TEAM TAO

In this issue (28 July 2005)

Editorials

India's just desserts p441
The nuclear technology transfer deal agreed by the United States and India makes some sense on its own merits — but it leaves international non-proliferation efforts in disarray.

Born to be in Brussels p441
The proposed European Research Council will be safest under the wing of the European Commission.

Independence at the top p442
The US Food and Drug Administration badly needs some strong and stable leadership.

Research Highlights

Research highlights p444

News

Nuclear deal riles India's researchers p446
Scientists say home-grown programme will suffer as international isolation ends.
Declan Butler and K. S. Jayaraman

Sighting of 'extinct' bird may have been a case of mistaken identity p447
Doubts raised over validity of ivory-bill woodpecker report.
Rex Dalton

Beta-blockers tackle memories of horror p448
Common drug might treat post-traumatic stress disorder.
Jim Giles

Sidelines p450

Senate hearings strengthen calls for US action over climate p450
Winds of change blow through Congress.
Andreas von Bubnoff

Snapshot: Fresh scope p451
Archive photograph shows Darwinism in the dock.

Australian scientists protest at loss of funding board p451
Academic community fears lobbying power will be lost to government.
Carina Dennis

Biodiversity and climate form focus of forest canopy plan p452
International rainforest project looks skywards.
Michael Hopkin

Neuroscientists see red over Dalai Lama p452
Researchers petition against meditation lecture.
David Cyranoski

News in brief p453

News Features

In defence of data p454
As spokesman for America's scientific élite, Ralph Cicerone will have to do some tough talking in Washington. Emma Marris asks him how he'll ensure that politicians will listen to the science.

Lymphatic system: Unlocking the drains p456
After centuries of playing second fiddle to the blood system, our lymphatic circulation is coming into its own as a key player in diseases ranging from cancer to asthma. Phyllida Brown reports.

50 and 100 years ago p473

Naturejobs

Futures

India's just desserts

The nuclear technology transfer deal agreed by the United States and India makes some sense on its own merits — but it leaves international non-proliferation efforts in disarray.

'Smiling Buddha' was the codename for India's first nuclear bomb test in 1974. He must be grinning widely this week after Manmohan Singh, India's prime minister, reached a nuclear cooperation deal with US President George W. Bush.

The agreement would end a three-decade-old ban on US exports of nuclear technology or fuel to India. The United States will also move to amend international rules that permit nuclear exports only to countries that allow International Atomic Energy Agency (IAEA) safeguards on all their nuclear facilities. These rules were created in response to India's 1974 test, which used plutonium supplied by Canada.

A joint statement issued by Bush and Singh (see page 446) said that "as a responsible state with advanced nuclear technology, India should acquire the same benefits and advantages as other such states". The substitution of the phrase "nuclear weapons state" for "state with advanced nuclear technology" would have rendered the language more accurate, as the agreement is tantamount to a *de facto* admission of India into the established nuclear-weapons club of the United States, Russia, China, France and the United Kingdom.

This is a very big deal. It flies in the face of the 1970 Nuclear Non-Proliferation Treaty (NPT). Under the treaty, states that were not already nuclear-weapons states agreed not to develop weapons, and in return are given access to nuclear-power technologies and fuels, under the oversight of the IAEA. As part of the bargain, the nuclear weapons states also agreed to progressively disarm.

India is not a member of the NPT, and it deserves some credit for the consistency of its nuclear policies. As early as 1955, it was calling unsuccessfully for a global ban on nuclear testing. Whereas Iraq and North Korea each signed the treaty and went on to surreptitiously attempt to develop nuclear weapons, India has refused to sign, arguing that existing nuclear powers should first demonstrate their commitment to reducing their own arsenals.

That is one reason why it would be simplistic to say that extending nuclear cooperation to India is fundamentally wrong. India has not acted outside international law. Arguably, if Britain and France can justify maintaining nuclear weapons, the world's largest democracy has a strong case for being accepted as a nuclear-weapons state.

Unlike its neighbour Pakistan, India has acted responsibly to prevent the proliferation of its own nuclear technology. The deal with the United States will also be a boon to the growth of India's nuclear power sector — although some Indian scientists fear that they may now become unnecessarily dependent on US technology. That is seen as a setback in a country that prides itself on its independent scientific and technical capacity (see India Outlook, page 477).

Nonetheless, the sight of Singh and Bush toasting India's nuclear future must have left Brazil, Argentina and South Africa — which each abandoned nuclear weapons and signed up to the NPT — feeling short-changed.

Indeed, the bilateral agreement between India and the United States could deal a hammer blow to the NPT. This May, a review conference of the treaty made little headway, and the latest agreement seems to reflect the Bush administration's desire to sideline the treaty altogether. In its place, the administration proposes non-proliferation à la carte, with the menu determined by the United States' ever-changing geopolitical interests.

Doubts about the feasibility of this approach should weigh heavily on the minds of US legislators as they consider the changes to existing law needed to implement the Bush–Singh agreement. The Indian deal may be defensible in its own terms — but the overall philosophy behind it is a recipe for further nuclear proliferation. ■

> **"The agreement between India and the United States seems to reflect the Bush administration's desire to sideline the Non-Proliferation Treaty."**

Born to be in Brussels

The proposed European Research Council will be safest under the wing of the European Commission.

Whatever happens over the next few months in budget negotiations for the European Union's Seventh Framework Programme for research, there now seems to be widespread political acceptance that the European Research Council (ERC) should be established as a funding body for the continent's researchers in the sciences and humanities.

The most visible step so far in the creation of the ERC occurred last week when the 22 members of its scientific board were named

(see www.europa.eu.int/comm/research/press/2005/pr1807en.cfm). The European Commission's selection committee for the board did a good job, combining strong scientific credibility with a reasonable geographical balance. The commission now needs to build on this success in its creation of a well-crafted agency.

Contrary to the wishes of some member states, the ERC's foot soldiers should not primarily be seconded from national agencies; they should be independently recruited. To keep the disciplinary programmes vital and responsive, they should be led by scientists from research institutions. And above the heads of the ERC executive staff, the agency should be placed firmly in the hands of the commission during its first few years, rather than be controlled more directly by member states.

Such a structure will be not unlike that of the US National Science

Foundation (NSF). It is not a perfect system, but it has served US science well for more than 50 years and it has the sort of autonomy to which the ERC should aspire. All would agree that, like the NSF, the ERC should be a competitive organization for bottom-up basic research, funding researchers from across all disciplines and from across the continent, solely on the basis of excellence.

But 'autonomy' means different things to different people. For some, it means as complete a dissociation as is possible from the European Commission. History might seem to support this argument, even though it will be the commission that will be footing the bill, likely to be up to €1.5 billion (US$1.8 billion) per year. The commission's customary bureaucracy has never failed to alienate researchers.

But the commission can in this case play a useful role in sheltering the new-born agency from interference by competing national interests. Opponents of management by the commission have latched on to a hitherto unused article in the rules governing commission research, which allows a separate agency to be set up to run a particular programme. At an informal meeting of research ministers of the European Union's 25 member states in Cardiff earlier this month, opinion was equally divided between the merits of this and a commission-led agency.

It may already be too late for the more independent agency to be in place at the start, if the ERC is to be launched in 2007 as planned. Its establishment and location would require approval by both the European council of ministers and the European parliament —

bodies that stretched out the decision to create the somewhat analogous European food agency over three years in an acrimonious and heavily politicized struggle.

Even more dangerously, there is a strong possibility that the agency would establish a decision-making board filled with national representatives. In contrast, a commission-run agency would have only a five-strong steering committee to ensure that it executes appropriately the will of its scientific council.

European experience demonstrates time and again that when the European Union's member nations get too closely involved, their diverse individual goals distort agendas and delay decisions. One of the commission's cardinal roles is to shelter the implementation of European policies from the distorting effects of national interests.

The best way forward is therefore to set up the ERC as a commission-led agency with the option of switching to a more independent operation later if it proves desirable. The commission would then have time to prove that it can run an agency capable of executing the desires of its scientific council without debilitating bureaucracy. Better to live in hope than to sacrifice the ERC to the certain buffeting between the conflicting fancies of Europe's 25 member states. ∎

> "Experience demonstrates time and again that when Europe's member nations become involved, their individual goals distort agendas and delay decisions."

Independence at the top

The US Food and Drug Administration badly needs some strong and stable leadership.

On 18 July, the US Senate finally confirmed Lester Crawford as commissioner of the Food and Drug Administration (FDA). Crawford has been acting commissioner since March 2004. This critically important agency has lacked a confirmed leader for more than half of President George W. Bush's tenure — and the cracks are beginning to show.

Crawford, a pharmacologist, veterinarian and member of the Institute of Medicine, has previously held senior positions at both the FDA and the Department of Agriculture. In 2002 and 2003 he was the FDA's deputy commissioner. Patient advocates praise him for trying to improve the predictability and efficiency of cancer-drug approvals. And he has expertise in food safety and animal health — relevant skills in an era of bioterrorism and bird flu.

On some issues, however, Crawford already stands accused of lacking the independence and leadership skills required of the position. One bone of contention, which held up his Senate confirmation for several months, concerns access to an emergency contraceptive. Despite the approval of a panel of outside scientific experts, as well as the FDA's own staff scientists, Crawford had failed to act on an application from the manufacturer to make the 'morning after' contraceptive known as Plan B available over-the-counter. This month, the agency finally promised to make a decision by 1 September. The

troubling implication is that until his nomination got bogged down in the Senate, Crawford bent to political pressure from an anti-abortion White House, instead of making a decision based on the best scientific evidence, which concluded that Plan B is safe and effective.

