similar to those of local dwarf disk galaxies. However, the detailed properties of young low-mass galaxies formed at early cosmic times are unknown; here we are assuming they are cold disks. If the smaller dwarf in our simulations was set up so as to satisfy the condition for resonant stripping even more precisely and over a larger portion of the disk, it is possible that more luminous material could be stripped, depleting the stellar surface density in the inner regions more significantly than in the example shown in Fig. 3. In that event, resonant stripping might yield ultra-faint dwarf galaxies or even systems that are nearly entirely dark. This leads us to speculate that, when placed in a proper cosmological framework, resonant stripping might explain the missing satellite problem in the Local Group[28,29]. Moreover, spectroscopic measurements indicate that the Virgo cluster is also missing satellite galaxies[30]. Resonant stripping acting in low mass groups when the Universe was younger may have caused haloes to lose their gas and stars, pre-processing them before these groups were accreted into galaxy clusters, providing a similar evolutionary mechanism to reproduce the luminosity functions of both galaxy groups and clusters.

1. Kleyna, J. T., Wilkinson, M. I., Evans, N. W., Gilmore, G. & Frayn, C. Dark matter in dwarf spheroidals – II. Observations and modeling of Draco. *Mon. Not. R. Astron. Soc.* **330**, 778–791 (2002).
2. Chapman, S. *et al.* Keck DEIMOS kinematic study of Andromeda IX: dark matter on the smallest galactic scales. *Astrophys. J.* **632**, L87–L90 (2005).
3. Wilkinson, M. I. *et al.* Kinematically cold populations at large radii in the Draco and Ursa minor dwarf spheroidal galaxies. *Astrophys. J.* **611**, L21–L24 (2004).
4. Mateo, M. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).
5. Grebel, E. K. in *The Stellar Content of Local Group Galaxies* (eds Whitelock, P. & Cannon, R.) 17–38 (IAU Symposium 192, Astronomical Society of the Pacific, 1999).
6. Gallagher, J. S., Madsen, G. J., Reynolds, R. J., Grebel, E. K. & Smecker-Hane, T. A. A search for ionized gas in the Draco and Ursa Minor dwarf spheroidal galaxies. *Astrophys. J.* **588**, 326–330 (2003).
7. Willman, B. *et al.* A new Milky Way dwarf galaxy in Ursa Major. *Astrophys. J.* **626**, L85–L88 (2005).
8. Zucker, D. B. *et al.* A new Milky Way dwarf satellite in Canes Venatici. *Astrophys. J.* **643**, L103–L106 (2006).
9. Peñarrubia, J., Navarro, J. F. & McConnachie, A. The tidal evolution of Local Group dwarf spheroidals. *Astrophys. J.* **673**, 226–240 (2008).
10. Strigari, L. E. *et al.* A common mass scale for satellite galaxies of the Milky Way. *Nature* **454**, 1096–1097 (2008).
11. Mayer, L., Kazantzidis, S., Mastropietro, C. & Wadsley, J. Early gas stripping as the origin of the darkest galaxies in the Universe. *Nature* **445**, 738–740 (2007).
12. Grebel, E. K., Gallagher, J. S. III & Harbeck, D. The progenitors of dwarf spheroidal galaxies. *Astrophys. J.* **125**, 1926–1939 (2003).
13. Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (Yale University Observatory, 1977).
14. Moore, B., Katz, N., Lake, G., Dressler, A. & Oemler, A. Galaxy harassment and the evolution of clusters of galaxies. *Nature* **379**, 613–616 (1996).
15. Bullock, J., Kravtsov, A. & Weinberg, D. H. Reionization and the abundance of galactic satellites. *Astrophys. J.* **539**, 517–521 (2000).
16. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303**, 39–55 (1986).
17. Mac Low, M. M. & Ferrara, A. Starburst-driven mass loss from dwarf galaxies: efficiency and metal ejection. *Astrophys. J.* **513**, 142–155 (1999).
18. Gnedin, O. Y., Hernquist, L. & Ostriker, J. P. Tidal shocking by extended mass distributions. *Astrophys. J.* **514**, 109–118 (1999).
19. Sales, L., Navarro, J. F., Abadi, M. G. & Steinmetz, M. Cosmic menage a trois: the origin of satellite galaxies on extreme orbits. *Mon. Not. R. Astron. Soc.* **379**, 1475–1483 (2007).
20. Lynden-Bell, D. Dwarf galaxies and globular clusters in high velocity hydrogen streams. *Mon. Not. R. Astron. Soc.* **174**, 695–710 (1976).
21. Libeskind, N. I. *et al.* The distribution of satellite galaxies: the great pancake. *Mon. Not. R. Astron. Soc.* **363**, 146–152 (2005).
22. Zentner, A. R., Kravtsov, A. V., Gnedin, O. Y. & Klypin, A. A. The anisotropic distribution of galactic satellites. *Astrophys. J.* **629**, 219–232 (2005).
23. D'Onghia, E. & Lake, G. Small dwarf galaxies within larger dwarfs: why some are luminous while most go dark. *Astrophys. J.* **686**, L61–L64 (2008).
24. Tully, R. B. *et al.* Associations of dwarf galaxies. *Astron. J.* **132**, 729–748 (2006).
25. Toomre, A. & Toomre, J. Galactic bridges and tails. *Astrophys. J.* **686**, L61–L64 (1972).
26. Ferguson, H. C. & Binggeli, B. Dwarf elliptical galaxies. *Annu. Rev. Astron. Astrophys.* **6**, 67–122 (1994).
27. Penny, S., Conselice, C., De Rijcke, S. & Held, E. V. Hubble Space Telescope survey of the Perseus Cluster – I: The structure and dark matter content of cluster dwarf spheroidals. *Mon. Not. R. Astron. Soc.* **393**, 1054–1062 (2009).
28. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where are the missing galactic satellites? *Astrophys. J.* **522**, 82–92 (1999).
29. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
30. Rines, K. & Geller, M. J. Spectroscopic determination of the luminosity function in the galaxy clusters A2199 and Virgo. *Astrophys. J.* **135**, 1837–1848 (2008).

Author Contributions E.D. designed and led the study and analysis, ran the simulations, and wrote the paper; G.B. was involved in study design and contributed to text writing; T.J.C. designed the initial condition programs; and L.H. was involved in study design and contributed to text writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.D. (edonghia@cfa.harvard.edu).

LETTERS

Saturn's rotation period from its atmospheric planetary-wave configuration

P. L. Read[1], T. E. Dowling[2] & G. Schubert[3]

The rotation period of a gas giant's magnetic field (called the System III reference frame) is commonly used to infer its bulk rotation[1]. Saturn's dipole magnetic field is not tilted relative to its rotation axis (unlike Jupiter, Uranus and Neptune), so the surrogate measure of its long-wavelength (kilometric) radiation is currently used to fix the System III rotation period[2]. The period as measured now by the Cassini spacecraft is up to ~7 min longer[3] than the value of 10 h 39 min 24 s measured 28 years ago by Voyager[2]. Here we report a determination of Saturn's rotation period based on an analysis of potential vorticity. The resulting reference frame (which we call System IIIw) rotates with a period of 10 h 34 min 13 ± 20 s. This shifted reference frame is consistent with a pattern of alternating jets on Saturn that is more symmetrical between eastward and westward flow. This suggests that Saturn's winds are much more like those of Jupiter than hitherto believed[4].

Saturn's bulk rotation period is not expected to have changed appreciably over the course of the last few decades (the interior is nearly isentropic and there is no plausible way to alter the moment of inertia). A good surrogate of the rotation period should therefore likewise remain constant. The case for this role to be played by periodicities in Saturn's kilometric radiation (SKR) has been considerably weakened now that the Cassini spacecraft has found a value[3,5] of 10 h 47 min 6 s (here designated 'System IIIa'; Table 1), which is markedly different to that determined by Voyager. Moreover, this period appears to be changing[3] even during the short time interval that Cassini has been in orbit around Saturn. It seems clear, therefore, that the SKR periodicities are affected by factors that partially decouple the external magnetic field from the deep interior

and allow it to 'slip', possibly in association with a centrifugally driven instability in Saturn's plasma disk[3,6].

In a recent study[7], a different approach was used to estimate Saturn's rotation period by combining measurements of its gravity field with radio occultation and zonal wind measurements from the Pioneer and Voyager missions. This resulted in a value of 10 h 32 min 35 ± 13 s (hereafter referred to as 'System IIIb'; Table 1), which is faster than any of the earlier estimates, yields peak winds on the equator that are more than 100 m s^{-1} slower than previously considered, and yields alternating jets in the eastward and westward directions of more or less equal magnitude, much more like the pattern seen on Jupiter. However, this estimate is not yet widely accepted as the 'true' interior rotation period pending further observational or theoretical evidence.

Here we introduce an independent approach to the estimation of Jupiter's and Saturn's bulk rotation periods, based on considerations of their dynamic meteorology. From analyses of Voyager cloud-tracked winds on Jupiter, the distribution of zonally averaged potential vorticity, $\bar{q}$, appears to approach marginal stability[8–10] with respect to a nonlinear stability theorem due to Arnol'd[11,12]; in particular, Arnol'd's second stability theorem implies that perturbations will be stable if

$$\left(\frac{\bar{u}-\alpha}{\partial \bar{q}/\partial y}\right) \geq L_D^2 \qquad (1)$$

where $\bar{u}$ is the zonal-mean zonal velocity, y is the northward coordinate, α is a constant that corresponds to a shift in the reference frame, and L_D is an intrinsic length scale that may vary with latitude (the first baroclinic radius of deformation). Rossby waves travel westward or eastward relative to the flow depending on whether $\partial \bar{q}/\partial y \equiv \bar{q}_y$ is positive or negative, respectively; on Earth, $\bar{q}_y$ is dominated by the planetary vorticity gradient, known as β, which is positive in both hemispheres, thus Rossby waves generally travel westward relative to the flow. However, on Jupiter and Saturn, the curvature of their jet streams is much stronger than on Earth, and $\bar{q}_y$ is negative at many latitudes. The longest Rossby waves, which are among the most deeply rooted of atmospheric disturbances[13], are also the fastest, and in the marginally stable case they propagate upstream with speed $-L_D^2 \bar{q}_y$ relative to the flow such that they just cancel their advection by the wind[9]. Thus, α represents the speed of these waves in the chosen reference frame.

The tendency towards marginal stability in the sense of the limiting value of equation (1) has recently been demonstrated in three-dimensional numerical models of gas-giant atmospheres[14] (the abyssal circulation underlying a gas giant's spherical-shell atmosphere, or 'weather layer', may not be governed by equation (1) because of the effects of convection and full spherical geometry in the interior[15]). On the observational front, much improved zonal-mean maps of

Table 1 | Saturn's rotation periods relative to Voyager System III

Name of reference frame	Rotation period (h min s)	Frame zonal velocity at the equator in System III (m s^{-1})
System III[1]	10 39 24 ± 7	0
System IIIa[3]	10 47 06 ± 40	−123 ± 11
System IIIb[7]	10 32 35 ± 13	112 ± 4
System IIIw	10 34 13 ± 20	85 ± 6

To determine System IIIw, a cross-section of zonally averaged quasi-geostrophic potential vorticity in latitude and pressure (height) was derived from the Cassini geostrophic thermal wind and temperature measurements, as described fully elsewhere[16,17]. For vertical levels located in the upper troposphere, bands of latitude up to 10° wide and consistent with marginal stability according to Arnol'd's second stability theorem were identified from the condition that d$\bar{q}$/d$\bar{\Psi}$ < 0, where $\bar{\Psi}$ is the geostrophic streamfunction for the mean zonal flow (obtained from a latitudinal integral of the geostrophic zonal wind). Within each marginally stable latitude band so identified, a linear regression line between $\bar{u}$ and $\bar{q}_y$ of the form $\bar{u} = L_D^2 \bar{q}_y + \alpha$ was obtained by a least squares fit. This enabled a profile of $\alpha(\phi)$ to be obtained, evaluated at the centre of each latitude band for each pressure level, which could then be converted to an angular velocity profile $\omega(\phi)$ as described in the text, taking full account of the oblate spheroidal shape of the planet. An average value for the angular velocity of System IIIw was obtained from a weighted mean of $\omega(\alpha_i)$ over the range of pressures 150–300 hPa and all relevant latitudes, using weights $1/s_i^2$, where s_i is the standard regression error of $\omega(\alpha_i)$. The quoted uncertainty on the rotation period is an estimate of the 1σ weighted standard error on the mean.

[1]Department of Physics, University of Oxford, Oxford OX1 3PU, UK. [2]Department of Physics & Astronomy, University of Louisville, Louisville, Kentucky 40292, USA. [3]Department of Earth & Space Sciences, University of California, Los Angeles, California 90095-1567, USA.

potential vorticity (quasi-geostrophic and Ertel) have recently been derived[16,17] in the upper tropospheres and lower stratospheres of both Jupiter and Saturn, making use of both cloud-tracked velocities and retrievals of thermal structure and composition from the Voyager and Cassini missions. These results confirm that substantial bands of latitude on both planets are broadly consistent with the marginally stable form of equation (1) (with an equality). Moreover, equation (1) was used to deduce the variation of L_D with latitude from the gradient of the linear regression between $\bar{u}$ and $\bar{q}_y$. The primary advance here is the recognition that this same linear regression can independently solve for α itself, and hence the unique reference frame in which the longest Rossby waves are stationary.

Figure 1 compares the latitudinal variation of α on Jupiter and Saturn, derived from the quasi-geostrophic potential vorticity profiles obtained from Cassini data[16,17], relative to their traditional System III reference frames, converted to the respective angular velocity (that is, $\omega(\alpha) = \alpha/(a(\phi)\cos\phi)$, where $a(\phi)$ is the radius of the planet as measured from its centre, allowing for its ellipticity, at planetocentric latitude ϕ). The quasi-geostrophic theory underlying equation (1) is not valid near the equator; taking a variance-weighted average of $\omega(\alpha)$ across all latitudes $>15°$ from the equator, excluding the equatorial jet, and over pressure levels from 150 to 300 hPa, leads to a reference frame on Jupiter (hereafter denoted 'System IIIw') that differs in rotation period from Voyager's System III by just 18 ± 1 s, which is statistically consistent with only a very small shift (about $-6.6\,\text{m s}^{-1}$ at the equator). For Saturn, if we restrict the coverage to latitudes $>20°$ from the equator, also excluding the equatorial jet, the mean difference in period between System III and System IIIw is 311 ± 20 s. Although the formal uncertainty is higher than for the same analysis applied to Jupiter, this does suggest that dynamical stability considerations on Saturn favour a reference frame that rotates over 5 min faster than Voyager System III, which corresponds to over $80\,\text{m s}^{-1}$ faster at the equator (Table 1). When viewed in

this frame, Saturn's pattern of alternating zonal jets adopts a more symmetrical appearance (Fig. 2a), with more or less equal amplitude eastward and westward jets at middle and high latitudes. Such a symmetric jet pattern resembles the pattern on Jupiter much more closely than when viewed in Saturn's traditional System III frame (compare Fig. 2b).

The notion that an atmospheric flow might adjust towards a state of marginal dynamical stability is not new. Such a scenario has been suggested, for example, as an explanation for the static stability structure of the Earth's atmosphere through the process of 'baroclinic adjustment'[18]. In the case of Jupiter and Saturn, barotropic processes and instabilities are likely to be at least as important as baroclinic ones. This would lead to a dynamical equilibration between upscale cascade processes, forcing zonal jets towards an ever stronger, dynamically unstable structure, and the action of large-scale instabilities in restoring the flow towards a marginally stable structure[9,19,20]. A physical interpretation[8,9,21] of α associates it with the speed of the longest-wavelength Rossby wave for almost barotropic waves. This is also consistent with the theoretical finding[21] that marginally unstable Jovian barotropic Rossby waves are approximately stationary. Such waves may 'sense' planetary-scale disturbances deep in the interior[9], although we do not yet know the precise nature of such a teleconnection.

On the basis of the atmospheric planetary-wave configuration (potential vorticity and zonal wind) observed on Jupiter, we have deduced that its longest Rossby waves are frequently coherent across its adjacent alternating jets and almost stationary relative to its bulk planetary rotation period. To the extent that this System IIIw method is thus verified, we have been able to mitigate Saturn's lower signal-to-noise ratio (for both its magnetic field and atmosphere) by applying the same approach. The main result is that the zonal-wind profiles for Saturn and Jupiter now appear to be qualitatively similar. Accurate knowledge of Saturn's rotation rate is necessary not only to understand the dynamics of its atmosphere, but it is also required to infer the planet's interior structure and dynamical state, deduce whether it has a heavy element core, and constrain theories of its origin and evolution. The new rotation period obtained here, for example, would indicate[22] a mean density for the planet of $\rho_0 = 686.895\,\text{kg m}^{-3}$ and a value for its J_2 gravitational moment of

Figure 1 | Variation of limiting planetary wave angular velocity with latitude for Saturn and Jupiter. Shown are latitudinal profiles of the 'long wave' angular velocity, $\omega(\alpha) = \alpha/(a(\phi)\cos\phi)$ (where $a(\phi)$ is the planetary radius and ϕ is planetocentric latitude), and corresponding increment in rotation period for Saturn (open diamonds) and Jupiter (stars) from the corresponding System III at pressure levels of 268 hPa (Jupiter) and 236 hPa (Saturn). Error bars on the points are representative of the formal (1σ) uncertainties in the linear regression coefficient, and indicate some significant variation in the degree to which the relationship between $\bar{u}$ and $\bar{q}_y$ is well represented by the linear form $\bar{u} \approx L_D^2 \bar{q}_y + \alpha$. Where the error bars are small, this assumed relationship is a relatively good fit, lending some confidence to the hypothesis that winds in this latitude band are close to marginal stability. At some other latitudes, however, the fit is much poorer, indicating that an assumption of marginal stability may not be valid[16,17]. The latitudes where the fit is relatively poor also seem[16,17] to correspond with dynamically active regions, with an increased presence of large-scale wave activity or the formation of oval vortices and other disturbances. The dashed line represents the angular velocity of System IIIw for comparison.
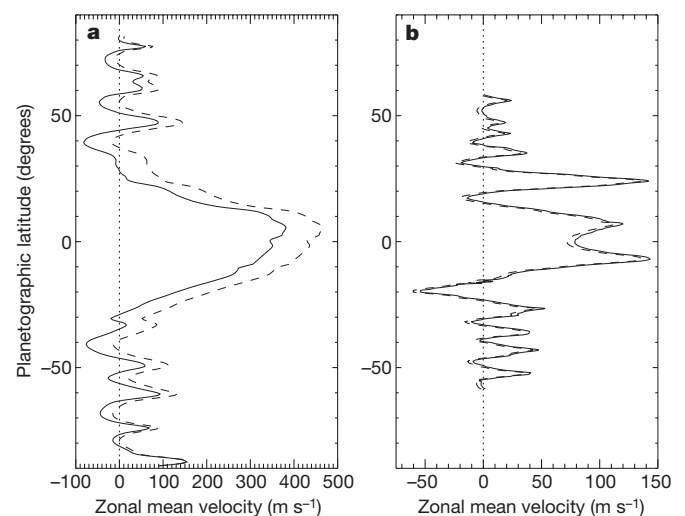
Figure 2 | Comparison of cloud-level zonal velocity profiles on Saturn and Jupiter as viewed in System IIIw and System III. a, b, Variations of cloud-level zonal velocity on Saturn (**a**) and Jupiter (**b**) with latitude, as viewed in the corresponding System IIIw reference frame (black solid line), and as viewed in the conventional Voyager System III frame (dashed line). Cloud-level velocities for Saturn were obtained from a combination of Voyager and Cassini data[17] whereas those for Jupiter are from the Cassini Imaging Sub-System (ISS)[23].

0.0162919, both of which may be amenable to accurate observational verification by future space missions.

1. Seidelman, P. K. & Devine, N. Evaluation of Jupiter's longitudes in System III (1965). *Geophys. Res. Lett.* **4**, 65–68 (1977).
2. Desch, M. D. & Kaiser, M. L. Voyager measurement of the rotation period of Saturn's magnetic field. *Geophys. Res. Lett.* **8**, 253–256 (1981).
3. Gurnett, D. A. *et al.* The variable rotation period of the inner region of Saturn's plasma disk. *Science* **316**, 442–445 (2007).
4. Aurnou, J. M. & Heimpel, M. H. Zonal jets in rotating convection with mixed mechanical boundary conditions. *Icarus* **169**, 492–498 (2004).
5. Giampieri, G., Dougherty, M. K., Smith, E. J. & Russell, C. T. A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* **441**, 62–64 (2006).
6. Goldreich, P. & Farmer, A. J. Spontaneous axisymmetry breaking of the external magnetic field at Saturn. *J. Geophys. Res.* **112**, A05225, doi:10.1029/2006JA012163 (2007).
7. Anderson, J. D. & Schubert, G. Saturn's gravitational field, internal rotation, and interior structure. *Science* **317**, 1384–1387 (2007).
8. Dowling, T. E. A relationship between potential vorticity and zonal wind on Jupiter. *J. Atmos. Sci.* **50**, 14–22 (1993).
9. Dowling, T. E. Dynamics of Jovian atmospheres. *Annu. Rev. Fluid Mech.* **27**, 293–334 (1995).
10. Dowling, T. E. Estimate of Jupiter's deep zonal wind profile from Shoemaker-Levy 9 data and Arnol'd's second stability criterion. *Icarus* **117**, 439–442 (1995).
11. Arnol'd, V. I. On an a priori estimate in the theory of hydrodynamical stability. [In Russian.] *Izv. Vyssh. Ucheb. Zaved. Matematika* **54**, 3–5 (1966).
12. McIntyre, M. E. & Shepherd, T. G. An exact local conservation theorem for finite-amplitude disturbances to non-parallel shear flows with remarks on Hamiltonian structure and on Arnol'd stability theorems. *J. Fluid Mech.* **181**, 527–565 (1987).
13. Charney, J. G. & Drazin, P. G. Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.* **66**, 83–109 (1961).
14. Lian, Y. & Showman, A. P. Deep jets on gas-giant planets. *Icarus* **194**, 597–615 (2008).
15. Ingersoll, A. P. & Pollard, D. Motion in the interiors and atmospheres of Jupiter and Saturn: scale analysis, anelastic equations, barotropic stability criterion. *Icarus* **52**, 62–80 (1982).
16. Read, P. L. *et al.* Mapping potential-vorticity dynamics on Jupiter. I: Zonal-mean circulation from Cassini and Voyager 1 data. *Q. J. R. Meteorol. Soc.* **132**, 1577–1603 (2006).
17. Read, P. L., Conrath, B. J., Fletcher, L. N., Gierasch, P. J. & Simon-Miller, A. A. Mapping potential vorticity dynamics on Saturn: zonal mean circulation from Cassini and Voyager data. *Planet. Space. Sci.* (in the press).
18. Stone, P. H. Baroclinic adjustment. *J. Atmos. Sci.* **35**, 561–571 (1978).
19. Vasavada, A. R. & Showman, A. P. Jovian atmospheric dynamics: an update after Galileo and Cassini. *Rev. Mod. Phys.* **68**, 1935–1996 (2005).
20. Read, P. L. *et al.* Dynamics of convectively driven banded jets in the laboratory. *J. Atmos. Sci.* **64**, 4031–4052 (2007).
21. Stamp, A. P. & Dowling, T. E. Jupiter's winds and Arnol'd's second stability theorem: slowly moving waves and neutral stability. *J. Geophys. Res.* **98**, 18847–18855 (1993).
22. Helled, R., Schubert, G. & Anderson, J. D. Empirical models of pressure and density in Saturn's interior: implications for the helium concentration, its depth dependence, and Saturn's precession rate. *Icarus* **199**, 368–377 (2009).
23. Porco, C. C. *et al.* Cassini imaging of Jupiter's atmosphere, satellites, and rings. *Science* **299**, 1541–1547 (2003).

Author Contributions P.L.R. obtained and processed the Cassini data, deriving potential vorticity and $\omega(\alpha)$ profiles, and conducted the statistical analysis. T.E.D. made the original suggestion to investigate $\alpha(\phi)$ and contributed to the interpretation of the results. G.S. provided additional insights into Saturn's rotation, and all authors contributed to the text and discussion of results.

LETTERS

A 'granocentric' model for random packing of jammed emulsions

Maxime Clusel[1]*, Eric I. Corwin[1]*, Alexander O. N. Siemens[1] & Jasna Brujić[1]

Packing problems are ubiquitous[1,2], ranging from oil extraction through porous rocks to grain storage in silos and the compaction of pharmaceutical powders into tablets. At a given density, particulate systems pack into a mechanically stable and amorphous jammed state[3,4]. Previous theoretical studies have explored a connection between this jammed state and the glass transition[4–8], the thermodynamics of jamming[9–12] and geometric modelling of random packings[13–15]. Nevertheless, a simple underlying mechanism for the random assembly of athermal particles, analogous to crystalline ordering, remains unknown. Here we use three-dimensional measurements of packings of polydisperse emulsion droplets to build a simple statistical model in which the complexity of the global packing is distilled into a local stochastic process. From the perspective of a single particle, the packing problem is reduced to the random formation of nearest neighbours, followed by a choice of contacts among them. The two key parameters in the model—the available space around a particle and the ratio of contacts to neighbours—are directly obtained from experiments. We demonstrate that this 'granocentric' view captures the properties of the polydisperse emulsion packing—ranging from the microscopic distributions of nearest neighbours and contacts, to local density fluctuations, to the global packing density. Application of our results to monodisperse and bidisperse systems produces quantitative agreement with previously measured trends in global density[16]. Our model therefore reveals a general principle of organization for random packing and may provide the foundations for a theory of jammed matter.

Previous studies of the jammed state include characterizations of the network of forces[7,17–20], the packing microstructure[10–12,21,22], and effects of particle shape[23] and packing history[24]. In the following experiments, we visualize three-dimensional random packings of frictionless emulsion droplets using confocal microscopy to characterize the geometry and connectivity of the packing. Figure 1a presents an image of a fluorescently labelled oil-in-water emulsion, creamed under gravity to form a mechanically stable random pack (Methods). The confocal data are analysed using a deconvolution technique, which extracts the radius and position of each droplet with an accuracy of 1% of the average particle size (Methods). A typical probability distribution of radii $P(r)$, shown in Fig. 1b, exhibits a width of 23% of the mean radius. To characterize the local neighbourhood of each particle, we tessellate the packing using the navigation map[25], an extension of the Voronoi map to polydisperse systems. Each point in space is allocated to the particle whose surface is closest (Fig. 1c). Two particles are said to be neighbours if their corresponding cells share a common interface in the navigation map. Each cell in Fig. 1c is coloured according to the number of neighbours n around the central particle, which ranges from 4 to 30. This wide range arises from the polydispersity of the sample and stands in contrast to the narrow range of nearest neighbours, typically from 12 to 17, of monodisperse systems[26].

A subset of the neighbours is in contact with the central particle and therefore capable of transmitting forces. The resulting network of forces gives rise to the mechanical stability of the packing. In the confocal images, points of contact between particles are self-consistently determined by the geometric overlap of spheres that reconstruct the particles and an intensity enhancement of the fluorescence of Nile Red dye[22] (Fig. 1a). We measure the mean number of contacts, also known as the coordination number, to be $\langle z \rangle = 6.3 \pm 0.3$, in good agreement with $\langle z_{\text{iso}} \rangle = 6$ required for isostatic mechanical equilibrium[27]. A slight discrepancy may be expected owing to the small deformation of the droplets.

While the global constraint of isostaticity is satisfied, the number of neighbours n and contacts z around each particle fluctuates significantly within the packing. Figure 1d shows that for each n, the number of contacts z can take any value between 0 and n, which suggests that z is randomly distributed. In addition, for particles with more neighbours the most likely number of contacts increases, indicated by the colour-coded probability map in Fig. 1d. Moreover, both n and z increase with the radius r of the central particle, as shown in the scatter plot in Fig. 1e. This makes sense, as larger particles have more surface area available on which to fit neighbours.

To understand these observations, we consider the packing problem from the granocentric point of view of a single particle in the bulk, exemplified by the particle marked by a star in the confocal image in Fig. 1a. We propose a granocentric model for random packing, initially achieved by the formation of a set of neighbours, followed by the creation of contacts. This model allows for a definition of a local cell, giving access to the local packing fraction.

The starred particle in Fig. 1a is surrounded by 13 nearest neighbours. The space that each neighbour occupies depends on its size relative to the central particle. The space occupied by each neighbour around a central particle is characterized by the solid angle ω it subtends (Fig. 2a). Using this geometric description, we remove a trivial dependence of local packing on the central particle radius, which represents an important simplification. As shown in Fig. 2b, the packing of the neighbours precludes the addition of another neighbouring particle because there is a limited amount of available solid angle around the central particle, of radius r_c. This suggests an upper limit $\Omega_{\text{max}}(r_c)$ on the available solid angle.

We therefore model the formation of a set of n nearest neighbours around a central particle of radius r_c as a random selection of neighbours and a summation of their respective solid angles up to a threshold $\Omega_{\text{max}}(r_c)$. It follows that n is the number of particles needed to reach, without exceeding, $\Omega_{\text{max}}(r_c)$. Following ref. 14, we ignore all neighbour–neighbour correlations by neglecting relative positions of neighbours on the surface of the central particle. The sizes of each

[1]Center for Soft Matter Research, Department of Physics, New York University, 4 Washington Place, New York, New York 10003, USA.
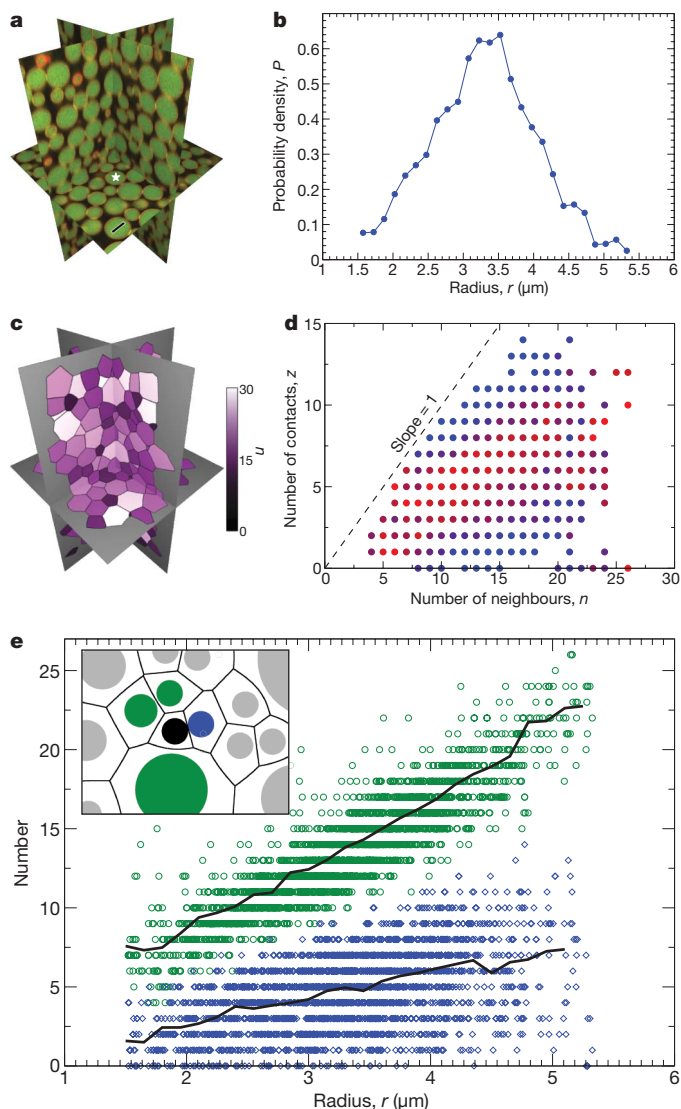*These authors contributed equally to this work.

Figure 1 | Microstructure of random polydisperse emulsions. a, Projection of a three-dimensional confocal image of a creamed emulsion shows the emission fluorescence of droplets in green (505–570 nm) and the droplet contacts in yellow (575–640 nm). Scale bar, 5 μm. **b,** The distribution of droplet radii in this emulsion exhibits a mean of 3.3 μm and a standard deviation of 0.74 μm. **c,** Tessellation of the image in **a** according to the navigation map defines a local cell for each droplet, while the colour map indicates the number of nearest neighbours n. **d,** For a given n, the number of contacting neighbours z is shown as a scatter plot. The colour range from blue to red indicates an increasing probability of finding a particle with z contacts. **e,** The scatter plot shows the fluctuations in the number of contacts z (blue diamonds) and neighbours n (green circles) with droplet size, while their average values are shown to increase almost linearly (black lines). Inset, diagram illustrating the contacting neighbours of a droplet (blue) and those sharing an interface (green).

successive neighbour are therefore statistically independent and identically distributed. The formation of a set of neighbours thus corresponds to the first passage of a directed one-dimensional random walk, with steps given by the solid angles of the nearest neighbours (Fig. 2c).

In order to describe all the properties of a packing, knowledge of nearest neighbours must be supplemented by knowledge of the contact network responsible for mechanical stability. As noted in Fig. 1d, only a fraction of the n nearest neighbours are in contact with the central particle. The fluctuations in z shown in Fig. 1e suggest that contacts are randomly chosen among neighbours. We model the selection of contacts by Bernoulli trials with success rate $p(r_c)$,

thereby neglecting all possible collective behaviour. Thus the distribution of the number of contacts z for a given number of neighbours n, $P(z|n)$, is a binomial distribution.

We have introduced two processes determining the connectivity of experimental packings. They are further characterized by fluctuations in the local packing fraction[11], which is relevant to global properties such as permeability and yield stress. The local packing fraction ϕ_{local} is defined as the ratio of the particle volume $V_{particle}$ to the cell volume V_{cell}. As our granocentric model does not include information on neighbour positions, it is not possible to use the experimental definition of a cell in this framework. Instead, we use an effective definition for the local cell incorporating two salient physical features of the experimental cell: first, a particle with more neighbours will have a higher ϕ_{local}; second, for a given number of neighbours a particle with more contacts will have a higher ϕ_{local}. Therefore, we approximate the cell volume as the sum of the volume of the central particle plus the volumes contributed by a portion of space between the central particle and each of its neighbours (see Fig. 3c inset; a more complete description is given in Methods). The differing contribution to the cell volume by the neighbours and contacts is described by an effective surface-to-surface distance δ, for non-contacting neighbours.

Thus, the statistical model reduces the packing problem complexity to two independent random processes at the single-particle level: first, the formation of a set of nearest neighbours is effected by assuming that the neighbours are chosen independently; second, the selection of contacts is assumed to be independent. The local definition of a cell provides further access to local packing fraction fluctuations. Whereas previous geometric models have predicted the average coordination number and density of discrete multicomponent systems[14,15], by mapping the packing problem onto the first passage of a random walk we are able to analytically study continuous distributions of radii and their influence on the full distributions of the number of nearest neighbours, coordination number and local density. Importantly, our model exploits the previously neglected observation that not all particles touch their neighbours.

Overall, the model introduces three parameters, $\Omega_{max}(r_c)$, $p(r_c)$ and δ, with different physical interpretations. The central parameter of our model, $\Omega_{max}(r_c)$, is the available solid angle around a particle of radius r_c. It is an effective parameter whose value encompasses the details of the packing structure, such as polydispersity, steric effects, and long range correlations. For example, the available solid angle of a face-centred-cubic crystal is 3.2π owing to the space between neighbours, whereas the total solid angle around a sphere is 4π. As described in Methods, the tools of random walks yield a functional relationship between the average number of neighbours around particles of radius r_c and the threshold $\Omega_{max}(r_c)$. Using the experimental average value of n for each radius r_c presented in Fig. 1e, the model estimates $\Omega_{max}(r_c)$. Remarkably, the experimentally determined values of $\Omega_{max}(r_c)$ are independent of r_c to within $\pm0.15\pi$ around a value of $\sim3.68\pi$. Thus, the ratio of the occupied surface area to the total surface area of the central particle does not depend on the size of the central particle in this polydisperse sample. The value of 3.68π is reasonable because it implies that there are unfilled spaces between the neighbours, observed in the example of Fig. 2b. The interpretation of Ω_{max} as an upper limit is tested by directly measuring whether the total solid angle Ω_{tot} around each particle lies below Ω_{max}. Figure 2d shows that 98% of the measured values are bounded by the estimated value of Ω_{max}, consistent with the proposed model assumptions.

To select contacts among neighbours, we introduced the success rate $p(r_c) = \langle z|r_c\rangle/\langle n|r_c\rangle$. There is no reason to suspect that this probability $p(r_c)$ should be independent of particle size. However, our measurements in Fig. 2e reveal that this average fraction of neighbours in contact is always $p(r_c) \approx 0.41$. This independence represents an important simplification and enhances the appeal of the model. We test the assumption that contacts are chosen independently by
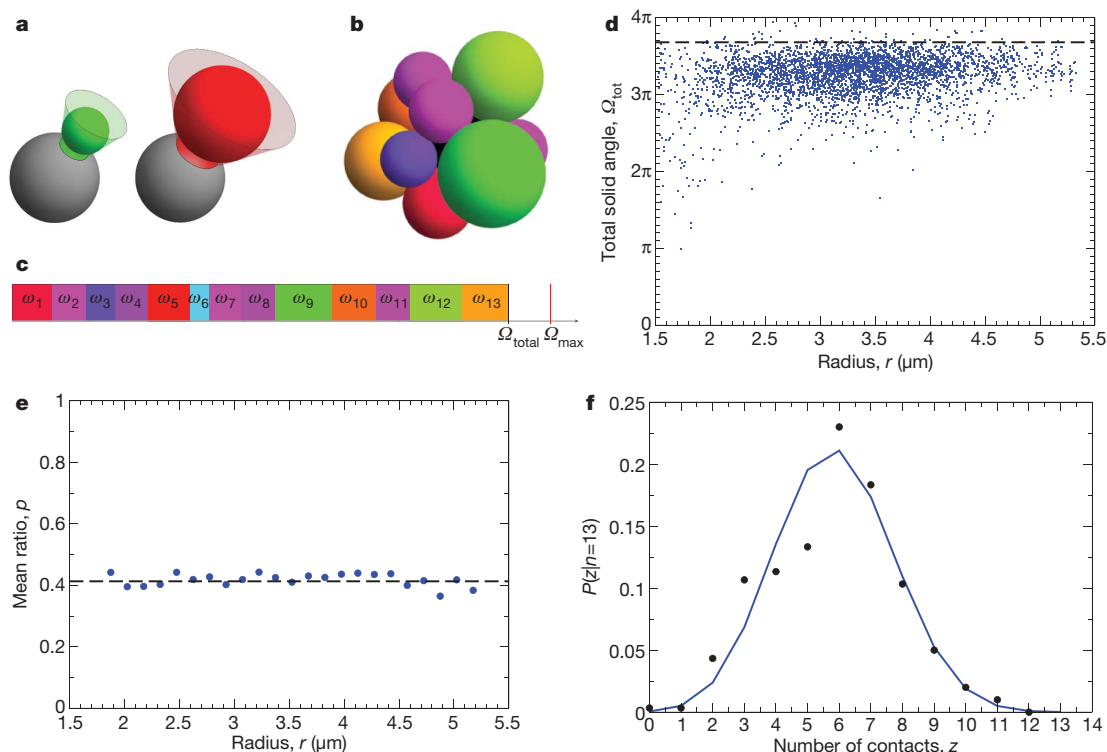
Figure 2 | Granocentric view of random packing. a, The space occupied by a neighbour around a central particle is measured by the solid angle it subtends, which is shown to depend on the neighbour size (green, smaller; red, larger). **b,** Nearest neighbours of the starred particle in Fig. 1a cover the surface of the central particle. **c,** Space-filling around the central particle in **b** is represented by a sum of the colour-coded solid angles ω occupied by each neighbour. **d,** A scatter plot of the total solid angle Ω_{tot} for each particle in the packing shows no dependence on the radius r of the central particle. Consistent with the model prediction, almost all the points lie below the maximal solid angle $\Omega_{max} = 3.68\pi$, shown by the dashed line. **e,** Mean ratio of the number of contacts to neighbours, $p = \langle z/n \rangle$, is shown to be independent of r. **f,** For all particles with $n = 13$, the distribution of contacts $P(z|n)$ is consistent with a binomial distribution with a probability $p = 0.41$.

examining the experimental distribution of z for a given n. As seen in Fig. 2f for $n = 13$, this distribution is captured by a binomial distribution with $p = 0.41$.

The final parameter in our model characterizes the local cell definition. The mean radius is a natural length scale in our system, which we use to quantify the effective distance δ. We introduce a dimensionless parameter α such that $\delta = \alpha\langle r \rangle$, assumed to be the same for every non-contacting particle. We choose $\alpha = 0.30$ to match the experimentally measured mean local packing fraction. Thus, gaps between the neighbours are significantly smaller than the average particle size, consistent with the notion that one cannot fit further particles between a particle and its neighbour.

Using this framework, we calculate distributions of n, z and ϕ_{local}, and compare them with experimental results to test our model. Using the equations presented in Methods, we demonstrate in Fig. 3c that the predicted distributions are in excellent agreement with the experimental data. Furthermore, the full distributions as a function of the radius of the central particle are also shown to be in agreement (Supplementary Figs A–F). These results validate the model as a tool to predict the effects of polydispersity on random packings. As shown in Fig. 3b, both the experiments and the model show that 18% of particles are mechanically unstable, with fewer than four contacts. These particles, known as 'rattlers', arise naturally from the random packing processes of our model, in contrast to existing models that exclude rattlers and only focus on the network of contacts[24]. The agreement between model and experiment for local quantities n, z and ϕ_{local} shows that the model quantitatively captures the local packing structure. The applicability of the model to packings with other size distributions is presented in Supplementary Information Section 4.0 and Supplementary Figs G–I. Numerical simulations show that the range of values for each parameter is narrow, but that the parameters are not universal for all polydispersities.

We can extend our local view of packing to predict the global density, a long-standing question in understanding random close packing[11,12,14,24,28,29]. The influence of the particle size distribution on global density has many industrial applications, such as predicting the density of dried paint or the density of porous rocks. The global packing fraction is defined as the ratio of the total volume of matter divided by the total volume of the sample. This is translated into our local model as:

$$\phi_{global} = \frac{\langle V_{particle} \rangle}{\langle V_{cell} \rangle} \quad (1)$$

It is important to note that the model value for the global packing fraction is relatively insensitive to the local fluctuations[14] and depends strongly only on α. Using the model definition of the local cell volume, we predict the global density of our packings to within 0.5% of the experimentally measured 66.4%. It is not surprising that this polydisperse packing is denser than its monodisperse counterpart ($\phi_{global} \approx 64\%$)[24], as small particles can pack in the interstices of larger ones.

We further probe the predictive power of our model for systems with other size distributions by considering bidisperse packings. In such packings, polydispersity is a function of the size ratio of the two species and the volume ratio of all small particles to all particles. Using the local cell definition with the same parameters as for our experimental system, the model reproduces published experimental data[16] for different volume fractions of small particles and size ratios, even in the limit of the monodisperse case (Fig. 3d). This demonstrates that a local model based on geometric considerations alone provides an effective description of the packing as a whole.

The model accurately characterizes random, polydisperse packings in terms of numbers of nearest neighbours and contacts per particle,
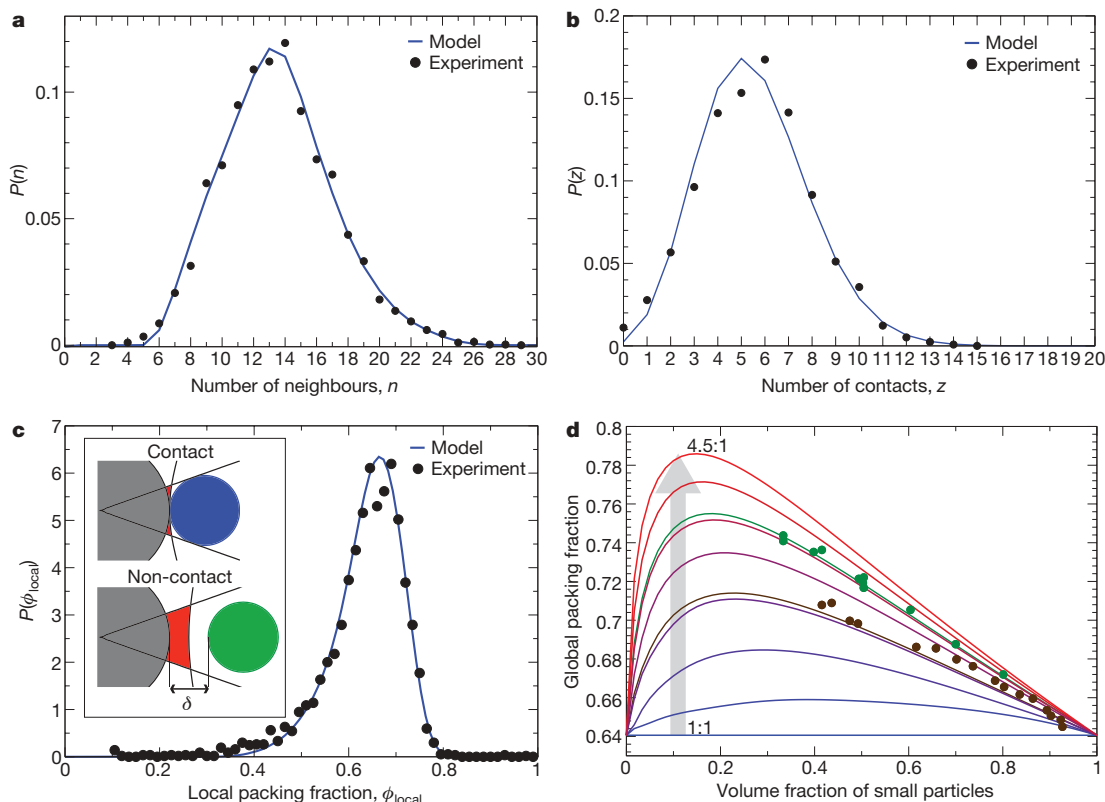
Figure 3 | Comparison of model predictions with experiments. a, b, The probability densities of the number of neighbours (**a**) and contacts (**b**) show excellent agreement between the experimental data and the statistical model. **c,** The probability density of the local packing fraction from the model successfully captures that obtained from the navigation map in Fig. 1c. Inset, volume contributions to the cell around each particle in the model. **d,** Model predictions for the global packing fraction of bidisperse systems are in good agreement with the experimental data reported[16] for size ratios of 2.58:1 (brown) and 3.41:1 (green). The lighter curves represent model predictions for size ratios ranging from 1:1 to 4.5:1 in increments of 0.5.

fluctuations in local density, and global density of the system. Our experiments reveal that the model parameters, Ω_{max} and p, are independent of the central particle size, leading to a simple physical interpretation of the mechanism of random packing. From the perspective of each particle, a stable random packing is achieved when (1) the surrounding space is filled with a set of nearest neighbours and (2) some of those neighbours touch the central particle to achieve mechanical equilibrium. This model offers a simplified view of the packing problem, and opens an analytical path to explore the industrially important effects of polydispersity on local and global properties of packings.

It is surprising that such a simple model should describe the random packing of polydisperse spheres. For monodisperse packings, the only source of disorder is positional. Such a system is an archetypal example of a complex system, where correlations between particle positions determine the physical behaviour. In our system, the polydispersity serves as a second source of randomness added to the positional disorder. Our model is able to describe a polydisperse packing because the local source of randomness coming from the distribution of radii, $P(r)$, dominates the positional disorder.

The granocentric model can be used to count the number of equivalent local configurations in disordered packings, and thus may provide a definition for entropy. Thus, we may probe the contention that jammed packings can be described using a thermodynamic approach[9]. Other open questions that extensions to the model could answer include the effects of spatial dimensions and particle shape on packings. We have so far considered purely random packings, but it would be interesting to examine how correlations influence the microstructure, both in the model and in the experiments. This work represents an effective medium model for random close packing, while the determination of model parameters from first principles may provide a route to a complete theory.

METHODS SUMMARY

Mathematical methods. The assumption that successive steps in the random walk are independent leads us to use the Laplace transform $\mathcal{L}$. If X and Y are two independent random variables of respective probability density P_X and P_Y, then the probability density of the sum P_{X+Y} satisfies $\mathcal{L}[P_{X+Y}] = \mathcal{L}[P_X]\,\mathcal{L}[P_Y]$. This basic observation, combined with standard probability tools, leads to the main results of this Letter.

Let $\omega = 2\pi\left(1 - \dfrac{1}{1+r/r_c}\sqrt{1+\dfrac{2r}{r_c}}\right)$ be the solid angle subtended by a particle of radius r on a given central particle of radius r_c, where r is drawn from the distribution $P(r)$. We compute $f_{r_c}(\omega)$, the probability density of solid angle around the central particle, by a change of variables from $P(r)$. The mean number of neighbours $\langle n|r_c\rangle$ around the central particle of radius r_c is then given by

$$\langle n|r_c\rangle = \mathcal{L}^{-1}\left[\frac{1}{s}\frac{\mathcal{L}[f_{r_c}](s)}{1-\mathcal{L}[f_{r_c}](s)}\right](\Omega_{max})$$

where $\mathcal{L}^{-1}$ is the inverse Laplace transform with respect to s, the conjugate variable of Ω_{max}. The probability density for a particle of radius r_c to have n neighbours given the maximal solid angle Ω_{max}, $P_{neighbour}(n;r_c,\Omega_{max})$, is given by:

$$P_{neighbour}(n;r_c,\Omega_{max}) = \mathcal{L}^{-1}\left[\frac{1-\mathcal{L}[f_{r_c}](s)}{s}\mathcal{L}[f_{r_c}](s)^n\right](\Omega_{max})$$

Likewise, the distribution of coordination number z can be computed as:

$$P_{contact}(z;r_c,\Omega_{max}) = \mathcal{L}^{-1}\left[\frac{1-\hat{f}_{r_c}(s)}{s}\frac{\left(p\,\hat{f}_{r_c}(s)\right)^z}{\left(1-(1-p)\hat{f}_{r_c}(s)\right)^{z+1}}\right](\Omega_{max})$$

Definition of local cell. The volume of a cell, V_{cell}, is defined as

$$V_{cell} = V_{particle} + \sum_{j=1}^{z} v_j + \sum_{j=1}^{n-z} v_j^*$$

where v_j is the volume contribution of the jth contacting neighbour and v_j^* is the volume contribution of the jth non-contacting neighbour. Let $\mathcal{C}$ be the cone subtended on the central particle by the neighbour particle in contact, and $\mathcal{S}$ be the surface of the central particle.