Crawford also seemed to be caught flat-footed last autumn, when the UK authorities shut down a Chiron plant in Liverpool, halving at a stroke last winter's US supply of flu vaccine. The FDA had failed to re-examine the plant after it had contamination problems in 2003 — even after the company reported, two months before it was shut down last October, that millions of doses of flu vaccine had been contaminated.

Additionally, consumer advocates are worried by a recent sharp decline in FDA enforcement actions against false or misleading advertising. FDA warning letters to pharmaceutical companies fell by 70% between January 2002 and December 2004, at a time when drug companies were pouring unprecedented sums into advertising directly aimed at consumers. Crawford was the FDA's top manager for much of this period.

Now his position has been confirmed, Crawford has more leverage, and he should have plenty of opportunities to demonstrate the leadership that the agency requires. Another flu-vaccine shortage looms this winter, and bird flu threatens the prospect of a 1918-style pandemic. Public suspicion of the pharmaceutical industry continues to grow in the wake of the Vioxx withdrawal (see Business, page 459), amid evidence that drug-makers are still slow to file evidence about the safety of established products with the regulator. In the face of these challenges, Crawford has to prove that he is the right man for the job. He could start by doing the right thing by Plan B. ∎

RESEARCH HIGHLIGHTS

Wee beasties

J. Exp. Biol. **208,** 2865–2872 (2005)

When male lobsters clash, the loser remembers who beat him and retreats from the winning male during future encounters.

Researchers at the Boston University Marine Program in Woods Hole, Massachusetts, have discovered how this recognition works. It depends on chemoreceptors on the odour-sensitive hairs along the lobsters' antennae detecting urine pheromones.

When Jelle Atema and Meg Johnson shaved off these hairs, the weaker lobsters (*Homarus americanus*) continued to pick fights with champions who had defeated them before.

J. ATEMA

NEUROBIOLOGY
Spindle cycle

Cell **122,** 119–131 (2005)

When neural stem cells in the embryonic mammalian brain divide, they yield either two new stem cells or a stem cell plus a nerve cell.

Before any cell divides, its scaffolding, or cytoskeleton, forms a shape like an American football called the mitotic spindle. Kamon Sanada and Li-Huei Tsai of Harvard Medical School in Boston, Massachusetts, show in mice that the orientation of this spindle in neural stem cells determines whether one daughter cell will become a nerve cell. They also identify intracellular signalling molecules that control the spindle orientation — these are the $\beta\gamma$ subunits of G proteins and their activator, AGS3.

COSMOLOGY
Parting of the waves

Phys. Rev. Lett. **95,** 031301 (2005)

A weak light is thought to leak from black holes. It originates when the black hole captures one of a pair of virtual particles — the other seems to be emitted from the same spot.

The process, known as Hawking radiation, could be mimicked in the laboratory by studying microwaves moving in an electronic structure known as a waveguide, propose Ralf Schützhold of Germany's Dresden University of Technology and William Unruh of the University of British Columbia, Vancouver.

The velocity of the microwaves can be reduced by sweeping a laser beam along the waveguide. An observer moving with the beam would see slower-moving microwaves

'trapped' to their left and faster-moving ones escaping to their right — a simulation of what happens to virtual particles at the edge of a black hole (see artist's impression, below).

MATERIALS
All white now

J. Am. Chem. Soc. **127,** 9986–9987 (2005)

Making white light generally requires the combination of sources of red and green (making yellow) and blue light. Phosphor-based light sources come in all these individual colours and can be mixed appropriately — to mimic natural light, for example.

Now, Sue-Lein Wang and her colleagues at the National Tsing Hua University in Taiwan have found the first single-material white phosphor. It is a zinc gallophosphate laced with nanoscale pores, made by mildly heating the raw inorganic ingredients in water and ethylene glycol. The white light seems to stem from blue fluorescence in highly ordered regions of the material, combined with yellow emission where it is more disordered.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

C. DARKIN/SPL

OPTICS
Speed writing

App. Phys. Lett. **87,** 031101 (2005)

Laser pulses can be used to write patterns into transparent objects by burning tiny holes, but the repetition needed to build a whole picture can be time consuming.

Yoshio Hayasaki and his colleagues from the University of Tokushima, Japan, have devised a method whereby a complete pattern is created with a single laser pulse.

The laser light passes through a computer-controlled liquid-crystal cell. And the different diffraction patterns that result allow the laser to etch entire characters. The technique could be used to carve tiny serial numbers into products.

DEVELOPMENTAL BIOLOGY
Keeping mum

J. Exp. Med. **202,** 231–237 (2005)

For a fetus to survive, the mother's immune system must be appropriately suppressed — otherwise her body could react against proteins expressed by fetal genes inherited from the father.

Levels of a regulatory molecule called programmed death ligand 1 (PDL1) are known to be elevated in the human placenta. Indira Guleria of Harvard Medical School, Boston, and her colleagues show that this molecule may play an important role in the development of immune tolerance during pregnancy.

They show that decreased levels of PDL1 in mice raises the number of immune cells that tackle infectious diseases, and heightens the risk of fetal rejection.

STRUCTURAL BIOLOGY
In the loop

Nature Chem. Biol. doi:10.1038/nchembio723 (2005)
The human *MYC* gene regulates cell growth, but is overexpressed in many cancers. Lengths of its promoter region, the gene's activation switch, contain multiple stretches of repeating guanine residues. Such stretches typically do not assume the classical double-helix structure, but a quadruplex structure where four DNA strands fold back on themselves.

Dinshaw Patel and his colleagues at the Memorial Sloan-Kettering Cancer Center in New York report the first structure of a DNA fragment with five guanine repeat stretches within the *MYC* gene promoter (pictured). They show that it forms a quadruplex with a unique combination of strand directions and propeller-like loops that may provide a target for anticancer drugs.
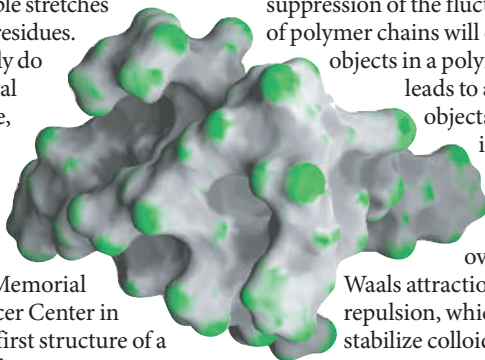
IMMUNOLOGY
Nosy parkers

PLoS Pathog. doi:10.1371/journal.ppat.0010001 (2005)
Some of the many species of bacteria that colonize mammalian noses can activate the innate immune systems of their hosts to destroy a competing species.

Elena Lysenko and her colleagues from the University of Pennsylvania School of Medicine, Philadelphia, studied the competition between two potentially pathogenic bacteria, *Haemophilus influenzae* and *Streptococcus pneumoniae*. When one strain was presented to mice in the absence of the other, each colonized a similar microenvironment in the nose and pharynx. But when both were present, *H. influenzae* stimulated immune cells called neutrophils to destroy its competitor.

The study indicates that the immune system can mediate competitive conflicts between species, influencing the composition of colonizing flora.

POLYMER PHYSICS
Suspended animation

Phys. Rev. Lett. **95,** 038305 (2005)
The mysterious Casimir force is an attraction found between two conducting plates when they are separated by only a few atomic diameters. The nearness of the plates suppresses a proportion of the electromagnetic fluctuations that naturally occur in space. With fewer such fluctuations there is 'less than nothing' between the plates, and this forces them together.

Sergei Obukhov and Alexander Semenov of France's Charles Sadron Institute in Strasbourg calculate that a similar suppression of the fluctuations in density of polymer chains will occur between two objects in a polymer solution. This leads to a force between the objects, but in this case the interaction is repulsive — an anti-Casimir force. The researchers say this force could overwhelm van der Waals attraction and lead to a net repulsion, which could be used to stabilize colloidal suspensions.

CANCER
Jagged path

Cancer Cell **8,** 13–23 (2005)
Tumour cells frequently release 'pro-angiogenic' proteins, which stimulate the growth of blood vessels needed to feed them.

But Cun-Yu Wang and his colleagues from the University of Michigan have found a molecular mechanism for such blood vessel stimulation that involves direct physical contact between tumour cells and the endothelial cells that line the vessel walls.

They found that a chemical called hepatocyte growth factor turns on a gene in tumour cells called *jagged1*. Blood-vessel growth is triggered when the Jagged1 protein, displayed in the tumour cells' outer membrane, binds to a protein called notch in the membrane of endothelial cells.

CELL BIOLOGY
Born again

Cell **122,** 85–95 (2005)
Peroxisomes are small, abundant organelles in cells that assist oxidative reactions. One theory holds that they evolved independently of eukaryotic cells, as mitochondria may have done. But, unlike mitochondria, their number in a living cell is constantly refreshed.

Henk Tabak from the University of Utrecht in the Netherlands and his colleagues have deciphered a key step in this regeneration process. By following the path of two proteins vital for the formation of peroxisome membranes, they show that the membranes are derived from another, highly complex organelle — the endoplasmic reticulum. This supports the notion that peroxisomes evolved in primitive eukaryotes, not independently.

A theoretical physicist hopes to see obese traits in skinny neutrinos, given cosmic patience.

The distinguished mathematical physicist Roger Penrose recently stirred up a debate that has been running for more than 75 years. The question is: why do the rules of quantum mechanics play no role in the macroscopic world?

In *The Road to Reality,* published in March, Penrose argues that the limit of quantum rules is rooted in a physicist's nightmare: the deep conflict between quantum mechanics and Einstein's general theory of relativity. He predicts that gravitational effects thoroughly destroy the quantum nature of anything heavier than an obese bacterium so quickly that we cannot detect the process.

But might we see the transition in objects that are just a little bit smaller? I got excited about Penrose's ideas because I follow the field of quantum computing. In the past, no one even came close to carrying out quantum experiments on objects as big as bacteria. Now experimentalists are perfecting fat quantum things in order to get quantum computers running, although messy environmental influences deteriorate their signals.

Neutrinos are isolated from such influences, but you probably wouldn't expect these particles to be much help because of their tiny mass — at 10^{-36} kg they are the skinniest things in the Universe. This should mean their quantum states remain unperturbed by gravity for longer than the lifetime of the Universe.

Yet Joy Christian from the University of Oxford, UK, argues that patience may pay off (http://arxiv.org/quant-ph/0503001; 2005). He shows that neutrino oscillations speed up Penrose collapse. Neutrinos created soon after the Big Bang may now have lost their quantum nature... and might be detected by telescopes that are currently being constructed.

NEWS

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The deal agreed by India's prime minister, Manmohan Singh, and President Bush could undermine the country's home-grown nuclear technology.

Nuclear deal riles India's researchers

India's nuclear-power programme, which has been cut off from foreign technology by international sanctions since the country exploded its first nuclear bomb test in 1974, looks set to come in from the cold. A historic deal reached last week in Washington by US President George W. Bush and Indian Prime Minister Manmohan Singh would end India's nuclear isolation, without requiring it to give up its nuclear arsenal.

Most of India's nuclear scientists are relieved that the accord will end fuel shortages and accelerate the country's nuclear-energy programme. But many in India's nuclear establishment, proud of having built up an independent programme despite international isolation, are concerned that it could now be supplanted by cheaper foreign technologies, and that India will lose its hard-won control over its civil and military nuclear future.

"All the self-reliance so relentlessly built over the years against heavy odds will go down the drain"

Nuclear power currently generates just 3% of India's electricity, but is an attractive alternative to burning coal or imported oil and gas to meet the country's burgeoning energy needs. According to the Washington-based Population Reference Bureau, India's population will reach 1.6 billion by 2050, surpassing the 1.4 billion predicted for China.

The deal must still be passed by the US Congress, but if it goes ahead, access to cheaper enrichment and fuel services would allow India to build or buy much larger light-water reactors than at present, says Per Peterson, a nuclear physicist at the University of California, Berkeley. "Due to the economies of scale, this has the potential to greatly reduce the cost of any large expansion of nuclear-energy use in India," he explains.

India has unsuccessfully sought technology for more light-water reactors from Russia, and would also be interested in obtaining

centrifuge enrichment technology, adds Frank von Hippel, professor of public and international affairs at Princeton University in New Jersey.

Conflicting interests

But the ability to buy such technologies may scupper India's home-grown plutonium fast-breeder reactors and its ambitious thorium fuel research programme.

India has only around 50,000 tonnes of natural reserves of uranium. "That's enough for its current programme but not for the large nuclear-energy programme its nuclear establishment dreams of," says von Hippel. But it has one of the world's largest reserves of thorium — 360,000 tonnes. The ultimate aim of its nuclear programme is to develop thorium fuels, which, along with plutonium fast-breeder reactors, could allow India to become self-sufficient in nuclear energy.

"Buying uranium and enrichment on the

446

world market would be much less costly. So if this agreement goes through, it would tend to undercut India's breeder reactor and reprocessing establishments," says von Hippel.

Many in the Indian nuclear establishment are also concerned at Singh's promise to accept the same restrictions as the five official nuclear weapons states. These include separating civilian nuclear facilities from military ones, allowing the International Atomic Energy Agency to inspect its civilian facilities, and maintaining a unilateral moratorium on nuclear testing.

Segregation of civilian and nuclear facilities would be expensive and impractical, says Padmanabha Krishnagopala Iyengar, former chairman of the Atomic Energy Commission and a key scientist in India's weapons development. He points out that at present both kinds of research are usually done at the same laboratories. "Nobody works full time in our weapons programme," he told *Nature*. "The moment we compartmentalize, our research and development will be crippled and creativity will end."

Segregation could leave the entire nuclear programme in a mess, agrees Annaswamy Narayana Prasad, former director of the Bhabha Atomic Research Center in Mumbai. Given the small scale of military activities, dedicating facilities for a single purpose is neither practical nor cost effective, he says.

U-turn required

Younger scientists in India's nuclear establishment are more open to the separation of civil and military facilities. "Those who oppose it belong to an older generation with a closed mind-set," says one reactor designer at the Bhabha research centre, who asked not to be named. "The responsibility for running the nuclear programme is now on new shoulders and the present reality calls for a U-turn. Segregation is better than camouflaging."

But if nuclear cooperation means that India will find it cheaper to buy reactors from abroad, this would be a heavy blow for India's nuclear establishment. "India has been outside the nuclear club for some time, and has had to rely on endogenous resources," says Ziad Haider of the Henry L. Stimson Center, a Washington-based security think-tank. "These scientists now see their programme opened to international supervision, and becoming reliant on US technology," he says.

Iyengar agrees: "I think the Indo-US deal, if implemented, would tie our hands. All the self-reliance so relentlessly built over the years against heavy odds will go down the drain."

Declan Butler and K. S. Jayaraman, New Delhi

Sighting of 'extinct' bird may have been a case of mistaken identity

It was touted as the conservation discovery of a lifetime — an ivory-billed woodpecker, a bird long thought to be extinct, spotted in the Arkansas swamps. But now the finding is being called into question, which could cast doubt on several prominent US conservation measures.

In April, biologists and bird-lovers were thrilled by a videotape, reported in *Science* by ornithologist John Fitzpatrick of Cornell University in Ithaca, New York, of what looked to be an ivory-billed woodpecker (*Campephilus principalis*). The species hadn't been confirmed in the United States since 1944, although there had been rare sightings in Cuba. The bird gradually vanished as its dense forest habitat was chopped down, making it a symbol of lost heritage.

So leading US officials, including the secretaries of the interior and agriculture, jumped at the chance to announce the bird's apparent rediscovery. They re-routed $10 million from other conservation measures to pay for efforts to save the bird's vanishing habitat. These changes came swiftly, even though some in the Bush administration and Congress have been working to reduce species protections under the US Endangered Species Act.

But the investment may be premature, a new study suggests. A team of ornithologists led by Richard Prum of Yale University in New Haven, Connecticut, plans to report what it thinks is a case of mistaken identity. The bird described in *Science*, the group says, is not an ivory-billed woodpecker after all, but a non-endangered relative, the pileated woodpecker (*Dryocopus pileatus*).

Prum's team includes the leading US authority on ivory-billed woodpeckers, Jerome Jackson of Florida Gulf Coast University in Fort Myers, who for decades has been unable to document a sighting. "I have serious questions about the *Science* report," Jackson told *Nature* in May.

Prum and his colleagues scrutinized the Cornell team's video, and believe that the bird's size and white markings suggest that it could be a pileated woodpecker, rather than an ivory-billed woodpecker. The Cornell team had considered this possibility and discounted it.

Prum declined to discuss details of his manuscript until it is published in *PLoS Biology*. The third author of the paper is Mark Robbins, an ornithologist at the University of Kansas in Lawrence.

Fitzpatrick and other co-authors of the *Science* paper also declined to comment. *PLoS Biology* plans to publish a response from the Cornell team, and a response to that from Prum's group. All three papers are expected to go online within a month.

Despite the challenge to the sighting, conservationists plan to meet in Little Rock on 2 August to discuss how to save the woodpecker's home forests. ■
Rex Dalton

Down the pecking order: a sighting of the ivory-billed woodpecker, thought to have been extinct, may instead have been a non-endangered relative.

Beta-blockers tackle memories of horror

Could a common drug be able to extinguish the trauma associated with horrific memories?

A group of US psychiatrists thinks it can. They claim that beta-blockers, which are widely used to treat high blood pressure and heart problems, interfere with the way that the brain stores memories. Administer these drugs at the right time, the psychiatrists suggest, and sufferers of post-traumatic stress disorder (PTSD) could snuff out the periods of panic that blight their lives.

The idea is timely, yet controversial. Studies of terrorist attacks suggest that the recent bombings in London and Egypt will cause a spike in the number of cases of PTSD (see 'The long-term toll of terrorism'). Treatments such as counselling have limited success, so alternatives are welcome. But some mental-health experts say that the treatment could be abused, perhaps by military commanders who want soldiers to become desensitized to terrible acts.

The beta-blocker idea was inspired by the discovery in the past decade that memories are surprisingly pliable. Some drugs, for instance, can interfere with a memory when it is recalled, such that an altered version is put back in storage. The beta-blocker propranolol is one example. It blocks the neurotransmitters involved in laying down memories, and studies have found that rats who have learned to fear a tone followed by an electric shock lose this fear if propranolol is administered after the tone starts[1].

A group of New York-based psychiatrists has produced similar results in humans, *Nature* has learned. They plan to submit their work for publication in September, but have already started work on a clinical trial of propranolol in PTSD patients. Roughly one in three people caught in events such as terrorist attacks will suffer PTSD symptoms: panic attacks and flashbacks that are triggered by events associated with memories of the trauma, such as sirens or bangs.

"The memory of the event is associated with the fear and they always occur together," says Margaret Altemus, a psychiatrist at Cornell University who is working on the clinical trial.

Breaking the cycle

Altemus and her colleagues ask their subjects to take a dose of propranolol whenever they feel the characteristic symptoms of PTSD, such as rapid heart rate or troubled breathing. If the theory is right, the trauma memory should be pliable at that point. And propranolol should break the link between recall and fear.