For contacting particles, let $\mathcal{H}$ be the surface of the hyperboloid defined by the navigation map of these two particles with surfaces in contact. The volume v is defined as the volume of a region between the central particle and the neighbour that is the portion of $\mathcal{C}$ between $\mathcal{S}$ and $\mathcal{H}$.

For non-contacting particles, let $\mathcal{H}'$ be the surface of the hyperboloid defined by the navigation map of these two particles with surfaces separated by a distance δ. The volume v^* is then the portion of $\mathcal{C}$ between $\mathcal{S}$ and $\mathcal{H}'$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids, and gases. *Rev. Mod. Phys.* **68,** 1259–1273 (1996).
2. de Gennes, P. G. Granular matter: a tentative view. *Rev. Mod. Phys.* **71,** S374–S382 (1999).
3. Cates, M. E., Wittmer, J. P., Bouchaud, J.-P. & Claudin, P. Jamming, force chains, and fragile matter. *Phys. Rev. Lett.* **81,** 1841–1844 (1998).
4. Liu, A. J. & Nagel, S. R. Nonlinear dynamics: jamming is not just cool anymore. *Nature* **396,** 21–22 (1998).
5. Barrat, A., Kurchan, J., Loreto, V. & Sellitto, M. Edwards' measures for powders and glasses. *Phys. Rev. Lett.* **85,** 5034–5037 (2000).
6. Coniglio, A., Fierro, A., Herrmann, H. J. & Nicodemi, M. (eds) *Unifying Concepts in Granular Media and Glasses: From the Statistical Mechanics of Granular Media to the Theory of Jamming* (Elsevier Science, 2004).
7. Corwin, E. I., Jaeger, H. M. & Nagel, S. R. Structural signature of jamming in granular media. *Nature* **435,** 1075–1078 (2005).
8. Parisi, G. & Zamponi, F. Mean field theory of the glass transition and jamming of hard spheres. Preprint at ⟨http://arXiv.org/abs/0802.2180⟩ (2008).
9. Edwards, S. & Oakeshott, R. Theory of powders. *Physica A* **157,** 1080–1090 (1989).
10. Lechenault, F., da Cruz, F., Dauchot, O. & Bertin, E. Free volume distributions and compactivity measurement in bidimensional granular packing. *J. Stat. Mech. Theory Exp.* P07009 (2006).
11. Aste, T. Volume fluctuations and geometrical constraints in granular packs. *Phys. Rev. Lett.* **96,** 018002 (2006).
12. Song, C., Wang, P. & Makse, H. A. A phase diagram for jammed matter. *Nature* **453,** 629–632 (2008).
13. Gotoh, K. & Finney, J. Statistical geometrical approach to random packing density of equal spheres. *Nature* **252,** 202–205 (1974).
14. Dodds, J. Simplest statistical geometric model of simplest version of multicomponent random packing problem. *Nature* **256,** 187–189 (1975).
15. Dodds, J. The porosity and contact points in multicomponent random sphere packings calculated by a simple statistical geometric model. *J. Colloid. Interface Sci.* **77,** 317–327 (1980).
16. Yerazunis, S., Cornell, S. & Wintner, B. Dense random packing of binary mixtures of spheres. *Nature* **207,** 835–837 (1965).
17. Liu, C. H. *et al.* Force fluctuations in bead packs. *Science* **269,** 513–515 (1995).
18. Brujić, J. *et al.* 3d bulk measurements of the force distribution in a compressed emulsion system. *Faraday Discuss.* **123,** 207–220 (2003).
19. Majmudar, T. & Behringer, R. Contact force measurements and stress-induced anisotropy in granular materials. *Nature* **435,** 1079–1082 (2005).
20. Zhou, J., Long, S., Wang, Q. & Dinsmore, A. Measurement of forces inside a three-dimensional pile of frictionless droplets. *Science* **312,** 1631–1633 (2006).
21. Bernal, J. Geometrical approach to the structure of liquids. *Nature* **183,** 141–147 (1959).
22. Brujić, J. *et al.* Measuring the coordination number and entropy of a 3d jammed emulsion packing by confocal microscopy. *Phys. Rev. Lett.* **98,** 248001 (2007).
23. Donev, A. *et al.* Improving the density of jammed disordered packing using ellipsoids. *Science* **303,** 990–993 (2004).
24. Torquato, S., Truskett, T. M. & Debenedetti, P. G. Is random close packing of spheres well defined? *Phys. Rev. Lett.* **84,** 2064–2067 (2000).
25. Richard, P., Oger, L. & Troadec, J.-P. A model of binary assemblies of spheres. *Eur. Phys. J. E* **6,** 295–303 (2001).
26. Finney, J. Random packings and structure of simple liquids. *Proc. R. Soc. Lond. A* **319,** 479–493 (1970).
27. Alexander, S. Amorphous solids: their structure, lattice dynamics and elasticity. *Phys. Rep.* **296,** 65–236 (1998).
28. Kamien, R. & Liu, A. Why is random close packing reproducible? *Phys. Rev. Lett.* **99,** 155501 (2007).
29. Aste, T., Saadatfar, M. & Senden, T. Local and global relations between the number of contacts and density in monodisperse sphere packs. *J. Stat. Mech. Theory Exp.* P07010 (2006).
30. Schmitt, V., Leal-Calderon, F. & Bibette, J. in *Topics in Current Chemistry* Vol. 227, *Colloid Chemistry II* (ed. Antonietti, M.) 195–215 (Springer, 2003).

Author Contributions M.C. and E.I.C. contributed equally to this Letter and are listed in alphabetical order.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B. (jb2929@nyu.edu).

METHODS

Emulsion preparation and imaging. The oil-in-water emulsion is prepared by mixing 3 vol.% Tergitol NP7 (nonyl phenolethoxy 7, Sigma) non-ionic surfactant with 3 wt% high viscosity Alginate HF120L (Promova) in 30 ml deionized water (Millipore), after the method of ref. 30. The disperse phase is composed of silicone oil (PDMS) droplets with varying viscosities, fluorescently labelled by Nile Red dye[18]. This mixture is injected into a narrow gap Couette mixer with a gap size of 100 μm and sheared between 18–22 r.p.m. A 16 mM solution of sodium dodecyl sulphate (SDS, Sigma) in a refractive index matched solution of water (50 wt%) and glycerol (50 wt%) infused with Nile Red dye is added to the creamed droplets.

The transparent emulsion is loaded into a sample cell, allowed to cream under gravity and analysed using a Leica SP2 confocal laser scanning microscope equipped with a high numerical aperture oil-immersion objective lens with a 100× magnification. The fluorescent dye is excited using a 488 nm argon laser and emission is detected using a photomultiplier behind a long-pass 500 nm filter with two channels.

Image analysis. The centre point and radius of each droplet was measured from the confocal images using a deconvolution technique that is well suited to measurements on jammed systems. Each droplet was considered as the convolution of a δ-function at its centre and a sphere of radius r. To make the measurements, we simply deconvolved the images to obtain the positions of all particles with a given radius r using the following procedure.

A test 3D volume was created with a sphere of radius r at its centre. Gaussian noise was added to this test particle with similar characteristics to the noise found in the confocal image. Both volumes were transformed into the frequency domain, where the Fourier transform of the confocal image was divided by the Fourier transform of the test volume. A tuned Wiener filter was applied to increase the signal relative to the noise. The resulting Fourier volume was then transformed back into the positional domain. In the resulting volume, numerical approximations to δ-functions marked the centre point of each droplet of radius r. This procedure was carried out for each radius r until every droplet in the volume was located.

LETTERS

Carbon respiration from subsurface peat accelerated by climate warming in the subarctic

Ellen Dorrepaal[1], Sylvia Toet[1]†, Richard S. P. van Logtestijn[1], Elferra Swart[1], Martine J. van de Weg[1]†, Terry V. Callaghan[2,3] & Rien Aerts[1]

Among the largest uncertainties in current projections of future climate is the feedback between the terrestrial carbon cycle and climate[1]. Northern peatlands contain one-third of the world's soil organic carbon, equivalent to more than half the amount of carbon in the atmosphere[2]. Climate-warming-induced acceleration of carbon dioxide (CO_2) emissions through enhanced respiration of thick peat deposits, centuries to millennia old, may form a strong positive carbon cycle–climate feedback. The long-term temperature sensitivity of carbon in peatlands, especially at depth, remains uncertain, however, because of the short duration or correlative nature of field studies[3–5] and the disturbance associated with respiration measurements below the surface *in situ* or during laboratory incubations[6,7]. Here we combine non-disturbing *in situ* measurements of CO_2 respiration rates and isotopic (^{13}C) composition of respired CO_2 in two whole-ecosystem climate-manipulation experiments in a subarctic peatland. We show that approximately 1 °C warming accelerated total ecosystem respiration rates on average by 60% in spring and by 52% in summer and that this effect was sustained for at least eight years. While warming stimulated both short-term (plant-related) and longer-term (peat soil-related) carbon respiration processes, we find that at least 69% of the increase in respiration rate originated from carbon in peat towards the bottom (25–50 cm) of the active layer above the permafrost. Climate warming therefore accelerates respiration of the extensive, subsurface carbon reservoirs in peatlands to a much larger extent than was previously thought[6,7]. Assuming that our data from a single site are indicative of the direct response to warming of northern peatland soils on a global scale, we estimate that climate warming of about 1 °C over the next few decades could induce a global increase in heterotrophic respiration of 38–100 megatonnes of C per year. Our findings suggest a large, long-lasting, positive feedback of carbon stored in northern peatlands to the global climate system.

The long-term effect of climate warming on soil carbon in terrestrial ecosystems has been debated for more than a decade[8–13]. This controversy has focused primarily on upland mineral soils[5,13], for which enzyme kinetic theory, laboratory studies and multi-pool models have established the short-term intrinsic temperature sensitivity of soil respiration[8,12,14]. However, acclimation[9], fast depletion of labile carbon pools[8,10,12] or secondary environmental or ecological constraints[13,15] have caused limited or transient responses of soil respiration to long-term *in situ* experimental warming of these upland ecosystems and have obscured the temperature sensitivity of old (that is, long-term), recalcitrant carbon pools. The apparent sensitivity of soil carbon respiration to long-term climate warming thus depends strongly on the relative and absolute availability of recalcitrant substrates and the suite of environmental constraints to decomposition[13], which both differ with soil types and climatic conditions.

It is therefore surprising that the long-term *in situ* sensitivity to climate warming of one of the largest global stocks of old, recalcitrant organic carbon, residing in northern peatlands, has received far less attention[5,13]. Although northern peatlands cover only 2% of the land, their thick peat deposits contain one-third of the world's soil organic carbon[2]. Peat formation has withdrawn large amounts of CO_2 from the atmosphere for centuries to millennia as organic carbon was stabilized by increasing decay-resistance, anoxia and decreasing temperatures with depth or permafrost (Supplementary Fig. 1). It has been suggested[9,10] that soil organic carbon in cold regions may be highly sensitive to above-average current and predicted climate warming at northern high latitudes[16] because of the higher temperature sensitivity of soil respiration at low temperatures[3,8,17]. Furthermore, recalcitrant organic matter at depth might be more temperature sensitive than labile carbon at the surface because of its higher catabolic activation energy[12,14], although this remains controversial[7,18]. Because of the short duration or correlative nature of field studies[3–5] and disturbance artefacts of respiration measurements below the surface[6,7], it remains unclear whether northern peatland soil carbon, especially at depth (25–50 cm), is susceptible to long-term climate warming, and is at risk of causing long-lasting increased carbon losses to the atmosphere. Here we provide evidence that this may well be true.

We investigated the climate-change response of ecosystem respiration rates and of respiration processes related to short-term (auto-trophic respiration and heterotrophic respiration of plant exudates) and medium- to longer-term (heterotrophic respiration of young, surface and older, subsurface peat) carbon cycles (Supplementary Fig. 1), using two whole-ecosystem climate-change experiments in a subarctic blanket-bog underlain by permafrost in Abisko, north Sweden (68° 21′ N, 18° 49′ E). We used open-top chambers (OTCs) to simulate changes in summer, winter and/or spring climate projected for the northern European subarctic during the coming decades[16] (Figs 1 and 2). The OTCs increased air[19] and soil temperatures (down to at least 20 cm depth) by up to about 1 °C in spring (May: between snow-melt and bud-break) and/or summer (June–September: growing season) and/or doubled the winter (October–April) snow depth[19], without consistently affecting soil moisture or active layer depth (Supplementary Information).

During the snow-free season (May–September) of years 4, 6, 7 and 8 of the long-term experiment, the doubled winter snow depth did not result in carry-over effects on spring or summer ecosystem respiration rates (Fig. 1). However, respiration rates increased strongly in response to warming by OTCs in spring (34–76%) or in summer

[1]Department of Systems Ecology, Institute of Ecological Science, VU University Amsterdam, De Boelelaan 1085, NL-1081 HV Amsterdam, The Netherlands. [2]Abisko Naturvetenskapliga Station, Royal Swedish Academy of Sciences, SE-98107 Abisko, Sweden. [3]Sheffield Centre for Arctic Ecology, Department of Animal and Plant Sciences, University of Sheffield, 26 Taptonville Road, Sheffield S10 5BR, UK. †Present addresses: Environment Department, University of York, Heslington, York YO10 5DD, UK (S.T.); School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK (M.J.v.d.W.).
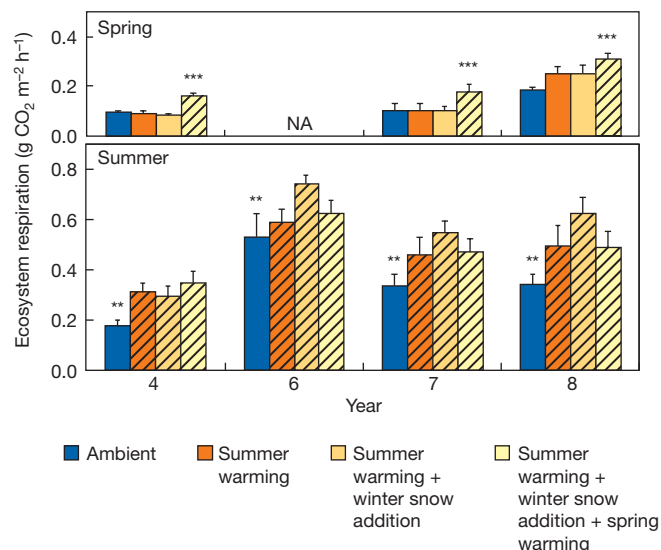
Figure 1 | Ecosystem respiration rates in a subarctic bog subjected to long-term experimental climate change or ambient conditions. Data were averaged separately for spring (May) and for summer (June–September). Hatched bars indicate passive warming with OTCs within the respective season. Error bars represent the standard error (s.e.) of residuals of treatments and blocks in spring, and the standard error of the mean (s.e.m.) of treatments in summer (one treatment replicate per block, $n = 5$ blocks). NA, no data available. Bars with asterisks differ significantly from bars without asterisks within the same season and year (planned orthogonal contrasts: $**P = 0.003$; $***P = 0.001$).

(23–80%). In contrast to long-term studies in forest, meadow and tundra ecosystems[10,15,20], the warming effects did not decline towards the eighth year of the study (Fig. 1; year × treatment: $P = 0.53$ for spring, $P = 0.50$ for summer), indicating that climate warming has a long-lasting stimulating effect on CO_2-emission in subarctic peatlands.

To partition the sources of increased ecosystem respiration rates, we compared the effects of OTCs during the snow-free season between intact and trenched-plus-clipped parts of plots in a companion experiment. Heterotrophic respiration of peat (trenched-plus-clipped treatment) accounted for 70% of the total ecosystem respiration rate (intact treatment) and both heterotrophic and plant-related (aboveground, roots, rhizosphere) respiration rates

responded equally strongly to the warming treatment (Fig. 2). No change in their relative contribution was observed over the first year of the warming treatment, consistent with longer-term observations in other ecosystems[10,21]. This strongly suggests that heterotrophic respiration was responsible for the largest part of the increase in ecosystem respiration rates observed in the long-term experiment.

We further analysed the relative contribution of heterotrophic respiration of subsurface peat (25–50 cm depth, that is, the lower part of the active layer) to increased respiration rates in response to warming with a non-disturbing, *in situ* method, using the natural increase with peat depth of the $\delta^{13}C$ signatures of bulk C and respired CO_2 (Supplementary Fig. 2) and the overall $\delta^{13}C$ signature of CO_2 emitted from the plots. Because isotope discrimination during decomposition is dominated by the relative use by microbes of substrates differing in isotopic composition rather than by metabolic fractionation[22], a shift towards greater contribution of deeper peat to total respiration would be reflected as an increase in the overall $\delta^{13}C$ signature of CO_2 emitted in the field. Experimental warming increased the $\delta^{13}C$ signature of the respired CO_2 on average by 0.77‰ (Fig. 3), indicating a shift towards respiration of less-depleted carbon substrates. The effect was similar for heterotrophic respiration and total ecosystem respiration (Fig. 3), and therefore did not result from changes in photosynthetic fractionation by plants in response to aboveground environment[23]. Laboratory studies have shown that warming may decrease $\delta^{13}C$ signatures of respired CO_2 by 0.12–0.42‰ per °C owing to shifts towards decomposition of more depleted, recalcitrant substrates caused by microbial community changes[17]. Such changes in substrate use within peat layers may thus not explain, but may instead have partially obscured, the opposite, positive shift in $\delta^{13}C$ signature of respired CO_2 in our warming treatments. The clear increase in $\delta^{13}C$ signature of respired CO_2 after 1 or 7 years of experimental climate change (Fig. 3), compared with the 1.7‰ increase in $\delta^{13}C$ signature of bulk C over the upper 50 cm of peat (Supplementary Fig. 2), therefore strongly suggests that both short-term and long-term warming stimulated subsurface respiration disproportionately. Applying the range of temperature-related fractionation[17] to our data, we calculated that increased subsurface peat respiration was responsible for at least 69% of the increase in respiration rates in the warming treatments (Supplementary Information).
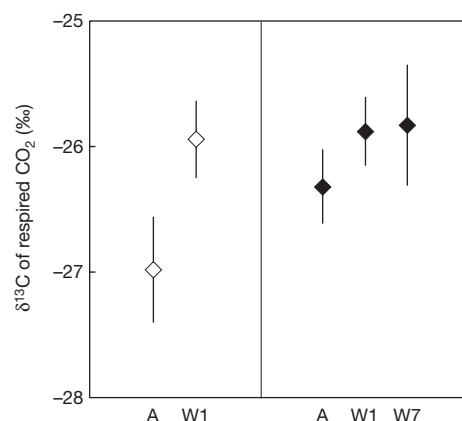

Figure 2 | Ecosystem respiration rates and their heterotrophic and plant-related components in a subarctic bog subjected to experimental warming or ambient conditions. Spring and summer warming (black bars) stimulated total ecosystem respiration ($P = 0.001$), and stimulated heterotrophic (R_h) and plant-related (R_a) respiration components equally ($P < 0.001$; warming × flux-component: $P = 0.65$) compared with ambient conditions (white bars). Response patterns remained unchanged over the first two experimental years (period × warming: $P = 0.41$ for ecosystem respiration, $P = 0.18$ for R_h and R_a), which were averaged. Error bars represent s.e.m. of treatments ($n = 5$ plots).

Figure 3 | Isotopic signatures of carbon respired as CO_2 in a subarctic bog subjected to experimental climate change or ambient conditions. Spring and summer warming (one year; W1) and spring and summer warming and winter snow addition (seven years; W7) increased $\delta^{13}C$ of respired CO_2 ($P = 0.032$) compared with ambient conditions (A). Climate-change duration did not affect $\delta^{13}C$ of respired CO_2 ($P = 0.93$). Climate-change effects on $\delta^{13}C$ of heterotrophic respiration (open symbols) and of total ecosystem respiration (filled symbols) did not differ (warming × flux-type: $P = 0.38$). Response patterns remained unchanged for four measurements (period × treatment: $P = 0.56$), which were averaged. Error bars represent s.e.m. of treatments ($n = 5$ plots).

Our data show for the first time that soil carbon respiration in a subarctic peatland, especially towards the bottom of the active layer, has a high, long-lasting sensitivity to climate warming. This contrasts with limited or transient soil respiration responses to *in situ* warming in other ecosystems, mostly on upland mineral soils[9,10,15,20], which were dominated by highly sensitive responses of plant-related respiration[24] and of fast-depleting, young, labile carbon[9,10,20,21]. In northern peatlands, fresh surface litter and plant-related respiration contribute proportionally less to total respiration than in most other ecosystems because of the large, old soil organic matter pool at depth[11,25]. This subsurface carbon sustained the accelerated respiration in our long-term warming treatments. Its vulnerability might be explained by a higher intrinsic temperature sensitivity of recalcitrant carbon[12,14] and be related to the presence of permafrost. Although our long-term climate manipulations did not alter the active layer depth (Supplementary Information), the progressively increasing thickness of unfrozen soil during the growing season continuously exposes deeper, highly temperature-sensitive carbon[3] to temperatures above freezing, a transition to which soil respiration is especially sensitive[3,8].

The net effect of climate warming on northern peatland carbon stocks will depend on the responses of both carbon losses from peat and carbon gains through new plant production (Supplementary Fig. 1). In the Arctic, ecosystem respiration is generally more sensitive to temperature than gross ecosystem productivity, especially in moss-dominated systems[26]. Biomass production of *Sphagnum* mosses, the primary peat-formers in peatlands, is also not enhanced in our and other ecosystem warming experiments[19,27], and expansion of productive woody shrubs in response to climate change, as observed in other tundra ecosystems[28], is hampered in peatlands by the fast vertical growth of *Sphagnum*[29]. We therefore expect that stimulation of carbon assimilation rates in peatlands will be modest compared with the observed strong increase in respiration rates.

The extensive reservoirs of organic carbon in northern peatlands are thus highly sensitive to climate warming. Indeed, large-scale, long-term observations across England and Wales have already shown dramatic declines in carbon content in peat soils, which are suggested to be linked to recent climate change[11]. Taking our data as indicative of the responses of undisturbed northern peatlands worldwide, we estimate that a mild growing-season temperature increase of about 1 °C during the coming decades[16], as simulated by our OTC treatments, may induce a global increase in heterotrophic respiration from northern peatlands of 38–100 megatonnes of C per year (1 megatonne = 10^{12} g) (Supplementary Information). This estimate is based on direct summer warming effects only, because the OTCs did not consistently affect other environmental conditions, for example, soil moisture or active layer depth (Supplementary Information). While permafrost thawing and, depending on initial moisture conditions, drying of peatlands might further increase peat respiration[5,13,30], the estimated potential extra global CO_2 emission based on direct warming alone is enough to offset much of the Kyoto Protocol target for greenhouse gas emission reductions for the whole European Union (92 megatonnes of C per year). In economic terms, this extra CO_2 emission would require an investment of €2.4–6.3 billion to compensate based on the current carbon allowance price of the EU Emissions Trading Scheme. Although our estimation may be crude, it clearly indicates the large, long-lasting potential feedback of increased respiration from northern peatlands to the global climate.

METHODS SUMMARY

Climate manipulations. In the long-term experiment (eight years), we combined changes in summer and/or spring temperatures, winter snow thickness, or ambient conditions (for the four treatment combinations used here, see Fig. 1). In the companion experiment (two years) we combined spring and summer warming or ambient conditions with aboveground manipulation of the vegetation of part of the plots, so as to be able to partition total ecosystem respiration (intact part) into heterotrophic respiration (trenched-plus-clipped part) and plant-related respiration (aboveground, roots, rhizosphere; difference between intact and trenched-plus-clipped parts). We used hexagonal OTCs that passively increased temperatures by up to approximately 1 °C in both air[19] and soil (to a depth of at least 20 cm, Supplementary Information) when used in spring and/or summer and passively doubled snow depth[19] when used in winter, without consistently affecting soil moisture or active layer depth (Supplementary Information). Each treatment was replicated five times.

Respiration measurements. Respiration rates were measured over 120-s intervals with an infrared gas analyser coupled to a non-transparent respiration chamber placed on permanently installed collars, during the daytime on six to twenty occasions per year in the snow-free season (May–September) in years 4, 6, 7 and 8 of the long-term experiment and in years 1 and 2 of the companion experiment. Isotopic ($\delta^{13}C$) analyses of respired CO_2 were made by sampling air with accumulating CO_2 eight times from similar closed chambers installed for 30–60 min, on four occasions during the first and second seasons of the companion experiment and the seventh and eighth seasons of the long-term experiment. Samples were analysed with a gas chromatography-isotope ratio mass spectrometer (GC-IRMS), after which we estimated the $\delta^{13}C$ of the respired CO_2 from Keeling plots. The relative contribution of subsurface peat to the respiration rate increase in the OTC treatments was calculated using two end-member mixing models (Supplementary Information).

1. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C^{4}MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
2. Gorham, E. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecol. Appl.* **1**, 182–195 (1991).
3. Goulden, M. L. *et al.* Sensitivity of boreal forest carbon balance to soil thaw. *Science* **279**, 214–217 (1998).
4. Updegraff, K., Bridgham, S. D., Pastor, J., Weishampel, P. & Harth, C. Response of CO$_2$ and CH$_4$ emissions from peatlands to warming and water table manipulation. *Ecol. Appl.* **11**, 311–326 (2001).
5. Lafleur, P. M., Moore, T. R., Roulet, N. T. & Frolking, S. Ecosystem respiration in a cool temperate bog depends on peat temperature but not water table. *Ecosystems* **8**, 619–629 (2005).
6. Hogg, E. H., Lieffers, V. J. & Wein, R. W. Potential carbon losses from peat profiles: effects of temperature, drought cycles, and fire. *Ecol. Appl.* **2**, 298–306 (1992).
7. Christensen, T. R., Jonasson, S., Callaghan, T. V. & Havström, M. On the potential CO$_2$ release from tundra soils in a changing climate. *Appl. Soil Ecol.* **11**, 127–134 (1999).
8. Kirschbaum, M. U. F. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biol. Biochem.* **27**, 753–760 (1995).
9. Luo, Y. Q., Wan, S. Q., Hui, D. F. & Wallace, L. L. Acclimatization of soil respiration to warming in a tall grass prairie. *Nature* **413**, 622–625 (2001).
10. Melillo, J. M. *et al.* Soil warming and carbon-cycle feedbacks to the climate system. *Science* **298**, 2173–2176 (2002).
11. Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M. & Kirk, G. J. D. Carbon losses from all soils across England and Wales 1978–2003. *Nature* **437**, 245–248 (2005).
12. Knorr, W., Prentice, I. C., House, J. I. & Holland, E. A. Long-term sensitivity of soil carbon turnover to warming. *Nature* **433**, 298–301 (2005).
13. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
14. Bosatta, E. & Ågren, G. I. Soil organic matter quality interpreted thermodynamically. *Soil Biol. Biochem.* **31**, 1889–1891 (1999).
15. Saleska, S. R. *et al.* Plant community composition mediates both large transient decline and predicted long-term recovery of soil carbon under climate warming. *Glob. Biogeochem. Cycles* **16**, 1055–1072 (2002).
16. Kattsov, V. M, *et al.* in *Arctic Climate Impact Assessment* (eds Symon, C., Arris, L. & Heal, B.) Ch. 4, 99–150 (Cambridge Univ. Press, 2005).
17. Biasi, C. *et al.* Temperature-dependent shift from labile to recalcitrant carbon sources of arctic heterotrophs. *Rapid Commun. Mass Spectrom.* **19**, 1401–1408 (2005).
18. Fang, C. M., Smith, P., Moncrieff, J. B. & Smith, J. U. Similar response of labile and resistant soil organic matter pools to changes in temperature. *Nature* **433**, 57–59 (2005).
19. Dorrepaal, E., Aerts, R., Cornelissen, J. H. C., Callaghan, T. V. & van Logtestijn, R. S. P. Summer warming and increased winter snow cover affect *Sphagnum fuscum* growth, structure and production in a sub-arctic bog. *Glob. Change Biol.* **10**, 93–104 (2004).
20. Rustad, L. E. *et al.* A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. *Oecologia* **126**, 543–562 (2001).
21. Zhou, X., Wan, S. Q. & Luo, Y. Q. Source components and interannual variability of soil CO$_2$ efflux under experimental warming and clipping in a grassland ecosystem. *Glob. Change Biol.* **13**, 761–775 (2007).

22. Šantrůčková, H., Bird, M. I., Frouz, J., Šustr, V. & Tajovský, K. Natural abundance of ^{13}C in leaf litter as related to feeding activity of soil invertebrates and microbial mineralisation. *Soil Biol. Biochem.* **32**, 1793–1797 (2000).

23. Ekblad, A. & Högberg, P. Natural abundance of ^{13}C in CO_2 respired from forest soils reveals speed of link between tree photosynthesis and root respiration. *Oecologia* **127**, 305–308 (2001).

24. Boone, R. D., Nadelhoffer, K. J., Canary, J. D. & Kaye, J. P. Roots exert a strong influence on the temperature sensitivity of soil respiration. *Nature* **396**, 570–572 (1998).

25. Subke, J. A., Inglima, I. & Cotrufo, M. F. Trends and methodological impacts in soil CO_2 efflux partitioning: a meta-analytical review. *Glob. Change Biol.* **12**, 921–943 (2006).

26. Grogan, P. & Chapin, F. S. Initial effects of experimental warming on above- and belowground components of net ecosystem CO_2 exchange in arctic tundra. *Oecologia* **125**, 512–520 (2000).

27. Weltzin, J. F., Harth, C., Bridgham, S. D., Pastor, J. & Vonderharr, M. Production and microtopography of bog bryophytes: response to warming and water-table manipulations. *Oecologia* **128**, 557–565 (2001).

28. Chapin, F. S. & Shaver, G. R. Physiological and growth responses of arctic plants to a field experiment simulating climatic change. *Ecology* **77**, 822–840 (1996).

29. Dorrepaal, E., Aerts, R., Cornelissen, J. H. C., Van Logtestijn, R. S. P. & Callaghan, T. V. *Sphagnum* modifies climate-change impacts on subarctic vascular bog plants. *Funct. Ecol.* **20**, 31–41 (2006).

30. Ise, T., Dunn, A. L., Wofsy, S. C. & Moorcoft, P. R. High sensitivity of peat decomposition to climate change through water-table feedback. *Nature Geosci.* **1**, 763–766 (2008).

METHODS

Site. We performed two whole-ecosystem climate-change experiments in a subarctic blanket-bog in Abisko, north Sweden (68° 21′ N, 18° 49′ E; ref. 19). Mean annual temperature is −0.5 °C with mean monthly temperatures ranging from −10.7 °C in January to 11.3 °C in July. Mean total annual precipitation is 323 mm (Abisko Scientific Research Station Meteorological Station, 1971–2000 averages). The vegetation of the site is dominated by a continuous and homogeneous carpet of the peat moss *Sphagnum fuscum* (Schimp.) H. Klinggr. Other bryophytes and lichens are sparse. The vascular plant community is low and open (maximum shrub height ~15 cm, average cover ~20%), and consists mainly of crowberry, *Empetrum nigrum* ssp. *hermaphroditum* (Hagerup) Böcher, cloudberry, *Rubus chamaemorus* L., bog rosemary, *Andromeda polifolia* L., small cranberry, *Vaccinium microcarpum* (Turcz. ex Rupr.) Schmalh., dwarf birch, *Betula nana* L., bog whortleberry, *Vaccinium uliginosum* L. and Lapland small-reed, *Calamagrostis lapponica* (Wahlenb.) Hartm. The thickness of the organic peat layer is unknown but it extends at least to the top of the permafrost, which is present at the site at a depth of approximately 55 cm. The water table closely follows the decreasing thaw front during the summer, and because of the gently sloping aspect of the terrain the site is relatively dry and a true water table on top of the frozen peat layer is rare and below 30 cm depth during most of the summer. Because of the exposed location of the site at the shore of Lake Torneträsk, the ambient snow cover is relatively thin, reaching its maximum thickness of about 15 cm in February–March[19].

Climate manipulations. The experimental site was divided into five blocks, perpendicular to the direction of the slope, and all treatments were replicated once in each of these blocks. The long-term experiment, initiated in June 2000, consisted of six climate-change scenarios, each a different combination of presence or absence of transparent OTCs[19,31] (2.2 m in diameter) in spring, summer and winter. The OTCs passively altered temperatures and/or snow conditions in the different seasons (Supplementary Information), corresponding to predicted changes in climate for the coming decades in the northern European subarctic[16]. For the present study, we focused on four treatments: 'ambient'; 'summer warming'; 'summer warming + winter snow addition'; 'summer warming + winter snow addition + spring warming'.

The companion experiment consisted of an ambient and an OTC treatment during spring and summer, in factorial combination with aboveground vegetation manipulation. We cut 20-cm-deep trenches around ten 384 cm² patches of vegetation for the plant removal treatment, and installed stainless-steel collars (22-cm diameter, 10-cm deep), identical to the ones used for the respiration measurements, in early summer 2005. All living, aboveground vascular vegetation was clipped off and removed at peak biomass by the end of July, and the active parts of the dense *Sphagnum fuscum* carpet (1–2 cm) in September 2005. Aboveground litter was left in place. We minimized potential effects of decomposing cut-off roots on CO_2 fluxes by allowing 11 months to pass before starting the OTC treatment and the flux measurements, while the collars served as a barrier against in-growing roots[32], concentrated in the upper 10-cm of peat in subarctic peatlands[33]. Regrowth of shoots within the removal patches was therefore extremely rare, but was consistently removed. We installed a second collar at a short distance from the removal patch in each plot and left the vegetation intact. In early July 2006, the warming treatment of the companion experiment was initiated by placing OTCs (1.1 m diameter) over half of the plots, each containing a pair of undisturbed vegetation and a plant removal patch, while leaving the other pairs as ambient controls.

Respiration measurements. We investigated the effects of the climate manipulations on respiration rates in years 4, 6, 7 and 8 after initiation in the long-term experiment (2003, 2005–2007), and in years 1 and 2 (2006–2007) of the companion experiment. Respiration rates were measured on six to twenty occasions per year during the snow-free season (May–September), one to twelve occasions each month except in June 2003, May and June 2005 and July–September 2007. The isotopic composition of respired CO_2 was measured in all four treatment combinations of the companion experiment and in the 'summer warming + winter snow addition + spring warming' treatment of the long-term experiment on four occasions in July, August and September 2006 and May 2007. For both types of measurements, complete blocks were sampled consecutively between 10:00 h and 16:00 h, and each measurement occasion lasted for 1–3 days.

Respiration measurements were made with an infrared gas analyser (EGM-4; PP Systems) attached to a custom-made, non-transparent, vented soil respiration chamber (22-cm diameter, 24-cm height), which was placed onto permanently installed collars as described above, equipped with a water-seal. Chamber volumes were corrected for differences in the air volume above the *Sphagnum* surface contained within the respiration collar of each plot. After an equilibration period of 65 s, the CO_2 concentration in the chamber was measured at 5-s intervals for 2 min and used to calculate the respiration rate by linear regression.

For the isotopic analysis, air with accumulating CO_2 was sampled eight times at increasing intervals from closed respiration chambers similar to those described above, but equipped with a rubber septum port and no vent, placed on each plot for 30–60 min. Air was transferred directly with double-ended needles into 12-ml septum-capped glass vials (Exetainers, Labco Ltd), which had been flushed with N_2 and pre-evacuated. Immediately afterwards we collected a 5-ml sample and measured the CO_2 concentration with the infrared gas analyser. The gas samples in the glass vials were transported to our laboratory in Amsterdam and analysed for their isotopic composition with a gas chromatograph coupled to a stable isotope ratio mass spectrometer (Gasbench II and Delta^plus, ThermoFinnigan) within five days of sampling. The isotopic composition of respired CO_2 was estimated for each plot and each occasion from Keeling plots[34]. Carbon isotopic composition is reported in δ-notation $\delta^{13}C\,(‰) = 1,000 \times [(^{13}C/^{12}C)_{sample} - (^{13}C/^{12}C)_{standard}]/(^{13}C/^{12}C)_{standard}$, relative to the Vienna PeeDee Belemnite standard.

31. Marion, G. M. *et al.* Open-top designs for manipulating field temperature in high-latitude ecosystems. *Glob. Change Biol.* **3**, 20–32 (1997).
32. Hanson, P. J., Edwards, N. T., Garten, C. T. & Andrews, J. A. Separating root and soil microbial contributions to soil respiration: a review of methods and observations. *Biogeochemistry* **48**, 115–146 (2000).
33. Wallén, B. Methods for studying below-ground production in mire ecosystems. *Suo* **43**, 155–162 (1993).
34. Keeling, C. D. The concentration and isotopic abundances of atmospheric carbon dioxide in rural areas. *Geochim. Cosmochim. Acta* **13**, 322–334 (1958).

LETTERS

Progressive mixing of meteoritic veneer into the early Earth's deep mantle

Wolfgang D. Maier[1,2], Stephen J. Barnes[3], Ian H. Campbell[4], Marco L. Fiorentini[2], Petri Peltonen[5,6], Sarah-Jane Barnes[7] & R. Hugh Smithies[8]

Komatiites are ancient volcanic rocks, mostly over 2.7 billion years old (from the Archaean era), that formed through high degrees of partial melting of the mantle and therefore provide reliable information on bulk mantle compositions[1]. In particular, the platinum group element (PGE) contents of komatiites provide a unique source of information on core formation, mantle differentiation and possibly core–mantle interaction[2–8]. Most of the available PGE data on komatiites are from late Archaean (~2.7–2.9 Gyr old) or early Proterozoic (2.0–2.5 Gyr old) samples. Here we show that most early Archaean (3.5–3.2 Gyr old) komatiites from the Barberton greenstone belt of South Africa and the Pilbara craton of Western Australia are depleted in PGE relative to late Archaean and younger komatiites. Early Archaean komatiites record a signal of PGE depletion in the lower mantle, resulting from core formation. This signal diminishes with time owing to progressive mixing-in to the deep mantle of PGE-enriched cosmic material that the Earth accreted as the 'late veneer' during the Early Archaean (4.5–3.8 Gyr ago) meteorite bombardment.

Komatiites are ultrabasic magmas that formed through high degrees of partial melting of the mantle and therefore provide the most reliable information on bulk mantle compositions. Komatiites have been subdivided into two main groups: Barberton-type komatiites are Al-depleted ($Al_2O_3/TiO_2 \approx 10$) and have trace element patterns that are undepleted in the most incompatible elements, whereas Munro-type komatiites are Al-undepleted ($Al_2O_3/TiO_2 \approx 20$) and show incompatible-element-depleted trace element patterns. It is widely accepted that both types form by melting of mantle plumes that are ultimately sourced in the deep lower mantle, probably in the thermal boundary layer above the core–mantle boundary[1]. The compositional differences between Barberton- and Munro-type komatiites reflect the conditions under which the melts separated from their plume sources. Barberton-type komatiites formed by ~30% batch melting of a mantle source enriched or slightly depleted in Ca–Al at a depth exceeding 300 km, and are depleted in Al owing to majorite garnet retention in the source, whereas Munro-type komatiites formed by ~50% fractional melting of a Ca–Al-depleted mantle source at a shallower depth (<300 km)[9].

Most of the available PGE data on komatiites to date have been obtained from 2.7–2.9-Gyr-old Al-undepleted lavas. The parental magmas to most of these rocks have Pd contents of around 10–12 parts per 10^9 (p.p.b.) (Fig. 1) at around 25% MgO, approximately twice that of the primitive upper mantle[10], which is broadly consistent with such magmas being the result of ~50% partial melting of undepleted or mildly depleted mantle[9]. Few PGE data have been generated until now for early Archaean komatiites, from the

Barberton greenstone belt of South Africa[6,11,12] and from chromitites in komatiites of the Pilbara craton of Western Australia[13].

Here we analysed 35 samples of komatiite from the type locality in the Barberton greenstone belt (Supplementary Table 1 and Supplementary
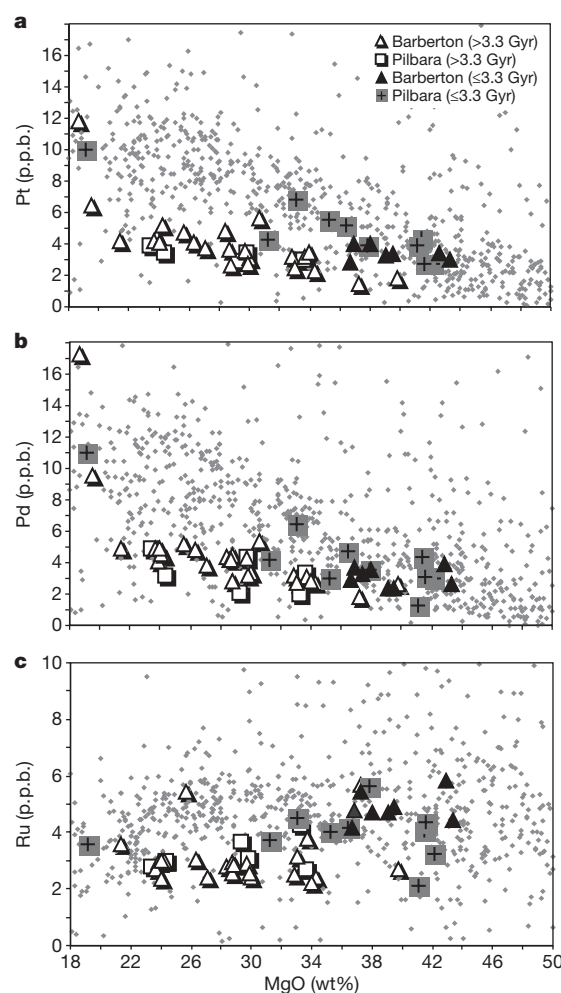

Figure 1 | PGE contents in Pilbara and Barberton komatiites. Larger data points are from the present study for Pt (**a**), Pd (**b**) and Ru (**c**) contents in the Pilbara and Barberton komatiites. For comparison, younger komatiites from the Yilgarn, Superior, Zimbabwe and Karelian cratons, and from Gorgona are shown as small grey data points from data compiled in ref. 8.

[1]Department of Geology, University of Oulu, Linnanmaa, 90014 Oulu, Finland. [2]Centre for Exploration Targeting, University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia. [3]CSIRO, 26 Dick Perry Avenue, Kensington 6151, Western Australia, Australia. [4]Research School of Earth Sciences, Australian National University, Canberra 0200, Australian Capital Territory, Australia. [5]Geological Survey of Finland, Betonimiehenkuja 4, Espoo 02151, Finland. [6]Department of Geology, University of Helsinki, Gustaf Hällströmin katu 2a, Helsinki 00014, Finland. [7]Sciences Appliques, Université du Québec à Chicoutimi, Chicoutimi G7H 2B1, Québec, Canada. [8]Geological Survey of Western Australia, 100 Plain Street, East Perth 6004, Western Australia, Australia.

Fig. 3). Most samples come from the 3.481-Gyr-old Al-depleted Komati formation, four samples are from the >3.53-Gyr-old Al-depleted Sandspruit formation, another four samples are from the 3.47-Gyr-old Al-depleted and Al-undepleted Hooggenoeg formation[14,15], three samples are from the ~3.3-Gyr-old Al-depleted and Al-undepleted Mendon formation and seven samples are from the 3.3-Gyr-old Al-undepleted Weltevreden formation. From MgO/FeO relationships we estimate that most Barberton lavas had about 25% MgO. Within the Pilbara craton, we analysed nine komatiites from the Al-depleted Ruth Well formation (with parental magma compositions of about 25–29% MgO). The Ruth Well komatiites are approximately 3.2 Gyr old, on the basis of SHRIMP dating of felsic tuff interlayers (M. J. van Kranendonk, personal communication). We also analysed one komatiite (and two komatiitic basalts) from the Al-undepleted 3.25-Gyr-old Sulphur Springs formation, two komatiites from the Al-undepleted 3.35-Gyr-old Euro basalt unit, one komatiite from the Al-undepleted 3.46-Gyr-old Apex basalt unit and four komatiites from the equally Al-undepleted 3.51-Gyr-old Coonterunah subgroup[16] (Supplementary Table 2 and Supplementary Fig. 4). We compare these data with analyses of younger komatiites from a variety of belts and cratons, from the work of Barnes and Fiorentini[8] and a number of other literature sources therein. We restrict these data to samples from komatiite units showing no association with nickel sulphide mineralization, in order to eliminate effects of crustal magmatic sulphide scavenging PGEs from the melts.

Most Barberton and Pilbara komatiites have ~2–5 p.p.b. Pt and Pd (Fig. 1). These metals show a good correlation, indicating that fractionation of Pt and Pd from one another is negligible in komatiites, except during alteration. The Barberton komatiites contain about 1–3.6 p.p.b. Ir, about 2–5 p.p.b. Ru and about 0.5–1 p.p.b. Rh. In the Komati formation, Ir content is sufficiently low that the magmas appear to be undersaturated in Ir metal, accounting for the unusual incompatible behaviour of Ir in this particular suite[8]. In the Pilbara komatiites, the Ir, Ru and Rh contents are slightly more variable than in the Barberton komatiites: 0.8–5 p.p.b. Ir, 2–5.5 p.p.b. Ru and 0.5–1.3 p.p.b. Rh. Over the same range of MgO, most Barberton and Pilbara komatiites are systematically depleted in all PGE relative to late Archaean and Proterozoic komatiites (Fig. 2). Because Pt and Pd behave in an incompatible manner during the crystallization of komatiites (Fig. 1), the concentration of these elements is affected by olivine fractionation. To neutralize this effect we correct the element concentrations to a constant MgO content of 25 wt%, by adding or subtracting olivine of an appropriate composition (~50–52% MgO, no Pt). Depletion of the Barberton and Pilbara komatiites in the Ir subgroup of the PGE is less clear than for Pt and Pd because the Ir-like PGE are effectively compatible with regard to the mantle. Ru shows the clearest signal because analytical precision for Ru is the best of all the PGEs and the effective partition coefficient, D, into cumulates/restites seems to be close to unity[8].

It is apparent that younger komatiites are progressively more enriched in Pt at a given MgO (Fig. 2), and that the most severe PGE depletion is found in the oldest rocks, from the ~3.5–3.55-Gyr-old Sandspruit formation. The correlation between PGE content and age is also evident within the individual cratons. Within the Barberton belt, the relatively older Sandspruit and Komati formations are more depleted than the younger Hooggenoeg, Mendon and Weltevreden formations, and within the Pilbara craton, the older Coonterunah subgroup is more PGE-depleted than the younger Euro, Sulphur Springs, and Ruth Well suites. These relationships are independent of whether the samples are from Al-depleted or Al-undepleted komatiites.

Several mechanisms may be considered to be responsible for the observed PGE depletion of the Earth's oldest komatiites. We discuss them below.

(1) Early Archaean komatiites could have reached sulphide saturation during ascent and/or emplacement, in response to contamination and/or differentiation. We have minimized this effect by excluding komatiite formations associated with sulphide mineralization.
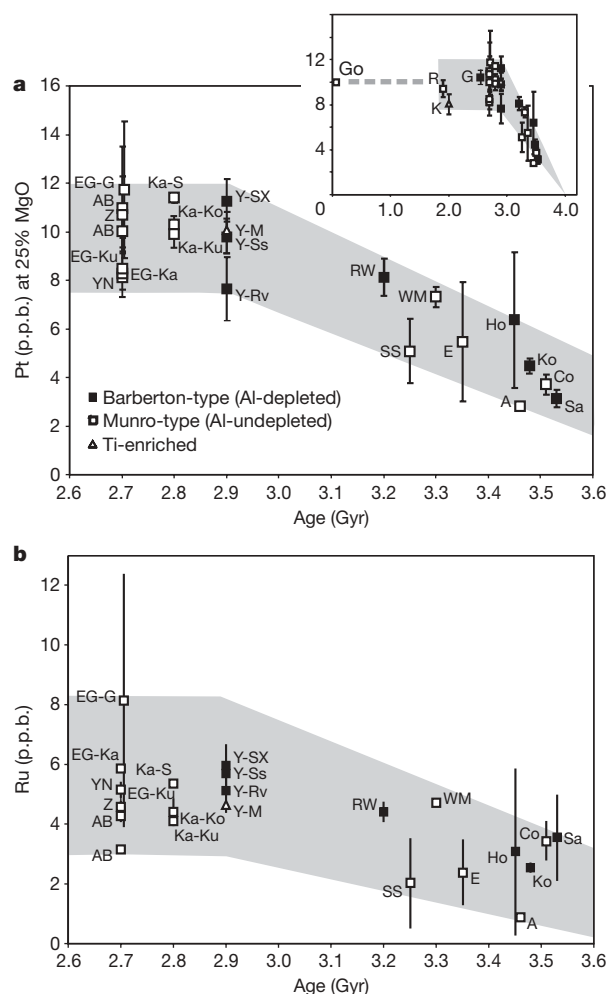

Figure 2 | **Diagram of Pt and Ru versus age of komatiites. a**, Pt versus age of komatiites. Data are normalized to 25% MgO. Data include those from this study, and those compiled in ref. 8. **b**, Ru versus age of komatiite. Data are anhydrous, but not normalized to 25% MgO because the normalization would require knowledge of the average D value with regard to the mantle. However, as the D value is likely to be around 0.5 to 1 (ref. 8), the correction would be small. EG, Eastern Goldfields superterrane (EG-G, Gindalbie terrane; EG-Ka, Kalgoorlie terrane; EG-Ku, Kurnalpi terrane). YN, Yilgarn North (Plutonic-Marymia belt). Ka, Karelian terrane (Ka-Ko, Kostamuksha belt; Ka-Ku, Kuhmo belt; Ka-S, Sumozero-Kenozero belt). Y, Youanmi terrane (Y-M, Murchison domain; Y-Ss, Sandstone belt; Y-Rv, Ravensthorpe belt). RW, Ruth Well formation. SS, Sulphur Springs formation. WM, Weltevreden-Mendon formation. E, Euro basalt unit. A, Apex basalt unit. Ho, Hooggenoeg formation. Ko, Komati formation. Co, Coonterunah subgroup. Sa, Sandspruit formation. The inset to **a** shows additional data for Proterozoic komatiites (Go, Gorgona island; R, Raglan belt; K, Karasjok belt; G, Gawler craton). Error bars represent one standard error of the mean for each belt, except for suites with fewer than five samples, in which error bars represent standard deviation.