The team wants 60 subjects, but has managed to recruit just one a month since work started in February. Altemus says many PTSD sufferers are reluctant to enrol in an experimental programme when established treatments are already available. In exposure therapy, for example, subjects recall the traumatic event and then try to calm the memory
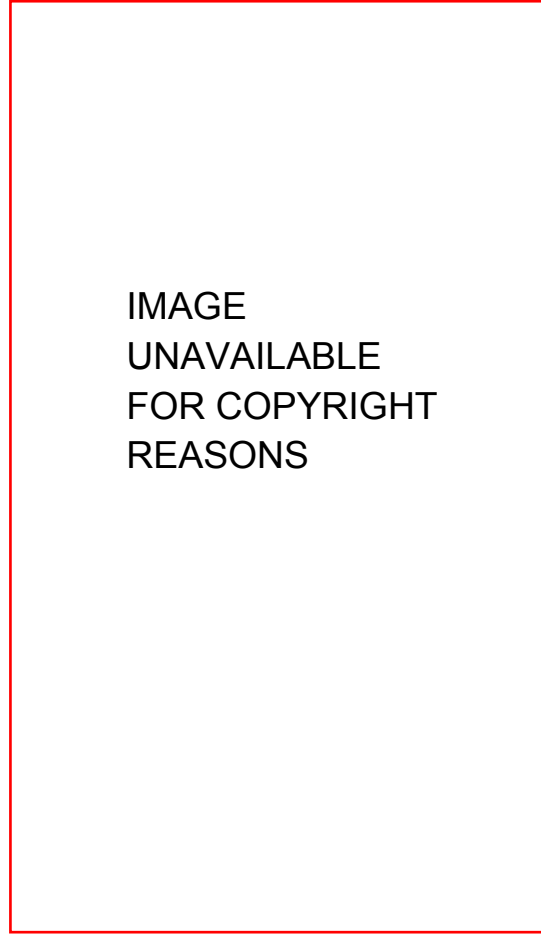
IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

by thinking of happy experiences.

This is probably the best treatment available at present, but such therapies benefit only two-thirds of patients. And some PTSD sufferers, notably those who experienced sexual abuse, are reluctant to actively recall their memories.

The long-term toll of terrorism

Just weeks after the terrorist attacks of 11 September 2001, psychiatrists were busy uncovering the psychological impact of the disaster. Thanks to their work, and to studies on Palestinian attacks in Israel and Islamic terrorism in France, scientists in Britain and Egypt know how this month's bombings are likely to affect people.

Terrorist attacks often take a greater toll than do other traumatic events, such as accidents, because there is more panic, and explosions expose people to gruesome disaster scenes. When psychiatrists in Israel compared survivors of terrorist attacks and traffic accidents, for

example, they found that almost 40% of the former developed post-traumatic stress disorder (PTSD), around twice the number of cases in the accident victims[2].

Few other studies have made such comparisons, but much is known about survivors of terrorism. Last year, psychiatrists in France published a study[3] of survivors of the Islamic terror attacks of 1995 and 1996. Around a third suffered from PTSD two to three years after the event. Women were more likely to be affected than men, as were those who suffered severe initial injuries.

Other PTSD studies are more controversial. After the 2001 attack on the World Trade Center

(pictured), US researchers used telephone interviews to assess the event's impact on university students[4]. They found a high level of PTSD — around 30% — on one campus close to New York City. This elevated figure was linked to the amount of television coverage of the event that individuals had watched.

Such results have since been criticized by researchers who feel that the diagnosis of PTSD should be restricted to those who suffer severe and disabling reactions to memories of an event. Some psychiatrists point out that many people temporarily experience isolated symptoms of PTSD after a traumatic event, such as feeling jumpy or wanting to avoid the affected place. PTSD sufferers, in contrast, suffer for prolonged periods of time.   **J.G.**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Traumatic events may trigger panic in the years that follow, but there is hope for a simple drug therapy.

Given these limitations, the propranolol study is welcomed by other PTSD researchers. But some caution that a focus on a single drug may obscure other problems. "We hope it will work, but it's a simple solution," says Berthold Gersons, a psychiatrist at the University of Amsterdam in the Netherlands.

He recounts how he treated the insomnia suffered by the survivor of a fire in a crowded party on New Year's Eve. After several sessions of therapy, Gersons discovered that his patient was subconsciously listening for a sound he had heard during the fire: the noise of air being sucked into the blaze. As the survivor was not aware he was doing this, and did not initially recall the memory, Gersons questions how propranolol would have helped him.

Altemus acknowledges that her treatment will not work in every case, and adds that all trial participants will be given other help with PTSD, such as breathing exercises.

Some researchers have objections that may be harder to deal with. They argue that aspects of PTSD are a natural response to traumatic events and should not be treated with drugs.

"If soldiers did something that ended up with children getting killed, do you want to give them beta-blockers so that they can do it again?" asks Paul McHugh, a psychiatrist at Johns Hopkins University in Baltimore, Maryland, and a member of the US President's Council on Bioethics. "Psychiatrists are once again marching in where angels fear to tread."

Altemus rejects such charges. She says that propranolol should lessen the panic and flashbacks associated with a memory, but will not remove it altogether. With the memory still in place but the trauma dimmed, individuals should be better equipped to make moral judgements about the event.

"People with PTSD are disabled — their communication and relationships can be crippled," she adds. "It's a serious illness." ∎
Jim Giles

1. Dębiec, J. & Ledoux, J. E. *Neuroscience* **129,** 267–272 (2005).
2. Shalev, A. Y. & Freedman, S. *Am. J. Psychiatry* **162,** 1188–1191 (2005).
3. Verger, P. *et al. Am. J. Psychiatry* **161,** 1384–1389 (2004).
4. Blanchard, E. B. *et al. Behav. Res. Ther.* **42,** 191–205 (2004); **43,** 143–150 (2005).

449

ON THE RECORD

"**We have a better than zero chance of successfully intercepting an inbound warhead. That confidence will improve over time.**"

Lieutenant-General Henry Obering, director of the US Missile Defense Agency, shows confidence in the US$93-billion warning system.

"**A hapless toad that, for reasons of its own, lives its entire life in California.**"

This description of an endangered creature from John Roberts, who has just been nominated to join the US Supreme Court, fails to reassure environmentalists of his caring nature.

SCORECARD

▲ **Blogs in space**
No audience for your blog? No problem. A US company will beam postings to outer space for extraterrestrials to read.

▲ **Ancient mariners**
Research is buoying the old sailors' tradition of using oil to calm a wave-tossed ocean — the oil prevents tiny water droplets from forming and worsening the storm.

▼ **Morning people**
Certain members of the US Congress want to extend daylight saving time by two months to conserve energy. Early risers are not amused at being left in the dark.

OVERHYPED

Shots in the dark
Last week, Britain invited industrial tenders for a contract for 2 million doses of vaccine against H5N1 avian flu. Such a jab might give limited protection to the first victims, but the order is still for an old strain of H5N1. Meanwhile, the United States announced its plan for rationing pandemic vaccines, putting the emergency workers at the head of the line. The problem? By the time the country has a vaccine available and gets around to distributing it, the disease will have long since ripped across the planet.

Senate hearings strengthen calls for US action over climate

WASHINGTON DC

The scepticism over global warming that has pervaded US politics is showing signs of abating. Two Senate hearings last week saw some of the strongest assertions yet, especially from Republicans, that climate change is a problem. And a congressman who recently launched a personal inquiry into the validity of certain climate studies has drawn fire from members of his own party for what they call his "intimidating" tactics.

Senator Pete Domenici (Republican, New Mexico), who chairs the Senate's energy committee, helped to convene a hearing on 21 July to discuss possible solutions to global warming. Leading US scientists, including the new president of the National Academy of Sciences (NAS), Ralph Cicerone (see page 454), and Mario Molina, a Nobel laureate at the Massachusetts Institute of Technology, told the assembled politicians that the world is warming at a dangerous rate.

"Nearly all" climate scientists believe that the observed warming is caused by increases in atmospheric levels of greenhouse gases, Cicerone told the hearing. This sentiment echoed his testimony the previous day, when he spoke at a hearing organized by a Senate subcommittee on climate change. "Carbon dioxide in the atmosphere is now at its highest level in 400,000 years," he said, "and continues to rise."

Several Republicans broke rank at the 21 July hearing to voice their concerns about global warming. "I don't need to be converted on whether or not we have climate change," said Senator Lisa Murkowski (Republican, Alaska), adding that she had seen evidence for global warming in her home state. Domenici backed her view: "I have come to accept that something is happening to the Earth's climate."

Senators agreed that direct action should be taken, although they could not agree on specific energy policies that might address the problem of greenhouse gases.

Meanwhile, Joe Barton (Republican, Texas), who chairs the House Committee on Energy and Commerce, has been roundly criticized for demanding information from various climate researchers about studies that show Earth's temperature rising dramatically over the past century.

Barton originally wrote to three climate scientists on 23 June, as well as to the heads of the Intergovernmental Panel on Climate Change and the National Science Foundation, asking for extensive information about their funding and methodology. The tone of the letter alarmed many climate researchers, who saw it as a bullying tactic (see *Nature* **436,** 1; 2005).

Now it seems that politicians agree. On 14 July, Sherwood Boehlert (Republican, New York), chairman of the House science committee, wrote to Barton calling his inquiry a "misguided and illegitimate" investigation that sought to "intimidate the scientists rather than to learn from them".

Researchers are continuing their own protests. On 15 July, 20 leading climate scientists sent Barton a letter pointing out that the study in question is just one among thousands of pieces of evidence contributing to the present consensus over climate change.

But Cicerone has offered a kind of olive branch, suggesting that the NAS could do an independent survey of climate-change research for Barton. "Science is one of the few human endeavours that's self-correcting and that we can rely on to get to the truth of the matter," he told the 20 July hearing. ■

Andreas von Bubnoff

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Ice pack: melting glaciers and other data may be convincing US senators that climate change is real.

APPETITE HORMONE MAY SQUELCH LEARNING
Biochemists call for study into obesity's effect on schoolwork.
www.nature.com/news

PUNCHSTOCK

W. DAVIS/SMITHSONIAN ARCHIVES

SNAPSHOT
Fresh scope

This photograph, taken on 20 July 1925, is one of 52 previously unpublished pictures of the 'monkey trial', in which the state of Tennessee found John Scopes guilty of teaching evolution. The negatives have just been restored by the Smithsonian Institution Archives in Washington DC.

Marcel LaFollette, an independent historian, stumbled across the images while doing unrelated research. They were among 475 boxes of material donated to the archive in 1971 by the Science Service, a Washington-based group that promotes science in the media. Watson Davis, later director of the service, took the informal shots while reporting on the case.

Scopes was being tried for teaching evolution after it had been banned earlier that year. This image was taken after hot weather caused the proceedings to move outdoors, and it shows prosecutor William Jennings Bryan (seated at the left) being interrogated by Scopes' defence lawyer Clarence Seward Darrow.

▶ www.siarchives.si.edu/research/scopes.html

Australian scientists protest at loss of funding board

SYDNEY

The Australian government has abolished the board of the nation's basic research agency, raising concerns that funding will become susceptible to political interference.

The Australian Research Council (ARC) administered AUS$556 million (US$420 million) in competitive grants this year. Under the current system, recommendations from peer reviews of grant applications are ranked by experts and passed to a board of leading community, industry and academic representatives. They are then sent to the council's chief executive, who passes them to the federal science and education minister, Brendan Nelson, for final approval.

When the board goes in 2006, only the chief executive, Peter Høj, will stand between the peer-review process and the minister.

The government says that removing the board will expedite the grant review process. The restructuring follows a government-commissioned review of all statutory authorities, which recommended that the ARC be governed using an executive management model, not a board.

But many scientists are concerned that the changes will eliminate a crucial buffer to political interference. "It might be administratively cleaner, but it will be a troubling loss of independence," says Snow Barlow, a plant biologist at the University of Melbourne, Victoria, and president of the Federation of Australian Scientific and Technological Societies, near Canberra.

"I think the government is tightening its control," agrees Frank Larkins, deputy vice-chancellor of research at the University of Melbourne. Recent reviews of the higher-education and research sector have underscored the government's drive for priority-driven research, closer collaborations between industry and academia and a tighter rein on universities (see *Nature* 429, 118; 2004).

Abolishing the board would leave the ARC vulnerable to political whim, according to John Mullarvey, chief of the Australian Vice-Chancellors Committee, based in Canberra. "The board is a powerful lobby," he says. "Unless that lobbying comes from somewhere else, this will have a detrimental impact on Australian research. We don't want projects rejected because they don't fit the ideology of the government."

He says his committee will work to ensure that the ARC's peer-review process is adequately protected by law.
Carina Dennis

Neuroscientists see red over Dalai Lama

A growing number of neuroscientists are calling for the cancellation of a special lecture to be given by the Dalai Lama in November. The Buddhist leader is due to speak at the annual meeting of the Society for Neuroscience (SfN) in Washington DC, but a petition against the talk has already gathered some 50 signatures.

The Dalai Lama has lived in exile in India since he fled Chinese troops in Tibet in 1959. Over the past decade he has increasingly encouraged researchers, sometimes at gatherings at his home, to study whether Tibetan Buddhist meditation can reshape the brain and increase mental well-being (see *Nature* **432,** 670; 2005). It was during one of these meetings that he was asked by a member of the society's executive committee, to give an inaugural lecture on 'the study of empathy and compassion, and how meditation affects brain activity'.

Some of the critics believe that the Dalai Lama's lecture should be ruled out because of his status as a political and religious figure. "One of the reasons for inviting him is that he has views on controlling negative emotions, which is a legitimate area for neuroscience research in the future," says Robert Desimone, director of the McGovern Institute for Brain Research at the Massachusetts Institute of Technology. But "the SfN needs to distance itself as much as it can from the Dalai Lama and his beliefs", adds Desimone, who opposes the lecture but has not yet signed the petition.

Many of the scientists who initiated the protest are of Chinese origin. But they insist

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Food for thought: the Dalai Lama is interested in the biological effects that meditation may have on the brain.

that their concerns are purely scientific. Yi Rao a neuroscientist at Northwestern University in Chicago, Illinois, helped to draft the petition, which says that the science of meditation is "a subject with hyperbolic claims, limited research and compromised scientific rigour".

The letter singles out one of the key publications in the analysis of meditation, in which Richard Davidson, a psychologist at the University of Wisconsin, Madison, and his colleagues claim that neural networks are better coordinated in people who are practised in meditation (A. Lutz *et al. Proc. Natl Acad. Sci. USA* **101,** 16369–16373; 2004).

Rao says that the study is flawed, especially in terms of the controls it used, because it compared practising monks in their thirties and forties with much younger university students. "Davidson is a respectable scientist," he says, "but he has put his respectability on the line with this."

Davidson defends his work as the first step in a new field. "Meditation research is in its infancy," he says. He helped to arrange the Dalai Lama's talk at the SfN meeting, to be held on 12–16 November. He says that criticism of the lecture on scientific grounds is misplaced, because the Dalai Lama is not claiming to be a scientist. "He merely wants to increase scientific attention on the topics that he thinks are important for human welfare," Davidson says.

The lecture is the first in a new series organized by the SfN, billed as "dialogues between neuroscience and society". The controversy has ensured that dialogue is already off to a rocky start.

The SfN's president, Carol Barnes, says that she is trying to find a resolution to the protest that will not involve cancelling the lecture. But one of the petition's organizers, Jianguo Gu of the University of Florida, says that he and several other scientists will cancel their lectures if the Dalai Lama's talk goes ahead. ■
David Cyranoski

Biodiversity and climate form focus of forest canopy plan

LONDON
Conservation researchers last week announced a plan to create a global network of rainforest research stations. If it secures funding, they say that the network would boost knowledge about biodiversity in the forest canopy, and help track the effects of climate change.

The project, run by the Global Canopy Programme at the University of Oxford, UK, would recruit local workers and give them the equipment and skills needed to study the canopies of their home forests. Results from the stations would be compiled to give a global picture of the state of forest canopies.

One focus will be biodiversity — ecologists want to refine their estimate that the canopies contain 40% of all terrestrial species. The stations will also look for changes in how much carbon dioxide is taken up by forests — an 'early warning' of the effects of global warming. Climatologists fear that as temperatures rise and forested regions become drier, forest growth may slow and the trees' capacity for soaking up carbon dioxide will diminish.

The project aims to set up five new stations in Brazil, Ghana, Madagascar,

Malaysia and India, to complement ten that already exist in forests around the world.

The plan is backed by the United Nations, which has already called for a world network of 20 canopy stations. Now the proponents, from 15 universities and research centres around the world, just need the money.

The stations should cost about US$18 million, the project's head, Andrew Mitchell, told a meeting on the Amazon Rainforest at the House of Commons in London on 21 July. "Compared to, say space exploration, it's a piffling amount of money," he said. ■
Michael Hopkin

Simpler funding rules offer hope to UK academics

The political manoeuvring can begin in earnest now that details of the research assessment exercise (RAE), the process by which research funds are distributed among 100 UK universities roughly every six years, have been released. Department heads are now trying to maximize their share of the £8 billion ($14 billion), available from 2008.

The new rules have some significant differences from previous years. The controversial seven-point scale used in the 2001 assessment, which allowed for big jumps in money awarded between some ratings classes, has been scrapped. Now researchers will submit just four publications to the assessment panel; other factors, such as research income from other sources, will also be considered. Each factor will be rated on a four-point scale, and these will be added together to create a profile for the entire department, which determines the amount of money awarded.

The rules are open for comment until 19 September at www.rae.ac.uk.

Deep Impact team seeks funding for next target

NASA's Deep Impact spacecraft, having scored a bullseye on one comet earlier this month, may get to visit another. The camera-equipped probe — minus its impactor, which vaporized on contact with Tempel 1 — will swing past Earth in December 2007. It could then be redirected to a close fly-by of comet Boethin in December 2008.

Principal investigator Michael A'Hearn of the University of Maryland says that this would cost about $32 million, or one-tenth of the original Deep Impact price. But it will have to compete with any other missions proposed to NASA's Discovery planetary exploration programme later this month.

Boethin would be the fifth comet that scientists have inspected up close. It was discovered in 1975, but scientists don't even know how big it is.

Mexico scours genes to combat local diseases

Mexico inaugurated some cutting-edge genomic research facilities on 25 July. These will be used to explore the genetic make-up of its population and to devise new drugs for certain diseases.

The facilities, at the National Institute of Genomic Medicine of Mexico in Mexico City, will tap the latest technologies for sequencing and genotyping, gene expression and bioinformatics. The Applied Biosystems

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Local knowledge: Mexico is set to investigate the genetic make-up of its diverse population.

Group, Affymetrix, and IBM are installing systems in three new labs.

Mexico's population includes those of Spanish descent and about 60 native Mesoamerican groups. The research findings will be used to tackle diseases that disproportionately strike Mexicans, such as diabetes, hypertension and asthma.

Breakaway ice sheet reveals secret community

An unusual ecosystem has been discovered on the floor of the Southern Ocean. But the collapse of the Larsen ice sheet that allowed researchers to find the community also threatens its future.

In 2002, a 3,250-square-kilometre piece of ice broke away from the Larsen ice shelf on the Antarctic Peninsula. In March this year, researchers from the US Antarctic Program found signs of life on a patch of sea bed that used to be covered by the ice. Images reveal a thick white mat that resembles the bacterial mats seen in other unusual deep-sea communities, and a video shows mud volcanoes with large collections of clams.