Furthermore, empirical observations show that sulphide saturation resulted in wide variation in PGE contents within individual komatiite units, for example at Kambalda, where Pt and Pd show variability over three orders of magnitude within individual sulphide-saturated flows[17].

(2) Because sulphide solubility in basic–ultrabasic magmas is significantly reduced at high pressure[18], and the older, PGE-depleted Barberton-type komatiites formed at greater depth than many of the younger komatiites, this could have resulted in some mantle sulphides (and their highly chalcophile PGE) being retained in the mantle[12]. However, PGE depletion is independent of composition and occurs in both Al-depleted and Al-undepleted Pilbara and Barberton komatiites (Fig. 2).

(3) The source of some early Archaean komatiites may contain an eclogite component[19]. This model would be consistent with Hf isotopic data suggesting a long-term garnet enrichment in the Barberton source[20]. Many eclogites have lower PGE contents than mantle peridotites[21], reflecting the PGE depletion of oceanic crustal protoliths. However, explaining the PGE depletion of the Early Archaean komatiites using this model would require an approximately 60% eclogite component in their mantle source (assuming simple batch mixing between eclogite and primitive upper mantle). Such high proportions of crust are inconsistent with the high MgO contents of the komatiites and their mantle-like trace element signatures. Furthermore, our data set includes 2.9-Gyr-old Ti-enriched komatiites from the Gabanintha and Sandstone localities in the Yilgarn craton, which are believed to have been derived from an eclogite-enriched source, yet they show no systematic PGE depletion relative to other 2.9–2.7-Gyr-old komatiites.

(4) The Early Archaean komatiites could have been derived from a mantle source in a subduction zone setting[6,22]. However, lithophile trace element signatures of Barberton komatiites bear no similarity to those of subduction magmas and therefore we follow Arndt et al.[9] in unanimously rejecting the hydrous komatiite model.

(5) We have considered the possibility that the observed trend of progressive enrichment in Pt at 25% MgO with time is due to the evolution of the incompatible-element-depleted mantle source. There are two possible causes of this depletion: extraction of the continental crust[23], or partial melting and removal of a dense eclogitic component that separates gravitationally from the depleted mantle[24]. The concentration of PGE in the continental crust is low and its mass is less than 0.6% of the mass of the mantle. Extraction of the continental crust will therefore have no detectable influence on the PGE content of the mantle. The PGE content of basalt is also low, so removal of a partial melt will concentrate PGE in the residual mantle. If we assume 15% partial melting and no PGE in the basalt, the increase in the PGE content of the depleted mantle is ~18%, less if the basalt contains some PGE. An increase of this level cannot explain the observed increase between 3.5 and 2.9 Gyr ago and, in any case, lies within our combined analytical and sampling errors. Larger fractions of partial melting may have occurred in the mantle but are likely to have left a residue that is too refractory to be sampled by later melting events. Furthermore both Al-depleted (not depleted in incompatible elements) and Al-undepleted (depleted mantle source) komatiites have the same PGE content at the same age. We therefore conclude that the trend observed in Fig. 2 cannot be due to the processes that produced the depleted mantle source.

(6) The early Archaean mantle could have been more reduced than late Archaean and Proterozoic mantle such that in the Archaean mantle all S was present as sulphide and a significant portion of the PGE were present in alloys. If this were the case alloys would have remained in the source even at high degrees of melting, resulting in the low PGE content of the old komatiites. However, a change in the oxidation state of the mantle through this time interval is unlikely[25]. First, the contents of redox-sensitive elements, such as V and Cr, in Archaean basalts are similar to those of modern basalts. Second, the mass of available O (that is, O bound to Fe(III)) in the crust and atmosphere is less than 1% of that in the mantle. Therefore, formation of the crust and atmosphere can have no significant influence on the oxidation state of the mantle.

(7) The early Archaean komatiites could have been derived from a relatively PGE-poor majorite cumulate layer near the base of the lower mantle, accumulated from the 4.5-Gyr-old magma ocean[6]. This model would imply that the older PGE-poor komatiites should preferentially display majorite source signatures, that is, they should all be Al-depleted Barberton-type komatiites. This is inconsistent with the fact that our data set encompasses both Al-depleted and Al-undepleted komatiites across the range of Archaean ages, and there is no systematic difference in their PGE contents. The >3.2-Gyr-old Al-undepleted komatiites, showing no evidence for majorite

in the source, are just as depleted as the Al-depleted komatiites of the same age.

Here we propose an alternative model to explain the PGE depletion of early Archaean komatiites. The model rests on the assumption that the Earth's mantle underwent wholesale bulk depletion in PGE during core formation (at about 4.55 Gyr ago), followed by progressive re-enrichment with PGE in response to addition of cosmic material from the Hadean to Early Archaean (4.5–3.8 Gyr ago) heavy meteorite bombardment, the so-called 'late veneer'[26]. It is widely accepted that komatiites are derived from the tails of mantle plumes sourced deep in the lower mantle, probably a thermal boundary layer that overlies the core[1]. As a result, the observed PGE signal of the komatiites should reflect that of the deep early Archaean lower mantle. ÓNeill et al.[27] suggested that if the late veneer had not been homogenized within the mantle, material rising from the lower mantle would produce magmas poor in highly siderophile elements.

Our data provide the first direct evidence for such magmas in the early Archaean. We suggest that the Pt content of the plume source region rose progressively throughout the Archaean. At 4.5 Gyr ago, after core formation and solidification of the mantle, the Pt content of the lower mantle was close to zero. During the following 1.5 Gyr, the PGE content of the lower mantle progressively increased in response to mixing-in of late veneer into the upper mantle and transport of this material to the core–mantle boundary. By 3.2 Gyr ago enough metal had trickled down to give some komatiite sources a near-modern composition. By 2.9 Gyr ago the lower mantle had largely equilibrated with the late veneer and the PGE content was broadly similar to that of modern primitive upper mantle[28].

This places a new constraint on the rate of convective mantle mixing during the first half of the history of the planet. Our hypothesis also provides an explanation for the chondritic Os isotopic systematics of most Archaean komatiites, and the decoupling of the Re–Os isotopic system from that of Sm–Nd[9]. The Os signature of the mantle source back to at least 3.8 Gyr ago[13] was dominated by the chondritic late veneer component. Addition of this component was an entirely independent process from that which produced fractionation of Nd isotopes in komatiite source regions.

METHODS SUMMARY

The lithophile geochemistry of the samples was characterized by X-ray fluorescence and inductively coupled plasma mass spectrometry (ICP-MS) at Geoscience Australia (Pilbara samples, using techniques described in ref. 29), the University of Pretoria (Barberton samples, major and selected trace elements using techniques described in ref. 12), and the Ultratrace laboratories, Perth (Barberton rare-earth elements, Ta, Nb, Hf, Y, Th, using techniques described by ref. 8). PGE and Au analyses were conducted at the University of Quebec at Chicoutimi, Canada. The PGE and Au were collected from 15 g of sample by Ni-sulphide collection and Te co-precipitation followed by ICP-MS solution analysis[30].

There has been a suggestion that the Ni-sulphide collection method underestimates Pt and Ir values by approximately 15 relative per cent[31]. We do not believe that this is the case for our analyses for a number of reasons. First, the results we obtain for international reference materials agree with certified values (Supplementary Table 3). Second, the suggestion of ref. 31 is based on a meta-analysis of literature values. It is known that if the ratio of S to sample is reduced in the bead, collection of the PGE is not complete[32]. This may be the reason for low results obtained by some laboratories. Third, Barnes and Fiorentini[8] have recently shown that the method used in this study gives similar results to Carius tube digestion with isotope dilution ICP-MS. Fourth, our results using a high-pressure asher and isotope dilution on the komatiite standard show that there is no statistically significant difference for the results obtained by Ni–Te collection and isotope dilution. Finally, even if the results presented here are 15% too low, increasing the Ir and Pt results by 15% would not materially change any of the conclusions.

1. Campbell, I. H. & Griffiths, R. W. The changing nature of mantle hotspots through time: implications for the chemical evolution of the mantle. *J. Geol.* **92**, 497–523 (1992).
2. Barnes, S.-J., Naldrett, A. J. & Gorton, M. P. The origin of the fractionation of the platinum-group elements in terrestrial magmas. *Chem. Geol.* **53**, 303–325 (1985).

3. Brügmann, G. E., Arndt, N. T. & Hofmann, A. W. Platinum-group element abundances in komatiitic basalts and komatiites. *Terra Cognita* 5, 288–289 (1985).

4. Rehkamper, M. *et al.* Ir, Ru, Pt, and Pd in basalts and komatiites: new constraints for the geochemical behavior of the platinum-group elements in the mantle. *Geochim. Cosmochim. Acta* 63, 3915–3934 (1999).

5. Puchtel, I. S., Brandon, A. D. & Humayun, M. Precise Pt-Re-Os isotope systematics of the mantle from 2.7-Ga komatiites. *Earth Planet. Sci. Lett.* 224, 157–174 (2004).

6. Puchtel, I. S., Walker, R. J., Anhaeusser, C. R. & Gruau, G. Re-Os systematics and HSE abundances of the 3.5 Ga Schapenburg komatiites, South Africa: hydrous melting or prolonged survival of primordial heterogeneities in the mantle? *Chem. Geol.* doi:10.1016/j.chemgeo.2009.02.006 (2009).

7. Hanski, E., Huhma, H., Rastas, P. & Kamenetsky, V. S. The Palaeoproterozoic komatiite-picrite association of Finnish Lapland. *J. Petrol.* 42, 855–876 (2001).

8. Barnes, S. J. & Fiorentini, M. L. Iridium, ruthenium and rhodium in komatiites: Evidence for iridium alloy saturation. *Chem. Geol.* 257, 44–58 (2008).

9. Arndt, N. T., Lesher, C. M. & Barnes, S. J. *Komatiite* (Cambridge Univ. Press, 2008).

10. Becker, H. *et al.* Highly siderophile element composition of the Earth's primitive mantle: constraints from new data on peridotite massifs and xenoliths. *Geochim. Cosmochim. Acta* 70, 4528–4550 (2006).

11. Keays, R. R. in *Gold '82—The Geology, Geochemistry and Genesis of Gold Deposits* (ed. Foster, R. P.) 17–51 (Balkema, 1983).

12. Maier, W. D., Roelofse, F. & Barnes, S.-J. The concentration of the platinum-group elements in South African komatiites; implications for mantle sources, melting regime and PGE fractionation during crystallization. *J. Petrol.* 44, 1787–1804 (2003).

13. Bennett, V. C., Nutman, A. P. & Esat, T. M. Constraints on mantle evolution from Os-187/Os-188 isotopic compositions of Archean ultramafic rocks from southern West Greenland (3.8 Ga) and Western Australia (3.46 Ga). *Geochim. Cosmochim. Acta* 66, 2615–2630 (2002).

14. Lowe, D. R. in *Geologic Evolution of the Barberton Greenstone Belt, South Africa* (eds Lowe, D. R. & Byerly, G. R.) 287–312 (Geological Society of America Special Paper 329, 1999).

15. Van Kranendonk, M. J., Kröner, A., Hegner, E. & Connelly, J. Age, lithology and structural evolution of the c. 3.53 Ga Theespruit Formation in the Tjakastad area, southwestern Barberton Greenstone Belt, South Africa, with implications for Archean tectonics. *Chem. Geol.* 261, 115–139 (2009).

16. Van Kranendonk, M. J., Smithies, R. H., Hickman, A. H. & Champion, D. C. in *Earth's Oldest Rocks: Developments in Precambrian Geology* (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V.C.) 303–357 (Elsevier, 2007).

17. Lesher, C. M., Burnham, O. M., Keays, R. R., Barnes, S. J. & Hulbert, L. Geochemical discrimination of barren and mineralized komatiites asssociated with magmatic Ni-Cu-(PGE) sulphide deposits. *Can. Mineral.* 39, 673–696 (2001).

18. Mavrogenes, J. A. & O'Neill, H. S. C. The relative effects of pressure, temperature and oxygen fugacity on the solubility of sulfide in mafic magmas. *Geochim. Cosmochim. Acta* 63, 1173–1180 (1999).

19. Sobolev, A. V. *et al.* The amount of recycled crust in sources of mantle-derived melts. *Science* 316, 412–417 (2007).

20. Blichert-Toft, J. & Arndt, N. T. Hf isotope compositions of komatiites. *Earth Planet. Sci. Lett.* 171, 439–451 (1999).

21. McDonald, L. & Viljoen, K. S. Platinum group element geochemistry of mantle eclogites; a reconnaissance study of xenoliths from the Orapa Kimberlite, Botswana. *Trans. Inst. Mining Metallurgy* 115, 81–93 (2006).

22. Grove, T. L., de Wit, M. J. & Dann, J. in *Greenstone Belts* (eds de Wit, M. J. & Ashwal, L. D.) 422–437 (Oxford Science Publications, 1997).

23. McCulloch, M. T. & Bennett, V. C. Progressive growth of the Earth's continental crust and depleted mantle: geochemical constraints. *Geochim. Cosmochim. Acta* 58, 4717–4738 (1994).

24. Chase, C. G. & Patchett, P. J. Stored mafic/ultramafic crust and early Archaean mantle depletion. *Earth Planet. Sci. Lett.* 91, 66–72 (1988).

25. Campbell, I. H. & Allen, C. M. Formation of supercontinents linked to increases in atmospheric oxygen. *Nature Geosci.* 1, 554–558 (2008).

26. Chou, C. L. Fractionation of siderophile elements in the Earth's upper mantle. *Proc. 9th Lunar Planet. Sci. Conf.* 9, 219–230 (1978).

27. O'Neill, H. S. C. *et al.* Experimental petrochemistry of some highly siderophile elements at high temperatures, and some implications for core formation and the mantle's early history. *Chem. Geol.* 120, 255–273 (1995).

28. Puchtel, I. S. *et al.* Platinum group element geochemistry of komatiites from the Alexo and Pyke Hill areas, Ontario, Canada. *Geochim. Cosmochim. Acta* 68, 1361–1383 (2004).

29. Smithies, R. H. & Champion, D. C. The Archaean high-Mg diorite suite: links to tonalite-trondhjemite-granodiorite magmatism and implications for early Archaean crustal growth. *J. Petrol.* 41, 1653–1671 (2000).

30. Richardson, T. & Burnham, O. M. *Summary of Field Work and Other Activities 2002. Precious Metal Analysis at the Geoscience Laboratories: Results from the New Low-Level Analytical Facility.* Open File Report 35–1-35–535 (Ontario Geological Survey, 2002).

31. Lorand, J.-P. *et al.* Abundance and distribution of platinum-group elements in orogenic lherzolites; a case study in a Fontete Rouge lherzolite (French Pyrénées). *Chem. Geol.* 248, 174–194 (2008).

32. Frimpong, A. *et al.* Recovery of precious metals using nickel fire assay collection—problems at nanogram per gram concentrations. *Analyst* 120, 1675–1680 (1995).

LETTERS

A viscosity-enhanced mechanism for biogenic ocean mixing

Kakani Katija[1] & John O. Dabiri[1,2]

Recent observations of biologically generated turbulence in the ocean have led to conflicting conclusions regarding the significance of the contribution of animal swimming to ocean mixing. Measurements indicate elevated turbulent dissipation—comparable with levels caused by winds and tides—in the vicinity of large populations of planktonic animals swimming together[1]. However, it has also been noted that elevated turbulent dissipation is by itself insufficient proof of substantial biogenic mixing, because much of the turbulent kinetic energy of small animals is injected below the Ozmidov buoyancy length scale, where it is primarily dissipated as heat by the fluid viscosity before it can affect ocean mixing[2]. Ongoing debate regarding biogenic mixing has focused on comparisons between animal wake turbulence and ocean turbulence[3,4]. Here, we show that a second, previously neglected mechanism of fluid mixing—first described over 50 years ago by Charles Darwin[5]— is the dominant mechanism of mixing by swimming animals. The efficiency of mixing by Darwin's mechanism is dependent on animal shape rather than fluid length scale and, unlike turbulent wake mixing, is enhanced by fluid viscosity. Therefore, it provides a means of biogenic mixing that can be equally effective in small zooplankton and large mammals. A theoretical model for the relative contributions of Darwinian mixing and turbulent wake mixing is created and validated by *in situ* field measurements of swimming jellyfish using a newly developed scuba-based laser velocimetry device[6]. Extrapolation of these results to other animals is straightforward given knowledge of the animal shape and orientation during vertical migration. On the basis of calculations of a broad range of aquatic animal species, we conclude that biogenic mixing via Darwin's mechanism can be a significant contributor to ocean mixing and nutrient transport.

Darwin described a mechanism of fluid mixing by a solid object that does not require the object to generate rotational or shearing motion in the flow[5]. As the solid body travels through fluid, a portion of the surrounding fluid is set into motion by the body's pressure field and propagates along with the body. The volume of fluid that drifts with the solid object is proportional to the volume of the object itself (see Supplementary Fig. 1 and Supplementary Videos 1 and 2). The ratio of drift volume (V_{drift}) to body volume (V_{body})—called the 'added-mass' coefficient—depends only on the shape of the body.

In the case of a vertically stratified fluid such as the ocean, the induced fluid drift caused by vertical motion of a solid body will result in a concomitant change in the total potential energy of the fluid, because higher-density fluid is raised above lower-density fluid during upward body motion (and vice versa during downward body motion). The resulting increased interface between the fluid masses of different density will be susceptible to further stirring by ambient fluid motions and by interaction with neighbouring solid bodies, ultimately leading to molecular mixing[7,8]. The process of potential energy increase via fluid mixing is essential to the maintenance of

global ocean circulation and can also facilitate transport of nutrients and other dissolved matter[9]. When the mixing process is entirely fluid-driven, for example, through winds, tides, and potentially the turbulent wakes of swimming animals, the efficiency of the mixing process (that is, the fraction of the fluid kinetic energy that is converted to potential energy) depends on the length scale of the fluid motions. Fluid motions that are smaller than the smallest naturally-occurring ocean length scales (for example, the Kolmogorov and Ozmidov scales) are primarily dissipated as heat by the fluid viscosity and therefore cannot have an impact on ocean mixing[10].

In contrast, Darwin's mechanism of mixing is enhanced by the presence of viscosity in the fluid[11,12]. In the limit of Stokes flow, the induced fluid drift becomes infinitely large[13]. We demonstrated the effect of viscosity by computing the flow around bodies of circular cross-section and diameter D migrating in a viscous fluid at velocity U. The relative effect of viscosity (v) was manipulated by varying the Reynolds number of the flow, Re = UD/v, from 1 to 100. As the drift volume increases monotonically with time in the presence of viscosity[13], we calculated a nominal drift after each body had migrated the same distance. Figure 1a indicates substantial enhancement of the induced drift at decreasing Reynolds numbers. The drift volume in each case is significantly larger than for an equivalent
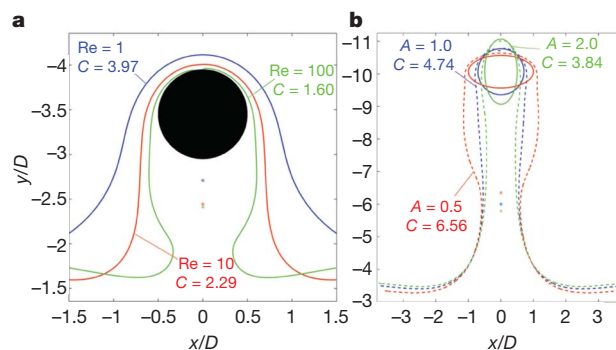

Figure 1 | Numerical simulations of induced vertical drift of initially horizontal layers in the presence of fluid viscosity. a, Simulation of induced fluid drift due to vertical migration of a circular cross-section at Reynolds numbers 1 (blue), 10 (red) and 100 (green). The cylinder moves vertically at unit velocity in each case. The area of induced drift bounded by each curve at the instant shown in the figure and its initial unperturbed horizontal position is indicated in the figure, normalized by the area of the circle (that is, C). Induced drift increases with decreasing Reynolds number. **b,** Simulation of induced fluid drift due to vertical migration of elliptical cross-sections at Reynolds number 10. The fineness (that is, length-to-diameter) ratio A and the normalized drift area C are indicated for each case. Induced drift increases with decreasing fineness ratio. Note that the values of C differ for the circle in panels **a** and **b** because of the difference in time at which the drift area is measured in each case.

[1]Bioengineering, [2]Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, California 91125, USA.
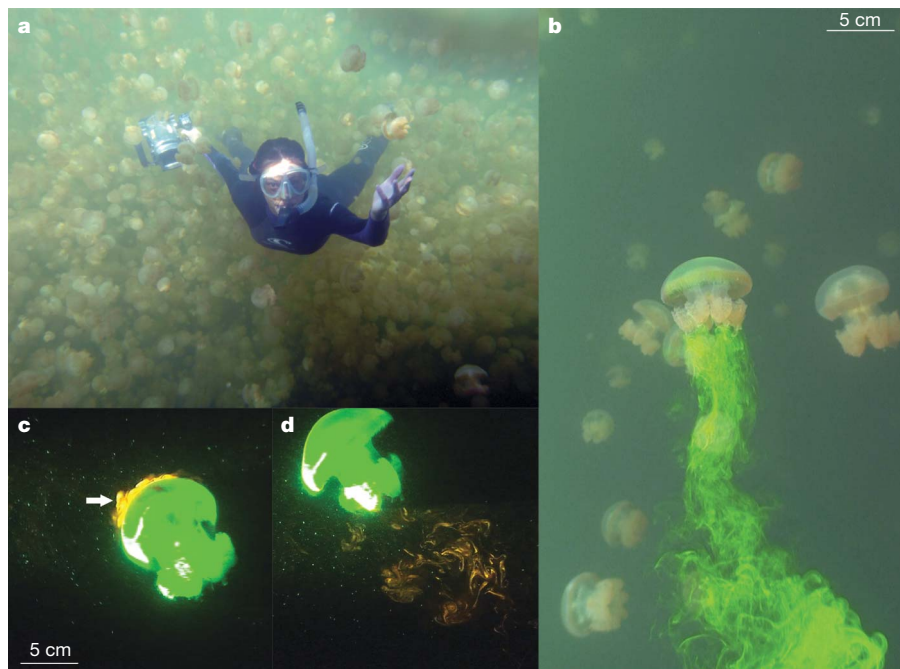
Figure 2 | Field measurements of induced drift during jellyfish swimming.
a, Image of diver (K.K.) in water column handling the camera for laser
velocimetry. (Photo provided by W. M. Graham.) **b**, Induced forward drift of
dye behind a swimming *Mastigias* sp. (see also Supplementary Videos 3 and
4). Image is rotated 90°, that is, gravity acts from left to right in the image.
c, *In situ* measurement of *Mastigias* sp. swimming through an artificially
created density stratification (indicated by white arrow). Red fluorescence
intensity is proportional to fluid density. **d**, Mixed density field following
passage of *Mastigias* sp. The centroid of the density field is shifted in the
direction of animal swimming (see Fig. 3). For a vertically migrating animal,
this will result in an increase in the potential energy of the fluid. Images are
rotated to vertical orientation. See also Supplementary Video 5.

body in an unbounded, inviscid fluid, for which the added-mass
coefficient $C = 1$ (ref. 14).

We determined the dependence of Darwin's mechanism on body
shape by computing the fluid drift induced by bodies of elliptical cross-
section at Re = 10. Figure 1b shows that the drift volume decreases with
increasing fineness (that is, length-to-diameter) ratio, consistent with
the known trends in inviscid flows. However, the magnitude of the drift
induced by even the most prolate body is substantial due to viscous
effects at the relatively low Reynolds number.

We experimentally determined the relative contributions of
induced drift and wake turbulence to the fluid mixing affected by a
swimming animal. *In situ* measurements of swimming jellyfish
(*Mastigias* sp., 1–10 cm body diameter) were conducted in a lake in
Palau during September 2008 (Fig. 2a). Video recordings of dye
injected upstream of individual jellyfish enabled empirical obser-
vation of the induced drift effect described above (Fig. 2b; see also
Supplementary Videos 3 and 4); dye is carried along behind each
animal for several swimming cycles, consistent with the simulations
of induced drift above. For quantitative measurements, we con-
structed a self-contained, underwater velocimetry apparatus[6] to
allow a scuba diver to measure directly the wake kinetic energy of
freely swimming jellyfish in the water column (see Methods). In
addition, planar laser-induced fluorescence measurements of an arti-
ficially created stratified layer in mid-water enabled measurement of
the change in fluid potential energy caused by a jellyfish swimming
through the layer (Figs 2c, d and 3; see also Supplementary Video 5).
As the Ozmidov length scale of the injected dye layer ($B \approx 4$ cm) was
significantly smaller than the natural buoyancy length scale of the
lake[15], the observed dynamics were decoupled from ambient fluid
motions. The measurements demonstrated that the mixing process
was dominated by the induced drift effects, with 90% of the potential
energy increase attributable to induced drift (see Methods).
Therefore, neglect of the contribution of induced drift in theoretical
models of the mixing efficiency (that is, flux Richardson number)
would result in an order-of-magnitude underestimate.

We determined that the contribution of induced fluid drift relative
to turbulent wake mixing can be predicted by considering the swim-
ming efficiency of the animal. The Froude propulsive efficiency—the
ratio of useable energy for overcoming drag to the total fluid energy
input during locomotion—can be approximated as (see Methods)

$$\eta_F \approx \frac{V_{drift}^{2/3} U^3}{V_{drift}^{2/3} U^3 + KN^2 V_{wake} + \left(1 - \frac{V_{drift}^{4/9}}{(\varepsilon/N^3)^{2/3}}\right)\varepsilon V_{wake}} \quad (1)$$

where we have formulated the swimming efficiency in terms of
properties of the animal (that is, its swimming speed U and kinetic
energy flux per unit time ε into the turbulent wake of volume V_{wake}),
the drift volume (V_{drift}), and the ocean (that is, turbulent diffusivity K
and Brunt–Väisälä buoyancy frequency

$$N = \sqrt{-g \, d\rho/\rho \, dz}$$

where g is gravitational acceleration, ρ is the fluid density and z is the
vertical coordinate direction). The two terms $KN^2 V_{wake}$ and

$$\left\{1 - \left[V_{drift}^{4/9} \Big/ \left(\varepsilon/N^3\right)^{2/3}\right]\right\}\varepsilon V_{wake}$$

in the denominator of equation (1) together represent the kinetic energy
flux per unit time (ε) lost to the turbulent wake as the animal swims; the
existing debate over biogenic mixing has focused on their relative mag-
nitudes[1,2]. However, the present simulations and experiments demon-
strate that this is only one, potentially small, component of biogenic
mixing. Using equation (1) and knowledge of the swimming efficiency,
we were able to draw conclusions regarding both the relative contribu-
tions of induced drift and turbulent wake mixing in animal swimming,
as well as the absolute magnitude of the effect of the Darwinian mech-
anism on ocean mixing.

Qualitatively speaking, a higher swimming efficiency leads to a
larger relative contribution of induced fluid drift to mixing com-
pared with turbulent wake energy. Indeed, in the limit of perfect
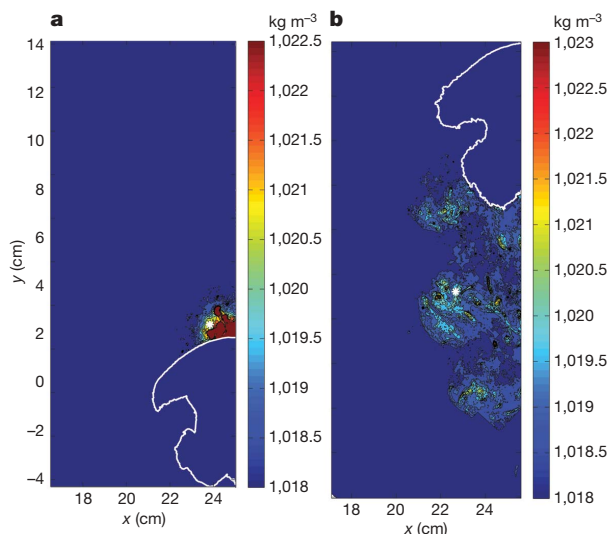swimming efficiency, no energy is lost to the wake and Darwin's

Figure 3 | Analysis of planar laser-induced fluorescence images.
a, Calibrated density field before passage of *Mastigias* sp. through density stratification. **b**, Calibrated density field following passage of *Mastigias* sp. through density stratification. Change in dye mass centroid is indicated by shift in white dot from panel **a** to panel **b**. Animal position is indicated by white outlines. Images are rotated to vertical orientation.

mechanism is the only one available for fluid mixing. Swimming efficiencies as high as 90% are reported in the literature on aquatic animals[16], consistent with our experimental observations that Darwin's mechanism can dominate over turbulent wake mixing. Even in the limit of low swimming efficiency, Darwin's mechanism will still contribute to mixing because induced drift will persist as long as the animal is moving.

For small animals, that is, those with length scale smaller than the Ozmidov scale *B*, turbulent mixing becomes inefficient due to the effects of fluid viscosity[2]. In these cases, Darwin's mechanism will be the dominant source of mixing irrespective of the swimming efficiency.

To determine the absolute contribution of Darwin's mechanism to ocean mixing, we note that previous studies of animal wake turbulence estimate kinetic energy production rates on the order of 10^{-5} to 10^{-4} W kg^{-1} for animals spanning a range from 10^{-3} to 10^{1} m in length and 10^{-5} to 10^{3} kg in mass[3]. As the swimming efficiency is of the order 10^{-1} to 1 across the same range of animals[3], we conclude based on equation (1) that the mixing power provided by induced fluid drift is at least of the order 10^{-5} to 10^{-4} W kg^{-1}. This is comparable with the turbulent dissipation observed due to physical processes[1–4] (for example, wind, tides and so on). By a similar argument, the estimated 10^{12} W of total turbulent kinetic energy input per unit time by marine animals[4] indicates that the global contribution to mixing by Darwin's mechanism is also in the range of 10^{12} W. This, too, is of the same order of magnitude as the global contributions of winds and tides. The local contributions of swimming animals to ocean mixing will depend on details of the migration behaviour, which can be heterogeneous even within a single school[17].

We consider these estimates to be conservative because they are based on the fluid transport induced by individual animals swimming in isolation. The drift volumes of neighbouring animals in an aggregation may interact via their boundary layers to enhance vertical transport further, thereby increasing the horizontal scale of the vertically drifting fluid mass to the horizontal extent of the entire aggregation of animals. Furthermore, we have not included in our estimates the effect of passively sinking particulates (for example, marine snow, faecal pellets and so on) that will also induce a drift of the surrounding fluid,

with potential implications for downward transport through the deep ocean (for example, carbon sequestration).

METHODS SUMMARY

The online-only Methods detail: (1) analytical and computational modelling of the induced drift effect in inviscid and viscous flows; (2) *in situ* measurement of wake kinetic energy of the swimming animals; (3) measurement of changes in the potential energy of the artificial density stratification due to animal swimming; (4) estimation of the mixing efficiency of a representative animal; (5) estimation of the contribution of induced drift to the change in potential energy; and (6) derivation of equation (1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Kunze, E., Dower, J. F., Beveridge, I., Dewey, R. & Bartlett, K. P. Observations of biologically generated turbulence in a coastal inlet. *Science* **313**, 1768–1770 (2006).
2. Visser, A. W. Biomixing of the oceans? *Science* **316**, 838–839 (2007).
3. Huntley, M. E. & Zhou, M. Influence of animals on turbulence in the sea. *Mar. Ecol. Prog. Ser.* **273**, 65–79 (2004).
4. Dewar, W. K. *et al.* Does the marine biosphere mix the ocean? *J. Mar. Res.* **64**, 541–561 (2006).
5. Darwin, C. Note on hydrodynamics. *Proc. Camb. Phil. Soc. Biol. Sci.* **49**, 342–354 (1953).
6. Katija, K. & Dabiri, J. O. In situ field measurements of aquatic animal-fluid interactions using a self-contained underwater velocimetry apparatus (SCUVA). *Limnol. Oceanogr. Methods* **6**, 162–171 (2008).
7. Eckart, C. An analysis of the stirring and mixing process in incompressible fluids. *J. Mar. Res.* **7**, 265–275 (1948).
8. Eames, I. & Bush, J. W. M. Longitudinal dispersion of bodies fixed in a potential flow. *Proc. R. Soc. Lond. A* **455**, 3665–3686 (1999).
9. Thorpe, S. A. *The Turbulent Ocean* (Cambridge Univ. Press, 2005).
10. Tennekes, H. & Lumley, J. L. *A First Course in Turbulence* (The MIT Press, 1972).
11. Eames, I. The concept of drift and its application to multiphase and multibody problems. *Phil. Trans. R. Soc. Lond. A* **361**, 2951–2965 (2003).
12. Dabiri, J. O. Note on the induced Lagrangian drift and added-mass of a vortex. *J. Fluid Mech.* **547**, 105–113 (2006).
13. Eames, I., Gobby, D. & Dalziel, S. B. Fluid displacement by Stokes flow past a spherical droplet. *J. Fluid Mech.* **485**, 67–85 (2003).
14. Lamb, H. *Hydrodynamics* 6th edn (Cambridge Univ. Press, 1993).
15. Hamner, W. M., Gilmer, R. W. & Hamner, P. P. The physical, chemical, and biological characteristics of a stratified, saline, sulfide lake in Palau. *Limnol. Oceanogr.* **27**, 896–909 (1982).
16. Fish, F. E. Comparative kinematics and hydrodynamics of odontocele cetaceans: morphological and ecological correlates with swimming performance. *J. Exp. Biol.* **201**, 2867–2877 (1998).
17. Parrish, J. K., Viscido, S. V. & Grunbaum, D. Self-organized fish school: an examination of emergent properties. *Biol. Bull.* **202**, 296–305 (2002).
18. Turner, J. S. The flow into an expanding spherical vortex. *J. Fluid Mech.* **18**, 195–208 (1964).
19. Eames, I., Belcher, S. E. & Hunt, J. C. R. Drift, partial drift and Darwin's proposition. *J. Fluid Mech.* **275**, 201–223 (1994).
20. Willert, C. E. & Gharib, M. Digital particle image velocimetry. *Exp. Fluids* **10**, 181–193 (1991).

Author Contributions K.K. and J.O.D. designed the study, analysed the data and wrote the manuscript. K.K. performed the experiments and viscous flow simulations. J.O.D. performed the inviscid flow simulations and derivation of equation (1).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.O.D. (jodabiri@caltech.edu) or K.K. (kakani@caltech.edu).

METHODS

Analytical and computational models. The motion of fluid particles induced by the vertical translation of a sphere was simulated by advecting a horizontal row of particles according to the theoretical solution for the flow around a sphere in inviscid, irrotational flow[18]:

$$v_r = \frac{dr}{dt} = -\left(1 - \frac{a^3}{r^3}\right)\cos\theta \qquad (2)$$

$$v_\theta = r\frac{d\theta}{dt} = \left(1 + \frac{a^3}{2r^3}\right)\sin\theta \qquad (3)$$

where r is the radial coordinate, θ is the angular coordinate measured from the direction of oncoming flow (in the reference frame of the propagating sphere), and a is the sphere radius. Particles in the flow were advected according to the local fluid velocity v, and the drift volume was calculated assuming axisymmetry. The computational domain is 50 sphere radii axially in both directions normal to the reference plane. A partial drift correction[19] was applied to determine the total induced drift volume because the radial extent of the horizontal row tracked in the simulation was finite.

The motion of fluid particles induced by the passage of two-dimensional bodies of circular and elliptical cross-section through viscous flow at Re = 1, 10 and 100 was computed using the Fluidica software tool (T. Colonius and K. Taira, California Institute of Technology).

Wake kinetic energy. Velocity fields were measured using digital particle image velocimetry[6,20]. The scuba diver operating the system remained motionless in the water column as animals swam freely into the plane of a 1-mm thick vertical sheet of light produced by a 250 mW, 532 nm diode-pumped solid-state laser (Wicked Lasers). The laser sheet illuminated natural particulates in the sea water (typically 10–100 μm), and the particle motion induced by the animal swimming was recorded at 30 Hz using a high-definition (1,920 × 1,080 pixels) complementary metal-oxide–semiconductor camera (Sony). Consecutive images of each video were interrogated with a digital particle image velocimetry algorithm, using an interrogation window size of 64 × 64 pixels and 50% window overlap. A representative case of swimming was selected for in-depth analysis of the wake kinetic energy.

The wake kinetic energy was computed at each node i of the velocity field according to the definition $KE_i = \frac{1}{2}\rho\forall_i\|\mathbf{u}_i\|^2$, where ρ is the ambient water density (1,020 kg m^{-3}), $\forall_i$ is the volume of the node (assuming radial symmetry about the animal body axis), and $\|\mathbf{u}_i\|$ is the magnitude of the velocity (that is, the speed) at each node. The total wake kinetic energy was determined by summing the energy at each of the nodes in the wake. The maximum increase in wake kinetic energy during each swimming cycle, ΔKE, was identified from the temporal trend of the wake kinetic energy measurements. The corresponding dissipation rate ε was estimated as $\Delta KE/mT_c$ where m is the mass of fluid encountered by the animal per swimming cycle and T_c is the duration of bell contraction.

Potential energy and mixing efficiency measurement. A sheet of fluorescent dye mixture of known density, ρ_{dye} (<1% above ambient $\rho_{ambient}$), was injected upstream of swimming *Mastigias* sp. and recorded using the optics described in the previous section. Fluorescence intensity was correlated with fluid density via an a priori laboratory calibration. The displacement, Δh, of the dye centroid along the direction of animal swimming (not necessarily vertical) was measured using an in-house image processing code and was used to infer the change in fluid potential energy for a vertically migrating animal as $\Delta PE = (\rho_{dye} - \rho_{ambient})V_{dye}g\Delta h$, where V_{dye} is the volume of injected dye. The total mixing efficiency (that is, flux Richardson number) was given by the ratio $\Delta PE/\Delta KE$ and was $24 \pm 18\%$ for a representative measurement, with the primary source of measurement uncertainty being the value of V_{dye} estimated from the dye images. The corresponding Brunt–Väisälä buoyancy frequency was computed according to the definition

$$N = \sqrt{-g\,d\rho/\rho\,dz},$$

where dz was estimated based on the length scale of the dye patch before interaction with the animal ($\approx$2 cm). The corresponding Ozmidov buoyancy length scale was computed according to the definition $B = \sqrt{\varepsilon/N^3}$ and was nominally 4 cm.

Induced drift estimation. As the shape of the induced drift region is approximately conical during the first few body lengths of travel through the stratified layer[11], the contribution of induced drift to the measured fluid displacement Δh was estimated based on the distance from the base to the centroid of a right circular cone with base and slant height equal to the characteristic length, L, of the animal. Hence, the contribution from induced drift to the measured change in dye centroid, Δh_{drift}, was estimated as $\sqrt{3/64}L$. For the *Mastigias* measurement, $\Delta h_{drift}/\Delta h \approx 0.9$, indicating a dominant contribution of induced drift to the change in fluid potential energy.

Froude efficiency derivation. The Froude propulsive efficiency is defined as the ratio of useful work to the total input of energy during swimming, and can be written as

$$\eta_F = \frac{F_D U}{F_D U + \varepsilon} \qquad (4)$$

where F_D is the drag force on the animal, U is the swimming speed, and ε is the flux of kinetic energy into the wake per unit time. As an order-of-magnitude estimate, the drag per unit fluid density can be approximated as $F_D \approx V_{drift}^{2/3}U^2$, where for the sake of convenience in what follows, the characteristic area over which hydrodynamic forces act is derived from the drift volume, that is, $V_{drift}^{2/3}$. The wake energy flux per unit time ε can be partitioned[2] into the portion that contributes to changes in the potential energy of the fluid, KN^2 and the portion that is dissipated as heat by the fluid viscosity, $(1 - \Gamma_\varepsilon)\varepsilon$, which depends on the mixing efficiency Γ_ε. For animals smaller than the characteristic buoyancy length scale of the stratified fluid, the mixing efficiency can be estimated[2] as

$$\Gamma_\varepsilon \approx \left[V_{drift}^{1/3}\Big/\left(\varepsilon/N^3\right)^{1/2}\right]^{4/3},$$

where the characteristic length scale of the body is again derived from its drift volume, that is, $V_{drift}^{1/3}$. Substituting these expressions into equation (4), we arrived at the relationship given in equation (1).

Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements

Christoph Lepper[1,2], Simon J. Conway[3] & Chen-Ming Fan[1]

Myogenic potential, survival and expansion of mammalian muscle progenitors depend on the myogenic determinants *Pax3* and *Pax7* embryonically[1], and *Pax7* alone perinatally[2–5]. Several *in vitro* studies support the critical role of *Pax7* in these functions of adult muscle stem cells[5–8] (satellite cells), but a formal demonstration has been lacking *in vivo*. Here we show, through the application of inducible Cre/*loxP* lineage tracing[9] and conditional gene inactivation to the tibialis anterior muscle regeneration paradigm, that, unexpectedly, when *Pax7* is inactivated in adult mice, mutant satellite cells are not compromised in muscle regeneration, they can proliferate and reoccupy the sublaminal satellite niche, and they are able to support further regenerative processes. Dual adult inactivation of *Pax3* and *Pax7* also results in normal muscle regeneration. Multiple time points of gene inactivation reveal that *Pax7* is only required up to the juvenile period when progenitor cells make the transition into quiescence. Furthermore, we demonstrate a cell-intrinsic difference between neonatal progenitor and adult satellite cells in their *Pax7*-dependency. Our finding of an age-dependent change in the genetic requirement for muscle stem cells cautions against inferring adult stem-cell biology from embryonic studies, and has direct implications for the use of stem cells from hosts of different ages in transplantation-based therapy.

Skeletal muscle regeneration is of clinical importance to muscular dystrophies and sport injuries, and depends on a resident reservoir of muscle stem cells called satellite cells[10]. Satellite cells are set aside to become quiescent postnatally from a pool of highly proliferative muscle progenitors[1,11]. The survival and expansion of muscle progenitors rely on the transcription factor *Pax7*. As such, *Pax7* germline mutant mice which survive to adulthood have few or no myofibre-associated cells resembling satellite cells[2–5], and display severely compromised muscle regeneration[3,4]. Although it remains controversial whether *Pax7* mutant myofibre-associated cells are myogenic *in vivo*[3,4], *in vitro* studies corroborate the role of *Pax7* in myoblast survival, proliferation and the establishment of myogenic potential[5,7,8]. Together with persistent *Pax7* expression in adult quiescent satellite cells[2,12], these data have raised the expectation that adult muscle stem-cell self-renewal and injury-induced muscle regeneration depend on *Pax7*. Thus, we examined these highly anticipated roles of *Pax7 in vivo*.

To determine *Pax7* function in adult satellite cells, we bypassed its earlier requirement via temporal gene inactivation in the adult. Previous analyses of *Pax7* mutants without lineage tracing have led to conflicting conclusions[3,4]. Therefore, we labelled conditionally inactivated *Pax7* cells and tracked their fate. Two new *Pax7* alleles were generated: *Pax7^f*, a conditional allele, which after Cre-mediated recombination becomes the null allele *Pax7^Δ* (Supplementary Fig. 1); and *Pax7^CE* which expresses a tamoxifen-inducible Cre recombinase–oestrogen receptor fusion protein, CreER^T2 (ref. 13), in place of *Pax7*,

and is therefore also null for *Pax7* (Supplementary Fig. 2). We assayed for inducible Cre activity by crossing *Pax7^CE* mice to *Rosa26* reporter (*R26R*) mice[14], which express *lacZ* after Cre-mediated recombination (Supplementary Fig. 3). Tamoxifen-dependent CreER^T2 activity was confirmed by β-gal activity (X-galactosidase reaction). Efficacy and specificity of tamoxifen-induced satellite cell labelling was confirmed by 99.8% co-localization of β-gal and Pax7 in tibialis anterior muscles (Supplementary Fig. 3).

The *Pax7^CE* and *Pax7^f* alleles were combined (*Pax7^CE/f;R26R^+/−*) for conditional inactivation of *Pax7* between postnatal days 60–90 (P60–90) by tamoxifen. After five daily tamoxifen injections (see Methods; Fig. 1a, b), no wild-type *Pax7* transcript, no Pax7 protein and no Pax7^+ cells were detected in tibialis anterior muscles (Supplementary Fig. 4). Conditional *Pax7* inactivation in adult satellite cells did not lead to loss of β-gal^+ cells (Supplementary Fig. 5), indicating that *Pax7* is not required for their survival. Conditional mutant cells maintain satellite cell characteristics as shown by M-cadherin (M-Cad), CD34, and integrin β1 and α7 expression[15,16] (Supplementary Fig. 6).

To determine if *Pax7* conditional mutant cells can support injury-induced myogenesis, we used cardiotoxin to induce injury. In *Pax7* heterozygotes (*Pax7^+/CE;R26R^+/−*), all regenerating fibres at days 5 and 10 after injury are β-gal^+ (Fig. 1c, d and Supplementary Fig. 7). This result provides unequivocal evidence that *Pax7* descendants are a major source of regenerating myofibres, extending data[15] that single *Pax7^+* cells can form myofibres after transplantation into tibialis anterior muscles. Contrary to expectation, regenerative myogenesis from *Pax7* descendants was not impaired in conditional *Pax7* mutants (*Pax7^CE/f;R26R^+/−*) as assessed by X-galactosidase reactions (Fig. 1e, f), regenerated fibre size (Fig. 1g, h) and number (Supplementary Fig. 7). No reappearance of wild-type *Pax7* mRNA or protein was detected in injured muscles 10 days after injury (Fig. 1i, j). Thus, *Pax7* is not required for injury-induced myogenesis in adulthood.

These results contrast with *Pax7* germline mutants where only a minimal number of small regenerative myofibres are found after injury (refs 3, 4 and Supplementary Fig. 1). These sparse regenerated fibres were proposed to be derived from alternative source(s), that is, non-*Pax7* lineage[4]. To test this, we assayed for injury-induced myogenesis in surviving *Pax7^CE/CE;R26R^+/−* adults using the experimental paradigm described above. Surprisingly, the rare regenerative myofibres (Fig. 1k, l) were all of *Pax7* descent. This result indicates that *Pax7* germline mutant cells surviving to adulthood do have myogenic potential and argues against alternative non-*Pax7* sources.

Isolated surviving cells associated with myofibres in adult *Pax7* germline nulls lack typical satellite cell characteristics[4]. Therefore, we examined whether conditional *Pax7* mutant cells also lose satellite cell characteristics after supporting one round of regeneration. For this,

[1]Department of Embryology, Carnegie Institution, 3520 San Martin Drive, Baltimore, Maryland 21218, USA. [2]Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. [3]Riley Heart Research Center, Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.
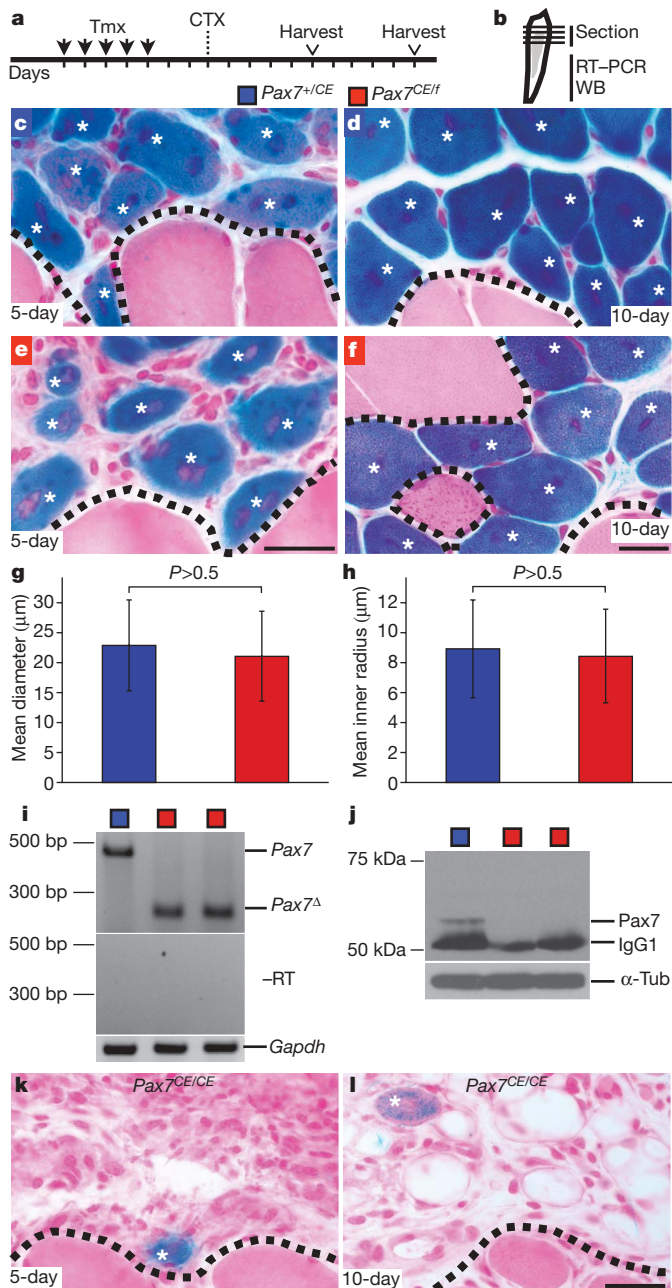
Figure 1 | Adult-specific *Pax7* mutant cells regenerate muscle efficiently.
a, Tamoxifen (Tmx) and cardiotoxin (CTX) regimen and regeneration assay
scheme. Vertical lines indicate daily intervals. **b**, Tibialis anterior muscle
diagram with injury in grey. Horizontal lines indicate cross-sections; WB,
western blot. **c–f**, β-gal (by X-galactosidase reaction) and nuclear fast red
(NFR) stained *Pax7*[+/CE] (**c**, **d**) and *Pax7*[CE/f] (**e**, **f**) muscles at 5 (**c**, **e**) and 10
(**d**, **f**) days after injury. Asterisks indicate regenerating fibres with central
nuclei; dashed lines indicate boundary of injury. **g**, **h**, Mean diameter (**g**) and
mean inner radius (**h**) of 10-day regenerated β-gal[+] fibres; 300 fibres per
animal (*n* = 3 per genotype); error bars indicate s.d.; *P*, two-tailed Student's
t-test. **i**, **j**, RT–PCR (**i**) and western blot (**j**) of examples of *Pax7*[+/CE] and
Pax7[CE/f] (two shown) muscles; wild-type *Pax7* and recombined (*Pax7*[Δ])
transcripts are indicated; −RT control, *Gapdh* control, endogenous IgG1
and α-tubulin (α-Tub, loading control) are indicated. **k**, **l**, β-gal and NFR
stained *Pax7*[CE/CE] muscles at 5 (**k**) and 10 (**l**) days after injury; regimen in
a. All animals were *R26R*[+/−]. Scale bars: 25 μm for **c**, **e** (shown in **e**), **d**, **f** (shown
in **f**) and **k**, **l** (shown in **l**).