But with the protective ice sheet gone, the ecosystem may soon be buried by organic debris, say the researchers in *Eos*, the American Geophysical Union's newsletter.

Transgenic rape may form hybrids with wild weeds

British researchers have found evidence that transgenic oilseed rape can interbreed with a wild weed species, raising fears that herbicide-tolerant genes may jump into weeds. But they have dismissed concerns that 'superweeds' will take over the fields.

The three-year study, carried out at the Centre for Ecology and Hydrology in Dorset, UK, suggests that oilseed rape may have hybridized with charlock, a related weed species. In 28 fields sown half with transgenic oilseed rape and half with a non-transgenic version, two plants were found that seemed to have characteristics of both oilseed rape and charlock.

Despite claims by some environmental groups, the plants do not provide evidence that herbicide tolerance is spreading to wild plants, experts say — and the plants are expected to be less viable than their competitors anyway.
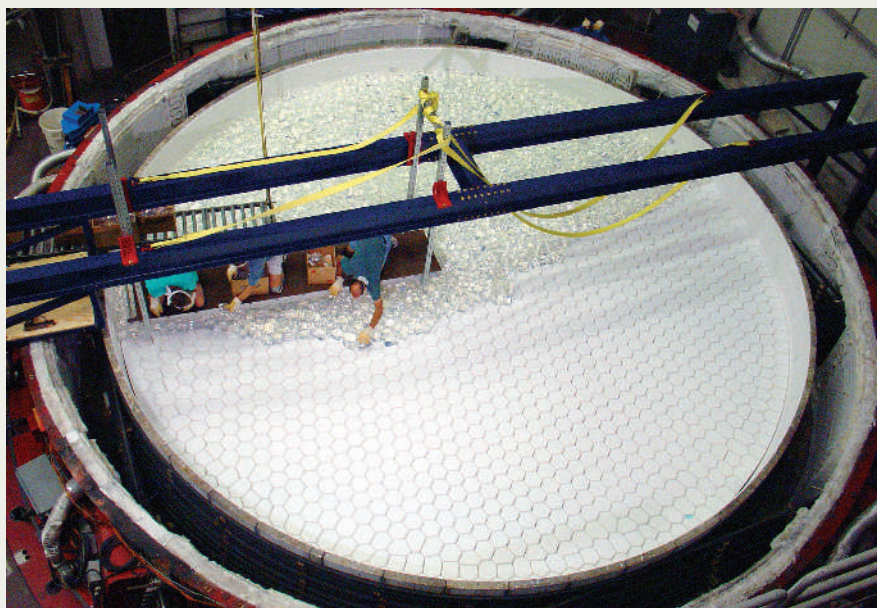
Giant mirror reflects scale of astronomers' vision

It took 18 tonnes of borosilicate glass and several days to melt it, but technicians have cast the first of the Giant Magellan Telescope's seven mirrors. The US project will be the biggest ground-based telescope ever built.

Eventually, the 8.4-metre diameter mirrors will nestle against each other to create a telescope with an effective collecting aperture of 21.4 metres — giving it four-and-a-half times the light-gathering power of any other optical telescope. The telescope is expected to start operating in northern Chile in 2016.

In defence of data

As spokesman for America's scientific élite, Ralph Cicerone will have to do some tough talking in Washington. **Emma Marris** asks him how he'll ensure that politicians will listen to the science.

Less than a month into his new job, Ralph Cicerone is already on the offensive on Capitol Hill. Taking over as the 21st president of the National Academy of Sciences (NAS) on 1 July, Cicerone has had to leave his office littered with unpacked boxes as he heads for Congress to do battle over climate change.

Cicerone earned his reputation not as a fighter but as a temperate manager and rational thinker. That level-headedness may serve him well in the volatile environment of Washington, but he clearly is not averse to standing his ground.

In his new role, Cicerone will have to work with a US administration that has often clashed with scientists over findings that oppose its approved policy. But he is clear that if the academy's reports come to unpopular conclusions, he won't be shy about letting the world know. "Once we come up with a position, I don't want to hide it in the closet," Cicerone told *Nature* last week. "I want it to be out there and useful and maybe a little bit aggressive."

Cicerone seems equally at home dealing with politics as he does defending scientific evidence. One of his early acts as NAS president was to engage Congressman Joe Barton (Republican, Texas) over a letter sent out last month. Barton had demanded that three climate researchers turn over reams of data on their climate-change research, a stance that has sparked disapproval among his fellow Republicans as well as other scientists (see page 450). Cicerone entered the fray to offer Barton an independent NAS report to stop individual scientists from being intimidated.

Balancing act

Balancing science and politics is a basic characteristic of Cicerone's new job. As president of the NAS, he represents the nation's most distinguished scientists. But he is also head of the National Research Council (NRC), which generates most of the influential reports advising the government on scientific issues.

Those reports don't always tell policy-makers what they want to hear, particularly when it comes to Cicerone's own field. As an atmospheric chemist, he is well equipped to deal with issues relating to climate change — a major scientific flashpoint for the Bush administration.

> **"I don't want to be part of an organization that just shoots off its mouth with opinions that are not as well justified as can be."**
> — Ralph Cicerone

Cicerone is probably best known for his contribution to work in the mid-1970s showing that human activities could damage Earth's protective ozone layer. He also played a prominent role in the debate over banning the use of chlorofluorocarbons to ease ozone depletion.

In Washington, scientific knowledge is just one of many factors taken into account when making a decision, and quite often not the most important. Cicerone says he is preparing himself to deal with that. "Being an engineer and a scientist, I tend to think that the facts and the data should dominate everything," he says. "I have to get used to the fact that a lot of people don't start that way."

Cicerone's career history spans US scientific hotspots, including the Massachusetts Institute of Technology and the Scripps Institution of Oceanography in La Jolla, California. Most recently he served as chancellor at the University of California, Irvine, where he founded the Department of Earth System Science.

Clear vision

The 62-year-old Cicerone has been involved with the national academies for years, and has served on 20 of their study committees. In 2001, he chaired an NAS panel on climate change. The resulting report, *Climate Change Science: An Analysis of Some Key Questions*, was widely praised as straightforward and easy to understand. It was also phenomenally fast. Academy reports normally take one or two years to complete; Cicerone got the job done in a month.

The report helped to put Cicerone's name near the top of the list to succeed Bruce Alberts as NAS president, says Peter Raven, director of the Missouri Botanical Garden in St Louis and chair of the nomination committee that put Cicerone's name forward. Cicerone had extensive Washington experience, is a skilled fundraiser, and knew the NAS so well "he could hit the ground running", says Raven.

Cicerone can expect to deal with issues ranging from stem cells to nuclear power — two topics he mentions as possible subjects for the relatively rare self-funded studies undertaken by the academies. Cicerone's fundraising skills may help pump up the $332-million NAS endowment and allow for a few more homegrown projects, his supporters say.

He will also have to balance government interests against scientific openness. Earlier this year, the NAS was caught in a dispute over a paper in its *Proceedings of the National Academy of Sciences*. The paper presented a model of a bioterrorist attack using botulinum toxin, and officials at the Department of Health and Human Services asked journal editors to pull it, citing security concerns. Alberts, Cicerone's predecessor, delayed the paper, but published it in the end. Cicerone was involved in the decision, and plans to ask the newly formed National Science Advisory Board for Biosecurity for opinions on what could have been done.

M. TEMCHINE

Ralph Cicerone, new president of the National Academy of Sciences, wants to reach out to the public.

adopt, but the only ones he is eager to talk about at the moment are politically acceptable topics such as energy efficiency. He argues that some simple fixes, such as making sure manufacturers use energy-efficient lighting and motors, aren't made because they are swamped by the charged atmosphere surrounding climate change. "I find it very frustrating," he says, "that because of the really difficult questions that cause political polarization, we are missing some of the cherry-pickings."

Sherwood Rowland, a climate scientist and longtime colleague of Cicerone's at Irvine, expects him to speak out strongly in his new role. "He spent almost his whole career in areas of science that have very strong policy implications, and he has always been upfront about the policies that ought to go along with that," he says.

Public figure

Rowland also mentions one of Cicerone's skills that is crucial for successful political manoeuvring: "Ralph has a phenomenal memory for people — who they are, what they do, what their strengths are, and what to look out for."

But any amount of political savvy on Cicerone's part won't bridge the gulf between when people want scientific answers and when the NRC can provide them, or between the massive academy reports and the one-pagers that make up the informational diet of most lawmakers. "It's a perpetual, continual source of tension here," says Cicerone. As one possible solution, he envisions making the reports' executive summaries shorter, and releasing interim reports more often. But he also shifts some of the onus onto those who request the reports, urging them to "ask questions that won't go away this budget cycle. Ask questions that have longer-lasting value."

Such old-fashioned rationalism drives Cicerone. When creating study committees, he prefers to look for unbiased thinkers rather than balancing an advocate from one side with an advocate from the other. He is aware that some say this is an impossible task. "Many people say that's hopelessly naive: that science is not objective, everything is relative, everything stems from an individual's philosophy of life, ideology and so forth. I don't believe that."

When he's not working for the benefit of scientists, Cicerone plans to set up a long-term project to improve the public understanding of, and enthusiasm for, science. He feels that interest in science is beginning to decline, partly because good popularizers are rare. "I miss Carl Sagan," he says wistfully.

But he is not willing to go along with the often-repeated tale that Sagan was denied NAS membership because of his fondness for appearing on television and writing popular books. During the arguments over whether or not to elect Sagan "there were good people on both sides of the debate", says Cicerone. Spoken like a true Washington insider. ■

Emma Marris is a Washington correspondent for *Nature*.

But how feisty on political matters is the soft-spoken Cicerone likely to be? "I'll probably be a bit more conservative than people want," he says. "I don't want to be part of an organization that just shoots off its mouth with opinions that are not as well justified as can be."