β-gal-marked cells were assessed for satellite niche occupancy using
anti-M-Cad (for satellite/muscle junction) and anti-laminin (for basal
lamina). We found β-gal[+]M-Cad[+]Pax7[+] satellite cells in the sublami-
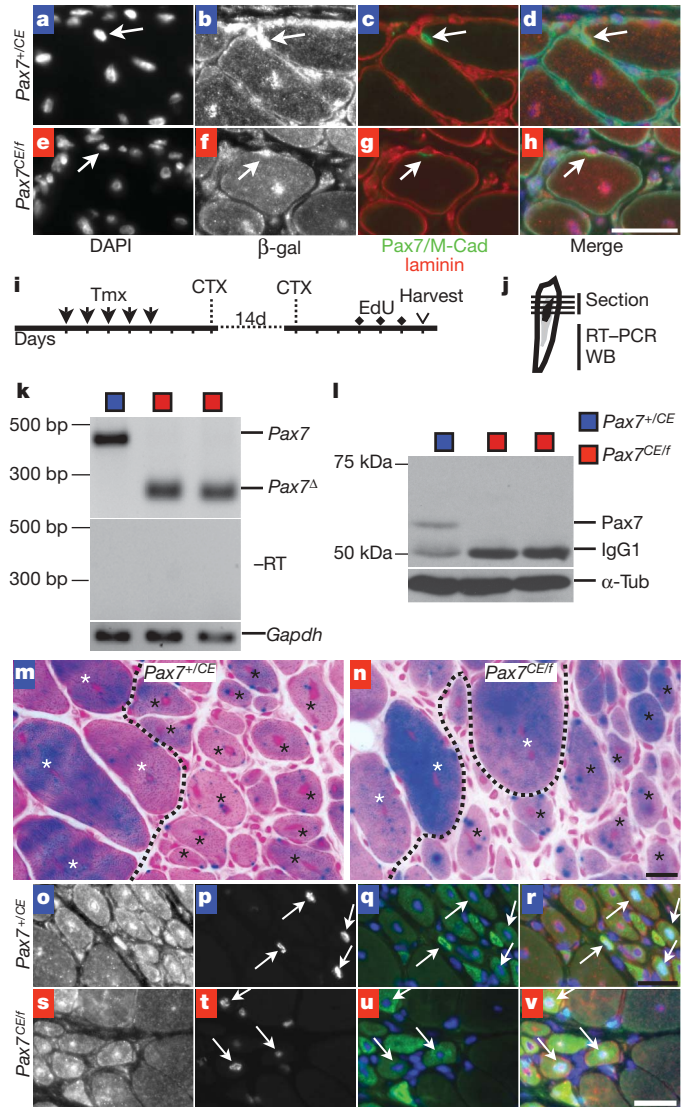nal space in control regenerated tissue (Fig. 2a–d and Supplementary

Figure 2 | Adult-specific *Pax7* mutant cells are functional satellite cells.
a–h, Fluorescent microscopy of *Pax7*[+/CE] (**a–d**) and *Pax7*[CE/f] (**e–h**) 10-day
regenerates. **a**, **e**, DAPI. **b**, **f**, β-gal. **c**, **g**, Laminin (red), Pax7 and M-Cad
(both in green: Pax7, nuclear; M-Cad, satellite–muscle junction). **d** and **h** are
merged from **a–c** and **e–g**, respectively; β-gal pseudo-coloured in red, DAPI
in blue, laminin in cyan, Pax7 and M-Cad in green; white arrows indicate
satellite cells. **i**, Double injury regeneration assay scheme with EdU
injections; labelling as in Fig. 1a with 14 days (14d) between two injuries.
j, Tibialis anterior muscle diagram with first injury in grey and second injury
in black. **k**, **l**, RT–PCR (**k**) and western blot (**l**) of tibialis anterior muscles
after second injury; labelling as in Fig. 1i, j. **m**, **n**, β-gal and NFR stained
Pax7[+/CE] (**m**) and *Pax7*[CE/f] (**n**) regenerates at day 6 after second injury;
dashed lines indicate boundaries between first and second regenerations
based on differences in fibre size; white and black asterisks indicate old and
new regenerated fibres, respectively. All fibres are β-gal[+] compared to
interstitial cells, but with varying staining signals. **o–v**, Fluorescent
microscopy of *Pax7*[+/CE] (**o–r**) and *Pax7*[CE/f] (**s–v**) regenerates. **o**, **s**, β-gal.
p, **t**, EdU. **q**, **u**, RNMy2/9D2 (green) and DAPI (blue). **r** and **v** are merged
from **o–q** and **s–u**, respectively; β-gal in red and other colours are the same as
panels to the left; EdU[+] nuclei are indicated by white arrows. Scale bars:
25 μm for **a–h** (shown in **h**), **m**, **n** (shown in **n**), **o–r** (shown in **r**) and
s–v (shown in **v**).

Fig. 8). Surprisingly, sublaminal β-gal[+]M-Cad[+]Pax7[−] cells were also
found in *Pax7* conditional mutants (Fig. 2e–h). When we used
5-ethynyl-2′-deoxyuridine (EdU) to monitor cells that had undergone
proliferation, we found EdU[+] myonuclei and EdU-retaining sublami-
nal β-gal[+]M-Cad[+]Pax7[−] cells in conditional mutants (Supplementary
Fig. 8). We did not detect any Pax7[+] or β-gal[−]M-Cad[+] sublaminal cells

in *Pax7*-inactivated regenerates, arguing against compensation by other lineages with or without *de novo* *Pax7* activation. These data indicate that adult-specific *Pax7* mutant descendants not only form muscle fibres but also proliferate and occupy the satellite niche.

To determine whether the β-gal$^+$M-Cad$^+$Pax7$^-$ sublaminal cells in conditional mutant regenerates still have myogenic potential, we induced a second round of injury (Fig. 2i, j). Twenty-eight days after the first tamoxifen injection, neither wild-type *Pax7* mRNA nor protein was detected in doubly injured conditional *Pax7* mutants (Fig. 2k, l). We found that small and large β-gal$^+$ fibres were comparable between controls (Fig. 2m) and conditional mutants (Fig. 2n) 6 days after the second injury. To identify second-round regenerated myofibres, we used the RNMy2/9D2 antibody, which labels new fibres only up to 6 days after injury[17]. Within the β-gal$^+$ regenerated area, small fibres were RNMy2/9D2$^+$ in control animals (Fig. 2o, q, r) and conditional mutants (Fig. 2s, u, v). To assess whether *Pax7* descendants had proliferated before forming new fibres, EdU was administered after the second injury. Indeed, we found EdU$^+$ nuclei in RNMy2/9D2$^+$ fibres of control animals (Fig. 2p, r) and conditional mutants (Fig. 2t, v). Although we cannot formally exclude contribution from *Pax7*-independent lineages, our data suggest that adult-inactivated *Pax7* mutant cells can repeatedly contribute to muscle regeneration.

The normal regenerative capacity of conditional *Pax7* mutants sharply contrasts with the severe germline *Pax7* mutant defects (refs 3, 4 and here), and indicates the possibility of adult-specific compensation by another *Pax* family member. *Pax3* is a likely candidate as *Pax3* and *Pax7* compensate for each other in embryonic myogenesis[1]. We detected *Pax3* transcripts in adult tibialis anterior muscles (Fig. 3a) but not Pax3 protein (not shown). To determine whether *Pax3* can compensate for *Pax7* in adult myogenesis, we conditionally inactivated both genes in adulthood. We used the *cre/Esr1* allele[18], which directs ubiquitous tamoxifen-inducible Cre activity, to conditionally recombine a *Pax3* floxed allele, *Pax3^f* (ref. 19). After the tamoxifen regimen in Fig. 1a, both *Pax3^f* and *Pax7^f* were recombined (Fig. 3a, b). Surprisingly, muscle regeneration was not impaired in *Pax3;Pax7* doubly inactivated tibialis anterior muscles as assessed by histology (Fig. 3d, compared to control in panel c) and regenerated fibre size (Fig. 3i, j). Furthermore, we found sublaminal M-Cad$^+$Pax7$^-$ cells within the regenerate (Fig. 3g, h; compare to the control M-Cad$^+$Pax7$^+$ cells, Fig. 3e, f). Thus, contrary to their essential roles for embryonic myogenesis, neither *Pax3* nor *Pax7* is required during adult tibialis anterior muscle regeneration.

We next determined if and when myogenic progenitors become independent of *Pax7* *in vivo*. To test this, we conditionally inactivated *Pax7* at different postnatal time points. Animals were then injured and regeneration was assessed 10 days later. When *Pax7* was inactivated between P7–11, regeneration was severely compromised (Fig. 4b). However, when inactivation was carried out at P14–18 and P21–25 (Fig. 4c, d), regenerative capacity gradually increased to levels similar to control mice (Fig. 4a). Notably, regenerated fibres were *Pax7* descendants (β-gal$^+$). Thus, *Pax7* function is critical before P21. To elucidate muscle progenitor behaviour at this postnatal period, we performed lineage tracing and found that myofibre fusion from *Pax7* descendants sharply declines by P21 (Supplementary Fig. 9), suggesting a transition to quiescent satellite cells. Intriguingly, myonuclei positioning in regenerated fibres also differs between P21 and adult animals, suggesting a coordinated transition of muscle biology around P21 (ref. 20). By contrast, myofibre incorporation by *Pax7*-mutant descendants occurs for a prolonged period (beyond P31, Supplementary Fig. 9), indicating that mutant cells have a greater propensity to differentiate. We therefore propose that in addition to survival and proliferation[3,5], *Pax7* also directs myogenic progenitors to withdraw from myogenic differentiation and transition into quiescent satellite cells, thereby acquiring maximum regenerative capacity.
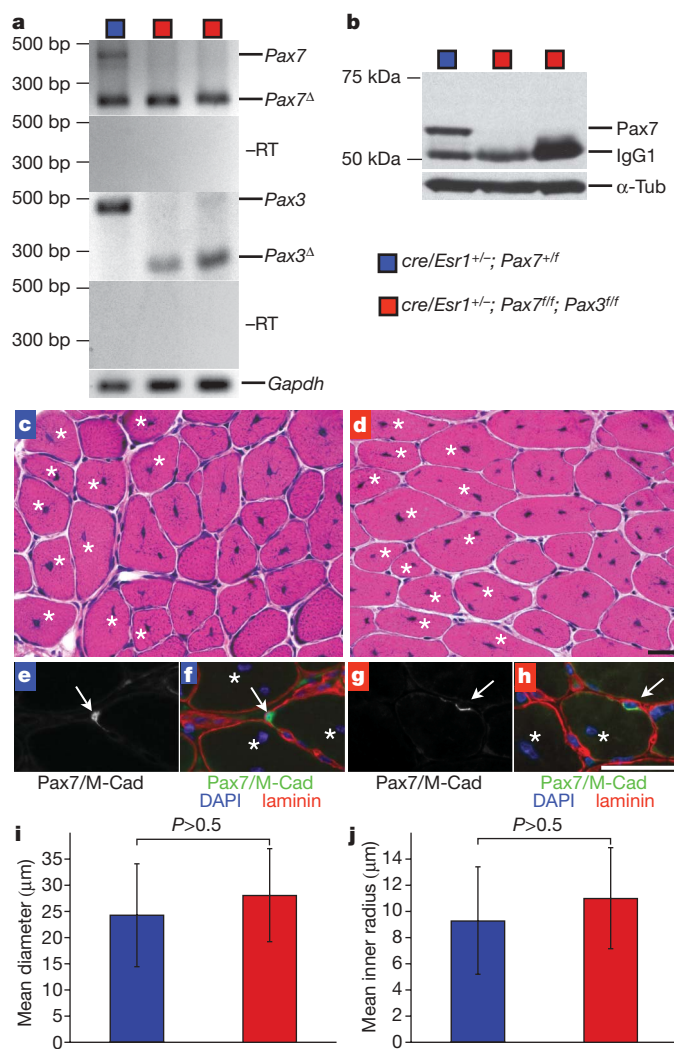
Figure 3 | *Pax3* and *Pax7* are dispensable for adult muscle regeneration. **a, b,** RT–PCR (**a**) and western blot (**b**) of indicated genotypes from the regeneration assay scheme in Fig. 1a. Labelling is the same as Fig. 1i, j, with additions of *Pax3* wild type and recombined (*Pax3^Δ*) transcripts. **c–h,** *cre/Esr1$^{+/-}$;Pax7$^{+/f}$* (**c, e, f**) and *cre/Esr1$^{+/-}$;Pax7$^{f/f}$;Pax3$^{f/f}$* (**d, g, h**) day-10 regenerated muscles. **c, d,** Haematoxylin and eosin stain. **e, g,** Fluorescent microscopy of Pax7 and M-Cad together (arrowheads, M-Cad$^+$ cells). **f** and **h** are coloured composites with co-stained laminin (red), DAPI (blue) and Pax7/M-Cad (green) of **e** and **g**, respectively; asterisks indicate central nuclei whereas arrows indicate satellite cells. **i, j,** Mean diameter (**i**) and mean inner radius (**j**) of regenerated fibres; 300 fibres per animal (*n* = 3 per genotype). Error bars indicate s.d.; *P*, two-tailed Student's *t*-test. Scale bars: 25 μm for **c, d** (shown in **d**) and **e–h** (shown in **h**).

To determine whether this temporal difference is a cell-intrinsic or environmental effect, we cultured myoblasts isolated from *Pax7$^{+/CE}$;R26R$^{+/-}$* and *Pax7$^{CE/f}$;R26R$^{+/-}$* tibialis anterior muscles at P0 (before quiescence) and P60 (after becoming quiescent), and assayed their properties after tamoxifen treatment. Compared to controls, conditional *Pax7*-inactivated P0 myoblasts (β-gal$^+$) were defective in expansion and myogenic potential (Fig. 4e–g). At passage 1, residual conditional mutant cells still expressed desmin (Fig. 4g, k), displayed fibroblastic morphology at a high frequency (Fig. 4g), and some were positive for the fibroblastic marker ER-TR7 (Fig. 4g, l); however, many cells are of unknown fate(s) as they do not express either marker. This is in contrast to controls (Fig. 4g), reminiscent of aged *Pax7*-descendant cells after losing *Pax7* expression *in vitro*[21]. After passage 2, mutant cells were no longer myogenic (not shown) and most β-gal$^+$ cells were fibroblastic. Thus, inactivation of *Pax7* in neonatal myoblasts *in vitro* decreases their expansive and myogenic capacity. In contrast, control and conditionally inactivated *Pax7*-descendant adult
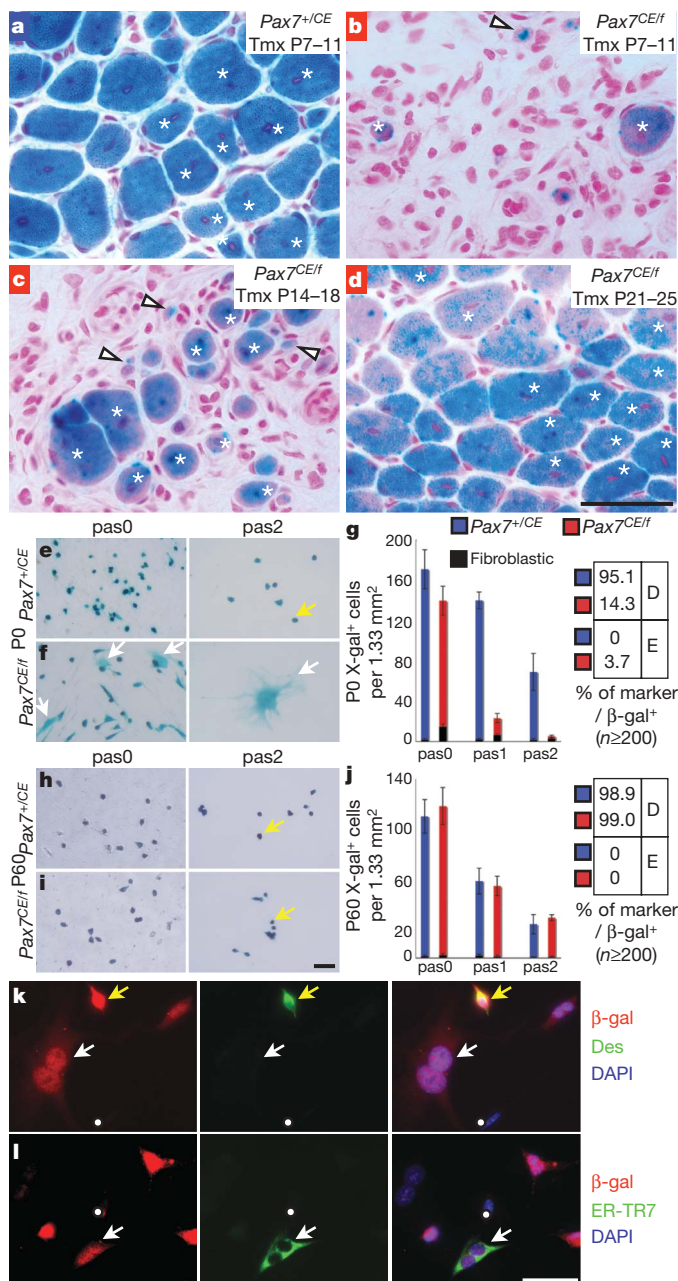
Figure 4 | Age-dependent intrinsic change of *Pax7* requirement.
a–d, Regeneration assessed at day 10 by β-gal and NFR staining. **a**, Control
(*Pax7+/CE*) regenerates of P7–11 tamoxifen-treated sample; essentially the
same for other time points (not shown). **b–d**, Conditional inactivation of
Pax7 (*Pax7CE/f*) by tamoxifen at P7–11 (**b**), P14–18 (**c**) and P21–25
(**d**), followed by injury at P21 for **b**, **c** and P26 for **d**. RT–PCR confirmed
complete *Pax7* inactivation (not shown). Asterisks indicate new fibres;
arrowheads indicate β-gal+ cells; *n* = 3 each; all are *R26R+/−*. **e–j**, P0
(**e–g**) and adult (P60, **h–j**) tamoxifen-treated (0.4 μM) and β-gal-stained
Pax7+/CE;R26R+/− (**e**, **h**) and *Pax7CE/f;R26R+/−* (**f**, **i**) myoblasts at passages 0
(pas0, left panels of **e**, **f**, **h**, **i**) and 2 (pas2, right panels of **e**, **f**, **h**, **i**). Yellow
arrows indicate round/short-spindled myoblasts; white arrows indicate
fibroblastic cells. **g**, **j**, Left: average cell numbers (≥9 fields) at each passage
(*x* axis); *n* = 3 per genotype per stage; error bars indicate s.d. Right:
percentages of desmin+ (D) and ER-TR7+ (E) of β-gal+ cells at passage 1
(*n* ≥ 200 each). **k**, **l**, Fluorescent microscopy of tamoxifen-treated P0
Pax7CE/f;R26R+/− myoblasts at passage 1. β-gal (left panels of **k**, **l**), desmin
(centre panel of **k**), ER-TR7 (centre panel of **l**) and merged with DAPI (right
panel of **k** is a merge of left and centre panels; right panel of **l** is a merge of left
and centre panels); yellow arrows indicate desmin+ cells; white arrows
indicate ER-TR7+ or fibroblastic cells; white dots indicate negative cells for
reference. Note β-gal staining and immunofluorescence in nuclei. Scale bars:
50 μm for **a–d** (shown in **d**), **e**, **f**, **h**, **i** (shown in **i**) and **k**, **l** (shown in **l**).

myoblasts had comparable expansive and myogenic potentials
(Fig. 4h–j) as well as myogenic differentiation capacity (Supplemen-
tary Fig. 10). No β-gal+ER-TR7+ cells were found despite the pre-
sence of rare fibroblastic cells (Fig. 4j). Tamoxifen-treated *cre/
Esr1+/−;Pax7f/f;Pax3f/f* adult myoblasts also had comparable prolifera-
tive and myogenic properties as mock-treated controls (Supplemen-
tary Fig. 11). Together, these data support a cell-autonomous change within
satellite cells compared to their progenitor state.

We have defined a critical period of *Pax7* dependency in the transi-
tion from muscle progenitor to adult stem-cell state, which ensures
that muscles achieve regenerative capacity. Given the essential roles
for *Pax3* and *Pax7* in embryonic and for *Pax7* alone in perinatal
myogenic progenitors, it was entirely unexpected that adult satellite
cells require neither *Pax7* nor *Pax3* for muscle regeneration. We
imagine that postnatal changes of muscle organization, mechanics
and physiology demand stem cells to alter their transcriptional
program as a means to adapt to these challenges. Changes in genetic
requirement for muscle stem cells from embryonic to juvenile to adult
stages elucidate the inadequacy of applying knowledge gained from
developmental studies to adult stem-cell biology. Our discovery
should encourage future investigations into how widespread genetic
transitions may occur in different adult stem-cell types. Age-depend-
ent differences in stem-cell properties should also urge careful con-
sideration of the age of stem cells used in transplantation-based
regenerative medicine.

METHODS SUMMARY

Animals. *Pax7f* and *Pax7CE* alleles are described in Supplementary Figs 1 and 2.
The *Pax3f* allele was described previously[19]. R26R (ref. 14) and cre/Esr1 (ref. 18)
mice were from the Jackson Laboratory. Tamoxifen (Sigma) was administered
intraperitoneally (animals >2 weeks) or subcutaneously (<2 weeks) at 3 mg per
40 g body weight per injection. Cardiotoxin (10 μM, Sigma) was injected into
tibialis anterior muscles (after anaesthesia) at 100 μl (animals >2 months) and
50 μl (<2 months). EdU (Invitrogen) was injected at 0.1 mg per 20 g bodyweight
per injection. All procedures were approved by IACUC.
PCR genotyping, RT–PCR and western blot. Primers used for genotyping and
RT–PCR are in Supplementary Tables 1 and 2. Antibodies used for western blots
are in Supplementary Tables 3 and 4, and ECL (Amersham) was used for detection.
X-galactosidase reactions, immunofluorescence and histology. Tibialis
anterior muscles were fresh frozen in isopentane/liquid nitrogen and cryo-sectioned
at 10 μm. X-galactosidase (Qiagen) reactions followed standard procedure[22];
counterstained by NFR (Lab Vision). Immunofluorescence was performed using
antibodies in Supplementary Tables 3 and 4. Histology was performed according to
manufacturer's instructions (Surgipath). Edu was detected by Alexa Fluor 649
(Invitrogen).
Cell culture. Myoblasts were isolated and cultured on collagen-coated dishes
(VWR) as described[23]. 4-OH tamoxifen (0.4 μM, Calbiochem) was added at the
beginning of culture. For quantitative studies, cells were passed (1:3 dilution)
every 4 days. For immunofluorescence, cells were plated onto 8-well chamber
slides.
Muscle fibre size. Digital images of X-galactosidase and histologically stained
sections were processed by Metamorph software (Molecular Devices). Parameters
were chosen to define muscle fibre boundaries, which were re-examined visually
for accuracy before quantification. Determination of fibre diameter and inner
radius was assisted by Metamorph. Measurements for 300 fibres per animal were
subjected to statistical analysis by Excel (Microsoft).

Full Methods and any associated references are available in the online version of
the paper at www.nature.com/nature.

1. Relaix, F., Rocancourt, D., Mansouri, A. & Buckingham, M. A. Pax3/Pax7-
dependent population of skeletal muscle progenitor cells. *Nature* 435, 948–953
(2005).
2. Seale, P. et al. Pax7 is required for the specification of myogenic satellite cells. *Cell*
102, 777–786 (2000).
3. Oustanina, S., Hause, G. & Braun, T. Pax7 directs postnatal renewal and
propagation of myogenic satellite cells but not their specification. *EMBO J.* 23,
3430–3439 (2004).
4. Kuang, S., Charge, S. B., Seale, P., Huh, M. & Rudnicki, M. A. Distinct roles for Pax7
and Pax3 in adult regenerative myogenesis. *J. Cell Biol.* 172, 103–113 (2006).

5. Relaix, F. *et al.* Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *J. Cell Biol.* **172**, 91–102 (2006).
6. Zammit, P. S. *et al.* Pax7 and myogenic progression in skeletal muscle satellite cells. *J. Cell Sci.* **119**, 1824–1832 (2006).
7. Olguin, H. C., Yang, Z., Tapscott, S. J. & Olwin, B. B. Reciprocal inhibition between Pax7 and muscle regulatory factors modulates myogenic cell fate determination. *J. Cell Biol.* **177**, 769–779 (2007).
8. Seale, P., Ishibashi, J., Scime, A. & Rudnicki, M. A. Pax7 is necessary and sufficient for the myogenic specification of CD45$^+$:Sca1$^+$ stem cells from injured muscle. *PLoS Biol.* **2**, E130 (2004).
9. Guo, Q., Loomis, C. & Joyner, A. L. Fate map of mouse ventral limb ectoderm and the apical ectodermal ridge. *Dev. Biol.* **264**, 166–178 (2003).
10. Charge, S. B. & Rudnicki, M. A. Cellular and molecular regulation of muscle regeneration. *Physiol. Rev.* **84**, 209–238 (2004).
11. Schienda, J. *et al.* Somitic origin of limb muscle satellite and side population cells. *Proc. Natl Acad. Sci. USA* **103**, 945–950 (2006).
12. Kuang, S., Kuroda, K., Le Grand, F. & Rudnicki, M. A. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**, 999–1010 (2007).
13. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757 (1997).
14. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
15. Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S. & Blau, H. M. Self-renewal and expansion of single transplanted muscle stem cells. *Nature* **456**, 502–506 (2008).
16. Cerletti, M. *et al.* Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134**, 37–47 (2008).
17. Bigard, A. X. *et al.* Changes in myosin heavy chain profile of mature regenerated muscle with endurance training in rat. *Acta Physiol. Scand.* **165**, 185–192 (1999).
18. Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* **244**, 305–318 (2002).
19. Koushik, S. V., Chen, H., Wang, J. & Conway, S. J. Generation of a conditional *loxP* allele of the *Pax3* transcription factor that enables selective deletion of the homeodomain. *Genesis* **32**, 114–117 (2002).
20. Pastoret, C. & Sebille, A. Age-related differences in regeneration of dystrophic (mdx) and normal muscle in the mouse. *Muscle Nerve* **18**, 1147–1154 (1995).
21. Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807–810 (2007).
22. Hogan, B., Beddington, R., Costantini, F. & Lacey, E. *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, 1994).
23. Springer, M. L., Rando, T. A. & Blau, H. M. Gene delivery to muscle. *Curr. Protoc. Hum. Genet.* Unit 13.4 doi:10.1002/0471142905.hg1304s31 (2002).

METHODS

Animals. The experimental outlines for making the *Pax7^f* and *Pax7^{CE}* alleles are in Supplementary Figs 1 and 2, respectively. The precise constructions for these alleles will be described elsewhere. The genetic crosses to remove the neo selection marker with actin-Flip mice are described (Supplementary Figs 1 and 2). Generation of the germline *Pax7$^\Delta$* allele with Nestin-Cre mice is also outlined (Supplementary Fig. 1). Methods for genotyping and characterization of these alleles are described in the Supplementary figure legends.

To obtain *Pax7$^{CE/f}$;R26R$^{+/-}$* and *Pax7$^{+/CE}$;R26R$^{+/-}$* progenies for analysis, *Pax7$^{+/CE}$* mice were mated to *Pax7$^{+/f}$;R26R$^{+/-}$* animals. To obtain *Pax7$^{CE/CE}$;R26R$^{+/-}$* animals, *Pax7$^{+/CE}$;R26R$^{+/-}$* mice were mated to *Pax7$^{+/CE}$* animals. Similar strategies of genetic crosses were used (excluding *R26R$^+$*) to obtain *cre/Esr1$^{+/-}$;Pax3$^{f/f}$;Pax7$^{f/f}$* and *cre/Esr1$^{+/-}$;Pax7$^{+/f}$* animals. Genotypes of these animals were determined by PCR using primers and conditions in Supplementary Table 1.

Tamoxifen. Tamoxifen (Sigma) was prepared by dissolving a freshly opened bottle of 5 g of tamoxifen in corn oil (Sigma) at 20 mg ml^{-1}. The mixture was incubated at 37 °C with periodic vortexing until tamoxifen was completely dissolved (~2–4 h). It was then divided into 10-ml conical tubes (VWR) and frozen in a −80 °C freezer. Each tube was thawed as needed and kept at 4 °C to be used for no more than 1 week. A 1-ml syringe with a 26-gauge needle was used for injection. Either intraperitoneal (for animals more than 3 weeks old) or subcutaneous (for animals less than 3 weeks old) delivery routes were used. The regimens are outlined in Figs 1a and 2i and Supplementary Fig. 8a.

Cardiotoxin. Cardiotoxin (Sigma) was prepared by dissolving a freshly opened tube in PBS at 10 μM. The solution was divided into Eppendorf tubes as 1-ml aliquot stocks, flash frozen, and stored at −80 °C. Each tube was thawed fresh before injection and not re-used. Animals were anaesthetized by intraperitoneal injection of Avertin (2,2,2-tribromoethanol from TCI America at 15 μl per gram body weight of 20 mg ml^{-1} solution). Right-leg tibialis anterior muscles were injected with 100 or 50 μl of cardiotoxin using an insulin needle (3/10cc Insulin Syringe from Becton Dickinson). For two rounds of injury, right-leg tibialis anterior muscles were initially injected with 100 μl of cardiotoxin and allowed to regenerate for 2 weeks. Following this, right-leg tibialis anterior muscles were re-injured via injection of 50 μl of cardiotoxin (see Fig. 2i). The volume and concentration of cardiotoxin are standard protocol used in published work[24–27]. Cardiotoxin animals were kept under a warming lamp until recovery before being returned to a normal cage rack.

EdU. For cultured cells, EdU supplied in the Click-iT EdU cell proliferation assay kit (Invitrogen) was used as instructed with minor modifications: EdU was used at 10 μM for 8 h in culture media; for detection, we performed immunostaining for primary and secondary antibodies (see below for a detailed protocol) first, then the click chemical reaction using Alexa Fluor 647 as a reactive fluorophore for detection, followed by DAPI staining (at 1 μg ml^{-1}).

For animals, EdU powder was dissolved in PBS at 0.5 mg ml^{-1} and frozen at −20 °C for long-term storage. EdU was injected at 0.1 mg per 20 g bodyweight intraperitoneally. The regimen for *in vivo* EdU injection for short-term proliferation and long-term retention assays are outlined in corresponding figures (Fig. 2i and Supplementary Fig. 8a, respectively). As for cultured cells, we performed the click chemical reaction using the same kit after immunostaining with primary and secondary antibodies and before DAPI staining (when applicable).

PCR genotyping and RT–PCR. For adult animals, tail DNA was used for genotyping by PCR. For newborns and animals younger than 21 days, toe DNA was used. DNA was extracted using the ExtractN'Amp kit (Sigma) following the manufacturer's instruction. One microlitre of neutralized DNA samples was used for PCR reaction using GoTaq polymerase (Promega) with buffers supplied by the manufacturer with 0.1 mM dNTPs and 2.5 mM MgCl$_2$. The PCR products were resolved in 2% agarose gel, stained with 0.5 μg ml^{-1} ethidium bromide (Gibco), and digitally imaged with a Bio-Rad Gel Doc system for record keeping.

Total RNA from injured tibialis anterior muscles was isolated using TRIzoL Reagent (Invitrogen) according to manufacturer's instructions. cDNA synthesis was performed using M-MLV Reverse Transcriptase (Invitrogen) also according to manufacturer's instructions. cDNA samples were then used for PCR as above.

Primer sequences, product sizes and PCR conditions used for genotyping and RT–PCR are in Supplementary Tables 1 and 2.

P0 animals to be used for satellite cell isolation were genotyped in the morning and killed in the afternoon according to their determined genotypes.

Western blot. Tibialis anterior muscles were harvested (time as indicated in each figure) and snap frozen in 1.5-ml Eppendorf tubes and stored at −80 °C until use. For protein extraction, frozen samples were weighed and suspended in 1:9 (weight to volume) of ice-cold RIPA buffer containing protease inhibitor cocktail (Roche) at 2× concentration and 1 mM PMSF, and ground with a sterile plastic mini-pestle (VWR) until no muscle bits were observable. The muscle homogenate was spun for 10 min in a microcentrifuge at maximum speed at 4 °C. 4× SDS sample buffer was added to the supernatant to reach 1× concentration. Samples were boiled for 5 min before being subjected to 7.5% SDS–PAGE. Kaleidoscope wide-range molecular mass marker (Bio-Rad) was used for size. Western transfer was performed with low molecular mass transfer buffer to an ECL membrane (Amersham) overnight at 30 V using a Bio-Rad mini-protein II transfer set up at 4 °C. Membrane was blocked in 5% low-fat carnation milk powder in TBS for 30 min, incubated with primary antibody in blocking solution overnight at 4 °C (mouse IgG1 anti-Pax7 at 1:50 and mouse IgG1 anti-α-tubulin (DM1α, from Sigma) at 1:10,000), washed in TBS, 3×, 10 min each, incubated with secondary antibody (HRP-conjugated anti-mouse IgG1 from Zymed at 1:3,000) in blocking solution for 1 h, washed in TBS, 3×, 10 min each, and then subjected to ECL reaction using the kit from Amersham (ECL Western Blotting Detection Reagents) and exposed to X-ray films. Standard solutions prepared for SDS–PAGE, western transfer, TBS and 4× sample buffer were as described previously[28].

X-galactosidase reaction and histology. Ten-micrometre frozen muscle sections (by Leica CM3000 Cryostat) were collected on Superfrost plus slides (VWR) and air dried for 30–90 min. Slides were either stored in a −20 °C freezer or used immediately.

For X-galactosidase reactions, sections were fixed with 0.2% glutaraldehyde/0.1 M phosphate buffer (pH 7.2) with 5 mM EGTA, 1 mM MgCl$_2$, for 5 min on ice, then washed and stained according to ref. 22 for 8–24 h to obtain satisfactory staining intensity. Stained sections were then washed in H$_2$O and stained with NFR (Lab Vision) for 5 min, rinsed with H$_2$O, dehydrated through graded EtOH series (25%, 50%, 75%, 95%, 100%), then in xylene and finally mounted in Permount mounting media (Fisher) with a coverslip (VWR).

For histology, sections were fixed in 4% paraformaldehyde/0.1 M phosphate (pH 7.2) for 30 min, rinsed with tap water, stained with Gill II haematoxylin for 5 min, washed with tap water, treated with Scott's tap water substitute for 10–15 s until staining colour turned blue, washed with tap water, stained with eosin for 1 min, destained in 95% EtOH, followed by 100% EtOH, xylene, then mounted in Permount with a coverslip. Staining reagents were purchased from Surgipath and used as instructed. Van Gieson stains were performed precisely as described[29].

Digital images of X-galactosidase stained and histological sections were taken using a SPOT camera and a Nikon Eclipse E800 vertical microscope.

Immunostaining. Freshly prepared 4% paraformaldehyde/0.1 M phosphate buffer (pH 7.2) was filtered through a 0.2-μm polyethersulphone filter (VWR) and chilled on ice. Paraformaldehyde solution was used within 12 h. Ten-micrometre frozen sections on Superfrost plus slides were fixed for exactly 12 min on ice, rinsed with 0.1 M phosphate buffer 3×, 5 min each, then permeabilized with 0.3% Triton X-100/PBS for 20 min. Sections were then incubated with mouse IgG blocking solution from the M.O.M. kit (Vector Lab) diluted in PBS/0.01% Triton X-100 following the dilution instructed in the manual. They were then incubated in blocking solution provided in the M.O.M. kit with an additional 15% goat serum in PBS/0.01% Triton X-100 for 20 min. Slides were then incubated with primary antibodies diluted (sources and dilution factors for each antibody are in Supplementary Table 3) in the blocking solution overnight at 4 °C. Next morning, the slides were washed 3× in PBS/0.01% Triton X-100 (all washes used this solution), 5 min each. Fluorescently conjugated secondary antibodies to each species-specific IgG in blocking solution was applied at dilutions in Supplementary Table 4 for 30 min. For Pax7 and M-Cad monoclonal IgG1 antibodies, the anti-IgG1 isotype-specific secondary antibody gave us the best result. The slides were washed again as after the primary antibody. If EdU was to be detected, we performed the click chemical reaction after this step according to manufacturer's instructions using the Click-iT kit components (Invitrogen). The slides were then washed, incubated with DAPI when applicable (at 1 μg ml^{-1}) for 10 min, washed two more times, and mounted in FluoromountG solution (Southern Biotechnology) with a coverslip. All images were taken under an Axioscope equipped with AxioCam with filters of non-overlapping spectra (according to the fluorophores used) for multiple fluorescent signals. Images were pseudo-coloured and superimposed (when necessary) using the MetaMorph program.

For cultured cells, 8-well chamber slides were fixed and processed the same way as sections except that the mouse IgG blocking (M.O.M. kit) step was omitted.

Myoblast cell culture. Genotyped animals were killed by cervical dislocation, and hindlimb muscles were removed and minced by razor blades. Each sample was digested with collagenase and dispase as described in ref. 23. Dissociated muscles were then transferred to 15 ml conical tubes with 10 ml myoblast media without FGF, spun down in the clinical centrifuge and re-suspended in culture media with 2.5 ng ml^{-1} bFGF.

For cell counting and marker analysis, 0.4 μM 4-OH tamoxifen (Calbiochem) was applied at the same time of cell plating. Each sample was divided into three parts. One part was plated on a 3-cm collagen-coated dish (prepared as described in ref. 23) for cell counting 4 days later. One part was plated on another 3-cm collagen-coated dish but used for sequential passaging at 1:3 dilution. The remaining part was plated into 8-well chamber slides (coated with collagen) for immunostaining. For passaging, cells were passaged at 1:3 dilution every 4 days. At each round, each sample was divided into three parts; the same routine as above for cell counting, continuous passaging and immunostaining was repeated. For quantification of β-gal$^+$ cell numbers, the 3-cm dish was fixed and subjected to X-galactosidase staining (see above), and nine random fields were imaged using the Canon EOS30D camera attached to the Zeiss SV11 APO microscope at the highest magnification (each frame is of 1.33 mm^2 area, 0.95 mm × 1.4 mm), and β-gal$^+$ cells (blue) counted and scored for their morphology. Cells that displayed large flattened multi-cellular processes were scored as 'fibroblastic'. For immunostaining quantification, 9–16 random fields per well were chosen for imaging and quantified for EdU labelling or marker expression in conjunction with anti-β-gal antibody staining.

For myogenic marker analysis, myoblast cultures were first treated with or without 4-OH tamoxifen for 2 days for gene inactivation. They were then assayed for myoblast markers (Pax3, Pax7, desmin or MyoD) by RT–PCR or immunostaining, or incubated with Edu (10 μM for 8 h) for proliferation assays. For differentiation, cells were switched to differentiation media (5% horse serum in DME; ref. 23) for 3–4 days and assayed for myogenin and MHC expression.

Muscle fibre quantification. Cross-sections (by cryostat sections at 10 μm) of tibialis anterior muscles were stained by X-galalactosidase reactions or by haematoxylin and eosin. Digital images were acquired with a CCD camera by routine brightfield microscopy. Image processing and morphometry were carried out using MetaMorph software (Molecular Devices). Specifically, myofibres were automatically delineated. Myofibre outlines were visually examined

to assure accuracy before further analysis. The centroid (geometric centre) of each myofibre was then determined to calculate the myofibre inner radius, the distance from the centroid to the closest edge. A circle of this radius will be completely contained (inscribed) within the myofibre outline. Therefore, this is the most conservative measure of fibre size.

For myofibre diameter measurements, we used the 'largest diameter of the lesser aspect of the myofibre' criterion[30]. This measurement is conventionally used to assess fibre size. Automatic measurement of this dimension for outlined myofibres was also implemented in MetaMorph (termed 'Breadth' within the software).

Statistical analysis. Quantitative data displayed as histograms are expressed as means ± standard error of the mean (represented as error bars). Quantification of total percentages is shown in tables. Regenerated muscle fibre numbers are displayed as a distribution plot. ANOVA was conducted using an Excel spreadsheet by using a t-test for two-tailed paired comparison. Statistical significance was set at a P value <0.05.

24. Hu, P., Geles, K. G., Paik, J.-H., DePinho, R. A. & Tijan, R. Codependent activators direct myoblast-specific MyoD transcription. *Dev. Cell* **15**, 534–546 (2008).
25. Melcon, G. *et al.* Loss of emerin at the nuclear envelope disrupts the Rb1/E2F and MyoD pathways during muscle regeneration. *Hum. Mol. Genet.* **15**, 637–651 (2006).
26. Iezzi, S. *et al.* Deacetylase inhibitors increase muscle cell size by promoting myoblast recruitment and fusion through induction of follistatin. *Dev. Cell* **6**, 673–684 (2004).
27. Meeson, A. P. *et al.* Cellular and molecular regulation of skeletal muscle side population cells. *Stem Cells* **22**, 1305–1320 (2004).
28. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
29. Bancroft, J. D. & Cook, H. D. *Manual of Histological Techniques and their Diagnostic Application* (Churchill Livingstone, 1994).
30. Dubowitz, V. *Muscle Biopsy: A Practical Approach* (Bailliere Tindall, 1985).

LETTERS

Presenilins are essential for regulating neurotransmitter release

Chen Zhang[1,2], Bei Wu[1], Vassilios Beglopoulos[1], Mary Wines-Samuelson[1], Dawei Zhang[1], Ioannis Dragatsis[3], Thomas C. Südhof[2] & Jie Shen[1]

Mutations in the presenilin genes are the main cause of familial Alzheimer's disease. Loss of presenilin activity and/or accumulation of amyloid-β peptides have been proposed to mediate the pathogenesis of Alzheimer's disease by impairing synaptic function[1–5]. However, the precise site and nature of the synaptic dysfunction remain unknown. Here we use a genetic approach to inactivate presenilins conditionally in either presynaptic (CA3) or postsynaptic (CA1) neurons of the hippocampal Schaeffer-collateral pathway. We show that long-term potentiation induced by theta-burst stimulation is decreased after presynaptic but not postsynaptic deletion of presenilins. Moreover, we found that presynaptic but not postsynaptic inactivation of presenilins alters short-term plasticity and synaptic facilitation. The probability of evoked glutamate release, measured with the open-channel NMDA (N-methyl-D-aspartate) receptor antagonist MK-801, is reduced by presynaptic inactivation of presenilins. Notably, depletion of endoplasmic reticulum Ca^{2+} stores by thapsigargin, or blockade of Ca^{2+} release from these stores by ryanodine receptor inhibitors, mimics and occludes the effects of presynaptic presenilin inactivation. Collectively, these results indicate a selective role for presenilins in the activity-dependent regulation of neurotransmitter release and long-term potentiation induction by modulation of intracellular Ca^{2+} release in presynaptic terminals, and further suggest that presynaptic dysfunction might be an early pathogenic event leading to dementia and neurodegeneration in Alzheimer's disease.

Conditional inactivation of presenilins in excitatory neurons of the mouse postnatal forebrain causes synaptic dysfunction, memory impairment and age-dependent neurodegeneration[3,6]. Before the onset of neurodegeneration, paired-pulse facilitation, long-term potentiation (LTP) and NMDA receptor (NMDAR)-mediated responses are altered[3], suggesting that synaptic defects caused by the loss of presenilins may be a cellular precursor of neuronal cell death. To determine the precise synaptic site of presenilin function, we performed a systematic genetic analysis by the restriction of presenilin inactivation to hippocampal CA1 or CA3 neurons. This strategy allowed selective examination of the effects of presenilin inactivation in either presynaptic or postsynaptic neurons of the Schaeffer-collateral pathway.

We crossed homozygous floxed presenilin 1 (*Psen1*), presenilin 2 (*Psen2*)-null mice (known as *fPsen1/fPsen1;Psen2⁻/⁻*) to *Camk2a-Cre*[7] and *Grik4-Cre* (also known as *KA1-Cre*)[8] transgenic mice to produce CA1- and CA3-restricted presenilin conditional double knockout (*Psen* cDKO) mice. *In situ* hybridization confirmed the selective loss of *Psen1* expression in CA1 and CA3 neurons of CA1- and CA3-*Psen* cDKO mice, respectively, at 2 months of age (Fig. 1a). We also crossed *Camk2a-Cre* and *Grik4-Cre* mice to *Rosa26-lacZ*

reporter transgenic mice, and observed the expected patterns of CA1- and CA3-restricted β-galactosidase expression (Fig. 1b).

We next examined the effect of selective presenilin inactivation in CA1 or CA3 neurons on theta-burst stimulation (TBS)-induced LTP, which is impaired in *Psen* cDKO mice lacking presenilin in both CA3 and CA1 neurons[3]. Surprisingly, TBS-induced LTP is normal in CA1-*Psen* cDKO mice but is markedly impaired in CA3-*Psen* cDKO mice (Fig. 1c). Thus, presynaptic but not postsynaptic presenilins are required for TBS-induced LTP. To determine whether postsynaptic NMDAR-mediated responses are affected in these mutant mice, we measured AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor (AMPAR)- and NMDAR-dependent synaptic responses, but detected no change in the NMDAR/AMPAR ratio in CA3- or CA1-*Psen* cDKO mice (Fig. 1d). Moreover, input–output curves of NMDAR-dependent responses are normal in CA3- and CA1-*Psen* cDKO mice (Fig. 1e). Thus, loss of presenilin in either presynaptic or postsynaptic neurons alone is insufficient to impair NMDAR-mediated responses. Similarly, input–output coupling (Supplementary Fig. 1) and the current–voltage (*I–V*) relationship (Supplementary Fig. 2) of AMPAR-mediated synaptic responses are normal in CA3-*Psen* cDKO mice. These results demonstrate that LTP deficits caused by presynaptic presenilin inactivation are not due to impaired postsynaptic receptor-mediated responses.

We thus investigated whether presynaptic activity is impaired during LTP induction, which could account for the observed LTP deficit. Indeed, we found that short-term depression during the initial stimulus train of TBS is increased in CA3-*Psen* cDKO mice (Fig. 2a). Paired-pulse facilitation and synaptic frequency facilitation are reduced in CA3-*Psen* cDKO mice but are normal in CA1-*Psen* cDKO mice (Fig. 2b, c), which are confirmed by whole-cell recordings (Supplementary Figs 3 and 4). Moreover, the deficit in synaptic facilitation in CA3-*Psen* cDKO mice is calcium-dependent and is rescued by higher external Ca^{2+} concentrations (Fig. 2d and Supplementary Fig. 5). Consistent with previous reports[3,9], inactivation of Psen1 or Psen2 alone is insufficient to alter frequency facilitation or paired-pulse facilitation (Supplementary Figs 6 and 7). The replenishment of the readily releasable pool after depletion is also normal in CA3-*Psen* cDKO mice (Supplementary Fig. 8), arguing against an impairment of synaptic vesicle recycling as a cause of the decreased synaptic facilitation.

To test directly whether presynaptic inactivation of presenilins alters the probability of glutamate release, we measured the overall release probability using the open channel blocker MK-801, which irreversibly blocks NMDARs after each synaptic release event[10,11]. Thus, during low-frequency stimulation in the presence of MK-801 and AMPAR blockade, the rate at which NMDAR-mediated synaptic responses decline reflects the average release probability of the
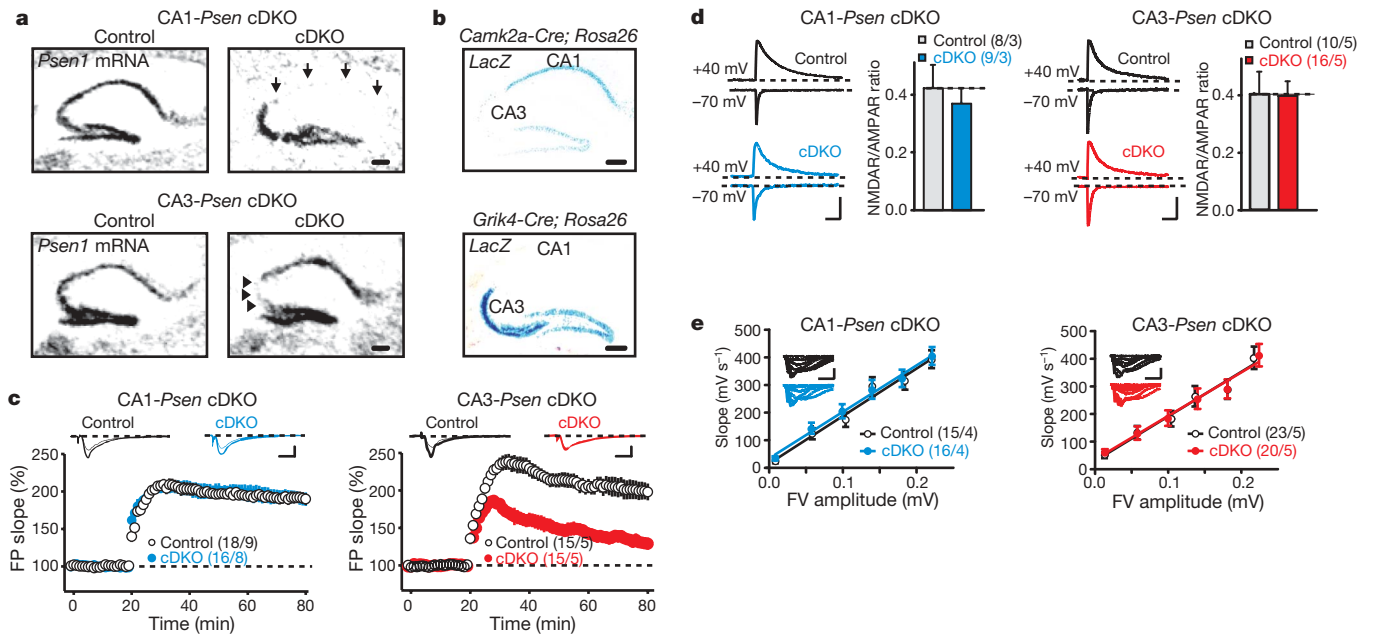
Figure 1 | Impaired LTP in CA3-*Psen* but not CA1-*Psen* cDKO mice. a, *In situ* hybridization shows loss of *Psen1* messenger RNA in CA1 (arrows) and CA3 (arrowheads) neurons in CA1-*Psen* and CA3-*Psen* cDKO mice, respectively. Scale bars, 200 μm. **b**, X-gal staining shows absence of Cre-mediated recombination in CA3 and CA1 neurons of *Camk2a-Cre; Rosa26-lacZ* and *Grik4-Cre; Rosa26-lacZ* mice, respectively. Scale bars, 200 μm. **c**, TBS-induced LTP in CA1-*Psen* cDKO (filled blue circles) and CA3-*Psen* cDKO (filled red circles) compared to their controls (open circles). Representative traces before (thin) and after (thick) LTP induction are shown. Superimposed traces are averages of four consecutive responses 1 min before and 60 min after TBS. FP, field potential. Scale bars, 10 ms, 1 mV. **d**, Normal ratio of NMDAR to AMPAR responses in CA3-*Psen* and CA1-*Psen* cDKO mice. Scale bars, 200 ms, 200 pA. **e**, NMDAR-mediated input–output curves. FV, fibre volley. Scale bars, 40 ms, 1 mV. All data represent mean and s.e.m. The values in parentheses indicate the number of hippocampal neurons or slices (left) and the number of mice (right) used in each experiment.
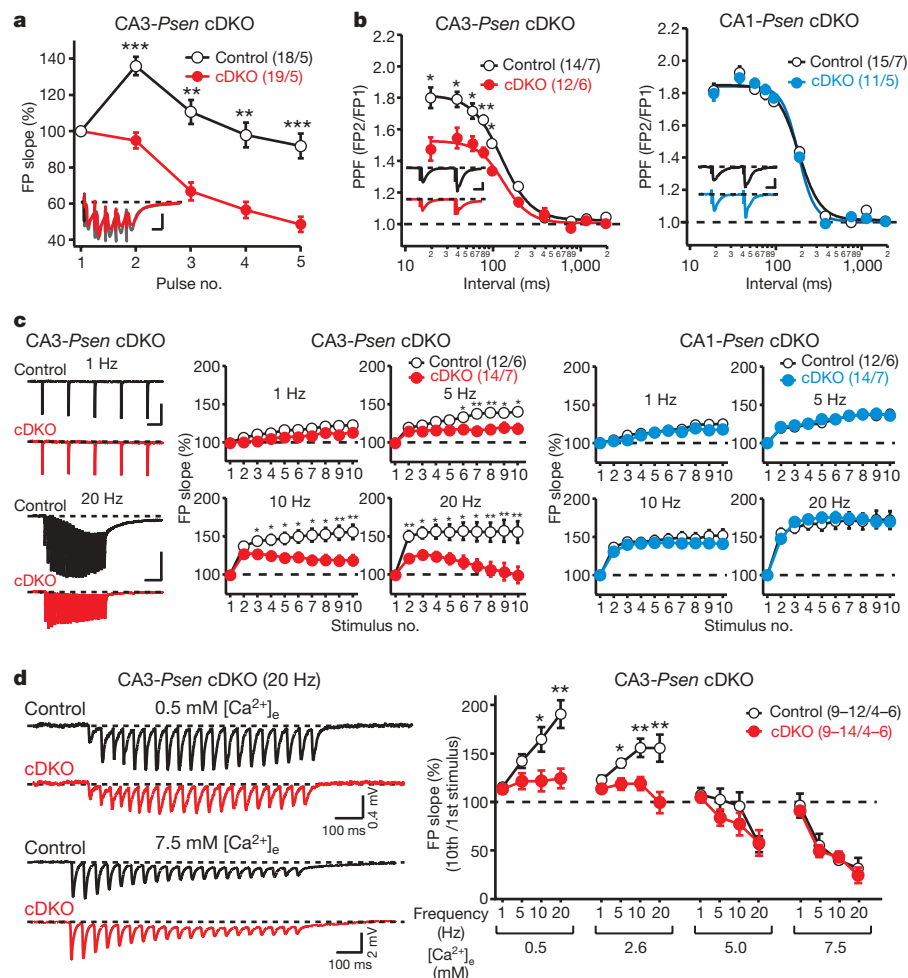

Figure 2 | Presynaptic defects in CA3-*Psen* but not CA1-*Psen* cDKO mice. a, Reduced facilitation of field excitatory postsynaptic potential (fEPSP) slope during single TBS in CA3-*Psen* cDKO mice. Inset shows representative traces of field responses during single TBS stimulus train. FP, field potential. Scale bar, 10 ms, 2 mV. **b**, Paired-pulse facilitation (PPF) in CA3- and CA1-*Psen* cDKO mice. Scale bars in insets, 10 ms, 0.5 mV. **c**, Synaptic facilitation elicited by stimulus trains of indicated frequencies in CA3- and CA1-*Psen* cDKO mice. Scale bars, 2 mV, 500 ms (top); 2 mV, 250 ms (bottom). **d**, Calcium-dependence of frequency facilitation defects in CA3-*Psen* cDKO mice. Scale bars, 0.4 mV, 100 ms (top); 2 mV, 100 ms (bottom). $[Ca^{2+}]_e$ denotes extracellular Ca^{2+} concentration. All data represent mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. The values in parentheses indicate the number of hippocampal slices (left) and the number of mice (right) used in each experiment.

synapses. We found that the decay rate of postsynaptic responses as a function of stimulus number is decreased in CA3-*Psen* cDKO mice (Fig. 3a). When these results were fitted to a single exponential as a rough measure of the average release probability, we observed an
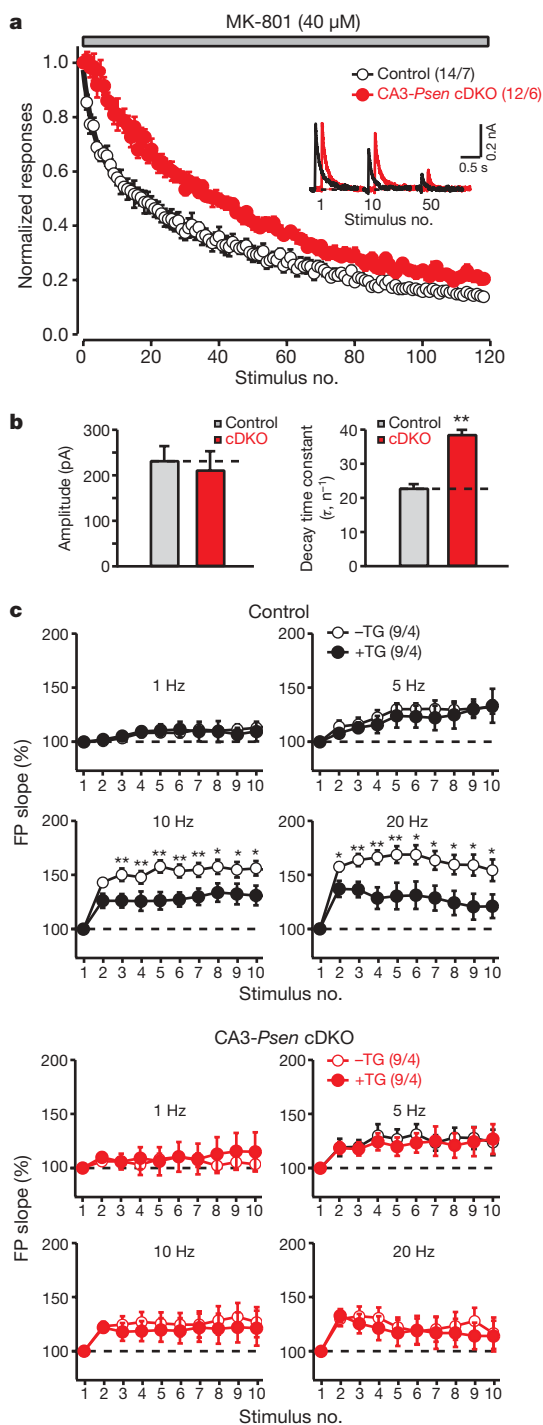

Figure 3 | Presynaptic presenilin regulates glutamate release by intracellular Ca^{2+} stores. a, Reduced decay rate of NMDAR-mediated responses in the presence of MK-801 in CA3-*Psen* cDKO mice. Representative traces of EPSCs after the first, tenth and fiftieth stimulus are shown in the inset. **b,** Amplitude of the first NMDAR-mediated response (left) and the decay time constant (right; fitted to a single exponential curve) of the NMDAR-mediated EPSC amplitude in the presence of MK-801. **c,** Effects of thapsigargin (TG) treatment on synaptic facilitation in control and CA3-*Psen* cDKO slices. All data represent mean and s.e.m. *$P < 0.05$, **$P < 0.01$. The values in parentheses indicate the number of hippocampal slices or neurons (left) and the number of mice (right) used in each experiment.

almost twofold increase in the decay constant in CA3-*Psen* cDKO mice (Fig. 3b). This result shows a major decrease in release probability in CA3-*Psen* cDKO mice, demonstrating a critical role for presenilins in regulating the probability of glutamate release. Spontaneous miniature excitatory postsynaptic currents (EPSCs), however, show a normal frequency and amplitude in CA3-*Psen* cDKO mice (Supplementary Fig. 9), suggesting that a defect in Ca^{2+}-dependent release may account for the observed presynaptic phenotypes.

Evoked neurotransmitter release is dependent on the local increase in intracellular calcium concentrations. Presynaptic Ca^{2+} increases are caused by Ca^{2+} influx by voltage-gated calcium channels (VGCCs) and by calcium release from intracellular stores[12]. Because changes in Ca^{2+} influx by VGCCs have been reported to affect release probability[10,11], we measured VGCC currents in the somata of CA3 neurons and found an unaltered $I–V$ relationship in CA3-*Psen* cDKO mice (Supplementary Fig. 10). Thus, the change in release probability in CA3-*Psen* cDKO mice is unlikely to be due to VGCC dysfunction. Because presenilins have been reported to be involved in the regulation of Ca^{2+} homeostasis in intracellular stores[13–16], we examined the effect of depletion of intracellular Ca^{2+} stores on synaptic facilitation in CA3-*Psen* cDKO mice. Thapsigargin, which irreversibly blocks Ca^{2+} pumps on the endoplasmic reticulum (ER), thereby abolishing intracellular Ca^{2+} release[17], suppresses synaptic facilitation during high-frequency stimulation in control synapses, but has no discernable effect in presenilin-deficient nerve terminals (Fig. 3c). Thus, thapsigargin treatment mimics and occludes the effect of presenilin inactivation on synaptic facilitation, suggesting that dysregulation of intracellular Ca^{2+} release underlies the presynaptic defects in CA3-*Psen* cDKO mice.

Calcium release from the ER is mediated by two main types of receptors: ryanodine receptors (RyRs), which mediate calcium-induced calcium release (CICR), and inositol-1,4,5-triphosphate receptors (InsP3Rs). We therefore tested the effects of specific inhibitors for RyRs or InsP3Rs on synaptic facilitation[18–20]. Blockade of RyRs by ryanodine (100 μM) or dantrolene mimics the effect of thapsigargin (Fig. 4a and Supplementary Fig. 11), whereas blockade of InsP3Rs by xestospongin C has no effect (Supplementary Fig. 12). Thus, a specific defect in RyR-mediated CICR probably underlies the presynaptic impairment in CA3-*Psen* cDKO mice.

To determine directly whether Ca^{2+} homeostasis is indeed affected by presenilin inactivation, we performed Ca^{2+} imaging in cultured hippocampal neurons, in which presenilin is acutely inactivated with a lentivirus expressing Cre recombinase. This postnatal culture system circumvents the requirement of presenilins in neurogenesis during embryonic development[21,22] and permits the direct measurement of Ca^{2+} concentrations in these neurons. Presenilin expression is abolished in Cre-infected (*Psen* cDKO) neurons, but their neuronal and synaptic morphology appear normal (Supplementary Fig. 13a, c). Similar to CA3-*Psen* cDKO mice, presynaptic short-term plasticity measured as paired-pulse ratio is altered in *Psen* cDKO hippocampal neurons (Supplementary Fig. 13b), confirming that this preparation recapitulates the presynaptic defect of the presenilin-deficient hippocampus. We then measured cytosolic calcium concentration ($[Ca^{2+}]_i$) changes elicited by depolarization (80 mM KCl), which are contributed by both Ca^{2+} influx through VGCCs and Ca^{2+} efflux from intracellular stores. The amplitude of $[Ca^{2+}]_i$ changes ($\Delta[Ca^{2+}]_i$) elicited by depolarization is reduced in *Psen* cDKO neurons (Fig. 4b and Supplementary Fig. 14). Blockade of RyRs with ryanodine (100 μM) in control neurons mimics the effect of presenilin inactivation, whereas ryanodine has no further effect in *Psen* cDKO neurons (Fig. 4b and Supplementary Fig. 14). Thus, blockade of RyRs mimics and occludes the effect of presenilin inactivation on depolarization-induced $[Ca^{2+}]_i$ changes. Blockade of InsP3R with xestospongin C, however, has no effect (Fig. 4b and Supplementary Fig. 14). These results directly show that presenilin inactivation in
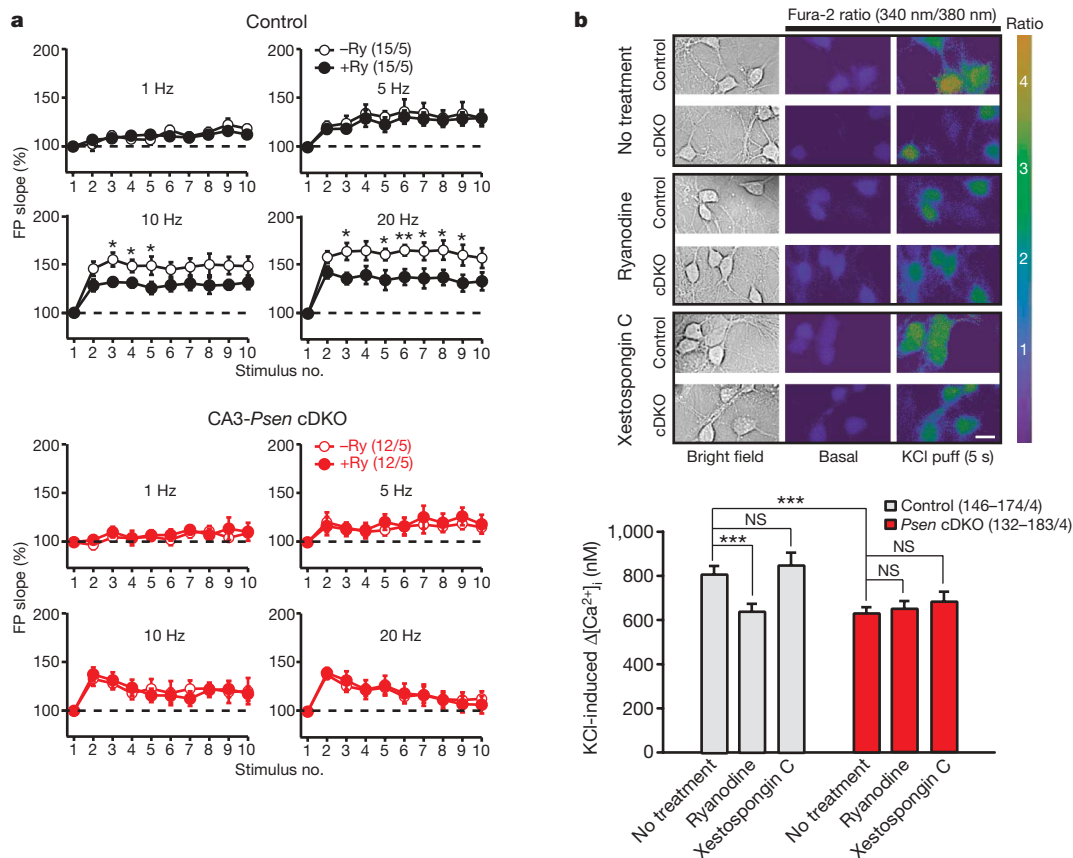
Figure 4 | Blockade of RyRs mimics and occludes the defects in synaptic facilitation and calcium homeostasis in CA3-*Psen* cDKO hippocampal slices and cultured *Psen* cDKO hippocampal neurons. a, Effect of ryanodine (Ry; $100\,\mu M$) treatment on synaptic facilitation in control and CA3-*Psen* cDKO mice. **b**, Effect of ryanodine ($100\,\mu M$) or xestospongin C ($1\,\mu M$) treatment on depolarization-induced $[Ca^{2+}]_i$ increases in cultured hippocampal neurons. Representative calcium images (top) show high potassium (80 mM)-induced Ca^{2+} responses in control and *Psen* cDKO neurons. Scale bar, $20\,\mu m$. All data represent mean and s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. The values in parentheses indicate the number of hippocampal slices or neurons (left) and the number of mice (right) used in each experiment.

neurons impairs depolarization-induced Ca^{2+} increases that involve RyR-dependent CICR.

Collectively, our studies demonstrate that the loss of presenilin impairs LTP induction and glutamatergic neurotransmitter release in mature neurons by a presynaptic mechanism (see model in Supplementary Fig. 15). Our pharmacological and imaging studies, coupled with electrophysiological analysis, further show that a specific impairment in RyR-mediated CICR underlies the presynaptic defects caused by the loss of presenilin. Therefore, the presynaptic function of presenilin unexpectedly acts, at least in part, on the RyR-mediated Ca^{2+} release from intracellular stores. Furthermore, our data suggest that short- and long-term plasticity in the hippocampus depend partly on intracellular Ca^{2+} release, which regulates neurotransmitter release.

Previous studies investigating synaptic dysfunction in the patho-physiology of Alzheimer's disease have uncovered defects in NMDARs and AMPARs, leading to the notion that postsynaptic impairment may be the early pathogenic change in this disease[3,23,24]. However, the possibility that impaired presynaptic function may be the primary synaptic defect in Alzheimer's disease was largely un-explored. The β-amyloid precursor protein (APP) and amyloid-β peptides were reported to be presynaptically localized and were implicated in vesicle recycling[25–27]. Our findings, which distinguish unequivocally between presynaptic and postsynaptic functions of presenilin, raise the possibility that presynaptic mechanisms are involved in Alzheimer's disease pathophysiology. This hypothesis is supported by the findings that presenilin is localized to presynaptic terminals (Supplementary Fig. 16), and that APP carboxy-terminal fragments, which are substrates of presenilin-dependent γ-secretase activity and precursors of amyloid-β, accumulate in presynaptic

terminals of *Psen1* cKO mice[28]. Intriguingly, gene products respons-ible for recessively inherited familial Parkinson's disease, such as DJ-1 (also known as PARK7) and PINK1, are required for evoked dopa-mine release from nigrostriatal terminals[29,30]. These findings suggest that defects in presynaptic neurotransmitter release may represent a general convergent mechanism leading to neurodegeneration.

METHODS SUMMARY

Electrophysiological analysis. Acute hippocampal slices ($400\,\mu m$) were prepared as described previously[3]. Synaptic strength was quantified as the initial slope of field potentials recorded with artificial cerebrospinal fluid (aCSF)-filled micro-electrodes ($1–2\,M\Omega$). Intracellular whole-cell recordings were performed using Multiclamp 700B in CA1 or CA3 pyramidal neurons. Data were analysed using Igor and Clampfit. Experimenters were blinded to the genotypes of the mice.

Hippocampal neuronal culture. *Psen* cDKO hippocampal neuronal cultures were derived from *fPsen1/fPsen1;Psen2$^{-/-}$* pups at postnatal day 1, followed by infection of lentiviral vectors expressing either a functional Cre–GFP or a mutant Cre–GFP fusion protein at 2 days *in vitro* (DIV) for 72 h. Whole-cell patch recordings from cultured hippocampal neurons at 13–15 DIV were performed at room temperature using a Multiclamp 700B amplifier with pCLAMP acquisi-tion software.

Ca^{2+} imaging. Hippocampal neurons were loaded with Fura-2 AM, and imaged with a Leica DMI6000 Microscope with a $\times40$ lens. Imaging processing and data analysis were performed using LAS AF software. High concentrations of pot-assium were applied using an eight-channel gravity perfusion system.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).

2. Hsia, A. Y. et al. Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. Proc. Natl Acad. Sci. USA 96, 3228–3233 (1999).
3. Saura, C. A. et al. Loss of presenilin function causes impairments of memory and synaptic plasticity followed by age-dependent neurodegeneration. Neuron 42, 23–36 (2004).
4. Selkoe, D. J. Alzheimer's disease is a synaptic failure. Science 298, 789–791 (2002).
5. Shen, J. & Kelleher, R. J. III. The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. Proc. Natl Acad. Sci. USA 104, 403–409 (2007).
6. Feng, R. et al. Forebrain degeneration and ventricle enlargement caused by double knockout of Alzheimer's presenilin-1 and presenilin-2. Proc. Natl Acad. Sci. USA 101, 8162–8167 (2004).
7. Zakharenko, S. S. et al. Presynaptic BDNF required for a presynaptic but not postsynaptic component of LTP at hippocampal CA1–CA3 synapses. Neuron 39, 975–990 (2003).
8. Nakazawa, K. et al. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 297, 211–218 (2002).
9. Yu, H. et al. APP processing and synaptic plasticity in presenilin-1 conditional knockout mice. Neuron 31, 713–726 (2001).
10. Hessler, N. A., Shirke, A. M. & Malinow, R. The probability of transmitter release at a mammalian central synapse. Nature 366, 569–572 (1993).
11. Rosenmund, C., Clements, J. D. & Westbrook, G. L. Nonuniform probability of glutamate release at a hippocampal synapse. Science 262, 754–757 (1993).
12. Emptage, N. J., Reid, C. A. & Fine, A. Calcium stores in hippocampal synaptic boutons mediate short-term plasticity, store-operated Ca^{2+} entry, and spontaneous transmitter release. Neuron 29, 197–208 (2001).
13. Chan, S. L., Mayne, M., Holden, C. P., Geiger, J. D. & Mattson, M. P. Presenilin-1 mutations increase levels of ryanodine receptors and calcium release in PC12 cells and cortical neurons. J. Biol. Chem. 275, 18195–18200 (2000).
14. Green, K. N. et al. SERCA pump activity is physiologically regulated by presenilin and regulates amyloid beta production. J. Cell Biol. 181, 1107–1116 (2008).
15. Stutzmann, G. E., Caccamo, A., LaFerla, F. M. & Parker, I. Dysregulated IP3 signaling in cortical neurons of knock-in mice expressing an Alzheimer's-linked mutation in presenilin1 results in exaggerated Ca^{2+} signals and altered membrane excitability. J. Neurosci. 24, 508–513 (2004).
16. Tu, H. et al. Presenilins form ER Ca^{2+} leak channels, a function disrupted by familial Alzheimer's disease-linked mutations. Cell 126, 981–993 (2006).
17. Treiman, M., Caspersen, C. & Christensen, S. B. A tool coming of age: thapsigargin as an inhibitor of sarco-endoplasmic reticulum Ca^{2+}-ATPases. Trends Pharmacol. Sci. 19, 131–135 (1998).
18. Gafni, J. et al. Xestospongins: potent membrane permeable blockers of the inositol 1,4,5-trisphosphate receptor. Neuron 19, 723–733 (1997).
19. Meissner, G. Ryanodine activation and inhibition of the Ca^{2+} release channel of sarcoplasmic reticulum. J. Biol. Chem. 261, 6300–6306 (1986).
20. Stutzmann, G. E. et al. Enhanced ryanodine receptor recruitment contributes to Ca^{2+} disruptions in young, adult, and aged Alzheimer's disease mice. J. Neurosci. 26, 5180–5189 (2006).
21. Handler, M., Yang, X. & Shen, J. Presenilin-1 regulates neuronal differentiation during neurogenesis. Development 127, 2593–2606 (2000).
22. Shen, J. et al. Skeletal and CNS defects in Presenilin-1-deficient mice. Cell 89, 629–639 (1997).
23. Kamenetz, F. et al. APP processing and synaptic function. Neuron 37, 925–937 (2003).
24. Snyder, E. M. et al. Regulation of NMDA receptor trafficking by amyloid-β. Nature Neurosci. 8, 1051–1058 (2005).
25. Buxbaum, J. D. et al. Alzheimer amyloid protein precursor in the rat hippocampus: transport and processing through the perforant path. J. Neurosci. 18, 9629–9637 (1998).
26. Lazarov, O., Lee, M., Peterson, D. A. & Sisodia, S. S. Evidence that synaptically released β-amyloid accumulates as extracellular deposits in the hippocampus of transgenic mice. J. Neurosci. 22, 9785–9793 (2002).
27. Yao, P. J. & Coleman, P. D. Reduced O-glycosylated clathrin assembly protein AP180: implication for synaptic vesicle recycling dysfunction in Alzheimer's disease. Neurosci. Lett. 252, 33–36 (1998).
28. Saura, C. A. et al. Conditional inactivation of presenilin 1 prevents amyloid accumulation and temporarily rescues contextual and spatial working memory impairments in amyloid precursor protein transgenic mice. J. Neurosci. 25, 6755–6764 (2005).
29. Goldberg, M. S. et al. Nigrostriatal dopaminergic deficits and hypokinesia caused by inactivation of the familial Parkinsonism-linked gene DJ-1. Neuron 45, 489–496 (2005).
30. Kitada, T. et al. Impaired dopamine release and synaptic plasticity in the striatum of PINK1-deficient mice. Proc. Natl Acad. Sci. USA 104, 11441–11446 (2007).

METHODS

Generation of CA1-*Psen* and CA3-*Psen* cDKO mice. CA1-*Psen* and CA3-*Psen* cDKO mice contain homozygous floxed *Psen1* alleles, homozygous *Psen2$^{-/-}$* alleles, and the *Camk2a-Cre*[7] and *Grik4-Cre*[8] transgene, respectively. Because *Psen2$^{-/-}$* mice have no detectable phenotypes[31], it was unnecessary to generate floxed *Psen2* mice. For each cDKO mouse line, *fPsen1/fPsen1;Psen2$^{-/-}$;Cre* mice were bred with *fPsen1/fPsen1;Psen2$^{-/-}$* mice to obtain more cDKO mice (*fPsen1/fPsen1;Psen2$^{-/-}$;Cre*) and *fPsen1/fPsen1;Cre* were bred with *fPsen1/fPsen1* to obtain control mice (*fPsen1/fPsen1*). The introduction of two *loxP* sites into *Psen1* introns 1 and 3 was previously confirmed not to affect transcription, splicing or translation[9]. The genetic background of these mice was similar in the C57BL6/129 hybrid background, with breeding carried out similarly for both groups. All procedures relating to animal care and treatment conformed to the Institutional and NIH guidelines.

In situ hybridization and LacZ staining. *In situ* hybridization was carried out as previous described using a 260-base-pair (bp) sense or antisense riboprobe specific for *Psen1* exons 2 and 3 (ref. 32). For X-gal staining, *Camk2a-Cre* and *Grik4-Cre* transgenic mice were bred to *Rosa26-lacZ* mice, and double transgenic offspring containing both the *Cre* and the *lacZ* transgenes were analysed.

Field and whole-cell electrophysiological analysis of acute hippocampal slices. All electrophysiological analysis was performed by experimenters who were blinded to the genotypes of the mice. Acute hippocampal slices (400 μm) were prepared as described before[3]. The slices were maintained in a storage chamber containing aCSF (124 mM NaCl, 5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 1.3 mM MgCl$_2$, 2.6 mM CaCl$_2$, 26 mM NaHCO$_3$, 10 mM dextrose, pH 7.4, 300–310 mOsM) at 30 °C. Stimulation (500 μs) pulses were delivered with a bipolar concentric metal electrode. Synaptic strength was quantified as the initial slope of field potentials recorded with aCSF-filled microelectrodes (1–2 MΩ). In LTP recordings, baseline responses were collected every 15 s with a stimulation intensity that yielded 60% of maximal response. LTP was induced by five episodes of TBS delivered at 0.1 Hz. Each episode contains ten stimulus trains (five pulses at 100 Hz) delivered at 5 Hz. Average responses (mean ± s.e.m.) are expressed as percentage of pre-TBS baseline response. Synaptic facilitations were measured as the percentage of the fEPSP slope versus the first fEPSP slope at a given stimulus train in individual slices.

Intracellular (whole-cell) recordings were performed using Multiclamp 700B (Molecular device) in CA1 or CA3 pyramidal neurons. Patch pipettes (3–5 MΩ) were filled with internal solution consisting of (in mM): 110 Cs-methanesulphonate, 20 tetraethylammonium-chloride, 8 KCl, 10 EGTA, 10 HEPES, 5 QX-314 (a derivative of lidocaine), 3 Mg-ATP, 0.3 Na$_2$GTP, pH 7.3; 275–285 mOsM. AMPAR responses were recorded in the presence of 50 μM DL-2-amino-5-phosphonovaleric acid and 100 μM picrotoxin to block NMDAR- and GABA (γ-aminobutyric acid) type A receptor-mediated responses, respectively. NMDAR responses were recorded in the presence of 10 μM 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and 100 μM picrotoxin to block AMPAR- and GABA type A receptor-mediated responses, respectively. For the MK-801 experiment, recordings of NMDAR-mediated EPSC were made every 20 s before and during exposure to MK-801 (40 μM). The NMDAR-mediated EPSC amplitude was plotted as a function of stimulus number. Decay curves were normalized to the amplitude of the first EPSC in the presence of MK-801 and were fitted to a single exponential curve to estimate the decay time course. NMDA-mediated EPSC was measured at +40 mV, and was elicited by focal stimulation in the presence of CNQX (10 μM) and picrotoxin (100 μM). To record calcium current through VGCCs, tetrodotoxin (extracellular, 500 nM) and QX-314 (intracellular, 5 mM) were used to block sodium current; Cs$^+$ (intracellular, 110 mM) and tetraethyl-lammonium (intracellular, 20 mM) were used to block potassium current. To measure the synaptic facilitation, values of the fEPSP slope (second to tenth responses in a 20-pulse stimulus train) were normalized to the slope of the first fEPSP of the stimulus train. Data were analysed using Igor (Wavemetrics) and Clampfit (Molecular device).

Psen cDKO hippocampal neuronal cultures. To circumvent the requirement of presenilin in neurogenesis during embryonic development[21,22], we established *Psen* cDKO hippocampal neuronal cultures derived from *fPsen1/fPsen1;*

Psen2$^{-/-}$ newborn pups, followed by infection of lentiviral vectors expressing either a functional Cre–GFP or a mutant Cre–GFP fusion protein. Hippocampi from *fPsen1/fPsen1;Psen2$^{-/-}$* pups were dissected and treated with 0.25% trypsin at 37 °C for 20 min. Cells were plated at a density of 65,000 cells cm^{-2} on poly-D-lysine-coated 35-mm dishes (Costar). Cultures were infected with lentiviruses (300 μl conditioned medium per well in a 24-well plate) expressing at 2 DIV for 72 h. Infected neurons were cultured until 13–15 DIV for further biochemical, morphological, electrophysiological and imaging analyses. Lentiviruses were produced by transfecting human embryonic kidney HEK293 cells (ATCC) with the respective pFUGW vectors and two helper plasmids (pVSVg and pCMVΔ8.9) using FUGENE 6 (Roche), as previously described[33]. Conditioned medium containing viruses was collected 48 h after transfection, and spun (800g for 5 min) to remove HEK cell debris before adding to the neuronal culture.

Morphological analysis of postnatal hippocampal cultures. Cultured neurons at 14 DIV were fixed with methanol (−20 °C). Fixed cultures were then incubated with primary antibodies against synaptophysin (monoclonal; 1:2,000; Sigma) and microtubule-activated protein 2 (MAP2; polyclonal; 1:2,000; Sigma). After rinsing three times with PBS, the neurons were incubated with fluorescent secondary antibodies for 30 min. After washing, cultures were mounted with Vectashield mounting medium (Vector labs). Confocal microscopic analysis was performed on a Zeiss LSM 510 microscope. Identical acquisition settings were applied to all samples of the experiment. Images of neurons were collected with a ×40 oil-immersion objective lens. Images were analysed in a genotype-blind manner using the NIH Image/Image J program.

Whole-cell electrophysiological analysis of postnatal hippocampal cultures. Whole-cell patch recordings from cultured hippocampal neurons at 13–15 DIV were performed at room temperature using a Multiclamp 700B amplifier (Molecular device) with pCLAMP acquisition software. Synaptic transmission was elicited with a concentric focal stimulus electrode, and EPSCs were recorded with a patch electrode (3–5 MΩ) in whole-cell recording mode and filtered at 2 kHz. Pipette solution contained (in mM): 136.5 K-gluconate, 0.2 EGTA, 10 HEPES, 9 NaCl, 17.5 KCl, 5 QX-314, 4 Mg-ATP and 0.3 Na-GTP (adjusted to pH 7.4 with KOH). The extracellular solution was HEPES-buffered saline containing (in mM): 145 NaCl, 3 KCl, 10 HEPES, 2 CaCl$_2$, 1 MgCl$_2$, 8 dextrose, pH 7.2.

Calcium imaging analysis. Hippocampal neurons were loaded with Fura-2 AM (5 μM, 45 min at 37 °C) (Molecular probes), and imaged with a Leica DMI6000 Microscope with ×40 lens (numerical aperture 0.75). The method and parameters for *in vitro* calibration (Invitrogen calibration kit) were as described previously[34]. Imaging processing and data analysis were performed using LAS AF software (Leica). High concentrations of potassium were applied using an 8-channel gravity perfusion system (ALA Scientific Instrument).

Subcellular fractionation analysis. For enrichment of synaptic vesicle (presynaptic) proteins, four adult cortices (3-month-old) were homogenized with a Dounce teflon homogenizer in ice-cold buffer containing 0.32 M sucrose, 4 mM HEPES, pH 7.3, and protease and phosphatase inhibitor cocktails. For the P2 fraction, crude homogenates were centrifuged at 800g twice to remove debris, the supernatant were centrifuged at 9,200g, and the pellet was resuspended in 0.32 M sucrose buffer. For the LP1 fraction, P2 synaptosomes were centrifuged at 10,200g, resuspended in 0.32 M sucrose buffer and hypotonically lysed in nine volumes of water. The lysate was centrifuged at 25,000g, and the pellet was resuspended in buffer containing 1% NP-40 to produce LP1. For the LP2 fraction, the supernatant from the LP1 purification step was centrifuged at 165,000g, and the pellet was resuspended in buffer containing 1% NP-40.

31. Steiner, H. *et al.* A loss of function mutation of presenilin-2 interferes with amyloid β-peptide production and notch signaling. *J. Biol. Chem.* **274**, 28669–28673 (1999).
32. Wines-Samuelson, M., Handler, M. & Shen, J. Role of presenilin-1 in cortical lamination and survival of Cajal-Retzius neurons. *Dev. Biol.* **277**, 332–346 (2005).
33. Watanabe, H. *et al.* Indirect regulation of presenilins in CREB-mediated transcription. *J. Biol. Chem.* **284**, 13705–13713 (2009).
34. Zhang, C. & Zhou, Z. Ca^{2+}-independent but voltage-dependent secretion in mammalian dorsal root ganglion neurons. *Nature Neurosci.* **5**, 425–430 (2002).

LETTERS

Macrophage elastase kills bacteria within murine macrophages

A. McGarry Houghton[1]*, William O. Hartzell[2]*, Clinton S. Robbins[1], F. Xavier Gomis-Rüth[3] & Steven D. Shapiro[1]

Macrophages are aptly positioned to function as the primary line of defence against invading pathogens in many organs, including the lung and peritoneum. Their ability to phagocytose and clear microorganisms has been well documented[1,2]. Macrophages possess several substances with which they can kill bacteria, including reactive oxygen species, nitric oxide, and antimicrobial proteins[3–9]. We proposed that macrophage-derived proteinases may contribute to the antimicrobial properties of macrophages. Macrophage elastase (also known as matrix metalloproteinase 12 or MMP12) is an enzyme predominantly expressed in mature tissue macrophages[10] and is implicated in several disease processes, including emphysema[11]. Physiological functions for MMP12 have not been described. Here we show that $Mmp12^{-/-}$ mice exhibit impaired bacterial clearance and increased mortality when challenged with both Gram-negative and Gram-positive bacteria at macrophage-rich portals of entry, such as the peritoneum and lung. Intracellular stores of MMP12 are mobilized to macrophage phagolysosomes after the ingestion of bacterial pathogens. Once inside phagolysosomes, MMP12 adheres to bacterial cell walls where it disrupts cellular membranes resulting in bacterial death. The antimicrobial properties of MMP12 do not reside within its catalytic domain, but rather within the carboxy-terminal domain. This domain contains a unique four amino acid sequence on an exposed β loop of the protein that is required for the observed antimicrobial activity. The present study represents, to our knowledge, the first report of direct antimicrobial activity by a matrix metallopeptidase, and describes a new antimicrobial peptide that is sequentially and structurally unique in nature.

To determine whether MMP12 has a role in host defence against bacteria, we subjected $Mmp12^{-/-}$ mice and wild-type littermates to intraperitoneal (i.p.) injection of either *Staphylococcus aureus* or *Escherichia coli*. In each case, there was a significant survival advantage for wild-type mice as compared to $Mmp12^{-/-}$ mice over a 3-day period (Fig. 1a–c, g). Lower titres of i.p. *S. aureus* injection (1×10^7
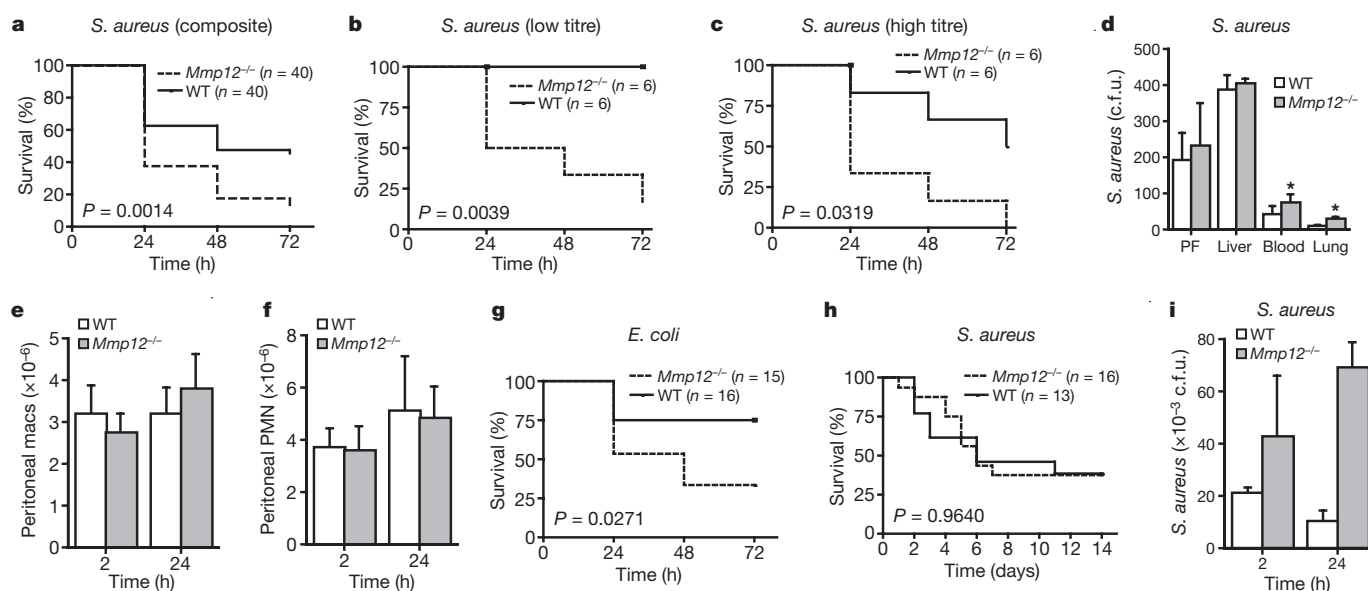

Figure 1 | MMP12 provides survival advantage in Gram-positive and Gram-negative infection. a, Kaplan–Meier curve for $Mmp12^{-/-}$ and wild-type (WT) mice after i.p. injection with 1×10^8 c.f.u. *S. aureus* ($n = 40$ for each group from six experiments); $P = 0.0014$, log-rank test. **b, c**, Kaplan–Meier curves for low (**b**; 1×10^7 c.f.u.) and high (**c**; 1×10^9 c.f.u.) *S. aureus* titres ($n = 6$ each group); $P < 0.05$. **d**, Bacterial c.f.u. 24 h after i.p. *S. aureus* injection in peritoneal fluid (PF), blood, liver and lungs ($n = 4$ each group); *$P < 0.01$. **e, f**, Peritoneal macrophage (macs; **e**) and neutrophil (**f**) counts 2 and 24 h after i.p. *S. aureus* injection ($n = 6$ each group). PMN, polymorphonuclear leukocytes; $P > 0.05$. **g**, Kaplan–Meier curve for $Mmp12^{-/-}$ ($n = 15$) and wild-type ($n = 16$) mice (from three experiments) after 1×10^8 c.f.u. i.p. *E. coli* injection; $P = 0.0271$. **h**, Kaplan–Meier curve for intravenous *S. aureus* (1×10^8 c.f.u.) to wild-type ($n = 13$) and $Mmp12^{-/-}$ ($n = 16$) mice (from three experiments); $P = 0.9640$. **i**, Lung bacterial burden (c.f.u.) 2 h after *S. aureus* TVI ($n = 6$ each group); *$P < 0.005$. All error bars represent s.d.

[1]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. [2]Division of Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [3]Proteolysis Lab, Molecular Biology Institute of Barcelona (CSIC), Barcelona Science Park, Helix Building, c/ Baldiri Reixac 15-21, 08028 Barcelona, Spain.
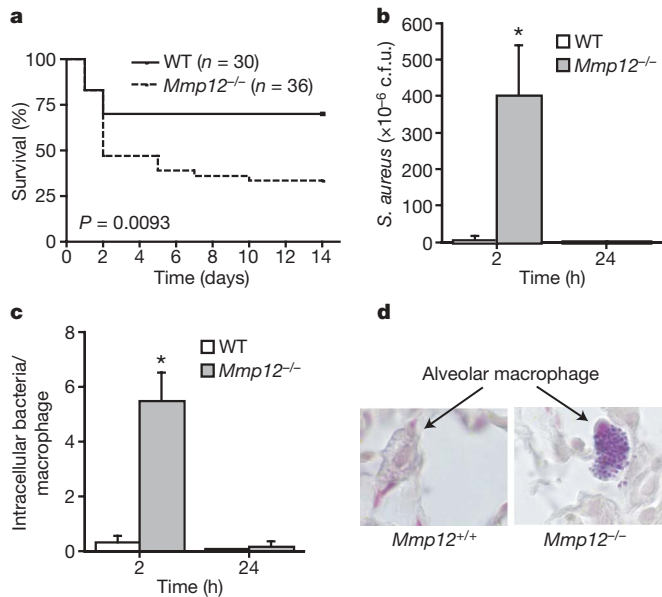*These authors contributed equally to this work.

Figure 2 | MMP12 improves bacterial clearance and survival in S. aureus pneumonia. a, Kaplan–Meier curve for $Mmp12^{+/+}$ ($n = 30$; wild-type, WT) and $Mmp12^{-/-}$ ($n = 36$) mice (from six experiments) administered with S. aureus i.t. (1×10^8 c.f.u.); $P = 0.0093$, log-rank test. **b**, Bacterial burden in the lung after sub-lethal dose of S. aureus i.t. (1×10^6 c.f.u.; $n = 8$ each group); *$P < 0.01$. **c**, Intramacrophage bacterial particle counts on Gram stain ($n = 8$ each group); *$P < 0.001$. **d**, $Mmp12^{+/+}$ and $Mmp12^{-/-}$ alveolar macrophages on Gram stain at the 2-h time point. Original magnification, $\times 100$. All error bars represent s.d.

colony-forming units (c.f.u.)), which did not result in mortality in wild-type mice, still proved >50% lethal in $Mmp12^{-/-}$ mice. Soon after i.p. injection of S. aureus, the local bacterial burden in the peritoneal fluid and liver was similar in both groups of mice. However, wild-type mice were significantly protected from bacterial dissemination to distal sites, such as the blood and lung (Fig. 1d). None of these observations was attributable to differences in inflammatory cell recruitment, because the macrophage and neutrophil counts were equivalent between the two groups at 2 and 24 h after i.p. S. aureus injection (Fig. 1e, f).

Bacteremia generated by tail vein injection (TVI) of 1×10^8 c.f.u. S. aureus in $Mmp12^{-/-}$ and wild-type control mice produced equivalent 2-week mortality in both genotypes (Fig. 1h). However, the bacterial burden in the lungs of $Mmp12^{-/-}$ mice was significantly greater than in wild-type counterparts at both 2 and 24 h after TVI (Fig. 1i). The bacterial burden was equivalent in kidney and spleen

homogenates between the two groups at the same time points (data not shown). Thus, it seems that MMP12 exerts its antimicrobial properties solely in organs, such as the lung, that are fortified with macrophages.

To validate this hypothesis, we subjected $Mmp12^{-/-}$ mice and wild-type littermates to an *in vivo* pneumonia model in which mice received the approximate lethal dose to 50% of animals tested (LD_{50} dose; 1×10^8 c.f.u.) of S. aureus intratracheally (i.t.). As with peritonitis, wild-type mice showed a decided survival advantage over $Mmp12^{-/-}$ mice (Fig. 2a). Repeating the experiment using a sublethal titre of S. aureus (1×10^6 c.f.u.) resulted in ~100-fold greater increased bacterial burden in $Mmp12^{-/-}$ lungs 2 h after infection (Fig. 2b). The bacteria were eliminated by 24 h in both groups, which correlated with an equivalent increase in lung neutrophils in both groups. Surprisingly, the bacterial burden in $Mmp12^{-/-}$ mice was predominantly located within macrophages (Fig. 2c, d).

MMP12, like other MMPs, is transcriptionally regulated[12]. The pro-enzyme is secreted and subsequently activated in the extracellular space where MMPs are thought to perform their biological functions[13]. Yet, preformed intracellular pools of MMP12 exist within quiescent macrophages[14], presumably for rapid secretion from the cell. To determine whether these intracellular stores of MMP12 are responsible for anti-microbial activity within macrophages, we performed macrophage intracellular bacterial killing assays. Wild-type and $Mmp12^{-/-}$ macrophages were co-incubated with S. aureus for 60 min in antibiotic-free medium, at which time ($t = 0$) the medium was changed and the cells were washed with PBS to remove non-phagocytosed bacteria. The bacterial content was equal at this time point, demonstrating equal phagocytic capacity in both genotypes (Fig. 3a). Macrophages were collected after further 30- and 90-min incubations, and intracellular bacterial burden was quantified. Wild-type macrophages efficiently cleared bacteria within 90 min. In contrast, bacterial content increased over time in $Mmp12^{-/-}$ macrophages. Overall, the bacterial burden was tenfold greater in $Mmp12^{-/-}$ macrophages than in wild-type macrophages. Scanning electron microscopy of wild-type macrophages demonstrated very few intracellular bacteria after 60 min (Fig. 3b). In contrast, $Mmp12^{-/-}$ macrophages phagocytosed the bacteria but were unable to kill them, resulting in the presence of several intracellular bacteria even after 60 min (Fig. 3b). Immuno-gold labelling and transmission electron microscopy demonstrated the presence of MMP12 in clusters on disrupted areas of bacterial cell walls within phagolysosomes of wild-type but not $Mmp12^{-/-}$ macrophages (Fig. 3c, d).