For instance, he is prepared to weigh in on arguments over whether science is being misused for political purposes. "I don't want to shy away from those disputes," he says, "because I have such a reverence for science. I don't want to see it twisted or distorted." But at the same time, he is not ready to say that there is an ongoing pattern of politicization of science that would justify putting a panel together to study the issue. Instead, he thinks the best strategy is to keep the NRC reports flowing, so that if someone tries to mischaracterize science, a study can be slipped off the shelf and presented for a quick refutation.

Cicerone does have ideas about specific science policies that the United States should

Unlocking the drains

After centuries of playing second fiddle to the blood system, our lymphatic circulation is coming into its own as a key player in diseases ranging from cancer to asthma. **Phyllida Brown** reports.

Sometimes it's hard to be a scientist. Just five years ago, Mihaela Skobe, a molecular biologist now at Mount Sinai School of Medicine in New York, and her team were struggling to publish their work on the lymphatic system. Editors and reviewers thought it boring. The polite but yawning rejection letters came. "They would say that everything in the paper was fine but that lymphatics were unimportant," says Skobe. "There was a complete lack of interest."

Not any more. Once dismissed as a mere drainage network, the body's 'second circulation' system (see Graphic, right) is emerging as a crucial player in numerous diseases from cancer to asthma, and as a vital part of the normal immune system. As a result of these discoveries, researchers are trying to intervene in its activities, for example to reduce the spread of tumours, to boost the efficiency of vaccines, or to treat the painful and disfiguring swelling known as lymphoedema.

"It's a hot field now," says Michael Detmar, a dermatologist and lymphatic system researcher at Harvard Medical School in Boston, Massachusetts. These days, he says, many laboratories are switching from the study of blood vessels to the lymphatic system. Detmar is one of several researchers showing how important the lymphatic system is for tumour spread. Other groups are finding it plays a key role in inflammatory diseases such as asthma or transplant rejection. "We are in the pioneering phase," says Detmar. "There is a feeling that there is still so much to discover."

Mystery vessels

It might seem strange that one of the body's major organs is *terra incognita*. Lymphatic vessels may be drains, but they are pretty sophisticated ones. Not only do they collect fluid that has leaked into tissues from the bloodstream and return it to the blood; they also process that fluid,

sending it through lymph nodes where key cells of the immune system, called dendritic cells, present fragments of foreign molecules to other cells to mount an immune response.

But the fact is that the lymphatic system has played second fiddle to the blood system for centuries. Only properly discovered in the 1600s (see box, overleaf), it faded into obscurity and, apart from a brief flurry of interest around 1900, was largely neglected until about ten years ago.

The turning point came in the early 1990s when Kari Alitalo, a cancer researcher at the University of Helsinki, began studying a family of proteins involved in generating new blood vessels. These proteins, called vascular endothelial growth factors, or VEGFs, stimulate the growth of cells lining blood vessels and enable new vessels to sprout. Tumours often subvert these signals to build a blood supply that nourishes their invading mass. Like other scientists, Alitalo reasoned that if these signals, or the receptor proteins

"There is now little doubt that the interplay between tumours and the lymphatic system is the main route used by solid cancers to spread."

C. DARKIN

that enable cells to receive them were blocked, blood vessels could be prevented from growing, and tumours could be starved to death.

But what Alitalo and his team discovered next was to lead their focus away from blood-vessel growth signals and into the backwater of the lymphatic system. He and his team happened upon a new VEGF receptor that was present mainly on the cells that line the insides of lymphatic vessels[1]. The receptor was similar to a known receptor for VEGFs but neither of the two members of this family, VEGF-A or VEGF-B, activated it. So the hunt for the signal was on.

High hopes

When they eventually found this signal, they discovered that it was an endothelial growth factor similar to the known VEGFs. They named it VEGF-C and the receptor VEGF receptor 3 (ref. 2). In mice that had been genetically engineered to express excessive amounts of VEGF-C, the lymphatic vessels proliferated but — crucially — the blood vessels did not. This was the first signal known to act specifically on the lymphatic system[3]. Meanwhile, on the other side of the world, Marc Achen and Steven Stacker at the Ludwig Institute for Cancer Research, Melbourne, were hunting for further VEGFs. Faxes flew from Melbourne to Helsinki and back. Together, the researchers soon identified another signal that also acted on VEGF receptor 3. This one was dubbed, not surprisingly, VEGF-D (ref. 4).

"It was very exciting," says Alitalo. He and the Australians hoped that their discoveries could eventually lead to treatments to help build new lymphatic vessels in people suffering from lymphoedema — for example after breast cancer surgery.

But progress was hampered by the fact that there were no 'markers' for the lymphatic system. These are proteins characteristic of the tissue being studied that scientists use to specifically target dyes, and so to see the tissue more easily. It was not until 1999, when David Jackson, a biochemist at the Weatherall Institute of Molecular Medicine in Oxford, discovered
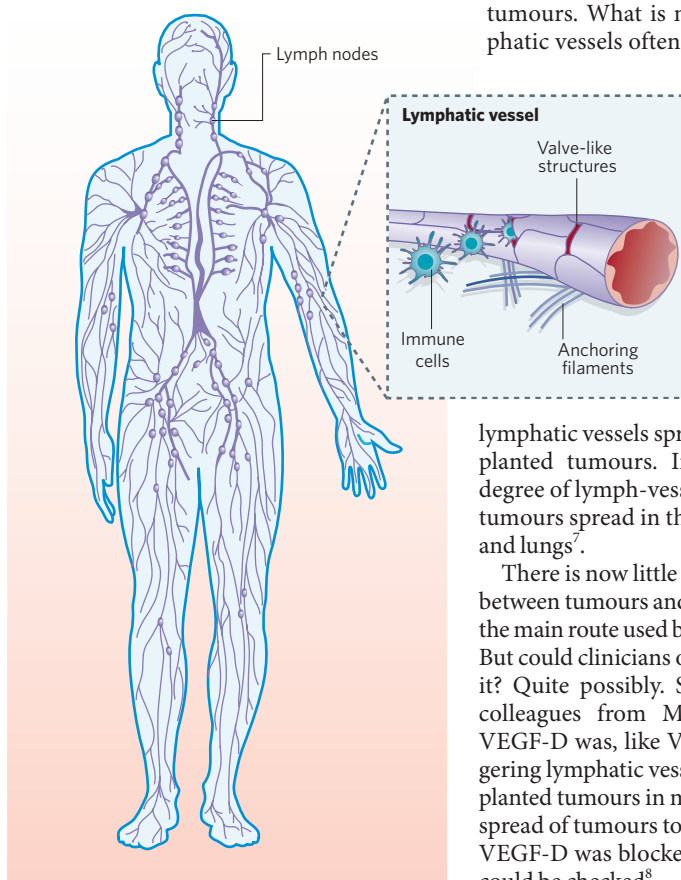
The lymphatic system's vessels have valve-like structures that let immune cells pass in and out.

such a marker, a protein called LYVE-1 (ref. 5), that lymphatic research really went into orbit.

For one thing, researchers could now probe the role of the lymphatic system in the spread of cancerous tumours. Many tumours, such as breast cancer and melanoma, spread from the original tumour via the lymphatic system to other organs, and a person's prognosis is worse if lymph nodes are involved. Yet, although the role of blood vessels in tumour spread had been well studied, researchers knew almost nothing about whether tumour cells actively persuaded lymphatic vessels to grow and assist their spread or whether lymphatics were just passive ducts.

Alitalo's group, together with Michael Pepper at the University of Geneva Medical Centre, and Gerhard Christofori at the Research Institute of Molecular Pathology in Vienna, created genetically engineered mice that developed tumours in the pancreas and had abnormally high levels of VEGF-C in the same organ. The team found lymphatic vessels sprouting in the animals'

tumours. What is more, the animals' lymphatic vessels often contained tumour cells that had originated in the pancreas[6].

Other teams also found evidence pointing the finger at the lymphatic system as an active agent in spreading tumours. Detmar and Skobe transplanted human breast cancers, engineered to make a lot of VEGF-C, into mice, and found lymphatic vessels sprouting within the transplanted tumours. Indeed, the greater the degree of lymph-vessel growth, the more the tumours spread in the animals' lymph nodes and lungs[7].

There is now little doubt that this interplay between tumours and the lymphatic system is the main route used by solid cancers to spread. But could clinicians one day intervene to stop it? Quite possibly. Stacker and Achen and colleagues from Melbourne showed that VEGF-D was, like VEGF-C, capable of triggering lymphatic vessels to grow inside transplanted tumours in mice. It also increased the spread of tumours to lymph nodes. But when VEGF-D was blocked, this increased spread could be checked[8].

Alitalo's group meanwhile, showed that mice genetically engineered to have VEGF-C and VEGF-D signalling blocked could not make new lymphatic vessels; their existing vessels even shrank[9]. The team then blocked VEGF-C and VEGF-D signalling in mice with human-breast-tumour transplants, and found it could reduce the amount of tumour spread by two-thirds[10]. The implications were clear: if VEGF-C and VEGF-D could be blocked in a mouse, then perhaps they could also be blocked in people with cancer to help prevent tumour spread.

Mice to men

Evidence shows that tumours in people behave like those in mice. Across a range of different human tumours, those that contain high levels of VEGF-C or VEGF-D are more likely to spread[11]. As a general rule, the more lymphatic vessels there are in the tumour, the greater the risk. Stacker and Achen are hopeful that trials testing agents that block VEGF-C and VEGF-D can begin soon. After all, points out Achen, an antibody called bevacizumab, or Avastin, that slows the growth of blood vessels in spreading colon cancer by blocking VEGF-A, has already prolonged some patients' lives.