To determine whether MMP12 possesses direct antimicrobial activity, as suggested by the transmission electron microscopy studies, we incubated full-length recombinant human MMP12 with S. aureus and *Klebsiella pneumoniae* in tryptic soy broth (TSB) buffer (Fig. 4a).
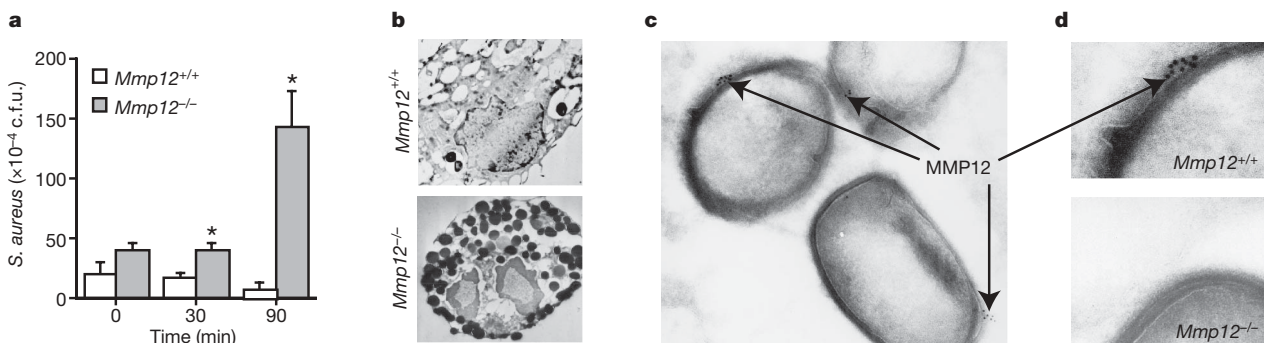

Figure 3 | MMP12 is required for early intracellular bacterial killing by macrophages. Peritoneal macrophages were incubated with S. aureus in RPMI plus 10% FCS without antibiotics for 60 min ($t = 0$). **a**, Intracellular bacterial c.f.u. from macrophage lysates. Data are from an experiment in triplicate. Error bars represent s.d., *$P < 0.05$. **b**, Scanning electron microscopy (original magnification, $\times 4,400$) of macrophages (from three

experiments) at $t = 60$ min. **c**, *Staphylococcus aureus* in phagolysosomes of wild-type macrophages. Immuno-electron-microscopy (original magnification, $\times 30,000$) shows MMP12 (gold particles) on S. aureus cocci (**c, d**) within wild-type macrophages. Gold particles were not detected within $Mmp12^{-/-}$ macrophages (**d**).
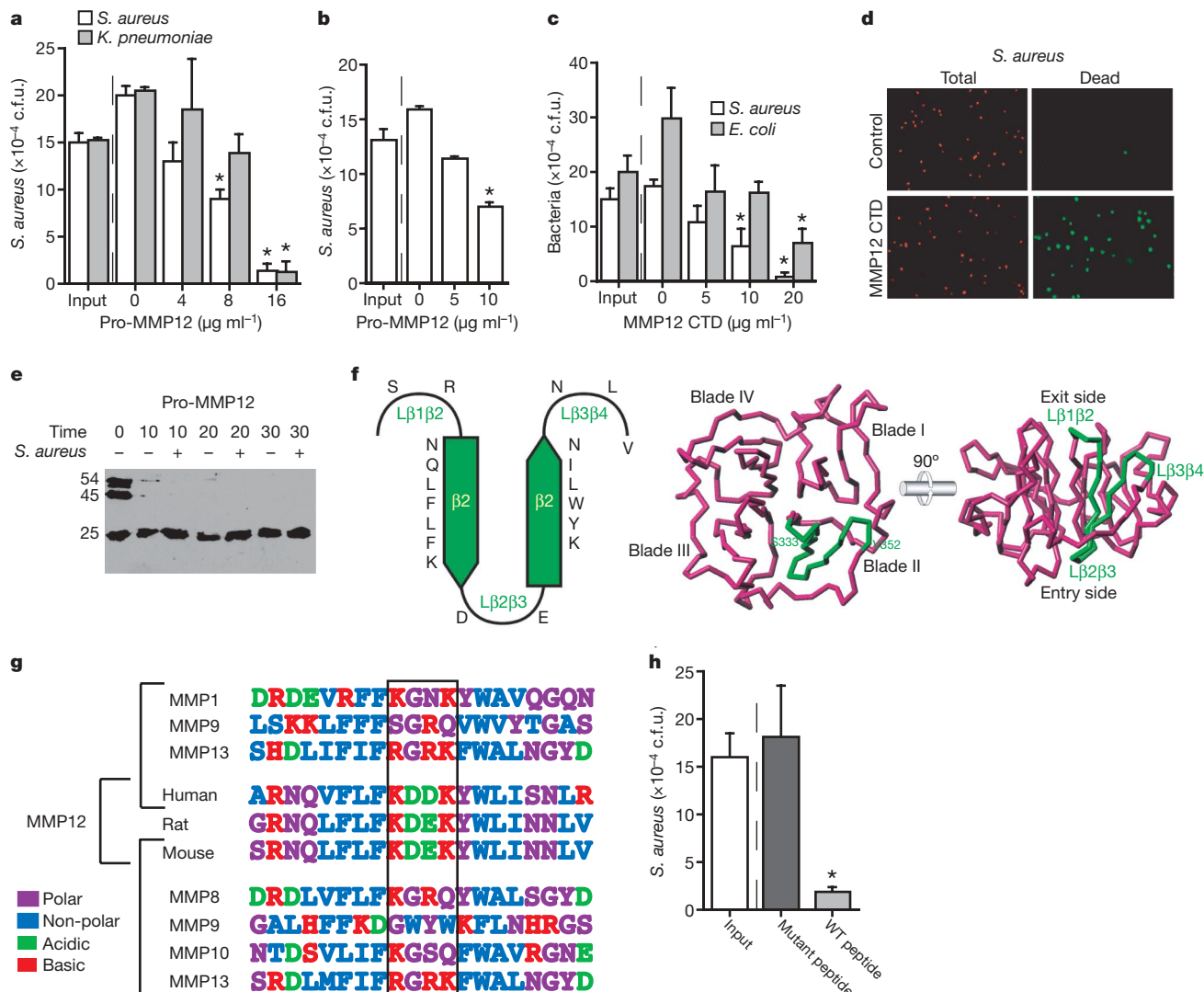
638

Figure 4 | MMP12 CTD possesses bactericidal activity. a–c, Bacteria were incubated with human pro-MMP12 in 5% TSB (**a**) or RPMI plus 10% FCS (**b**), or with mouse MMP12 CTD (**c**) for 60 min at 37 °C. Data are expressed as c.f.u. (from three experiments); $*P < 0.005$. **d,** Fluorescent propidium iodide exclusion assay after co-incubation of *S. aureus* and MMP12 CTD or control for 30 min. All bacteria stain positive for Syto 59 (red), but only those with disrupted membranes stain with S-7020 (green). **e,** Human pro-MMP12 incubated with and without *S. aureus* to study CTD processing in the presence of bacteria using a CTD antibody. The 54 kDa band represents full-length MMP12, the 45 kDa band represents shedding of pro-domain,

and the 25 kDa band represents the CTD. The CTD is not processed to smaller fragments. **f,** Computational three-dimensional model of mouse MMP12 CTD. The SR-20 sequence is located within CTD blade II including β strands β2 and β3, as well as the connecting and flanking loops (green trace). **g,** SR-20 shows a high degree of homology among MMP12 orthologues but not among other MMPs. **h,** *Staphylococcus aureus* was incubated with either wild-type (WT) MMP12 SR-20 peptide or mutant peptide (Lys-Asp-Glu-Lys replaced by Ser-Gly-Arg-Gln) (both at $20\,\mu g\,ml^{-1}$), in RPMI plus 10% FCS for 60 min. Data are expressed as c.f.u., $*P < 0.001$. All error bars represent s.d.

Staphylococcus aureus killing assays were repeated in RPMI with 10% FCS (Fig. 4b) in an attempt to produce a physiological environment. In both conditions we observed a dose-dependent inhibition of *S. aureus* growth by human MMP12. We suspect that the concentration of MMP12 within intracellular granules is much greater than the effective concentrations in these assays, although the exact concentrations are unknown. Bacterial content (c.f.u.) after incubation with $>5\,\mu g\,ml^{-1}$ MMP12 was less than the input amount, suggesting bactericidal activity. None of the other MMPs tested, including MMP3, 7, 8, 9 and 14, showed direct antimicrobial activity at concentrations above $100\,\mu g\,ml^{-1}$ (data not shown). The antimicrobial properties of MMP12 were not dependent on catalytic activity, as determined by continued activity with the co-administration of a hydroxamic acid MMP inhibitor and the lack of activity of the catalytic domain alone. The catalytic domain may contribute to host defence by processing/degrading other proteinases or bacterial toxins. However, MMP12 does not degrade *S. aureus* alpha-toxin (data not shown).

Full-length MMP12 is a 54 kDa pro-enzyme consisting of three common MMP domains: an amino-terminal pro-domain, a zinc- and calcium-binding catalytic domain, and a haemopexin-like carboxy-terminal domain (CTD)[10]. CTD functions include enhancement of TIMP binding and improved catalysis of both collagen and chemokines[15–18]. The CTD is not required for MMP12-mediated substrate catalysis, and is commonly shed during MMP12 activation *in vitro*[19].

To isolate region(s) of antimicrobial activity, we generated recombinant mouse MMP12 CTD (residues 280–473) in *E. coli*. Co-incubation of MMP12 CTD (but not MMP9 CTD) with either *S. aureus* or *E. coli* demonstrated a dose-dependent inhibition of bacterial growth equivalent to full-length MMP12 (Fig. 4c). In addition to *S. aureus*, *K. pneumoniae* and *E. coli*, MMP12 also kills *Salmonella enteriditis*. Notably, MMP12 is not active against *Listeria monocytogenes*, an organism that is able to survive within macrophages (data not shown).

To expand on the immuno-electron-microscopy results, which suggest that MMP12 is directly bactericidal by disrupting bacterial outer

membrane integrity, we performed a propidium iodide exclusion assay. Mouse MMP12 CTD killed 98% of *S. aureus* in this assay (Fig. 4d). We also generated *S. aureus* cell-walls loaded with fluorescent dextran and performed a liposomal release assay demonstrating that MMP12 CTD can disrupt liposomal membranes (data not shown).

To more precisely locate the region of antimicrobial activity in mouse MMP12 CTD, we generated three recombinant fragments corresponding to the initial (Ser 280–Ile 342), middle (Glu 343—Glu 404), and last (Trp 405–Cys 473) third of the CTD. Only the middle fragment showed antimicrobial activity. Synthetic peptides were generated from this region of mouse MMP12 that narrowed the antimicrobial properties of the MMP12 CTD to a 20 amino acid sequence, designated SR-20 (344-SRNQLFLFKDEKYWLINNLV-363). SR-20 proved bactericidal against *S. aureus* in a propidium iodide exclusion assay, whereas the comparable region in mouse MMP13 (343-SRDLMFIFRGRKFWALNGYD-362) did not (data not shown). Bacterial killing assays also demonstrated that MMP12 CTD killed bacteria without further processing to smaller fragments, suggesting that SR-20 preferentially killed bacteria within the scaffold of the entire CTD (Fig. 4e).

We generated a three-dimensional homology model of mouse MMP12 CTD (Fig. 4f). The SR-20 antimicrobial peptide resides within blade II of the CTD, which is an all β structure featuring a fourfold β propeller in all experimental structures reported. Interactions with either the catalytic domain or the pro-domain are not predicted. The only exposed portions of SR-20 that could theoretically interact with bacteria are the surface accessible loops that flank and connect the two central β strands (Lβ2β3 on the entry side, and Lβ1β2 and Lβ3β4 on the exit side of the CTD disk; Fig. 4f). Lβ2β3 is of particular interest given its unique sequence of acidic amino acids flanked by basic residues. Although this amino acid sequence is homologous in rabbit, rat, mouse and human MMP12, it is not found in any other MMP (Fig. 4g). The sequence Lys-Asp-Glu-Lys is commonly found in various proteins, but a search of three-dimensional structures that contain this motif protruding from the surface of a β loop confirmed that it is sequentially and structurally unique to MMP12.

To prove that this motif is responsible for the observed antimicrobial properties of MMP12, we constructed a mutant peptide identical to SR-20 except that the Lys-Asp-Glu-Lys motif found in mouse MMP12 was replaced by the human MMP9 sequence, Ser-Gly-Arg-Gln. Incubation of *S. aureus* with wild-type and mutant peptide confirmed that the Lys-Asp-Glu-Lys motif is essential for the antimicrobial properties of mouse MMP12 CTD (Fig. 4h). Shorter four-amino-acid peptides (Ser-Gly-Arg-Gln, Lys-Asp-Asp-Lys and Lys-Asp-Glu-Lys) did not show antimicrobial activity (data not shown), suggesting that the loop structure of the protein is required for bacterial killing.

Previously, we, and others, have shown that neutrophil elastase is able to kill bacteria within neutrophil phagolysosomes[20–22]. Other serine-proteinases, including cathepsin G, proteinase 3, and the inactive azurocidin have been shown to possess antimicrobial activity, independent of catalytic activity[23]. However, sequence analysis of these serine-proteinases did not reveal the presence of the Lys-Asp-Glu-Lys motif (data not shown). Matrilysin (also known as MMP7), the only other MMP known to participate in bacterial killing, does so indirectly by cleaving and activating alpha-defensins within Paneth cells before their secretion into the gut[24].

MMPs participate in physiological functions, including embryonic growth and development, generation and termination of the cytokine/chemokine gradients, and regulation of cell growth, differentiation and angiogenesis by cleavage of non-matrix proteins[17,18,25,26]. When aberrantly or excessively expressed, MMPs cause tissue destruction leading to emphysema, vascular disease and arthritis[11,27,28]. Although virtually all MMP functions had been thought to occur extracellularly or on the cell surface, MMPs have recently been implicated in intracellular functions including apoptosis and cell cycle regulation[29,30].

This study extends this concept to include intracellular (within the phagolysosome) antimicrobial activity.

The present study supports the participation of the macrophage in the earliest stages of the host's defence against microorganisms, and adds to our understanding of MMP biology by demonstrating that MMP12 possesses direct antimicrobial activity. After engulfment of microorganisms by macrophages, intracellular stores of MMP12 are mobilized to phagolysosomes, where MMP12 uses a new antimicrobial peptide within the CTD to disrupt the integrity of cell wall structures in the invading microorganisms. The continued discovery of new antimicrobial mechanisms may prove essential to combat the increasing incidence of resistance to antibiotics now in use.

METHODS SUMMARY

Mmp12$^{-/-}$ and wild-type mice were maintained on a 129/SvJ background and housed in a sterile barrier facility. *In vivo* infection models were accomplished by i.p. (peritonitis), intratracheal (pneumonia) or i.v. (bacteremia) administration of bacteria. *Staphylococcus aureus*, *S. enteritidis* and *K. pneumoniae* (clinical isolates) and *E. coli* (K1 strain) were grown in TSB and isolated on Luria-Bertani broth agar (LB) plates. Organ c.f.u. were determined by plating limiting dilutions of organ homogenates on LB plates. Inflammatory cell counts were determined using a haemocytometer and Wright-stained cytospins. Macrophage killing assays were performed using thioglycollate-induced peritoneal macrophages. After 60 min incubation with bacteria, macrophages were washed with PBS and incubated for a further 30 or 90 min. Colony-forming units were determined from macrophage lysates. Recombinant protein constructs were generated in pet 20b vectors and grown in *E. coli*. All synthetic peptides were generated by the University of Pittsburgh protein core facility. Protein and peptide killing assays were performed in either TSB or RPMI plus 10% FCS at 37 °C for 60 min. Propidium iodide exclusion assay was performed using Syto 59 (red) to stain total bacteria and S-7020 (green) to stain bacteria with disrupted membranes (both from Molecular Probes). A three-dimensional homology model of mouse MMP12 CTD was constructed on the basis of crystal structures of related full-length MMPs. The programs CPHmodels 2.0 (www.cbs.dtu.dk/services/CPHmodels) and Swiss Model (swissmodel.expasy.org) were used for construction.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Green, G. M. & Kass, E. H. The role of the alveolar macrophage in the clearance of bacteria from the lung. *J. Exp. Med.* 119, 167–176 (1964).
2. Gordon, S. The macrophage: Past, present and future. *Eur. J. Immunol.* 37, S9–S17 (2007).
3. Shiloh, M. U. et al. Phenotype of mice and macrophages deficient in both phagocyte oxidase and inducible nitric oxide synthase. *Immunity* 10, 29–38 (1999).
4. Ganz, T. Oxygen-independent microbicidal mechanisms of phagocytes. *Proc. Am. Assoc. Physicians* 111, 390–395 (1999).
5. Selsted, M. E. & Ouellette, A. J. Defensins in granules of phagocytic and non-phagocytic cells. *Trends Cell Biol.* 5, 114–119 (1995).
6. Biggar, W. D. & Sturgess, J. M. Role of lysozyme in the microbicidal activity of rat alveolar macrophages. *Infect. Immun.* 16, 974–982 (1977).
7. del Cerro-Vadillo, E. et al. Cutting edge: a novel nonoxidative phagosomal mechanism exerted by cathepsin-D controls Listeria monocytogenes intracellular growth. *J. Immunol.* 176, 1321–1325 (2006).
8. Hiemstra, P. S., van den Barselaar, M. T., Roest, M., Nibbering, P. H. & van Furth, R. Ubiquicidin, a novel murine microbicidal protein present in the cytosolic fraction of macrophages. *J. Leukoc. Biol.* 66, 423–428 (1999).
9. Hiemstra, P. S. et al. Antimicrobial proteins of murine macrophages. *Infect. Immun.* 61, 3038–3046 (1993).
10. Shipley, J. M., Wesselschmidt, R. L., Kobayashi, D. K., Ley, T. J. & Shapiro, S. D. Metalloelastase is required for macrophage-mediated proteolysis and matrix invasion in mice. *Proc. Natl Acad. Sci. USA* 93, 3942–3946 (1996).
11. Hautamaki, R. D., Kobayashi, D. K., Senior, R. M. & Shapiro, S. D. Macrophage elastase is required for cigarette smoke-induced emphysema in mice. *Science* 277, 2002–2004 (1997).
12. Brinckerhoff, C. E. & Matrisian, L. M. Matrix metalloproteinases: a tail of a frog that became a prince. *Nature Rev. Mol. Cell Biol.* 3, 207–214 (2002).
13. Raza, S. L., Nehring, L. C., Shapiro, S. D. & Cornelius, L. A. Proteinase-activated receptor-1 regulation of macrophage elastase (MMP-12) secretion by serine proteinases. *J. Biol. Chem.* 275, 41243–41250 (2000).
14. Lijnen, H. R. et al. Temporal and topographic matrix metalloproteinase expression after vascular injury in mice. *Thromb. Haemost.* 81, 799–807 (1999).

15. Knauper, V. *et al.* The role of the C-terminal domain of human collagenase-3 (MMP-13) in the activation of procollagenase-3, substrate specificity, and tissue inhibitor of metalloproteinase interaction. *J. Biol. Chem.* **272**, 7608–7616 (1997).
16. Tam, E. M., Moore, T. R., Butler, G. S. & Overall, C. M. Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): the differential roles of the MMP hemopexin c domains and the MMP-2 fibronectin type II modules in collagen triple helicase activities. *J. Biol. Chem.* **279**, 43336–43344 (2004).
17. McQuibban, G. A. *et al.* Matrix metalloproteinase activity inactivates the CXC chemokine stromal cell-derived factor-1. *J. Biol. Chem.* **276**, 43503–43508 (2001).
18. McQuibban, G. A. *et al.* Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3. *Science* **289**, 1202–1206 (2000).
19. Gronski, T. J. Jr *et al.* Hydrolysis of a broad spectrum of extracellular matrix proteins by human macrophage elastase. *J. Biol. Chem.* **272**, 12189–12194 (1997).
20. Belaaouaj, A. *et al.* Mice lacking neutrophil elastase reveal impaired host defense against gram-negative bacterial sepsis. *Nature Med.* **4**, 615–618 (1998).
21. Belaaouaj, A., Kim, K. S. & Shapiro, S. D. Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase. *Science* **289**, 1185–1188 (2000).
22. Weinrauch, Y., Drujan, D., Shapiro, S., Weiss, J. & Zychlinsky, A. Neutrophil elastase targets virulence factors of enterobacteria. *Nature* **417**, 91–94 (2002).
23. Lehrer, R. I. & Ganz, T. Antimicrobial polypeptides of human neutrophils. *Blood* **76**, 2169–2181 (1990).
24. Wilson, C. L. *et al.* Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense. *Science* **286**, 113–117 (1999).
25. Vu, T. H. *et al.* MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes. *Cell* **93**, 411–422 (1998).
26. Li, Q., Park, P., Wilson, C. & Parks, W. Matrilysin shedding of syndecan-1 regulates chemokine mobilization and transepithelial efflux of neutrophils in acute lung injury. *Cell* **111**, 635–646 (2002).
27. Curci, J. A., Liao, S., Huffman, M. D., Shapiro, S. D. & Thompson, R. W. Expression and localization of macrophage elastase (matrix metalloproteinase-12) in abdominal aortic aneurysms. *J. Clin. Invest.* **102**, 1900–1910 (1998).
28. Heymans, S. *et al.* Inhibition of plasminogen activators or matrix metalloproteinases prevents cardiac rupture but impairs therapeutic angiogenesis and causes cardiac failure. *Nature Med.* **5**, 1135–1142 (1999).
29. Golubkov, V. S. *et al.* Membrane type-1 matrix metalloproteinase (MT1-MMP-exhibits an important intracellular cleavage function and causes chromosome instability. *J. Biol. Chem.* **280**, 25079–25086 (2005).
30. Limb, G. A. *et al.* Matrix metalloprteinase-1 associates with intracellular organelles and confers resistance to laminin a/c degradation during apoptosis. *Am. J. Pathol.* **166**, 1555–1563 (2005).

METHODS

Reagents. Pro- and active forms of human MMP2, 3, 7, 8, 9 and 13, and MT1-MMP (also known as MMP14) were obtained from Chemicon. Synthetic peptides (University of Pittsburgh protein core) with >95% purity were generated for: mouse MMP12 CTD I, 344-SRNQLFLFKDEKYWLINNLV-363 (known as SR-20); mouse MMP12 CTD II, 370-RSIYSLGFSASVKKVDAAVF-389; human SR-20, ARNQVFLFKDDKYWLISNLR; MMP13 CTD, 343-SRDLMFIFRGR KFWALNGYD-362; SR-20 mutant peptide, 344-SRNXLFLF*SGRQ*YW LINNLV-363, with italic residues replacing KDEK. Recombinant mouse MMP12 CTD was generated in pET 20b vector with a 6-histidine tag. Three fragments of mouse MMP12 CTD were generated: I, 280-SSPPSTFCHQSLSF DAVTTVGEKIFFFKDWFFWWKLPGSPATNITSISSIWPSIPSGIQAAYEI-342; II, 343-ESRNQLFLFKDEKYWLINNLVPEPHYPRSIYSLGFSASVKKVDAAVFD PLRQKVYFFVDKHY-404; and III, 405-WRYDVRQELMDPAYPKLISTHFPG IKPKIDAVLYFKRHYYIFQGAYQLEYDPLFRRVTKTLKSTSWFGC-473. Vectors were transformed into *E. coli* BL2(DE3) LysE (Novagen). Protein was resuspended in 6 M urea, 300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 at pH 8.0, and purified using Talon binding resin (Clontech). After dialysis against 0.75 M urea, 300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 at pH 7.4, recombinant protein was verified by western blotting using 6-histidine residue antibody (Invitrogen).

Mice. $Mmp12^{-/-}$ mice previously generated in our laboratory have been described elsewhere[10]. All mice were maintained on a 129/SvJ background and housed in a sterile barrier facility. All experiments were approved by the Harvard Standing Committee for Animal Research and used age- and sex-matched controls.

Bacteria. *Staphylococcus aureus*, *K. pneumoniae* and *S. enteriditis* (clinical isolates) and *E. coli* (K1 strain) were grown in TSB for 18 h at 37 °C. Bacteria in mid-log phase growth were centrifuged at 1,800*g* for 10 min, washed, and resuspended in PBS. Bacterial concentration was determined by counting c.f.u. from limiting dilutions after an 18-h incubation at 37 °C on antibiotic-free LB plates.

***In vivo* models of infection.** Peritonitis: i.p. injection of either 1×10^8 c.f.u. *S. aureus* (*n* = 40, each group from six experiments) or 1×10^8 c.f.u. *E. coli* ($Mmp12^{-/-}$, *n* = 15; wild type, *n* = 16 from three experiments). Peritoneal fluid, liver, blood and lungs were collected from a subset of mice (*n* = 4 each group) and c.f.u. were determined from homogenates. In a separate experiment, peritoneal fluid macrophage and neutrophil counts were measured from peritoneal lavage using a haemocytometer and Wright-stained cytospins (*n* = 6 each group).

Haematogenous infection: TVI of 1×10^8 c.f.u. *S. aureus* into $Mmp12^{-/-}$ (*n* = 16) and wild-type control (*n* = 13) mice (from three experiments). Lungs, spleen and kidneys were removed to determine c.f.u.

Pneumonia: 1×10^8 c.f.u. *S. aureus* i.t. into $Mmp12^{-/-}$ (*n* = 36) and wild-type control (*n* = 30) mice (from six experiments). The mortality was recorded over a 14-day time course, as described earlier. $Mmp12^{-/-}$ and wild-type mice (*n* = 8, each group) were also subjected to a sub-lethal dose i.t. (1×10^6 c.f.u.) of *S. aureus*. Lung homogenates were generated at 2 and 24 h after instillation. Organ homogenates were made in 1 ml of 50 mM Tris-HCl, 150 mM NaCl, 10 mM $CaCl_2$ at pH 7.4 containing a proteinase inhibitor cocktail (Sigma). Haematoxylin and eosin, Brown and Breen-modified Gram stain, and macrophage-specific mac-3 immunostaining (Invitrogen) were performed on 5-mm thick sections.

***In vitro* bacterial killing assays.** Bacteria were incubated with MMP3, MMP7, MMP8, MMP9, MMP12, MMP12 CTD, MMP13, MMP13 CTD, and MT1-MMP in either 5% TSB or RPMI plus 10% FCS, in separate experiments. $CaCl_2$ (100 mM) was added to reactions with peptides, which were solubilized in 10% dimethylsulphoxide (DMSO). Direct bactericidal activity of MMP12 CTD was assessed using fluorescent probes Syto 59 and S-7020 (Molecular Probes) at concentrations of 5 μM and 20 μM, respectively. Quantification of dead versus total cells was performed using Metamorph image analysis software from three experiments.

***In vitro* bacterial killing by murine peritoneal macrophages.** Thioglycollate-stimulated peritoneal macrophages were resuspended in DMEM with 5% FCS (without antibiotics), and plated at a concentration of 2.5×10^5 cells per well in a 24-well plate. Cytospin analysis confirmed >95% pure macrophages, with the remainder being comprised of neutrophils and fibroblasts. *Staphylococcus aureus* (1×10^6 c.f.u.) was added for a 60-min incubation. At this time (*t* = 0), the cells were washed with PBS to clear non-adherent bacteria. The cells were lysed 30 and 90 min later with 0.2% Triton X-100 and c.f.u. were measured. Experiment was performed three times. Teflon-coated wells (Costar) were used for both scanning electron microscopy and transmission electron microscopy. Immuno-electron-microscopy used an MMP12 antibody generated in our laboratory (1:100) with a gold-labelled (2-nm particles) secondary to rabbit.

Liposomal release assay. *Staphylococcus aureus* was grown to mid-log phase, centrifuged, pelleted and freeze-fractured with dry ice. Chloroform/methanol (2:1) was added to a final volume of 5 ml. The mixture was agitated for 20 min in an orbital shaker at 25 °C, and centrifuged at 700*g* for 10 min before removing the lipid phase. The chloroform was evaporated under vacuum. Bacterial membrane lipids were hydrated in 1 mM $CaCl_2$, 10 mM MOPS, 100 mM KCl at pH 7.2. Membranes were then freeze-fractured and incubated in the presence of fluorescent Calcium Green-1 Dextran conjugates 3000 MW (Molecular Probes). Bacterial membrane vesicles were incubated in the presence and absence of MMP12 C-terminal protein (20 μg ml^{-1}) for 60 min. Results are from three experiments.

MMP12 processing. Full-length pro-MMP12 was incubated with or without *S. aureus* in RPMI for 10, 20 or 30 min. The reactions were centrifuged at 8,000*g* and the supernatant was subjected to western blotting using an antibody directed against the MMP12 CTD (Santa Cruz).

Structural studies. In the absence of an experimental structure of mouse MMP12 CTD, a homology model was constructed on the basis of the crystal structures of related full-length MMPs. This strategy was followed instead of considering only CTD structures to get further information on the relative orientations of the pro-, catalytic domain and the CTD. An initial search with PSI-Blast (http://www.ncbi.nlm.nih.gov/blast/blast.cgi) of the mouse pro-MMP12 sequence (UniProt database entry P34960) against the Protein Data Bank (PDB) using standard values identified human pro-MMP1 (PDB accession 1su3) as the closest homologue with an E-value of 9E-110 and 48% sequence identity. On the basis of this experimental structure, models for mouse pro-MMP12 covering residues 55–462 were generated by using programs CPHmodels 2.0 (http://www.cbs.dtu.dk/services/CPHmodels) and Swiss Model (http://swissmodel.expasy.org). In addition, human pro-MMP1, the catalytic domain of human MMP12 (PDB accession 1JK3), together with other reported experimental MMP structures encompassing the catalytic and/or the haemopexin-like domain, that is, human pro-MMP2 (PDB accession 1CK7), human pro-MMP2 in complex with TIMP2 (PDB accession 1GXD), porcine MMP1 (PDB accession 1FBL), and human MMP13 (PDB accession 1PEX) were further superimposed on the homology model for structure assessment.

Statistics. Kaplan–Meier life survival curve analysis was performed using the Prism Data Analysis software, which uses the log-rank test for curve comparison analysis. For all other experiments, the data are expressed as the mean value and s.d. Statistical significance was determined using the Student's *t*-test (two-tailed distribution with a two sample equal variance). $P < 0.05$ was considered significant.

LETTERS

MicroRNA-mediated switching of chromatin-remodelling complexes in neural development

Andrew S. Yoo[1,2,3], Brett T. Staahl[1,2], Lei Chen[1,2,3] & Gerald R. Crabtree[1,2,3]

One of the most distinctive steps in the development of the vertebrate nervous system occurs at mitotic exit when cells lose multipotency and begin to develop stable connections that will persist for a lifetime[1,2]. This transition is accompanied by a switch in ATP-dependent chromatin-remodelling mechanisms that appears to coincide with the final mitotic division of neurons. This switch involves the exchange of the BAF53a (also known as ACTL6A) and BAF45a (PHF10) subunits within Swi/Snf-like neural-progenitor-specific BAF (npBAF) complexes for the homologous BAF53b (ACTL6b) and BAF45b (DPF1) subunits within neuron-specific BAF (nBAF) complexes in post-mitotic neurons. The subunits of the npBAF complex are essential for neural-progenitor proliferation, and mice with reduced dosage for the genes encoding its subunits have defects in neural-tube closure similar to those in human spina bifida[3], one of the most serious congenital birth defects. In contrast, BAF53b and the nBAF complex are essential for an evolutionarily conserved program of post-mitotic neural development and dendritic morphogenesis[4,5]. Here we show that this essential transition is mediated by repression of BAF53a by miR-9* and miR-124. We find that BAF53a repression is mediated by sequences in the 3′ untranslated region corresponding to the recognition sites for miR-9* and miR-124, which are selectively expressed in post-mitotic neurons. Mutation of these sites led to persistent expression of BAF53a and defective activity-dependent dendritic outgrowth in neurons. In addition, overexpression of miR-9* and miR-124 in neural progenitors caused reduced proliferation. Previous studies have indicated that miR-9* and miR-124 are repressed by the repressor-element-1-silencing transcription factor (REST, also known as NRSF)[6]. Indeed, expression of REST in post-mitotic neurons led to derepression of BAF53a, indicating that REST-mediated repression of microRNAs directs the essential switch of chromatin regulatory complexes.

The ATP-dependent chromatin-remodelling complexes, typified by the yeast Swi/Snf complex, regulate chromatin assembly and accessibility[7,8]. The mammalian genome encodes nearly 30 different Swi2/Snf2-like ATPases, two of which, BRG1 and BRM are alternative subunits in complexes of 11 subunits termed BAF or mammalian SWI/SNF (mSWI/SNF)[3,4,9–11]. To understand the essential switch in subunit composition of these complexes during neural development, we examined 180 kilobases (kb) around the BAF53a gene for transcriptional regulatory regions by replacing the first exon of BAF53a at the start codon with a destabilized nuclear enhanced green fluorescent protein (d2nucEGFP) in a BAF53a-containing bacterial artificial chromosome (BAC). The d2nucEGFP is followed by the stop codon and a 3′ untranslated region (UTR) so that EGFP expression reflects BAF53a transcription (Fig. 1a, diagram). The BAF53a BAC reporter with a SV40 3′ UTR gave persistent expression of EGFP in post-mitotic neurons in the neural tube (Fig. 1a, lower-left panel) and, hence, failed to mimic the expression pattern of the endogenous

gene (Fig. 1b). However, when the 3′ UTR of BAF53a was used in place of the SV40 UTR, the d2nucEGFP reporter faithfully reproduced the downregulation of the endogenous BAF53a gene in post-mitotic neurons (Fig. 1a, lower-right panel, and Supplementary
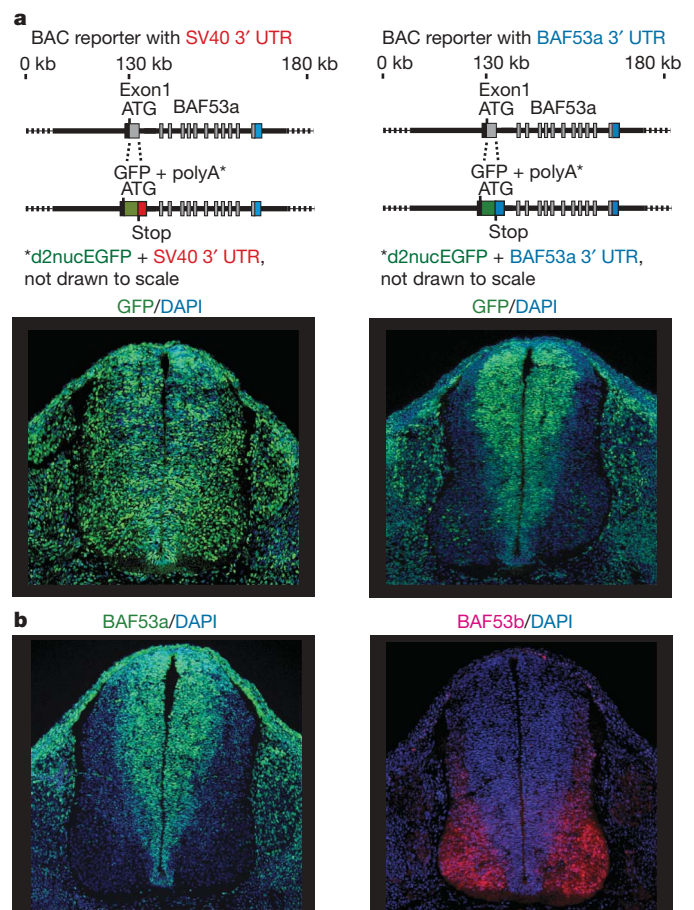

Figure 1 | BAF53a repression is mediated by sequences within its 3′ UTR. a, Schematic diagrams of BAC clones of approximately 180 kb containing BAF53a and the BAC-based reporter constructs. Exon 1 from the ATG of BAF53a to the end of the exon was replaced by d2nucEGFP and 3′ UTR. Whereas EGFP failed to be repressed in neurons with SV40 UTR (lower-left panel), the BAF53a 3′ UTR allowed efficient downregulation of the reporter in neurons (lower-right panel). Photographs in the lower panels show transverse sections of the neural tubes of E11.5 transgenic embryos. DAPI, 4,6-diamidino-2-phenylindole. **b,** BAF53a and BAF53b are expressed selectively in progenitors (green) and neurons (red), respectively, in the neural tubes of E11.5 embryos.

---

[1]Howard Hughes Medical Institute, [2]Department of Developmental Biology, [3]Department of Pathology, Stanford University, Stanford, California 94305, USA.

Fig. 1), implicating the evolutionarily conserved 3′ UTR (Supplementary Fig. 1) in post-mitotic suppression of BAF53a.

The 3′ UTR of BAF53a has predicted binding sites for microRNAs (miRNAs) miR-9, miR-9* and miR-124 (refs 12, 13; see Fig. 2a and Supplementary Fig. 2), small, 20–22-base-pair (bp), non-coding RNAs that inhibit expression of their target genes by transcript degradation or translation inhibition[14]. These miRNAs are enriched in

neurons[15] and mouse brain tissues[16]. There appeared to be three sites for miR-9* and single sites for miR-9 and miR-124 (Fig. 2a and Supplementary Fig. 2). Short, 6-bp, sequences complementary to the 5′ ends of miRNAs are important for target recognition by miRNAs[17,18]. To determine whether the predicted binding sites for miRNAs are responsible for BAF53a repression, we inserted point mutations into the 3′ UTR targeting sequences for miR-9, miR-9* and miR-124 (Fig. 2b, left panel). The mutant BAF53a 3′ UTR failed to repress its reporter in post-mitotic neurons (Fig. 2b and Supplementary Fig. 3), indicating that the predicted binding sites in the 3′ UTR mediate BAF53a downregulation.

To distinguish the relative roles of the different miRNAs targeting BAF53a, we designed a construct to 'sense' and quantify the miRNA-mediated repression of the reporter through the 3′ UTR of BAF53a. This sensor construct contains the d2nucEGFP with test 3′ UTRs and tomato fluorescent protein (nuc mTom) with the invariant SV40 3′ UTR (Supplementary Fig. 4). Stable cell lines containing miR-9, miR-9* and miR-124 overexpression and sensor constructs were established (Supplementary Fig. 4). The function of 3′ UTRs in EGFP expression was scored as the ratio of EGFP/Tomato signals in individual cells. When the BAF53a 3′ UTR was tested, EGFP expression was significantly reduced relative to EGFP expression with SV40 3′ UTR (Fig. 2c). Mutating the miR-9* and miR-124 sites abolished the downregulation of the EGFP signal, and removal of the miR-9 site gave no additional derepression (Fig. 2c). Mutating only the miR-9* sites or the miR-124 site was not sufficient to abolish the repression of the EGFP signal. These results suggest that miR-9* and miR-124 are both required to inhibit BAF53a expression. Furthermore, in situ analysis showed that miR-9* was expressed mostly in post-mitotic neurons in the neural tube of E11.5 embryos (Fig. 2d). Published studies show that miR-124 was expressed in a nearly identical pattern to miR-9* (refs 19–21), indicating that both miR-9* and miR-124 are expressed just before BAF53a repression in post-mitotic neurons.

MicroRNAs typically target multiple genes[22]. To define the contribution of miRNA-mediated repression of BAF53a to neural development, we prepared transgenic embryos expressing the BAC containing mutations in the 3′ UTR of BAF53a. In contrast to the wild-type BAC-transgenic embryos (Fig. 3a), embryos containing the mutant BAC transgene displayed a loss of the clear boundary of BAF53a downregulation, and instead showed irregular expansion of BAF53a expression into β-tubulin-III-positive neurons (Fig. 3b). Because BAF53a and BAF53b are expressed in a mutually exclusive pattern (Fig. 1b), we then investigated whether BAF53b expression would be affected by the prolonged expression of BAF53a. Embryos transgenic with the wild-type BAF53a BAC showed strong and specific expression of BAF53b in β-tubulin-III-positive neurons (Fig. 3c). However, continued expression of BAF53a in neurons repressed BAF53b expression (Fig. 3d, e). These results suggest that BAF53a directly or indirectly antagonizes BAF53b expression, and provide a possible explanation of how the mutually exclusive BAF switching is achieved during neural development.

Because BAF53b is essential for dendritic morphogenesis[4,5], this result suggests that miRNA-mediated repression of BAF53a is essential for the onset of post-mitotic neuronal function. We tested this hypothesis by overexpressing BAF53a in cultured hippocampal neurons and assayed activity-dependent dendritic outgrowth (stimulated by KCl treatment[4]). Persistent expression of BAF53a in neurons inhibited stimulated dendritic outgrowth (Fig. 3f), elucidating the importance of BAF53a switching in neurons. We then tested the contribution of miRNA-mediated regulation of BAF53a in dendritic outgrowth. When BAF53a was expressed with the SV40 3′ UTR, we observed strong inhibition of dendritic outgrowth in neurons; expression with the wild-type BAF53a 3′ UTR did not produce such inhibition (Fig. 3g). The dendritic phenotype reappeared when we mutated the binding sites for miR-9* and miR-124 in the BAF53a 3′
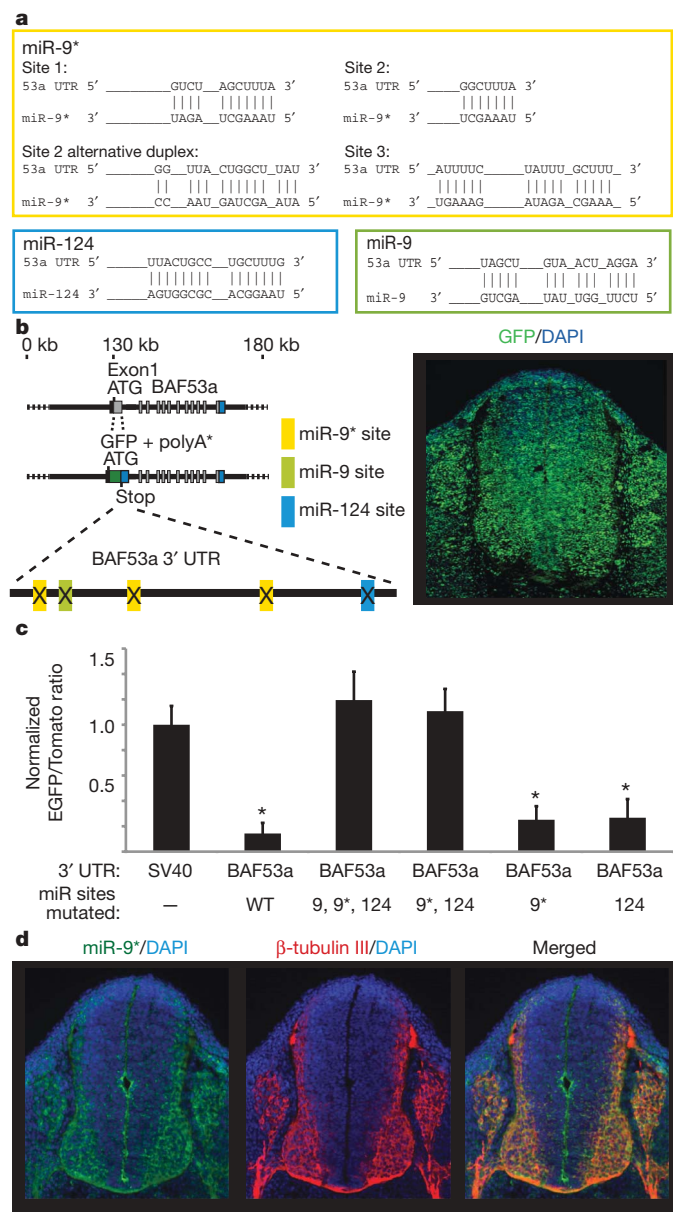
Figure 2 | BAF53a is a target of miR-9* and miR-124. a, Predicted configurations of duplex formations between miRNAs and their target sites. **b,** Expression of BAF53a BAC reporter, containing mutations at the miRNA-binding sites in the 3′ UTR of d2nucEGFP. Point mutations were made to 4–6 bp predicted to bind to the 5′ ends of miR-9, miR-9* and miR-124. The photograph shows a representative expression pattern seen in three individual transgenic lines. **c,** BAF53a 3′ UTR sensor experiments in stable Chinese hamster ovary (CHO) cells expressing miR-9, mirR-9* and miR-124 and sensor reporter constructs. The quantification of sensor expression was expressed as the ratio of EGFP/Tomato signals and normalized to the ratio values of the SV40 UTR sensor. Error bars, s.d. Signal intensities were averaged in ten random fields totalling approximately 500–1,000 cells. *P < 0.001, Student's t-test. WT, wild type. **d,** Selective expression of miR-9* in post-mitotic neurons. Photographs show fluorescent in situ hybridization analysis of miR-9* expression (in green) and β-tubulin-III staining (in red) in the neural tubes of E11.5 embryos.
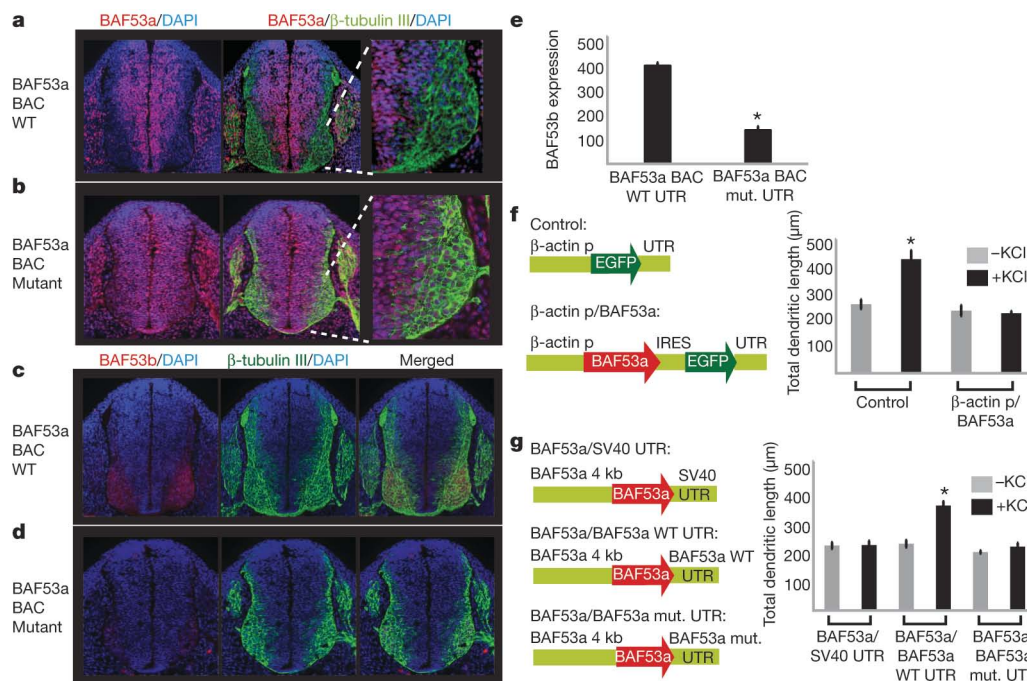
Figure 3 | BAF53a repression is essential for activity-dependent dendritic outgrowth in neurons. **a**, Normal downregulation of BAF53a in post-mitotic neurons in transgenic embryos with wild type BAF53a BAC. The rightmost panel shows the lower-right quadrant of the neural tube. **b**, Persistent expression of BAF53a in neurons seen with BAF53a BAC containing point mutations in the miRNA-binding sites. **c**, Normal expression of BAF53b (red) in β-tubulin-III-positive (green) neurons in transgenic embryos with wild-type BAF53a BAC. **d**, Reduced BAF53b expression with persistent expression of BAF53a in neurons. **e**, Quantification of BAF53b expression: ratio of BAF53b level (arbitrary units) and β-tubulin-III-positive neurons. Average values are from eight sections of the neural tube. Error bars, s.e. *$P < 0.005$, Student's *t*-test.

f, Constructs to overexpress BAF53a in cultured hippocampal neurons and quantification of dendritic outgrowth of GFP-positive neurons upon stimulation using KCl. The average values are from five individual coverslips from two independent experiments, with each coverslip containing 50–100 scored neurons. Error bars, s.e. *$P < 0.005$, Student's *t*-test. p, promoter; IRES, internal ribosome entry site. **g**, Schematic diagrams of BAF53a expression constructs using different 3′ UTRs and quantification of dendritic outgrowth of transfected neurons upon stimulation using KCl. In independent experiments, we found that the 4-kb upstream region of BAF53a (illustrated) was sufficient to drive expression of GFP reporters that could be repressed by endogenous miR-9* and miR-124. Error bars, s.e. *$P < 0.001$, Student's *t*-test.

UTR (Fig. 3g). These results underline the essential role of miR-9*- and miR-124-mediated BAF53a repression in dendritic outgrowth.

To elucidate further the function of miR-9* and miR-124, we created transgenic embryos in which miR-9* and -124 were ectopically expressed under the control of the nestin promoter in neural progenitors (Fig. 4a). Control transgenic embryos in which the nestin promoter drove EGFP expression only showed the expected co-localization between EGFP-positive cells and BAF53a expression. When the nestin promoter drove expression of miR-9* and miR-124 ectopically in progenitors, transgenic embryos showed significantly reduced expression of BAF53a in progenitors (Fig. 4b, top-right graph). This reduction in BAF53a expression was probably not due to precocious differentiation of the progenitors into neurons, because the miR-9* and miR-124 expressing progenitors did not show staining for BAF53b (Supplementary Fig. 5).

As BAF53a has previously been shown to promote proliferation of progenitors[3], we asked whether the reduction in BAF53a expression would be associated with the proliferation state of progenitors. As expected, control transgenic embryos showed complete co-localization between K_i-67 (a marker for proliferating cells) and EGFP-positive progenitors (Fig. 4b). We found that overexpressing miR-9* and miR-124 hindered the proliferation of progenitors, evidenced by the significantly reduced expression of K_i-67 in EGFP-positive progenitors (Fig. 4b). This proliferation was significantly restored when miRNA-resistant BAF53a (BAF53a$^{\Delta\Delta}$) was expressed with miR-9* and miR-124 (Fig. 4b). The results are consistent with the previous finding that one of the functions of BAF53a and other components of the npBAF complex in progenitors is to promote proliferation, and suggest a role for miR-9* and miR-124 in combination to prevent proliferation in post-mitotic neurons by removing BAF53a expression.

We wondered whether prolonging BAF53a expression in already committed neurons would be sufficient to trigger re-entry into the cell cycle in neurons. In mutant BAF53a BAC-transgenic embryos, BAF53a expression was extended into neurons (Fig. 3b). However, K_i-67 expression eventually disappeared in the neuronal zone in these embryos, suggesting that continued BAF53a expression in committed neurons was not sufficient to cause neurons to re-enter the cell cycle (Supplementary Fig. 6). Consistent with this finding, expressing BAF53a in neuronal cultures was insufficient to reactivate proliferation (data not shown).

The miRNAs miR-9* and miR-124 have been shown to be targets of REST[6]. REST, with its co-repressor, restricts the activation of neuronal genes to neurons by silencing its target genes in non-neuronal cells; REST activity is reduced as progenitors differentiate into neurons, allowing the expression of neuronal genes[23–28]. All three loci of miR-9* and miR-124 contain REST-binding sites, which inhibit miR-9* and miR-124 expression[6], suggesting that REST expression would activate BAF53a in post-mitotic neurons. Neurons transfected with an REST-expression plasmid reactivated BAF53a expression, which contrasted with the absence of BAF53a expression in control-transfected neurons (Fig. 4c). We reasoned that this BAF53a expression in neurons should be suppressed if we co-introduced a plasmid that overexpresses miR-9* and miR-124 when driven by a promoter whose activity is independent of REST. Indeed, co-transfection of miR-9* and miR-124 along with REST-expression plasmids suppressed BAF53a reactivation in neurons (Fig. 4c), suggesting an antagonism between REST and miRNAs in repressing BAF53a in post-mitotic neurons.

Our studies indicate that the mechanism underlying the essential switch in subunit composition of Swi/Snf-like BAF complexes
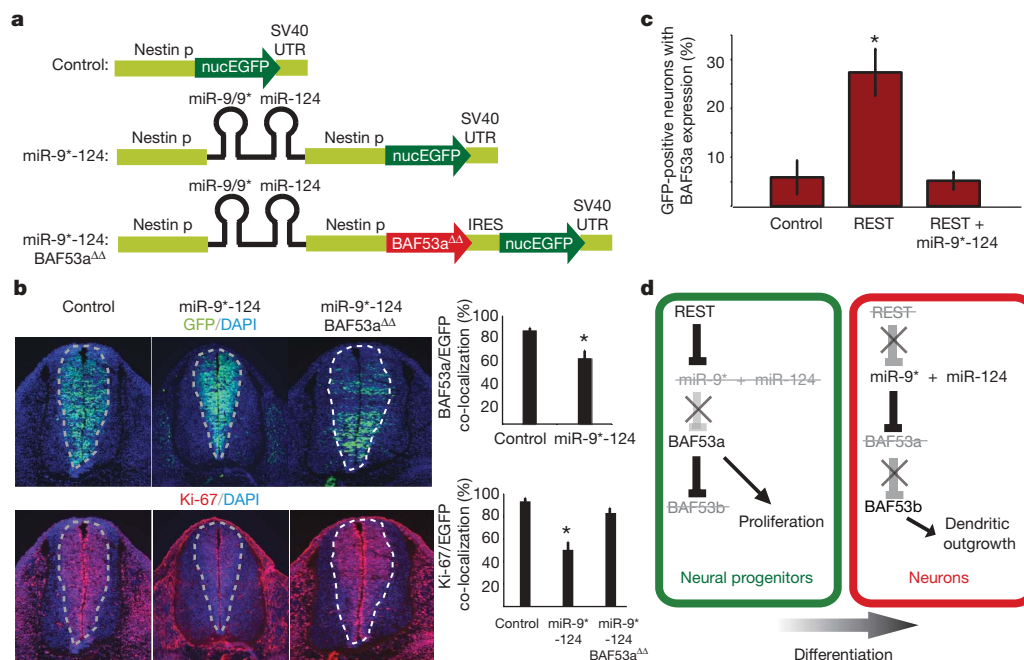
Figure 4 | Effect of miR-9* and miR-124 overexpression in progenitors and REST in neurons. a, miR-9* and miR-124 overexpression constructs (miR-9*-124) expressed in progenitors. nucEGFP, nuclear enhanced green fluorescent protein. **b,** Proliferative defect caused by miR-9* and miR-124 overexpression in progenitors. Overexpression of miR-9* and miR-124 reduced BAF53a expression in the EGFP-positive progenitors relative to control (upper-right graph). Photographs show transgenic embryos containing either control or miR-9*-124 overexpression constructs and K_i-67 staining, quantified in the lower-right graph on the basis of nine individual tissue sections. Error bars, s.e. *$P < 0.001$, Student's t-test.

c, BAF53a regulation is mediated by REST. Hippocampal neurons cultured from E18.5 embryos were transfected with control (EGFP only), REST (EGFP + REST expression constructs) or REST + miR-9*-124 (EGFP + REST + miRNA overexpression constructs). The quantification represents average percentages of transfected GFP-positive neurons that had BAF53a expression from five coverslips. We analysed 15–25 GFP-positive neurons per coverslip. Error bars, s.e. *$P < 0.05$, Student's t-test. **d,** Genetic pathway leading to BAF53-subunit switching during neuronal differentiation.

involves miRNA-mediated repression of BAF53a as neural progenitors differentiate into neurons (Fig. 4d). The repression is accomplished through the 3′ UTR of BAF53a and mediated by the simultaneous activities of miR-9* and miR-124. REST participates in this switch by repressing miR-9* and miR-124, thereby permitting BAF53a expression in neural progenitors. In neurons, the absence of REST allows the expression of miR-9* and miR-124 to repress BAF53a (Fig. 4d). Repression of BAF53a allows BAF53b expression, which is critical for dendritic development[4,5]. Failure of miRNA-mediated repression of BAF53a results in reduced BAF53b expression and defective dendritic outgrowth (Fig. 4d). Because repression of BAF53a reduces progenitor proliferation, miRNA-mediated repression of BAF53a may be essential in regulating the number of cell divisions in the vertebrate nervous system. Mice with reduced BAF155 and BRG1 levels have defects in neural-tube closure similar to spina bifida in humans[3,29,30], indicating a genetically dominant role of npBAF in progenitor proliferation. Thus, studies of the genes involved in the npBAF-to-nBAF switch might yield insights into the pathogenesis of spina bifida and perhaps also the mechanisms controlling the number cell divisions in the central nervous system.

METHODS SUMMARY

Transgenic mouse embryos. We made BAC-based reporters by recombineering using galK-positive and -negative selections. Detailed protocols for constructing BAC reporters and other related transgene constructs are provided in the full Methods. Transgenic mouse embryos were generated by pronuclear injection of fertilized oocytes from CB6-F1 mice, and analysed 11.5 days after fertilization in cryostat sections immunostained with appropriate antibodies.

Tissue culture. We cultured E18.5 hippocampal neurons on coverslips coated with poly-L-ornithine, fibronectin and laminin in 24-well plates for four days in vitro (div), and transfected with appropriate plasmids after 4 div. Cells were fixed and stained 48 h later, after 6 div. Immunocytochemistry was done with affinity-purified BAF53a and Tuj-1 (BD Biosciences) primary antibodies and goat

anti-rabbit or mouse secondary antibodies (Invitrogen) with DAPI (Vector) DNA staining. Detailed protocols are provided in the full Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Noctor, S. C., Martinez-Cerdeno, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nature Neurosci. **7,** 136–144 (2004).
2. Temple, S. The development of neural stem cells. Nature **414,** 112–117 (2001).
3. Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron **55,** 201–215 (2007).
4. Wu, J. et al. Regulation of dendritic development by neuron-specific chromatin remodeling complexes. Neuron **56,** 94–108 (2007).
5. Parrish, J. Z., Kim, M. D., Jan, L. Y. & Jan, Y. N. Genome-wide analyses identify transcription factors for proper morphogenesis of Drosophila sensory neuron dendrites. Genes Dev. **20,** 820–835 (2006).
6. Conaco, C., Otto, S., Han, J. J. & Mandel, G. Reciprocal actions of REST and a microRNA promote neuronal identity. Proc. Natl Acad. Sci. USA **103,** 2422–2427 (2006).
7. Laurent, B. C., Treich, I. & Carlson, M. The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. Genes Dev. **7,** 583–591 (1993).
8. Peterson, C. L. & Herskowitz, I. Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. Cell **68,** 573–583 (1992).
9. Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature **366,** 170–174 (1993).
10. Lemon, B., Inouye, C., King, D. S. & Tjian, R. Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. Nature **414,** 924–928 (2001).
11. Wang, W. et al. Diversity and specialization of mammalian SWI/SNF complexes. Genes Dev. **10,** 2117–2130 (1996).
12. Krek, A. et al. Combinatorial microRNA target predictions. Nature Genet. **37,** 495–500 (2005).
13. Miranda, K. C. et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell **126,** 1203–1217 (2006).

14. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).

15. Krichevsky, A. M., Sonntag, K. C., Isacson, O. & Kosik, K. S. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem Cells* **24**, 857–864 (2006).

16. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).

17. Didiano, D. & Hobert, O. Molecular architecture of a miRNA-regulated 3′ UTR. *RNA* **14**, 1297–1317 (2008).

18. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).

19. Cao, X., Pfaff, S. L. & Gage, F. H. A functional study of miR-124 in the developing neural tube. *Genes Dev.* **21**, 531–536 (2007).

20. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).

21. Visvanathan, J., Lee, S., Lee, B., Lee, J. W. & Lee, S. K. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes Dev.* **21**, 744–749 (2007).

22. Flynt, A. S. & Lai, E. C. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nature Rev. Genet.* **9**, 831–842 (2008).

23. Lunyak, V. V. *et al.* Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**, 1747–1752 (2002).

24. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121**, 645–657 (2005).

25. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20**, 136–142 (1998).

26. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).

27. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).

28. Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K. & Gage, F. H. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* **116**, 779–793 (2004).

29. Bultman, S. *et al.* A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* **6**, 1287–1295 (2000).

30. Kim, J. K. *et al.* Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Mol. Cell. Biol.* **21**, 7787–7795 (2001).

Author Contributions A.S.Y. and G.R.C. generated the hypotheses, designed experiments and wrote the manuscript. A.S.Y. performed experiments and generated Figs 1, 2, 3 and 4 and supplementary data. B.T.S. and A.S.Y. performed experiments for Figs 3 and 4. L.C. performed the pronuclear injections used in generating transgenic mouse embryos.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.R.C. (crabtree@stanford.edu).

nature

METHODS

Recombineering of BAC containing BAF53a and mutagenesis. BAC (RP23-370O17, approximately 180 kb around BAF53a) was recombined at the first exon of BAF53a with d2nucEGFP, using the *galK*-positive and -negative selection method as previously described[31]. SW106 cells were transformed with RP23-370O17, selected and transformed with *galK* PCR product amplified by primers with 50-bp arms homologous to the sites to be recombined: *galK* FWD, TGTTATCTTTCGTCCGGTAGTCTTCGGCCAGTCCCCGCCAGACAGTAG CC CCTGTTGACAATTAATCATCGGCA; *galK* REV, GGCCGGTCGTTCC GGTCCGCACGCCCCGAGGCGCCGCTGCGCTCACTCACTCAGCACTGTC CTGCTCCTT.

Recombination was performed by heat-shocking the transformed SW106 cells, and colonies containing the recombined products were selected on minimal agarose plates plus galactose. The *galK*-containing RP23-370O17 was then recombined to replace *galK* with d2nucEGFP gene by transforming SW106 cells with PCR product (below) of d2nucEGFP/3′ UTR and homology arms. Different versions of 3′ UTRs were subcloned into a d2nucEGFP-containing plasmid, and used as templates for PCR amplification. d2nucEGFP FWD, TGTTATCTTTCGTCCGGTAGTCTTCGGCCAGTCCCCGCCAGACAGTAG CC ATGGTGCACGTGGATCCAAAAAAG; d2nucEGFP REV (for SV40 UTR): GGCCGGTCGTTCCGGTCCGCACGCCCCGAGGCGCCGCTGCGCT CACTCAC ATCGCCCTGATAGACGGTTTTTCG; d2nucEGFP REV (for BAF53a UTR): GGCCGGTCGTTCCGGTCCGCACGCCCCGAGGCGCCG CTGCGCTCACTCACGGGAAGGATCTGAGGAGGGCTCAG.

Point mutations in BAF53a 3′ UTR were generated by QuikChange mutagenesis (Stratagene) using the following primers (bold fonts denote mutated base pairs): miR-9* site1, CACCTCAACGTCTGTA**CGAA**TAGTACACTCAGGAA; miR-9* site2-1, CCACTTAAATTTTTTAAG**CGAA**TAACTGGCTCTAT AAATTAAATG; miR-9* site2-2, CTTAAATTTTTTAAGCGAATAA**GACCGC** CTATAAATTAAATGAG; miR-9* site3, GAATGTCTATTTTCTCTAAATAT TTT**CGAA**TCAGTAAGTACTCTCC; miR-124 site, GTACAAGTTACTGC CTAA**ACGAAAC**TACCTTAGGCTACAAAAC; miR-9 site, TGTACGAATAG TACAC**ACTCCT**AAAGATGACCATCTTTTG.

SW106 cells transformed with d2nucEGFP PCR product were grown on 2-deoxy-D-galactose-containing minimal plates for negative selection. Colonies with correctly recombined BACs were identified by PCR using primers surrounding the recombined 5′ and 3′ junctions. BAC DNA was prepared from the isolated colonies, rechecked by PCR and further confirmed by sequencing.

BAF53a BAC containing mutations in the 3′ UTR of BAF53a was generated similarly by recombining the 3′ UTR of BAF53a with the PCR product of mutant 3′ UTR plus homology arms using the following primers: BAF53a UTR H1 *galK* FWD, AGGAATATGAAGAAGGAGGGAAGCAGTGTGTAGAAAGAAAATG CCCTTGACCTGTTGACAATTAATCATCGGCA; BAF53a UTR H2 *galK* REV, CAGCAAAGTCAGTTTGGAAATGCTGTTCTCACTAGTGTTAAACTG TGAGGTCAGCACTGTCCTGCTCCTT; BAF53a UTR H1 FWD, AGG AATATGAAGAAGGAGGGAAGCAGTGTGTAGAAAGAAAATGCCCTTGAG GGCTCCACCCTGCCTGCC; BAF53a UTR H2 REV, CAGCAAAGTCA GTTTGGAAATGCTGTTCTCACTAGTGTTAAACTGTGAGGGGGAAGGAT CTGAGGAGGGCTCAG.

Transgenic embryos. All BAC-transgenic embryos were generated by pronuclear injections. More than three E11.5 embryos were examined for each BAC-reporter transgene.

Data analysis. Images for embryonic sections were captured using a Leica DM5000B microscope with Leica Application Suite (LAS) Advanced Fluorescence 1.8.0. Intensity measurements for BAF53a expression in REST expression (Fig. 4c) were performed with LAS 1.8.0. Images for stable Chinese hamster ovary lines containing sensor reporters were made using a Leica DMI4000B microscope with LAS 2.8.1. Image intensity analyses for sensor experiments (Fig. 2c) were performed with METAMORPH (version 7, Molecular Devices) using the images captured with the same exposure parameters. Co-localization analyses for Fig. 4b were also performed using METAMORPH.

Fluorescent *in situ* hybridization. Fluorescent *in situ* analysis was performed using the fluorescein-labelled locked nucleic acid (LNA) oligonucleotide for

miR-9* (Exiqon). Detailed protocols for LNA oligonucleotide *in situ* staining are provided on the manufacture's website (www.exiqon.com). The same sections were also stained for β-tubulin III using Tuj-1 antibody (BD Biosciences).

MicroRNA cluster construct. A cluster of miR-9* and miR-124 precursors was made by amplifying approximately 300 base pairs around mmu-miR-9*-3 and mmu-miR-124-2, which were together cloned into pSilencer (Ambion). mmu-miR-9*-3 precursor was amplified by the following primers containing appropriate restriction sites (not shown): mmu-miR-9-3 OX FWD, TGTCG GTCCCCTCTGGCTCTC; mmu-miR-9-3 OX REV, AGGAGAGCCGAGG TCGGATACC. mmu-miR-124-2 precursor was amplified using the following primers containing appropriate restriction sites: mmu-miR-124-2 OX FWD, AGGTGGGAGTACTGCTCAGAGCTA; mmu-miR-124-2 OX REV: GTGT CGCACAGGATGGGTTC.

The ability of the cluster construct to generate mature miR-9* and miR-124 was assayed by quantitative real-time PCR using a TaqMan miR-9* and miR-124 microRNA assay kit (Applied Biosystems). RNA was extracted from COS7 cells transfected with the miR-9*-124 overexpression construct using a mirVana miRNA isolation kit (Ambion). The same cluster of miR-9* and miR-124 was used for the constructs containing nestin promoters (Fig. 4a). Proliferation of progenitors in nestin p/miR-9*-124-transgenic embryos was assayed by immunostaining with Ki-67 antibody (BD Biosciences).

Hippocampal neuronal culture and REST overexpression experiments. The pActin–IRES–EGFP plasmid was constructed by inserting the chicken β-actin promoter followed by the IRES–EGFP module from the pIRES2–EGFP plasmid (Clontech) into pUC19 backbone. The pActin–REST–IRES–EGFP construct was generated by cloning the coding region of mouse REST gene into the pActin–IRES–EGFP vector. E18.5 murine hippocampal cells were cultured as previously described[32]. Dissociated cells were cultured on coverslips coated with poly-L-ornithine, fibronectin and laminin in 24-well plates. AraC was added after 1 div. Culture media contained DMEM/F12 with putrescine, 2-mercaptoethanol, transferrin, insulin, selenium, progesterone, MEM vitamin additive and 5% FBS. After 4 div, cells were transfected with GFP or other constructs using Lipofectamine (Invitrogen) as per the manufacturer's suggestions. Each transfection had 0.6 μg total DNA: pCMV–Bcl2 (to prevent occasional apoptosis of neurons associated with transfection), pActin–EGFP (control) or pActin–REST–EGFP, with or without pSil–mir-9*-124. Cells were fixed (4% PFA) and stained 48 h later, after 6 div. Immunocytochemistry was done with affinity-purified BAF53a and Tuj-1 (BD Biosciences) primary antibodies and goat anti-rabbit or mouse secondary antibodies (Invitrogen) with DAPI (Vector) DNA staining.

Dendritic outgrowth analysis. Hippocampal neurons were prepared from P0 mouse brains and transfected with BAF53a overexpression constructs on day one. Activity-dependent dendritic outgrowth analysis was performed as described[4]. Briefly, after 4 div, cells were treated with KCl (30 mM) in half of the cultures. Cells were fixed and stained for GFP (Molecular Probes) and MAP2 (to mark dendrites; Sigma-Aldrich) after 5 div. A GFP construct was co-transfected to visualize transfected neurons for the experiments shown in Fig. 3f, g. The total dendritic length was measured from dendritic structures defined by GFP and MAP2 signals in individual neurons using the neurite outgrowth extension of METAMORPH.

Prediction of miRNAs targeting BAF53a. We used PICTAR[12] and RNA22[13] to predict miRNAs that target BAF53a 3′ UTR. Because BAF53a downregulation is specific for neurons, we sought miRNAs enriched in neurons and the brain[15,16]. miR-9* was the highest-scored miRNA when the BAF53a 3′ UTR was searched in PICTAR. We further examined the 3′ UTR for additional sites for other miRNAs enriched in the brain using RNA22 (number of unpaired bases allowed, 1; seed/nucleus ratio, 7; minimum number of paired bases in heteroduplex, 14; maximum folding energy, $-18\,\mathrm{Kcal\,mol^{-1}}$) and identified the binding sites for miR-124 and miR-9 as outlined in Fig. 2.

31. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic Acids Res.* **33**, e36 (2005).
32. Graef, I. A. *et al.* L-type calcium channels and GSK-3 regulate the activity of NF-ATc4 in hippocampal neurons. *Nature* **401**, 703–708 (1999).

Evidence of *Xist* RNA-independent initiation of mouse imprinted X-chromosome inactivation

Sundeep Kalantry[1,2,3], Sonya Purushothaman[1,2,3], Randall Bryant Bowen[1,2,3], Joshua Starmer[1,2,3]
& Terry Magnuson[1,2,3]

XX female mammals undergo transcriptional silencing of most genes on one of their two X chromosomes to equalize X-linked gene dosage with XY males in a process referred to as X-chromosome inactivation (XCI). XCI is an example of epigenetic regulation[1]. Once enacted in individual cells of the early female embryo, XCI is stably transmitted such that most descendant cells maintain silencing of that X chromosome[2]. In eutherian mammals, XCI is thought to be triggered by the expression of the non-coding *Xist* RNA from the future inactive X chromosome (Xi)[3–5]; *Xist* RNA in turn is proposed to recruit protein complexes that bring about heterochromatinization of the Xi[6,7]. Here we test whether imprinted XCI, which results in preferential inactivation of the paternal X chromosome (Xp), occurs in mouse embryos inheriting an Xp lacking *Xist*. We find that silencing of Xp-linked genes can initiate in the absence of paternal *Xist*; *Xist* is, however, required to stabilize silencing along the Xp. Xp-linked gene silencing associated with mouse imprinted XCI, therefore, can initiate in the embryo independently of *Xist* RNA.

In the mouse, *Xist* RNA expression is first detected at the late two-cell stage of embryogenesis exclusively from the Xp[8]. During subsequent stages of pre-implantation development, *Xist* RNA spreads from its site of synthesis to eventually coat most of the Xp[4,5,9], concomitant with transcriptional silencing of Xp-linked genes[4,10,11]. Imprinted XCI of the Xp is then maintained in extra-embryonic tissues of the post-implantation embryo[12,13]. The embryonic lineage, on the other hand, selectively reactivates the Xp during peri-implantation stages and subsequently undergoes random XCI[11]. As in imprinted XCI, *Xist* expression is upregulated from the Xi-elect before X-linked gene silencing during random XCI[14].

Despite the currently accepted role of *Xist* RNA in controlling XCI, two observations led us to question whether *Xist* is required to initiate imprinted XCI. In the course of a previous study, we noticed that mouse trophoblast stem cells, which undergo imprinted XCI of the Xp[15], do not proliferate if they reactivate their previously inactivated Xp[16]. We therefore reasoned that if *Xist* RNA were required for the initiation of imprinted XCI, then female embryos that inherit a paternal *Xist* mutation should lack trophoblast-derived structures. However, Xp–*Xist* mutant embryos can develop extensive trophectodermal derivatives[17]. We thus hypothesized that *Xist* RNA may not be required for the initiation of imprinted XCI in early mouse embryos. In the present study, we provide several lines of evidence that substantiate the notion that imprinted XCI can initiate in the absence of *Xist*.

To determine when silencing of X-linked genes associated with imprinted XCI first occurs and whether it is *Xist* RNA dependent, we compared the expression of eleven X-linked genes by RNA fluorescence *in situ* hybridization (FISH) in multiple two-, four-, eight- and 16-cell pre-implantation-stage female embryos that inherited either a wild-type (WT) Xp or an Xp harbouring a null allele of *Xist* (Xp–*Xist*[1lox], a deletion of the promoter and first three exons resulting in lack of *Xist* RNA expression[18]) (Fig. 1). If both the maternal (Xm) and Xp alleles are expressed, the expectation is that RNA FISH would yield biallelism; if one of the two alleles is silenced, RNA FISH would result in monoallelism (a third category is 'no signal', which applies if neither allele is detectably expressed).

In WT two-cell embryos, nine of the X-linked genes—*Ube1x* (also known as *Uba1*), *Fmr1*, *G6pdx*, *Chic1*, *Rnf12* (*Rlim*), *Abcb7*, *Atrx*, *Atp7a* and *Smcx* (*Kdm5c*)—displayed a high level of biallelism (83–100%) and a low level of monoallelism (0–8%) (Fig. 1d and Supplementary Fig. 1). The remaining two genes—*Mecp2* and *Pgk1*—were expressed biallelically at relatively lower levels (59% and 44%, respectively) and monoallelically at higher levels (15% and 22%, respectively), but the embryos nevertheless harboured more biallelically expressing nuclei. In Xp–*Xist*[1lox] embryos, the genes showed a similar distribution of nuclei (biallelism, monoallelism and no signal) to that observed in WT embryos. The paucity of mono-allelism and a preponderance of biallelism in two-cell embryos is consistent with previous reports of both alleles of X-linked genes being expressed at the two-cell stage[19,20], and argues against Xp alleles being inherited as transcriptionally inert owing to silencing of the Xp during meiotic sex-chromosome inactivation (MSCI) in the paternal germ line[4,21].

We next assayed X-linked gene expression by RNA FISH in WT and Xp–*Xist*[1lox] four-cell embryos, to assess the degree of silencing at this stage. As in two-cell embryos, the percentage of biallelically expressing nuclei in both genotypes was again greater than that of monoallelically expressing nuclei for most genes (Fig. 1d and Supplementary Fig. 2). These data reiterate that imprinted XCI is not a consequence of MSCI-mediated silencing of Xp-linked genes.

We then surveyed expression of the X-linked genes in WT eight- and 16-cell morula-stage embryos by RNA FISH. In comparison with two- and four-cell embryos, a significantly higher percentage of WT eight-cell nuclei expressed the X-linked genes from only one allele (Fig. 1d and Supplementary Fig. 3). An examination of Xp–*Xist*[1lox] eight-cell embryos unexpectedly also showed the presence of a significant percentage of monoallelically expressing nuclei. If *Xist* RNA were required for the initiation of imprinted XCI, the prediction would be that in its absence Xp-linked gene silencing should not occur; this should result in a higher percentage of biallelism and a lower percentage of monoallelism in Xp–*Xist*[1lox] embryos relative to WT embryos. However, nine of the 11 genes assayed (all except *Rnf12* and *Abcb7*) displayed statistically indistinguishable distributions of the three classes of nucleus (biallelism, monoallelism and no signal) in eight-cell WT and Xp–*Xist*[1lox] embryos (Supplementary Fig. 3). In 16-cell embryos, seven of the 11 genes (all except *Ube1x*, *Rnf12*,
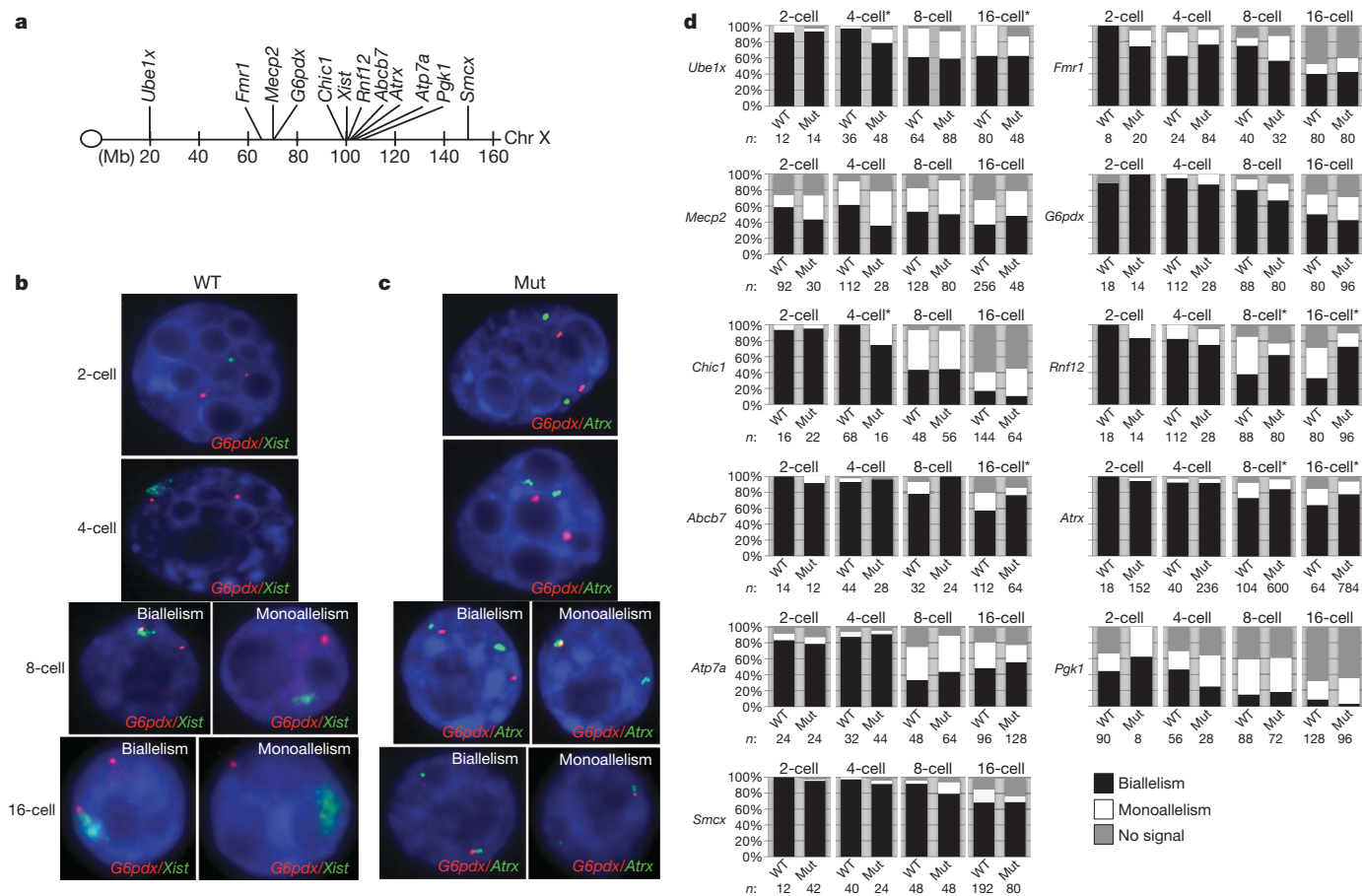
Figure 1 | Dynamics of X-linked gene expression assayed by RNA FISH in two-, four-, eight- and 16-cell WT and Xp–Xist[1lox] (Mut) female mouse embryos. **a**, Physical map of the eleven X-linked genes assayed. Mb, megabase; Chr, chromosome. **b**, **c**, Single representative nuclei from WT and Xp–Xist[1lox] embryos probed for expression of G6pdx (red punctate signal) and Xist (green; in WT embryos) or Atrx (green; in Xp–Xist[1lox] embryos). In WT embryos, Xist is expressed exclusively from, and marks, the Xp[19,30]. In Xp–Xist[1lox] embryos, which lack Xist expression, Atrx expression is used to

mark the two X chromosomes. In nuclei from two- and four-cell embryos, most genes are expressed predominantly biallelically. DNA is stained blue with 4′,6-diamidino-2-phenylindole (DAPI). **d**, Comparison of the distribution of nuclei displaying biallelism, monoallelism and no signal in WT and Xp–Xist[1lox] embryos. n, number of nuclei. Only those genes/stages marked with an asterisk show significant differences in the distribution of the three classes of nuclei (Supplementary Figs 1–4).

Abcb7, and Atrx) were expressed similarly in Xp–Xist[1lox] and WT embryos (Fig. 1d and Supplementary Fig. 4). The parallel decrease in biallelism and increase in monoallelism of most genes in WT and Xp–Xist[1lox] eight- and 16-cell embryos suggests that silencing of Xp-linked genes can occur in morulae in the absence Xp–Xist.

To confirm the silencing of Xp-linked genes in Xp–Xist[1lox] 8–16-cell embryos observed by RNA FISH, we assayed expression of X-linked genes by allele-specific PCR with reverse transcription (RT–PCR) in WT and Xp–Xist[1lox] 8–16-cell stage morulae (Fig. 2a, c and Supplementary Table 1). Silencing of Xp-linked genes during imprinted XCI contributes to the higher expression of maternal alleles of most X-linked genes relative to their paternal counter-parts[4,10,20]. If Xist RNA were required for the initiation of silencing of Xp alleles, the expression of both the Xm and the Xp alleles would be expected to be equal in Xp–Xist[1lox] morulae. We therefore assayed and quantified the relative expression of Xm and Xp alleles of seven X-linked genes distributed along the X chromosome—Ddx3x, Ube1x, Zfx, Rnf12, Atrx, Pdha1 and Utx (also known as Kdm6a)—in multiple individual WT and Xp–Xist[1lox] F[1]-hybrid 8–16-cell female morulae.

Although the relative levels of the two alleles varied between embryos, on average Ddx3x, Ube1x, Zfx, Rnf12, Atrx and Pdha1 were all expressed preferentially from the Xm allele in WT female morulae. Utx appeared not to undergo silencing at this stage, as indicated by a lack of difference in the expression of the two alleles. Notably, in

Xp–Xist[1lox] female morulae the expression of Ddx3x, Ube1x, Zfx and Pdha1 also showed a maternal-allele bias, at levels similar to that in WT embryos; Rnf12 was also predominantly expressed from the Xm, but the difference in expression was not as great as in WT morulae. Atrx appeared to require Xist for silencing of the Xp allele; in contrast to WT embryos, Atrx expression, on average, did not favour the maternal allele in Xp–Xist[1lox] females—although there clearly are morulae that do express the Xm Atrx allele more strongly than they do the Xp (Fig. 2a). Both Rnf12 and Atrx also showed significant differences between WT and Xp–Xist[1lox] 8–16-cell embryos as observed by RNA FISH (Fig. 1d and Supplementary Figs 3 and 4). As in WT embryos, Utx expression did not show an Xm-allele bias in Xp–Xist[1lox] morulae.

We excluded the possibility of strain-specific, as opposed to parent-of-origin-specific, transcriptional inhibition of the paternal Mus domesticus alleles by conducting allele-specific expression analysis of the seven X-linked genes in WT F[1]-hybrid morulae obtained from the reciprocal cross (Supplementary Fig. 5a, c and Supplementary Table 3). In total, of the six genes whose Xp alleles are subject to transcriptional silencing in WT morulae, five also undergo silencing in the absence of Xp–Xist.

We also examined expression of the X-linked genes in extra-embryonic tissues of WT and Xp–Xist[1lox] embryos at embryonic day 6.5 (E6.5), to determine the profile of allele-specific expression of the seven genes during post-implantation development (Fig. 2b, d,
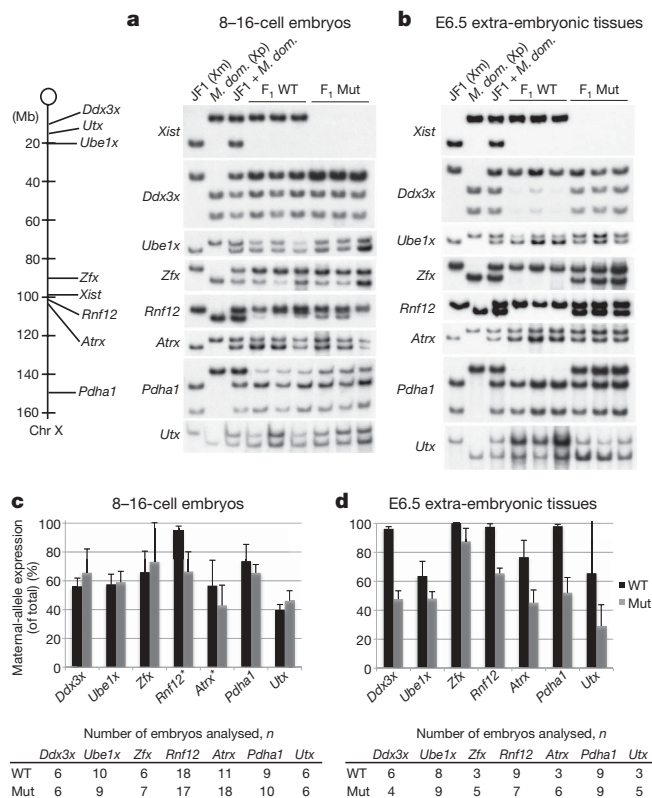
648

Figure 2 | Dynamics of X-linked gene expression assayed by allele-specific RT–PCR in WT and Xp–Xist[1lox] (Mut) 8–16-cell mouse embryos and embryonic-day-6.5 extra-embryonic tissues. Allele-specific RT–PCR expression analysis of the X-linked genes Xist, Ddx3x, Ube1x, Zfx, Rnf12, Atrx, Pdha1 and Utx in individual F_1-hybrid WT and Xp–Xist[1lox] 8–16-cell embryos (morulae) (**a, c**) and embryonic-day-6.5 (E6.5) extra-embryonic tissues (**b, d**). F_1-hybrid embryos were generated by a cross between the *Mus molossinus* strain-JF1 females and largely *Mus domesticus* (*M. dom.*)-derived males from WT and Xp–Xist[1lox] laboratory mice strains. Expression of the two alleles (maternal, Xm; paternal, Xp) was distinguished by single nucleotide polymorphisms between the strains resulting in a strain-specific pattern of fragments after restriction-enzyme digestion (Supplementary Methods and Supplementary Table 6). In **a** and **b**: lane 1, JF1 allele (Xm); lane 2, *M. dom.* allele (Xp); lane 3, equal amounts of JF1 and *M. dom*; lanes 4–6, representative F_1-hybrid WT samples; lanes 7–9, representative F_1-hybrid Xp–Xist[1lox] samples. Xist is expressed exclusively from the Xp in WT morulae and in WT E6.5 extra-embryonic tissues and is absent in Xp–Xist[1lox] morulae and E6.5 extra-embryonic tissues. Only those genes marked with an asterisk (**c**) show differences in the degree of silencing of the paternal alleles in Xp–Xist[1lox] morulae relative to WT embryos (Supplementary Table 1); all genes show differences in E6.5 extra-embryonic tissues (Supplementary Table 2). Error bars, s.d.

and Supplementary Table 2). Extra-embryonic tissues of post-implantation embryos maintain imprinted XCI that initiates during pre-implantation development[12,13]. In agreement, in WT E6.5 extra-embryonic tissues the Xm alleles of all seven genes were expressed at higher levels than were the Xp alleles, some almost exclusively. However, of the five genes that preferentially expressed the Xm allele in Xp–Xist[1lox] morulae (Ddx3x, Ube1x, Zfx, Rnf12 and Pdha1), Ddx3x, Ube1x and Pdha1 lacked biased expression of the Xm allele in Xp–Xist[1lox] E6.5 extra-embryonic tissues. Zfx and Rnf12 preferentially expressed the Xm allele in Xp–Xist[1lox] E6.5 extra-embryonic tissues, at levels similar to or greater than in Xp–Xist[1lox] morulae but not as great as in WT E6.5 extra-embryonic samples. We again ruled out strain-dependent silencing of the Xp alleles in E6.5 extra-embryonic samples through a reciprocal cross (Supplementary Fig. 5b, d and Supplementary Table 4). Thus, in the absence of Xp–Xist the silencing of paternal alleles of some, but not all, X-linked genes observed in morulae is not maintained in E6.5

extra-embryonic tissues. The silencing of Zfx and Rnf12 paternal alleles in Xp–Xist[1lox] E6.5 extra-embryonic samples, however, suggests that even in the absence of Xp–Xist long-term silencing of Xp-alleles of some genes can be maintained.

We also tracked the expression of an Xp-linked green fluorescent protein transgene (Xp–GFP) to visualize the dynamics of X-linked gene expression in pre-, peri- and post-implantation stage WT and Xp–Xist[1lox] embryos (Fig. 3). The Xp–GFP transgene is normally subject to imprinted XCI and is silenced in extra-embryonic tissues of post-implantation embryos but is expressed in the embryonic lineage owing to random XCI[16,22,23]. If Xist RNA is required for the initiation of imprinted XCI, then Xp–Xist[1lox] extra-embryonic cells should uniformly express the Xp–GFP transgene.

In pre-implantation-stage E3.5 embryos, the Xp–GFP transgene is expressed in both the embryonic and the extra-embryonic lineages (the inner cell mass and the trophectoderm, respectively) in both WT and Xp–Xist[1lox] embryos (Fig. 3a, b). At peri-implantation-stage E4.5 embryos, the transgene appears silenced in a majority of the cells of the extra-embryonic trophectoderm lineage in both genotypes (Fig. 3c, d). In post-implantation-stage E5.75, E6.25 and E7.5 WT embryos, green fluorescence due to transgene expression is undetectable in the extra-embryonic compartment (Fig. 3e, g, i). In corresponding
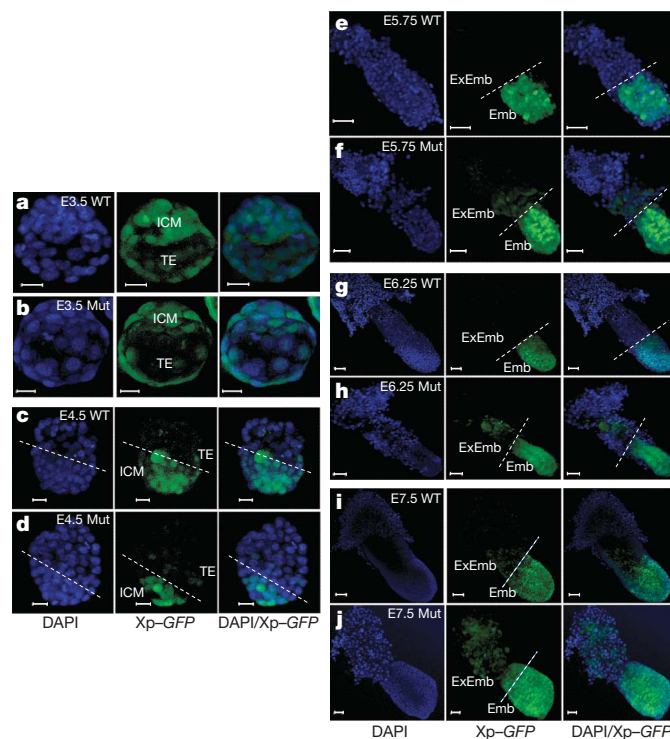

Figure 3 | Expression of paternal X-linked green fluorescent protein (Xp–GFP) transgene in WT and Xp–Xist[1lox] (Mut) pre- and post-implantation-stage female mouse embryos. a, b, WT and Xp–Xist[1lox] blastocyst (E3.5) female embryos expressing Xp–GFP in both the trophectoderm (TE) and inner cell mass (ICM) lineages. **c, d,** WT and Xp–Xist[1lox] E4.5 female embryos silencing Xp–GFP in the TE but showing continued expression of Xp–GFP in the ICM-derived cells. The dashed line marks the boundary between the TE and the ICM derivatives. **e, g, i,** WT E5.75, E6.25 and E7.5 female embryos devoid of Xp–GFP expression in the extra-embryonic ectoderm and its derivatives (ExEmb; derived from the TE) owing to imprinted XCI. The embryonic ectoderm (Emb), descended from the ICM, fluoresces green because of Xp–GFP expression due to random XCI that results in Xp activity in approximately half of the cells. The dashed line marks the boundary between the ExEmb and the Emb. **f, h, j,** Xp–Xist[1lox] E5.75, E6.25 and E7.5 female embryos with a mosaic extra-embryonic compartment that comprises cells that both silence and express Xp–GFP. DNA is stained blue with DAPI. Scale bars: 20 μm (**a–d**); 50 μm (**e–h, j**); 100 μm (**i**).

649

Xp–*Xist*[1lox] embryos, the transgene appears to be expressed in an increasing number of extra-embryonic cells, correlating with the increasing severity of the mutant phenotype (Fig. 3f, h, j).

To confirm the capability of the transgene to be expressed throughout the extra-embryonic compartment, we generated embryos that inherited the transgene on the maternal X chromosome (Xm–*GFP*); the Xm–*GFP* transgene in fact appears to be expressed in most, if not all, of the extra-embryonic cells of pre- and post-implantation embryos (Supplementary Fig. 6). Thus, the gradual increase in the number of extra-embryonic cells expressing the Xp–*GFP* transgene during post-implantation development reiterates a requirement of Xp–*Xist* in maintaining silencing of Xp genes during imprinted XCI. Nevertheless, the presence of extra-embryonic cells that do not express Xp–*GFP* in peri- and post-implantation Xp–*Xist*[1lox] embryos indicates that silencing of Xp-linked genes can initiate and persist well into development in the absence of *Xist* RNA.

Our data are consistent with the hypothesis that X-linked genes have evolved to become dependent on *Xist* RNA for their silencing; however, silencing can initially occur independently of *Xist* RNA and is a consequence of differentiation of the mammalian sex chromosomes (see below). Of the X-linked genes whose expression we have analysed, three (*Rnf12*, *Abcb7* and *Atrx*) appear to undergo silencing to a lesser degree in Xp–*Xist* mutants than in WT morulae. These genes map to the region of the X chromosome that is thought to have undergone dosage compensation first during the evolution of the mammalian sex chromosomes[24]. Thus, the degree to which a gene requires *Xist* RNA during initiation of silencing may depend on the length of time that the gene has been subject to dosage compensation: the longer a gene has been dosage compensated, the more it may be dependent on *Xist* RNA for its silencing.

Our results show that although silencing of Xp-linked genes can initiate in the absence of *Xist*, Xp–*Xist* is clearly required for imprinted XCI; in its absence, after initial silencing Xp-linked genes become expressed in progressively higher numbers of cells and/or at increased levels. Thus, *Xist* RNA appears to be required to stabilize imprinted XCI. *Xist* RNA preferentially decorates the fraction of the X chromosome that is rich in repetitive sequences[14,25]. The silencing of these regions may serve to form a heterochromatin scaffold that ensures long-term silencing of the X chromosome as a whole[25]. By participating in the stable silencing of repetitive sequences, *Xist* RNA may thus bring about enduring silencing of X-linked genes.

XCI is part of a dosage-compensation mechanism that is thought to be a by-product of sexual differentiation of a pair of homologous chromosomes into an X chromosome and a Y chromosome[26]. The gradual loss of genes from the proto-Y chromosome during X–Y differentiation is believed to have driven silencing of homologous X-linked genes in a piecemeal fashion[24,26]. Furthermore, evidence suggests that *Xist* has arisen after the advent of X-linked gene silencing[24]. Metatherian mammals (marsupials), which diverged from eutherians (placental mammals) about 150 million years ago[27], undergo imprinted XCI in both embryonic and extra-embryonic tissues without the aid of *Xist* RNA, although the silencing appears not to be stringent[28,29]. Combined with the results of this study, the evolution of XCI to encompass individual segments and the relatively recent appearance of *Xist* RNA argue that the initial silencing of X-linked genes may be a local or regional, rather than a chromosome-wide, event; therefore, multiple mechanisms may mediate the initiation of imprinted XCI.

METHODS SUMMARY

RNA FISH. We permeabilized embryos by means of sequential transfers into ice-cold cytoskeletal extraction buffer (CSK; 100 mM NaCl, 300 mM sucrose, 3 mM MgCl$_2$, and 10 mM PIPES buffer, pH 6.8) for 30 s, ice-cold CSK containing 0.4% Triton X-100 buffer for 1.5–5 min, followed twice with ice-cold CSK for 30 s each time. All CSK buffers were supplemented with RNase-inhibitor RNasin (Promega) at a concentration of 400 units per millilitre. Embryos were then mounted on a glass coverslip coated with 1× Denhardt's solution in a small drop of ice-cold solution containing 1% paraformaldehyde and 20% CSK in 1×

PBS buffer. Excess solution was aspirated off and the embryos air dried for 15 min and fixed in cold (4 °C) 3% paraformaldehyde for 10 min in a humid chamber. After fixation, the coverslips were submerged in 70% ethanol. After three changes of 70% ethanol, the cells or embryos were either stored in 70% ethanol at −20 °C or processed for RNA FISH (Supplementary Methods).
RT–PCR. Messenger RNA was prepared using the Dynabeads mRNA DIRECT Micro Kit (Invitrogen, catalogue no. 610.21). Individual morula-stage embryos and extra-embryonic tissues from E6.5 embryos were lysed in 100 μl lysis/binding buffer. SuperScript III One-Step RT–PCR Platinum *Taq* enzyme mixture (Invitrogen, catalogue no. 12574-035) was used to prepare and amplify the complementary DNA. A portion of the final RT–PCR product was used to perform one round of amplification in the presence of a trace amount of ^{32}P-dCTP to minimize heteroduplexes and to radioactively label the RT–PCR product. The final PCR product was digested with the appropriate restriction enzyme (Supplementary Table 6) and run on a 7.5% acrylamide gel, vacuum dried and exposed to film. Band intensities were quantified using NIH IMAGEJ software. In case of multiple maternal- or paternal-specific bands, the quantified band intensities were added to yield a single value for each allele.

1. Heard, E., Chaumeil, J., Masui, O. & Okamoto, I. Mammalian X-chromosome inactivation: an epigenetics paradigm. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 89–102 (2004).
2. Payer, B. & Lee, J. T. X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
3. Brown, C. J. et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
4. Huynh, K. D. & Lee, J. T. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. *Nature* **426**, 857–862 (2003).
5. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644–649 (2004).
6. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
7. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
8. Nesterova, T. B., Barton, S. C., Surani, M. A. & Brockdorff, N. Loss of Xist imprinting in diploid parthenogenetic preimplantation embryos. *Dev. Biol.* **235**, 343–350 (2001).
9. Sheardown, S. A. et al. Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell* **91**, 99–107 (1997).
10. Latham, K. E. & Rambhatla, L. Expression of X-linked genes in androgenetic, gynogenetic, and normal mouse preimplantation embryos. *Dev. Genet.* **17**, 212–222 (1995).
11. Mak, W. et al. Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666–669 (2004).
12. Takagi, N. & Sasaki, M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**, 640–642 (1975).
13. West, J. D., Frels, W. I., Chapman, V. M. & Papaioannou, V. E. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell* **12**, 873–882 (1977).
14. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–2237 (2006).
15. Mak, W. et al. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive X chromosome in trophoblast stem cells. *Curr. Biol.* **12**, 1016–1020 (2002).
16. Kalantry, S. et al. The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nature Cell Biol.* **8**, 195–202 (2006).
17. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156–166 (1997).
18. Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. & Jaenisch, R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nature Genet.* **22**, 323–324 (1999).
19. Okamoto, I. et al. Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438**, 369–373 (2005).
20. Singer-Sam, J., Chapman, V., LeBon, J. M. & Riggs, A. D. Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **89**, 10469–10473 (1992).
21. Cooper, D. W. Directed genetic change model for X chromosome inactivation in eutherian mammals. *Nature* **230**, 292–294 (1971).

22. Hadjantonakis, A. K., Gertsenstein, M., Ikawa, M., Okabe, M. & Nagy, A. Non-invasive sexing of preimplantation stage mammalian embryos. *Nature Genet.* **19**, 220–222 (1998).
23. Kalantry, S. & Magnuson, T. The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* **2**, e66 (2006).
24. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
25. Clemson, C. M., Hall, L. L., Byron, M., McNeil, J. & Lawrence, J. B. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc. Natl Acad. Sci. USA* **103**, 7688–7693 (2006).
26. Jegalian, K. & Page, D. C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776–780 (1998).
27. Wallis, M. C., Waters, P. D. & Graves, J. A. Sex determination in mammals–before and after the evolution of SRY. *Cell. Mol. Life Sci.* **65**, 3182–3195 (2008).
28. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653–1655 (2006).
29. Samollow, P. B., Ford, A. L. & VandeBerg, J. L. X-linked gene expression in the Virginia opossum: differences between the paternally derived Gpd and Pgk-A loci. *Genetics* **115**, 185–195 (1987).
30. Kay, G. F. *et al.* Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* **72**, 171–182 (1993).

CORRIGENDUM

Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia

Chen Zhao, Alan Chen, Catriona H. Jamieson, Mark Fereshteh, Annelie Abrahamsson, Jordan Blum, Hyog Young Kwon, Jynho Kim, John P. Chute, David Rizzieri, Michael Munchhof, Todd VanArsdale, Philip A. Beachy & Tannishtha Reya

Nature **458**, 776–779 (2009)

During resubmission of this work, another paper linking CML and Hedgehog signalling was published[1]. The studies use distinct approaches but come to similar conclusions.

1. Dierks, C. *et al.* Expansion of Bcr-Abl-positive leukemic stem cells is dependent on Hedgehog pathway activation. *Cancer Cell* **14**, 238–249 (2008).

PROSPECTS

Balancing belief and bioscience

Can religious belief really be reconciled with a life in science? Gene Russo contemplates the contradictions.

The recent nomination of Francis Collins to direct the US National Institutes of Health does more than raise the question of the agency's future direction. It poses another question. Can a scientific career go hand in hand with religious belief? Put another way, can a great scientist be deeply religious?

Collins is well known as one of the architects of the Human Genome Project. He has also achieved notoriety as a highly respected scientist with deep-seated religious beliefs. An evangelical Christian since the age of 27, Collins detailed his views on reconciling religion and science in his 2006 book *The Language of God: A Scientist Presents Evidence for Belief*.

The obvious reply to the suggestion of conflict is that religious belief can coexist perfectly well with scientific laws and methods — God simply exists outside nature. In Collins's book, for example, he calls himself a believer in "theistic evolution" — which espouses, in part, that "once evolution got under way no special supernatural intervention was required".

Yet theistic evolution, writes Collins, also implies that humans are unique in ways

that "defy evolutionary explanation and point to our spiritual nature". And in a 2006 *Time* magazine debate with noted atheist and evolutionist Richard Dawkins[1], Collins suggested that God could on rare occasions choose to "invade the natural world in a way that appears miraculous" — a peculiar outlook for a biologist.

How rare is a 'dual practitioner' such as Collins? In 1997, University of Georgia history professor Edward Larson and journalist Larry Witham reported survey results suggesting that 39% of scientists believe in God[2]. A 9 July 2009 survey, conducted by the American Association for the Advancement of Science and the US non-profit Pew Research

Center, found that figure to be 33%.

But in a second survey[3], Larson and Witham asked a slightly different question. Do leading scientists have religious faith of equal measure to scientists in general? They used members of the US National Academy of Sciences for their so-called "greater scientist" subgroup. Just 7% of respondents professed a belief in God.

University of Oxford chemist Peter Atkins assessed the Larson and Witham survey results in this way: "You clearly can be a scientist and have religious beliefs. But I don't think you can be a real scientist in the deepest sense of the word because they are such alien categories of knowledge." Eminent Harvard University psychologist Steven Pinker, in criticizing the Collins nomination, said much the same thing[4]. They may be right. Or perhaps it just takes a rare person to advance a scientific career while balancing belief and bioscience — without corrupting either. ■

Gene Russo is editor of *Naturejobs*.

1. www.time.com/time/magazine/article/ 0,9171,1555132,00.html.
2. Larson, E. J. & Witham, L. *Nature* **386,** 435–436 (1997).
3. Larson, E. J. & Witham, L. *Nature* **394,** 313 (1998).
4. *Nature* **460,** 310–311 (2009).

California scrambles to find money for pay

With draconian budget cuts already driving out top researchers at the University of California (UC), senior scientists are scrambling to find new revenue sources in the face of a university plan to impose unpaid leave and cut pay.

On 16 July the UC board of regents voted to empower university president Mark Yudof to force employees to take unpaid leave. The furloughs would range from 11 to 26 days and amount to salary cuts of 4–10%, intended to accommodate cuts in state funding. The university expects these cuts to total US$813 million, thanks to the state's $26-billion budget deficit. At press time, several faculty members were known to be leaving as a result (see *Nature* **460,** 441; 2009). UC has 108,000 full-time employees.

Solutions, however, are hard to come by. According to Sandra Faber, chair of the astronomy department at the University of California, Santa Cruz, the university could

RECESSION WATCH

accept more out-of-state students. "We have to cut loose from the state of California and come up with a new revenue model," she says, pointing to the University of Michigan as a successful example of an institution that does not rely overwhelmingly on in-state students for tuition revenue: about a third of that university's undergraduates came from outside Michigan. Conversely, in the 2007–08 school year, 91% of UC students were from California, according to university website data.

George Blumenthal, chancellor of the University of California, Santa Cruz, suggests raising in-state student fees as a means of boosting revenue. "To the extent we are supported by the state, we have an obligation to provide access [to in-state students]," he says. "To the extent that support is drying up, we have to [change that model]."

Continuing with the current funding strategy for more than a year will imperil

the reputation and calibre of UC, according to Faber. "The faculty will evaporate and the university will collapse," she says. "The university has one year's grace period. After that, these [early- and mid-career] people will leave, and replacing them with people of the same calibre is hopeless. We will have lost our brand name."

Despite the departures of faculty members, some say things could be worse. "Recruitment and retention would be more of a concern if the rest of the world weren't also in a terrible economic recession," says Kevin Woolfork, budget policy coordinator at the California Postsecondary Education Commission, the planning and coordinating body for higher education in the state.

UC receives about $3 billion annually from the state and earns much of the rest of its revenue from its research grants and hospitals. Its total annual budget is about $19 billion, according to the university website. ■

Karen Kaplan

Q&A

Lucila Ohno-Machado is the first chief of the new Division of Biomedical Informatics at the University of California, San Diego.

As a student, you were interested in both medicine and mathematics, how did you pursue the two?

At school, mathematics was my stronger subject and the one I enjoyed the most. I wanted to find a way to use mathematics in health care, initially with the idea of pursuing bioengineering. But growing up in Brazil I had to pursue a medical or an engineering degree and then find ways to combine my interests. I decided it would be easier to start with a medical degree and build on that.

How did you navigate career opportunities in the emerging field of biomedical informatics?

You must be a risk taker. You can't possibly be too cautious when pursuing a career that is not yet seen as mainstream. My self-confidence helped me overcome the barriers created by people's inability to understand the potential of an emerging field. Twenty years ago, the United States had the right combination of programmes available and

was open to interdisciplinary careers, so I came to Stanford University to get a PhD in medical information science.

Will the new graduate programme at the Division of Biomedical Informatics differ from other similar programmes?

Yes. Several programmes funded by the US National Institutes of Health do bioinformatics or medical informatics. But only a few of those do both, and even fewer provide trainees with access to clinical information systems. The San Diego programme will tightly combine bioinformatics and medical informatics while making use of data from the medical-centre hospitals, which will be available for our students and trainees as their laboratory.

What drives your interest in finding patterns in clinical data?

We are past the age when the experience of one clinician can lead to significant medical advances. We need to find practical ways to

access all the data from a larger collection of patients being treated. If we can extract patterns that tell us what we are doing right and what we can do better in the clinic, it will help patients and health-care providers. For example, data registries exist for certain procedures, such as interventional cardiology, and these may be helpful to the clinician assessing the complications associated with a new device.

What is the biggest challenge in biomedical informatics?

Achieving the same status as an established science. That status will attract more young scientists who will shape the future of biomedical informatics. Well-established programmes, developed since the 1980s, have paved the way for the dissemination of biomedical informatics around the world. It will not take long for every major academic institution to develop its own department in this field. ∎

Interview by Virginia Gewin

POSTDOC JOURNAL

A lack of funding

Times are tough for many principal investigators in Canada. Our lab is one of several to be turned down for a major operating grant recently.

According to our rejection letter, our grant application was among the 83% that were turned down for funding. Of course, many research projects go unfunded — in Canada and elsewhere — not because they lack excellence and merit, but because resources are scarce or the government's commitment is insufficient, especially in this dire economic climate. Nonetheless, I fear that, as a

young investigator starting my own research programme, I am not a top priority for reviewers at the Canadian Institutes of Health Research.

Moreover, restructuring of priorities at the Alberta Heritage Foundation for Medical Research — the province's main supporter of health-science research — has left many labs at my institute without operating funds for the next year or more.

For the first time, I hear hushed talk among senior staff that well-established principal investigators may have to let lab members go, or even shut

their doors, in the coming months. In our lab, some of us may have to move on if we can't soon acquire fellowships and other research grants to cover our salaries.

I worry that Canada's provincial and federal governments are sending the wrong message to young researchers.

And that means a more tenuous future for young health scientists in Canada — myself included. ∎

Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.

IN BRIEF

Germany seeks applicants

Germany's main awarder of science grants, the DFG, is calling for applications for the third phase of its Excellence Initiative, a programme aimed at promoting top research and raising the calibre of German universities and research institutions.

This phase will award some €2.7 billion (US$3.8 billion) in research funding to German universities and institutions; the funding is part of an overall €18-billion investment in science and education approved in June, the largest such single amount in Germany's post-war history (see http://tinyurl.com/l768c6). Awards will start in 2012 and will run for five years.

UK bioscience boost

Sixteen new bioscience fellowships worth up to £1.7 million (US$2.8 million) each have been created by the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC).

Eight of the fellowships have been awarded to early-career researchers, and five Research Development Fellowships and a Professorial Fellowship will support established scientists. The awards also include the BBSRC's first Industrial Impact Fellowship, under which a pharmaceutical researcher will work in an academic research lab, and a new Diamond Fellowship, to be based near the national synchrotron facility near Harwell.

The fellowships will run for between three and five years.

Stimulus for US science

In the United States, up to $85 million in federal stimulus funds is up for grabs by early-career researchers at academic institutions and Department of Energy national labs. US energy secretary Steven Chu announced the awards, to be made in early 2010, in mid-July. They will support at least 50 researchers for five years apiece.

Eligible researchers must be tenure-track assistant professors who have earned a PhD within the past 10 years. Each university award will be for at least $150,000 per year, and the national lab awards will be for at least $500,000 per year. Research topics are limited to scientific computing research; biological and environmental research; basic-energy sciences; fusion-energy sciences; high-energy physics; and nuclear physics.

Hard man to surprise

A birthday treat.

David Marusek

On Wednesdays after work, Adam meets with Vera and Pete at a bar across the street from his office.

"My man," he says, sliding into the booth. "Where's Vera?"

"Running late," Pete says. "No, wait, here she is."

Vera exacts a kiss from each of them and squeezes in next to Pete. "What a day!" she exclaims. "First I'm late for a meeting, then I lock myself in the stairwell. I had to climb down *20 floors* to find an exit!"

The men guffaw, and Adam says: "Kinda like that weekend we spent trapped on the Prudential roof, right?"

"Yes!" Vera shrieks. "Like that, only much shorter." She waves her hand to summon the waitress. "You know, I haven't thought of that in years."

Later, as they part company, Pete asks Adam about plans for his birthday, and Adam says he enjoyed the so-called surprise dinner they threw for him at Chili's last year.

Pete says, "Sounds like a plan. I'll make secret arrangements for three tables."

Vera winks. "I'll quietly handle the guest list."

That weekend Adam runs into Hector and Sylvester at Starbucks. He asks them if they're coming to Chili's on Friday.

"No doubt," Sylvester says, but Hector gives him a blank look.

"My birthday?" Adam prompts him. "My 'surprise' dinner?"

Hector pats his jacket pockets and hands Adam a blue card on which is printed:

Hello. It may seem strange that I don't recognize you, but I have recently undergone a memory extinction treatment to selectively erase a traumatic event from my mind. Quite possibly, you were also involved in that event, and the procedure has inadvertently wiped you as well. If this is the case, I apologize and wish you well.

Hector waits for Adam to finish reading. "All right then," he says and leaves the coffee shop.

Adam is floored. "What the hell just happened?"

"I have no idea," Sylvester says. "I'll go find out."

Adam is left holding the blue card. He flips it over and finds the logo of Clean Slate Salons.

The thing is, a few years back Adam handed out one of these cards himself. One night, on what must have been the world's worst first date, Adam and his date were mugged on their way back to his car after a show in the city. They were not physically hurt, but for weeks afterwards Adam's bowels would loosen each time a gun was shoved in his face, which was every time he closed his eyes, and all through the night.
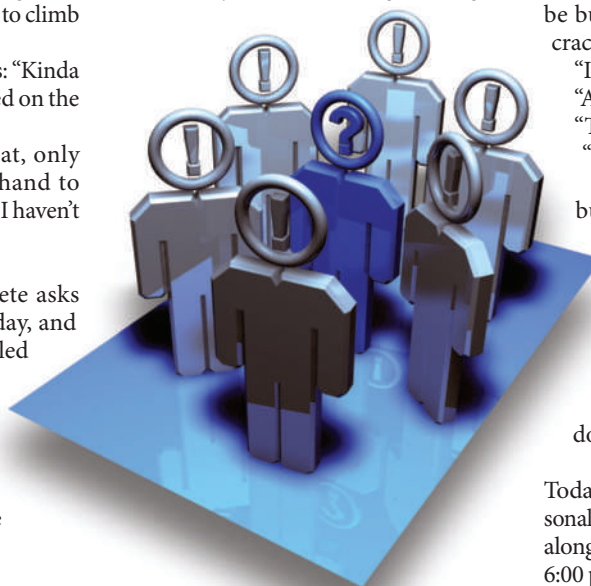
So Adam went to Clean Slate where he drank a carton of Protatter and lay on a couch in a booth with his head resting on a microtrode-encrusted pillow. A certified facilitator in Mumbai talked him through a complete 'narrative' of his 'incident', prompting him for every traumatic detail. Adam scrubbed the entire humiliating evening from his memory and tossed his date out with the bathwater. Hence the eventual blue card.

On Wednesday, Adam is waiting in the usual booth, but Pete and Vera are no-shows, and their icons fail to pop up on any of his maps. When he calls them he gets voicemail.

While he waits for them to call him back, he calls Sylvester to ask about Hector. It has occurred to him that if something bad happened to Hector, why hasn't he heard about it? He gets Sylvester's voicemail. He calls Rosemary to see what's up with Sylvester and gets voicemail. He tries Frank, Claudia and Conor. Finally, a little ticked off, he twitters: WHERE THE BLEEP IS

EVERYONE? HELLO? Then he notices that no one is following him anymore. Followers: 0. He stares at the hollow digit in astonishment. How can such a number even be possible? He has a sinking feeling and calls Chili's to confirm his reservation for Friday and learns there is no such reservation. Or, rather, there was one, but it was cancelled.

Adam is standing outside Pete's building. He has pressed the bell and is waiting to be buzzed in, but instead the intercom crackles. "Yes?"

"It's Adam."

"Adam who?"

"That's not funny."

"Just a sec."

Adam waits for the heavy door to buzz, but a moment later Pete peers at him through the glass, opens the door a crack and says, "You Adam?"

"Stop that!"

Pete hands him something and shuts the door. It's a blue card. "What did I do?" Adam shouts at the door. "Just tell me what I did!"

Today is Adam's birthday. He takes a personal day from work, sleeps in, rides his bike along the river, catches a matinee. Around 6:00 p.m. he returns home with a couple of DVDs, a pizza and a six-pack of beer.

No sooner does Adam get through the door than all the lights come on and a crowd of people spring from the furniture shouting "Surprise!"

Vera is there, and Pete, Sylvester, Hector and Rosemary. Frank is there and Claudia, Conor and a dozen more, all in party hats. "Surprise!" they shout. Streamers and balloons deck the living room, and the countertop is stacked with gifts.

"Did we surprise you?" Vera asks. "You look surprised."

"Here, let me help with that," Pete says and takes the beer from Adam. "Sorry about punking you, but you're a hard man to surprise."

"It was a bet, actually," Hector says.

"So are you surprised?" Vera demands.

Adam peers around at the expectant faces. "Oh, I'm surprised all right," he says and reaches for something in his pocket. "I just hope I have enough of these." ■

Science-fiction author David Marusek lives in a constant state of mild surprise in Fairbanks, Alaska. Visit him at www.marusek.com.

Biased reptilian palaeothermometer?

Arising from: J. J. Head et al. Nature 457, 715–717 (2009)

Palaeotemperatures can be estimated from characteristics of fossils if their living relatives represent the full evolutionary potential of the larger taxon to which the fossil belongs. By drawing on observations[1,2] that the body size of poikilotherms decreases globally with ambient temperature, Head et al.[3] used the 13 m length of the newly described fossil boid *Titanoboa cerrejonensis* to estimate that the Palaeocene neotropical mean annual temperature (MAT) was 30–34 °C. I question the validity of this palaeotemperature estimate by using the same data and approach as Head et al.[3] to show that *Varanus* (*Megalania*) *prisca*[4], a large extinct lizard that lived in eastern Australia during the Late Pleistocene, was 3–4 times longer than predicted by the largest lizard species in the tropics today. This suggests that the scarcity of large predatory reptiles today may primarily be a function of competition with mammalian carnivores, rather than a function of modern temperatures.

Makarieva et al.[1,2] established that the largest living species within 24 poikilotherm taxa from tropical (ambient temperature (T_a) = 26 °C), temperate (T_a = 14–18 °C) and polar (T_a = 2 °C) climates decrease in length 2–3 times with every 10 °C decrease in temperature. Using this relationship, Head et al.[3] estimated the Palaeocene neotropical MAT from the ratio of the *Titanoboa* total body length (TBL; 10.6–14.9 m) to that of the longest extant neotropical snake—the 7.3-m-long boine *Eunectes murinus*. I tested this approach by using palaeotemperatures experienced by *Varanus prisca*, and its inferred body length, to calibrate curves that predict T_a from body length (see Methods). I then compared these with a curve derived from the TBL of the world's largest extant lizard, equatorial *Varanus komodoensis*, which the analysis of Makarieva et al.[1,2] assumed is the largest lizard possible at current tropical T_a, to make predictions of plausible varanid body lengths at different temperatures.

Varanus prisca occurred across eastern Australia during the Late Pleistocene, but is most abundantly recorded in the Darling Downs district of southeast Queensland[5,6], where luminescence dating demonstrates it survived until at least 40,000 years ago[7], during the last glacial interval when regional temperatures were, conservatively, at least 2 °C lower, and probably much lower than present[8–10]. Modern representative T_a for the Darling Downs (at Warwick, Queensland, T_a = 21.5 °C) therefore indicates that *V. prisca* tolerated $T_a \leq 19.5$ °C. The TBL of *V. prisca* is uncertain[11,12], but Head et al.[4] accepted a range of 5–7 m for large individuals. Curves derived from these values (Fig. 1, large filled red squares) predict that the modern tropical MAT ($\approx T_a$) of 27 °C is capable of supporting a 10–14-m-long varanid (Fig. 1, shaded area L_1), which is much larger than any known fossil or extant lizard. Conversely, the 3.1-m length[1,2] of *V. komodoensis* at a MAT of 27 °C (Fig. 1, large filled black circle) predicts that eastern Australian glacial $T_a \leq 19.5$ °C should have supported varanids no longer than 1.55 m (Fig. 1, L_2), ~3–4 times shorter than *V. prisca*. Moreover, the *V. komodoensis* curve predicts that 5–7-m *V. prisca* would have required $T_a \geq 32–36$ °C to survive (Fig. 1, shaded T_2), which overestimates glacial temperature by 13–16 °C. Conversely, the *V. prisca* curve predicts that 3.1 m lizards could be expected to occur wherever $T_a \geq 10.7–14.3$ °C (Fig. 1, shaded T_1), as cool as or cooler than the United Kingdom where the largest lizard (46 cm) is less than one-sixth of this length[2].

These comparisons show that the 3.1-m length of *V. komodoensis* is not the maximum possible for a varanid at modern tropical temperatures, so its position within the analysis of Makarieva et al.[1,2] as the longest possible modern lizard probably biases the palaeotemperature estimate of Head et al.[3]. Moreover, the *V. prisca* curves indicate that few, if any living lizards are near maximum body size for their ambient temperature. Given the similar body shape of lizards and
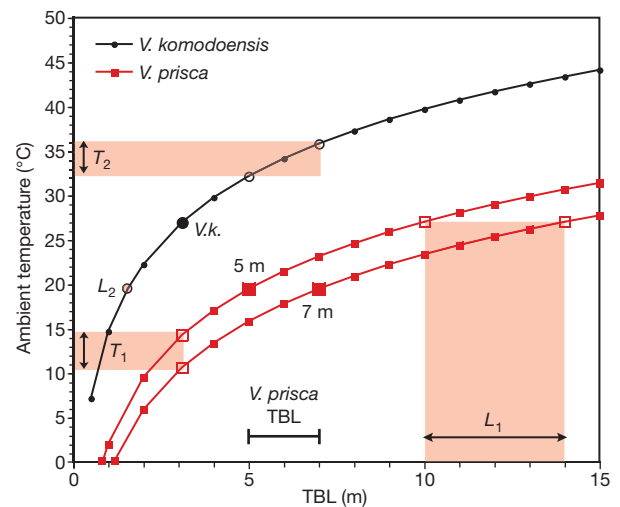

Figure 1 | Ambient temperature and body length relationships predicted from TBL of the komodo dragon *V. komodoensis* (black) and from TBL estimates of *V. prisca* (red). Curves represent model body size increase with temperature in varanids, based alternatively on a maximum TBL for *V. komodoensis* (3.1 m) at a modern tropical MAT of 27 °C (large filled black circle), and on TBL estimates for extinct Pleistocene *V. prisca* (5 and 7 m, representing the range accepted by Head et al.[4]) at glacial ambient temperature of 19.5 °C (large filled red squares). See text for explanation of open circles and squares.

snakes, a snake as large as *Titanoboa* might therefore have experienced tropical temperatures no warmer than today. If so, the lack of such large terrestrial reptiles today, and their scarcity during greenhouse intervals of the Cenozoic, may reflect non-climatic impediments to their evolution. One impediment may be competition with mammals for the large predatory niche, so that reptiles can approach size limits set by metabolic constraints only where this competition is reduced, as it may have been in the Australian Pleistocene[12,13] and more widely in the Palaeocene before the global emergence of large mammalian carnivores[14].

METHODS

Because outside of the tropics poikilotherms may only be active during part of the year, Makarieva et al.[2] compared ambient temperatures (T_a), equivalent to the mean temperature of the warm half-year. I therefore estimated T_a for Warwick, Queensland, as the mean temperature of the six warmest months (Australian Bureau of Meteorology, http://www.bom.gov.au). Curves relating TBL to T_a were calculated from equations (4) and (5) of Head et al.[2], using mean representative values of $Q_{10} = 2.3$, and $\alpha = 0.3$ (ref. 2), based on 3.1-m *V. komodoensis* at T_a = 27 °C; and 5- and 7-m *V. prisca* at T_a = 19.5 °C.

J. M. Kale Sniderman[1]

[1]School of Geography and Environmental Science, Monash University, Victoria 3800, Australia.

e-mail: kale.sniderman@arts.monash.edu.au

1. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Temperature-associated upper limits to body size in terrestrial poikilotherms. *Oikos* 111, 425–436 (2005).
2. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Gigantism, temperature and metabolic rate in terrestrial poikilotherms. *Proc. R. Soc. Lond. B* 272, 2325–2328 (2005).
3. Head, J. J. et al. Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* 457, 715–717 (2009).
4. Head, J. J., Barrett, P. M. & Rayfield, E. J. Neurocranial osteology and systematic relationships of *Varanus* (*Megalania*) *prisca* Owen, 1859 (Squamata: Varanidae). *Zool. J. Linn. Soc.* 155, 445–457 (2009).

5. Hecht, M. K. The morphology and relationships of the largest known terrestrial lizard, *Megalania prisca* Owen, from the Pleistocene of Australia. *Proc. R. Soc. Vic.* **87**, 239–250 (1975).
6. Price, G. J. & Webb, G. E. Late Pleistocene sedimentology, taphonomy and megafauna extinction on the Darling Downs, southeastern Queensland. *Aust. J. Earth Sci.* **53**, 947–970 (2006).
7. Roberts, R. G. *et al.* New ages for the last Australian megafauna: continent-wide extinction about 46,000 years ago. *Science* **292**, 1888–1892 (2001).
8. Miller, G. H., Magee, J. W. & Jull, A. J. T. Low-latitude glacial cooling in the Southern Hemisphere from amino-acid racemization in emu eggshells. *Nature* **385**, 241–244 (1997).
9. Pelejero, C., Calvo, E., Barrows, T. T., Logan, G. A. & De Deckker, P. South Tasman Sea alkenone palaeothermometry over the last four glacial/interglacial cycles. *Mar. Geol.* **230**, 73–86 (2006).
10. Calvo, E., Pelejero, C., De Deckker, P. & Logan, G. A. Antarctic deglacial pattern in a 30 kyr record of sea surface temperature offshore South Australia. *Geophys. Res. Lett.* **34**, doi:10.1029/2007GL029937 (2007).
11. Molnar, R. E. *Dragons in the Dust: the Paleobiology of the Giant Monitor Lizard Megalania* (Indiana Univ. Press, 2004).
12. Wroe, S. A review of terrestrial mammalian and reptilian carnivore ecology in Australian fossil faunas, and factors influencing their diversity: the myth of reptilian domination and its broader ramifications. *Aust. J. Zool.* **50**, 1–24 (2002).
13. Flannery, T. The case of the missing meat eaters. *Nat. Hist.* **102**, 41–45 (1993).
14. Gingerich, P. D. Environment and evolution through the Paleocene-Eocene thermal maximum. *Trends Ecol. Evol.* **21**, 246–253 (2006).

doi:10.1038/nature08222

Re-calibrating the snake palaeothermometer

Arising from: J. J. Head *et al.* Nature 457, 715–717 (2009)

In a recent study[1] a new proxy for palaeoclimate reconstructions was proposed on the basis of a theoretical approach linking the largest body sizes to ambient temperature in extant taxa of air-breathing poikilotherms[2,3]. The value of the largest fossil snake's body length was used to estimate the mean annual temperature (MAT) for the Palaeocene neotropics of $\Delta T = 3.8–7.2\,^\circ$C above the modern value[1]. Here we argue that the reported temperature difference is a twofold overestimate and obtain a corrected estimate of $\Delta T = 1.9–3.7\,^\circ$C using the taxon-specific metabolic scaling exponent $\alpha = 0.17$ for boid snakes. The importance of using relevant taxon-specific information in case of one-taxon-based temperature reconstructions[1] while leaving the theoretically derived generic α values (such as $\alpha = 0.33$ used by Head *et al.*[1]) for broad inter-taxonomic analyses[2,3] is emphasized.

It was proposed and tested against diverse sets of data[2-7] that there exists a minimum level $q = q_{min}$ of mass-specific metabolic rate q (energy spent per unit body mass per unit time), the fall below which is not compatible with successful biological and/or ecological performance of species in a given taxon. In poikilotherms, q declines with increasing body mass M but grows with increasing ambient temperature T, $q(M,T) = q_0(M/M_0)^{-\alpha}Q_{10}^{(T-T_0)/10\,^\circ C}$, in which q_0 is the value at a reference body mass M_0 and temperature T_0, Q_{10} and α (the metabolic scaling exponent) are constants. Species living in warmer climates can afford larger maximum body sizes, thus offsetting the size-related drop of metabolic rate by a higher temperature.

On a log–log scale, metabolic rates (q) of species living at different temperatures plotted against body length ($L \propto M^{1/3}$) form slanting parallel lines bounded by a common horizontal bottom-line $q = q_{min}$, Fig. 1. Metabolic scaling exponent (α) determines the tangent of the angle of slope ϑ, tg $\vartheta = 3\alpha$. The steeper the allometric lines go, the greater temperature difference ($\Delta T > 0$) is needed to extend body length from L_2 (length of smaller taxon) to L_1 (length of larger taxon) at constant $q = q_{min}$, $\Delta T = 3\alpha \times (10\,^\circ C) \times \log_{10}(L_1/L_2)/\log_{10}Q_{10}$. In comparative analysis of a large number of diverse taxa, a representative value of $\alpha = 0.3$ was used[2] reflecting the debate about possible universality of α at 0.33 or 0.25 (for example, ref. 8). Boid snakes, however, fall on the lower end of the observed range of α values, for them $\alpha = 0.17 \pm 0.04$ (± 2 s.e.m.)[9]. Using the estimated maximum body length of *Titanoboa cerrejonensis* $L_1 = 10.4–14.9$ m, maximum body length $L_2 = 7.3$ m for the anaconda *Eunectes murinus*[1] and $Q_{10} = 2.65$ for boid snakes[9], we conclude that the Palaeocene neotropics were $\Delta T = 1.9–3.7\,^\circ$C warmer than at present, Fig. 1. The use of $\alpha = 0.33$ instead of $\alpha = 0.17$ in ref. 1 resulted in a twofold overestimate of ΔT.

Interestingly, comparison of the largest extinct frog *Beelzebufo ampinga* found on Madagascar[10] ($L_1 = 42.5$ cm), and the largest extant frog *Conraua goliath* ($L_2 = 32$ cm) gives a similar estimate for the neotropical MAT of the Late Cretaceous (70.6–65.5 Myr). For anurans $\alpha = 0.20 \pm 0.07$ and $Q_{10} = 2.21$ (ref. 9), which gives $\Delta T = 2.1\,^\circ$C.

Generally, the approach used by Head *et al.*[1] can be considered a significant extension of the nearest living relative method widely used in palaeoclimate studies[11,12]; for example, when the presence of plant fossils with extant freeze-intolerant relatives is interpreted as indicative of mild winters. The proposed animal framework suggests that if the extant species are consistently smaller than their fossil relatives then the fossil species had probably evolved in a warmer-than-present climate. For example, among the few Palaeocene insect fossils[13] some very big lacewings[12] and giant ants[14] were found in temperate latitudes, providing further evidence for a warmer Palaeocene climate. Apparently, joint consideration of several 'palaeothermometers' built for different taxa will enhance the reliability of climate reconstructions, so the possibility of there being a universal scaling exponent α preferable for a broad inter-taxonomic analysis[3,4] needs to be further scrutinized. In the meantime, for temperature reconstructions focusing at one particular taxon as the
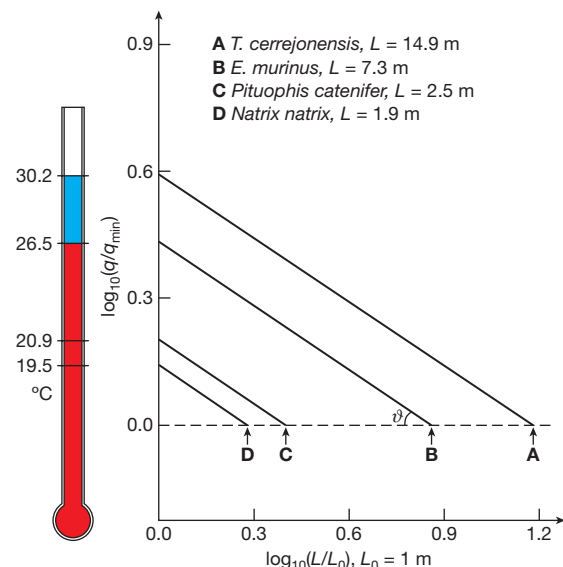

Figure 1 | The snake palaeothermometer. Allometric dependencies of mass-specific metabolic rate q on body length L, $q \propto L^{-3\alpha}$, $\alpha = 0.17$, are shown at four different temperatures. Points at which the allometric lines cross the horizontal dashed-line $q = q_{min}$ correspond to body lengths of the largest snakes[1,2] in Palaeocene neotropics (A), South America (B), Colorado (C) and the United Kingdom (D). Temperatures reconstructed from the metabolic allometry (assuming that the anaconda lives at 26.5 °C) are marked on the thermometer, with the difference between the modern and Palaeocene neotropics shown in blue. Note that the reconstructed temperature differences pertain to the differences between typical 'lifestyle' temperatures of the considered species that can differ significantly from MAT in seasonal climates.

one performed by Head *et al.*[1] it is plausible to use as accurate taxon-specific information on metabolic allometry as available.

Anastassia M. Makarieva[1,2]**, Victor G. Gorshkov**[1,2] **& Bai-Lian Li**[2]

[1]Theoretical Physics Division, Petersburg Nuclear Physics Institute, 188300, Gatchina, St Petersburg, Russia.
e-mail: elba@peterlink.ru
[2]Ecological Complexity and Modeling Laboratory, Department of Botany and Plant Sciences, University of California, Riverside, California 92521-0124, USA.

1. Head, J. J. *et al.* Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* **457**, 715–717 (2009).
2. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Gigantism, temperature and metabolic rate in terrestrial poikilotherms. *Proc. R. Soc. Lond. B* **272**, 2325–2328 (2005).
3. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Temperature-associated upper limits to body size in terrestrial poikilotherms. *Oikos* **111**, 425–436 (2005).
4. Gorshkov, V. G. The distribution of energy flow among the organisms of different dimensions. *Zh. Obshch. Biol.* **42**, 417–429 (1981).
5. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. A note on metabolic rate dependence on body size in plants and animals. *J. Theor. Biol.* **221**, 301–307 (2003).
6. Makarieva, A. M. *et al.* Size- and temperature-independence of minimum life-supporting metabolic rates. *Funct. Ecol.* **20**, 83–96 (2006).
7. Makarieva, A. M. *et al.* Mean mass-specific metabolic rates are strikingly similar across life's major domains: Evidence for life's metabolic optimum. *Proc. Natl Acad. Sci. USA* **105**, 16994–16999 (2008).
8. White, C. R., Phillips, N. F. & Seymour, R. S. The scaling and temperature dependence of vertebrate metabolism. *Biol. Lett.* **2**, 125–127 (2006).
9. Chappell, M. A. & Ellis, T. M. Resting metabolic rates in boid snakes: allometric relationships and temperature effects. *J. Comp. Physiol. [B]* **157**, 227–235 (1987).
10. Evans, S. E., Jones, M. E. H. & Krause, D. W. A giant frog with South American affinities from the Late Cretaceous of Madagascar. *Proc. Natl Acad. Sci. USA* **105**, 2951–2956 (2008).
11. Greenwood, D. R. *et al.* Fossil biotas from the Okanagan Highlands, southern British Columbia and northeastern Washington State: climates and ecosystems across an Eocene landscape. *Can. J. Earth Sci.* **42**, 167–185 (2005).
12. Archibald, S. B. & Makarkin, V. N. Tertiary Giant Lacewings (Neuroptera: Polystoechotidae): revision and description of new taxa from western North America and Denmark. *J. Syst. Paleontol* **4**, 119–155 (2006).
13. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* Ch. 2 (Cambridge Univ. Press, 2005).
14. Rust, J. & Andersen, N. M. Giant ants from the Paleogene of Denmark with a discussion of the fossil history and early evolution of ants (Hymenoptera: Formicidae). *Zool. J. Linn. Soc.* **125**, 331–348 (1999).

Can the giant snake predict palaeoclimate?

Arising from: J. J. Head *et al. Nature* 457, 715–717 (2009)

In their report on *Titanoboa cerrejonensis*, Head *et al.*[1] propose that the great size of this 58 to 60 million-year-old snake (estimated length = 13 m, mass = 1,135 kg) indicates a mean annual neotropical temperature (MAT) of 30–34 °C, substantially higher than previous estimates for that period. They argue that the high MAT was necessary to compensate for the decreased mass-specific metabolic rate intrinsic to a snake of this size. However, the relationship on which Head *et al.*[1] base their conclusion does not account for the scope of behavioural control over body temperature available to *Titanoboa* due to its huge mass. Our calculations suggest that because of its ability to behaviourally control its body temperature, *Titanoboa* cannot serve as an accurate palaeothermometer.

The metabolic rate of poikilothermic (ectothermic) animals such as snakes depends on both body temperature (T_b) and body mass (m): at a given temperature, mass-specific metabolic rate (M) decreases with increasing size (approximately as $m^{-0.25}$); for a given mass, M increases with increasing body temperature (governed by metabolic Q_{10}). Thus, if there is a minimal M required for cellular maintenance, there must be an upper limit to body size for a given body temperature, and the ratio of maximum body sizes in similar animals can be related to their temperatures[2]. Using a comparison to the mass and known MAT for the largest extant snake (an anaconda), Head *et al.*[1] suggest that the MAT of *Titanoboa*'s environment must have been 1–8 °C higher than earlier estimates.

However, Head *et al.*[1] implicitly assume that the relationship between body temperature and air temperature is constant across body size. This assumption is reasonable for small animals (such as those dealt with by Makarieva *et al.* in the paper on which Head *et al.* base their calculations[2]), but is questionable for the massive *Titanoboa*. Animals exchange heat with their surroundings across their surface area, whereas metabolic heat is generated by the animal's entire volume[3,4]. The relatively large surface area to volume ratio (SVR) for small organisms allows them to reach thermal equilibrium (metabolic heat produced = heat shed) with a body temperature only marginally higher than air temperature (Fig. 1). In contrast, the relatively low SVR for an animal as large as *Titanoboa* requires T_b to be increased substantially above air temperature to establish thermal equilibrium. For example, let us assume that *Titanoboa* had a

mass-specific metabolic rate of 0.021 W kg^{-1} at 28 °C, and a metabolic Q_{10} of 2.63 (values estimated from extant boid snakes[5]). If the snake were to coil itself into a hemispherical mound as it lay on an insulating substratum in a mild breeze (0.5 ms^{-1}), its equilibrium body temperature would be 4.3 °C above ambient. If the breeze were slower, or the snake more massive (the estimates of Head *et al.*[1] range as high as 1,819 kg), body temperature would be even higher. Indeed, if air temperature was 34 °C (the peak of the range estimated by Head *et al.*[1]), a coiled *Titanoboa* could dangerously overheat. When uncoiled, *Titanoboa*'s body temperature would still be >0.5 °C above ambient. In contrast, even when coiled, temperature increase in an anaconda would be less than half that of *Titanoboa*. In short, *Titanoboa* could have regulated its body temperature by varying its posture to a much greater degree than extant snakes, potentially changing its relationship to ambient air temperature.

The calculations made here for body temperature are themselves uncertain. For example, we have not taken into account any heat input
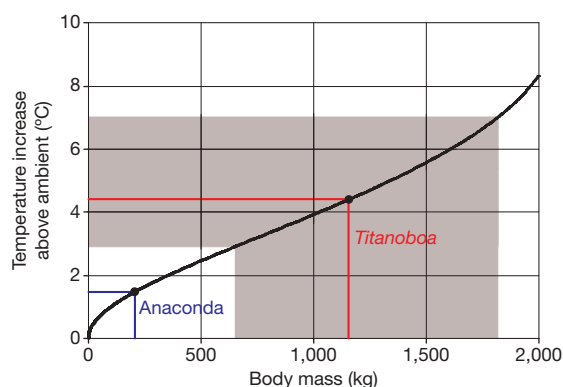

Figure 1 | Body temperature of a coiled snake at thermal equilibrium is increased above ambient air temperature. In this example, wind velocity is 0.5 ms^{-1}. The value for the green anaconda *Eunectes murinus* is calculated for an individual 7.3-m long—the length used by Head *et al.*[1] in their comparisons. The shaded area depicts the range of estimated mass for *Titanoboa* and the associated range of temperature increase.

into the snake from absorbed sunlight, or heat lost by conduction to the substratum. The former depends at least in part on the snake's behaviour, the latter on both the thermal properties of the substratum and the time for which the snake rested in a single spot, both of which would be difficult to quantify accurately for an extinct snake. Furthermore, an aquatic lifestyle would complicate things further, as entry into water would have led to major changes in thermoregulatory relationships. But these uncertainties serve to reinforce our main point: body size alone is an imprecise index on which to base reliable estimates of average habitat temperature. *Titanoboa* is a fascinating discovery, but a palaeothermometer of uncertain accuracy.

METHODS

Heat transfer coefficients (h_c) for a hemisphere and a cylindrical rod (axis parallel to flow, aspect ratio equal to that of *Titanoboa*) were measured in a wind tunnel at varying air speeds (see ref. 6). Objects rested on insulating substrata. Results were expressed as Nusselt/Reynolds number relationships (rod: $Nu = 0.316Re^{0.670}$, hemisphere: $Nu = 0.218Re^{0.587}$) from which h_c of *Titanoboa* and *Eunectes* are

estimated for given wind speeds. Overall heat loss at a given increase in body temperature (ΔT) is then estimated using Newton's law of cooling[3,4]. Metabolic heat production was calculated as $MmQ^{\Delta T/10}$, with $m = 1{,}135$ kg and $Q = 2.63$.

Mark W. Denny[1], Brent L. Lockwood[1] & George N. Somero[1]
[1]Hopkins Marine Station of Stanford University, Ocean View Boulevard, Pacific Grove, California 93950, USA.
e-mail: mwdenny@stanford.edu

1. Head, J. J. *et al.* Giant boid snake from the Paleocene neotropics reveals hotter past equatorial temperatures. *Nature* **457**, 715–717 (2009).
2. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Gigantism, temperature and metabolic rate in terrestrial poikilotherms. *Proc. R. Soc. Lond. B* **272**, 2325–2328 (2005).
3. Denny, M. W. *Air & Water: The Physics of Life's Media* 145–173 (Princeton, 2000).
4. Gates, D. M. *Biophysical Ecology* 12–24; 382–427 (Dover, 1980).
5. Chappell, M. A. & Ellis, T. M. Resting metabolic rates in boid snakes: allometric relationships and temperature effects. *J. Comp. Physiol. [B]* **157**, 227–235 (1987).
6. Denny, M. W. & Harley, C. D. G. Hot limpets: predicting body temperature in a conductance-mediated system. *J. Exp. Biol.* **209**, 2409–2419 (2006).

Head et al. *reply*

Denny *et al.*[1] and Sniderman[2] question our use of body size in *Titanoboa cerrejonensis* as an equatorial temperature proxy during the Palaeocene[3], and Makarieva *et al.*[4] suggest an adjustment to our temperature estimates. Denny *et al.*[1] misinterpret the physiological model of our study[5,6] and the implications of their body temperature (T_b) estimates relative to the thermal ecology of extant snakes. They assert that our model is inappropriate for large-bodied animals because the relationship between T_b and air temperature (mean annual temperature; MAT[3]) is not constant across different body sizes in poikilotherms. In fact, the model does not assume constancy of T_b relative to MAT. Changes in T_b (and thus body size) relative to MAT resulting from increasing thermal inertia with size are inherent in the model, as evidenced by the scaling of body length and MAT (see Fig. 3 in ref. 3). The model is accurate for the longest extant snake, *Python reticulatus*[6], with a maximum body length ~70% that of *Titanoboa*[3], indicating that it is appropriate to use with animals approximating the sizes of the largest known snakes.

Denny *et al.*[1] model the T_b for a coiled *Titanoboa* at 4.3 °C above ambient temperature at 28 °C, concluding that increased MAT was not necessary to maintain metabolic efficiency and high T_b could be behaviourally maintained. Increased T_b is not unique to *Titanoboa*—large extant snakes combine high thermal inertia with behavioural thermoregulation to maintain increased T_b relative to ambient temperature at similar magnitudes[7]. Similarly, the *Titanoboa* T_b values predicted by Denny *et al.*[1] for the MATs we estimate fall within the T_b ranges of extant boids between 30–38 °C[8]. In short, Denny *et al.*[1] provide a first-order approximation of T_b, but their results do not indicate a unique thermal ecology for *Titanoboa*, and do not falsify our palaeoclimatic reconstructions or the use of body size to estimate palaeotemperature.

The analysis of Sniderman[2] illustrates several fundamental assumptions of our model: body sizes in both extant and fossil taxa are maximized relative to MAT for a given habit and environment, and mass-specific metabolic rates are equivalent to estimate palaeotemperature. We assume this is the case for extant *Eunectes murinus* in the absence of any additionally recognized ecological regulators (resource availability, predation pressure) on body sizes, and conclude that it is an accurate comparison for *Titanoboa* given strong aquatic

habitat similarities between the two[3], a close phylogenetic relationship, and a known mass-specific metabolic rate for boid snakes[9]. The same assumptions cannot be made when comparing *Varanus prisca* to extant varanids or other lizards, as predation pressures are known to restrict body size ranges in extant *Varanus*[10] independently of MAT, and the largest living species, *Varanus komodoensis*, is an island endemic with limited habitat availability. Comparison of *V. prisca* to the semi-fossorial, limbless *Anguis fragilis* from the United Kingdom (and subsequent extension of these results to the rest of Squamata), is inappropriate because these species are not closely related, have different body plans, live in different habitats, and *Varanus* has significantly higher mass-specific metabolic rates than other squamates[11].

We also find the mechanics of Sniderman's analysis[2] problematic. He overestimates body length for *V. prisca* by relying on secondary reports of size maxima[12], despite empirical studies indicating smaller body lengths (snout–vent length = 2.19 m; total body length = 3.5–4.7 m)[13,14]. These smaller lengths reduce his reported temperature to body size discrepancies by half. A smaller than predicted body size in *V. komodoensis* cannot bias our analysis, as neither the theoretical model of Makarieva *et al.*[4–6], or our derivation of it to estimate palaeotemperature, is based on that datum. We consider Sniderman's results indicative of a complex relationship between size and environment in extant *Varanus*, and not an indictment of our methods.

Makarieva *et al.*[4] argue for a taxon-specific metabolic scaling exponent in calculating palaeotemperatures based on *Titanoboa*, in contrast to their previous argument that the exponent should be a universal value of 0.33 (ref. 6). If their adjustment is correct, then the estimated MAT for the Palaeocene neotropics would be 28–31 °C—a 2–3 °C decrease from our original analysis. Regardless of which metabolic scaling exponent is most appropriate, we note that the implications for palaeoclimate, related to the hypothesis that equatorial latitudes were much warmer 60-million years ago than today, are the same.

Jason J. Head[1], Jonathan I. Bloch[2], Alexander K. Hastings[2], Jason R. Bourque[2], Edwin A. Cadena[2,3], Fabiany A. Herrera[2,3], P. David Polly[4] & Carlos A. Jaramillo[3]

¹Department of Biology, University of Toronto, Mississauga, Mississauga, Ontario L5L 1C6, Canada.
e-mail: jason.head@utoronto.ca
²Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611-7800, USA.
³Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Ancon Republic of Panama.
⁴Department of Geological Sciences, Indiana University, Bloomington, Indiana 47405-1405, USA.

1. Denny, M. W., Lockwood, B. L. & Somero, G. N. Can the giant snake predict palaeoclimate? *Nature* **460**, doi:10.1038/nature08224 (2009).
2. Sniderman, J. M. K. Biased reptilian palaeothermometer? *Nature* **460**, doi:10.1038/nature08222 (2009).
3. Head, J. J. *et al.* Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. *Nature* **457**, 715–717 (2009).
4. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Re-calibrating the snake palaeothermometer. *Nature* **460**, doi:10.1038/nature08223 (2009).
5. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Gigantism, temperature and metabolic rate in terrestrial poikilotherms. *Proc. R. Soc. Lond. B* **272**, 2325–2328 (2005).
6. Makarieva, A. M., Gorshkov, V. G. & Li, B.-L. Temperature-associated upper limits to body size in terrestrial poikilotherms. *Oikos* **111**, 425–436 (2005).
7. Slip, D. J. & Shine, R. Thermoregulation of free-ranged diamond pythons, *Morelia spilota* (Serpentes, Boidae). *Copeia* **1988**, 984–995 (1988).
8. Avery, R. A. in *Biology of the Reptilia, Physiology C* Vol. 12 (eds Gans, C. & Pough, F. H.) 93–166 (Academic, 1982).
9. Chappell, M. A. & Ellis, T. M. Resting metabolic rates in boid snakes: allometric relationships and temperature effects. *J. Comp. Physiol. [B]* **157**, 227–235 (1987).
10. Sweet, S. S. & Pianka, E. R. Monitors, mammals and Wallace's Line. *Mertensiella* **16**, 79–99 (2007).
11. Thompson, G. Goanna metabolism: different to other lizards, and if so, what are the ecological consequences? *Mertensiella* **11**, 79–90 (1999).
12. Head, J. J., Barrett, P. M. & Rayfield, E. J. Neurocranial osteology and systematic relationships of *Varanus (Megalania) prisca* Owen, 1859 (Squamata: Varanidae). *Zool. J. Linn. Soc.* **155**, 445–457 (2009).
13. Wroe, S. A review of terrestrial mammalian and reptilian carnivore ecology in Australian fossil faunas, and factors influencing their diversity: the myth of reptilian domination and its broader ramifications. *Aust. J. Zool.* **50**, 1–24 (2002).
14. Erickson, G. M., De Ricqles, A., De Buffrénil, V., Molnar, R. E. & Bayless, M. K. Vermiform bones and the evolution of gigantism in *Megalania*—how a reptilian fox became a lion. *J. Vert. Paleo* **23**, 966–970 (2003).