But Alitalo warns that such trials may be a long time coming. One problem, he says, is that tumour spread can be a

A brief history of our second circulation

Despite its erstwhile obscurity, the existence of the lymphatic system has been known for millennia. It was first described by Hippocrates, who saw vessels containing "white blood" around 400 BC. Then, in the 1600s, the Italian physician Gasparo Aselli identified lymphatic vessels as "milky veins" in the gut of a "well-fed" dog.

In 1628, William Harvey published his detailed description of the blood circulatory system in animals, and from then on the blood circulation largely overshadowed any understanding of the lymphatic system. At the start of the twentieth century, anatomists injected ink into pigs and began to explore the development of the lymphatic network, but questions raised by their studies waited another century to be answered.

Research into the role of the lymphatic system in various immune responses also languished in obscurity for decades. Back in 1937, Howard Florey, a pathologist at the University of Oxford, UK, better known for his Nobel prizewinning contribution to the development of penicillin, found that lymphatic vessels proliferate in inflammatory conditions[15]. Only now are most immunologists beginning to ask how.

The lymphatic system seems to have evolved around the time that aquatic organisms moved on to land, says Kari Alitalo of the University of Helsinki. Amphibia have lymphatic vessels, whereas in those fish that have been studied, there is no well-developed lymphatic system. Alitalo suggests that the vessels allowed organisms to maintain fluid balance on land. In tadpoles of the species *Xenopus laevis*, the lymphatic vessels can be easily seen and share key genes that govern human lymphatic growth factors. These features, together with the ease of reproducing large numbers of tadpoles, make *X. laevis* a useful model 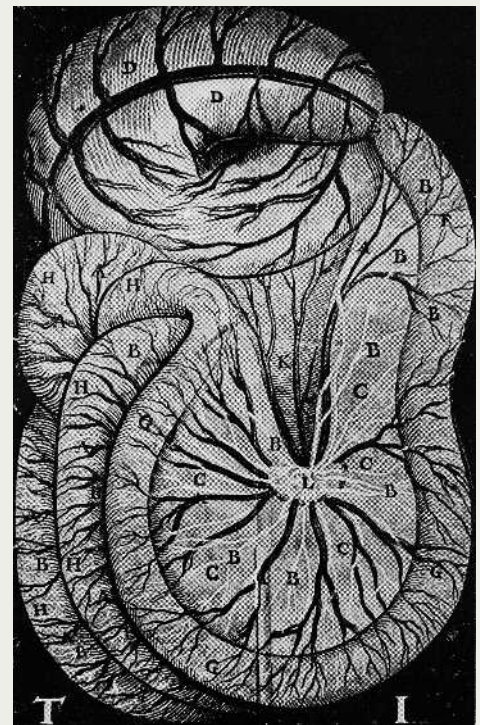for research, according to Peter Carmeliet at the University of Leuven in Belgium. Carmeliet reported his work on the tadpole at the Fifth European School of Haematology Euroconference on Angiogenesis, at Sitges, Spain, in May.

A 1627 drawing by Gasparo Aselli, who first noted lymphatic vessels as 'milky veins' in dogs.

slow process, and pharmaceutical companies are wary of embarking on costly trials that take as long as three years to reveal results.

As if all this were not news enough, it seems that the lymphatic system has been hiding an even bigger surprise up its sleeve. Over the past year, several research teams have begun to uncover evidence that the lymphatic system could be a major culprit in unwanted inflammation. Inflammation plays a key role in asthma, which affects an estimated five million people in Britain alone. It is also associated with other common conditions including psoriasis and rheumatoid arthritis, and with some medical problems such as transplant rejection.

Dontscho Kerjaschki, a pathologist at the Medical University of Vienna, has been studying what happens when kidney transplants are rejected. Normally, the kidney cortex, the part that filters blood, has hardly any lymphatic vessels. But in about a third of biopsies from transplanted kidneys, Kerjaschki found a 50-fold increase in the number of lymphatic vessels[12]. In most cases, such biopsies came from patients with chronic rejection, a condition in which the graft continues to be attacked by the host's immune system after initial immunosuppression treatment, until the transplant breaks down. Kerjaschki thinks the unusually extensive lymphatics could be involved, speculating that they may bring a continual supply of immune cells into the graft. "Maybe the lymphatics are organizing the whole catastrophe," he says.

Hints that the lymphatics can mastermind an immune response come from the discovery of a protein called podoplanin by Kerjaschki's team. Podoplanin sticks to a signalling molecule called CCL21, which is found mainly in lymphatic vessels. CCL21 is a powerful attractant to immune cells such as dendritic cells and macrophages[13]. In kidney grafts, where some inflammatory cells are already present, the complex of CCL21 and podoplanin breaks down, releasing CCL21 into the vessel, and so attracting further inflammatory cells.

Key controllers

Another piece of the jigsaw implicating lymphatic vessels in harmful immune responses comes from Donald McDonald, a vascular biologist at the University of California, San Francisco, and his colleagues. They have studied a mouse model of asthma. The animals' lungs are chronically infected with *Mycoplasma pulmonis*, a bacterial infection that produces swelling of the mucous membranes, alterations to vessel linings and scarring — all symptoms that resemble those of human asthma. McDonald's group found that infected mice grew additional lymphatic vessels in their tracheas, and their airway blood vessels also proliferated. When the team treated the mice with antibiotics, the blood vessels shrank but the extra lymphatic vessels persisted[14].

"This was a surprise," says McDonald. He speculates that the new lymphatic vessels help to set the lung up for more rapid and accentuated immune responses to subsequent infections, exacerbating inflammation. His team is now looking at the signalling molecules involved, hoping that it may eventually be possible to manipulate the inflammatory immune responses to help control asthma.

Back in Helsinki, where it all began, Alitalo is upbeat. He is happy that there are still fundamental questions to answer, including how cancer cells move beyond the lymph nodes as they travel through the vessels to distant organs, and how exactly tumour cells enter lymphatic vessels in the first place. And Skobe? She and her team are getting to grips with the molecules in lymphatic vessels that are involved in tumour spread. Skobe is not giving away details of her latest work yet. But when she does have new results to report, it seems unlikely, this time around, that the journal editors will need convincing. ■

Phyllida Brown is a science writer based in Exeter.

1. Kaipainen, A. *et al. J. Exp. Med.* **178,** 2077–2088 (1993).
2. Joukov, V. *et al. EMBO J.* **15,** 290–298 (1996).
3. Jeltsch, M. *et al. Science* **276,** 1423–1425 (1997).
4. Achen, M. G. *et al. Proc. Natl Acad. Sci. USA* **95,** 548–553 (1998).
5. Banerji, S. *et al. J. Cell Biol.* **22,** 789–801 (1999).
6. Mandriota, S. J. *et al. EMBO J.* **20,** 672–682 (2001).
7. Skobe, M. *et al. Nature Med.* **7,** 192–198 (2001).
8. Stacker, S. *et al. Nature Med.* **7,** 186–191 (2001).
9. Mäkinen, T. *et al. Nature Med.* **7,** 199–205 (2001).
10. Karpanen, T. *et al. Cancer Res.* **61,** 1786–1790 (2001).
11. Achen, M. G., McColl, B. K. & Stacker, S. A. *Cancer Cell* **7,** 121–127 (2005).
12. Kerjaschki, D. *et al. J. Am. Soc. Nephrol.* **15,** 603–612 (2004).
13. Yoneyama, H., Matusuno, K. & Matsushimaa, K. *Int. J. Haematol.* **81,** 204–207 (2005).
14. Baluk, P. *et al. J. Clin. Invest.* **115,** 247–257 (2005).
15. Pullinger, B. D. & Florey, H. W. *J. Pathol. Bact.* **45,** 157–170 (1937).

LETTERS

Temporal targeting of tumour cells and neovasculature with a nanoscale delivery system

Shiladitya Sengupta[1]*, David Eavarone[1]*, Ishan Capila[1], Ganlin Zhao[1], Nicki Watson[3], Tanyel Kiziltepe[2] & Ram Sasisekharan[1]

In the continuing search for effective treatments for cancer, the emerging model is the combination of traditional chemotherapy with anti-angiogenesis agents[1] that inhibit blood vessel growth. However, the implementation of this strategy has faced two major obstacles. First, the long-term shutdown of tumour blood vessels by the anti-angiogenesis agent can prevent the tumour from receiving a therapeutic concentration of the chemotherapy agent. Second, inhibiting blood supply drives the intra-tumoural accumulation of hypoxia-inducible factor-1α (HIF1-α); overexpression of HIF1-α is correlated with increased tumour invasiveness and resistance to chemotherapy[2–5]. Here we report the disease-driven engineering of a drug delivery system, a 'nanocell', which overcomes these barriers unique to solid tumours. The nanocell comprises a nuclear nanoparticle within an extranuclear pegylated-lipid envelope, and is preferentially taken up by the tumour. The nanocell enables a temporal release of two drugs: the outer envelope first releases an anti-angiogenesis agent, causing a vascular shutdown; the inner nanoparticle, which is trapped inside the tumour, then releases a chemotherapy agent. This focal release within a tumour results in improved therapeutic index with reduced toxicity. The technology can be extended to additional agents, so as to target multiple signalling pathways or distinct tumour compartments, enabling the model of an 'integrative' approach in cancer therapy.

Although the inhibition of angiogenesis is an elegant concept in cancer therapy[1], there are concerns about its implementation. Anti-angiogenesis therapy-induced tumoural hypoxia upregulates genes[2] that can significantly reduce the pro-apoptotic effect of chemotherapy[3] leading to 'reactive resistance'[4,5]. Furthermore, the activation of the hypoxic response can enhance the metastatic and invasive potential of tumour cells[6,7], although this has not been resolved clinically[8]. We reasoned that the optimal strategy to overcome these limitations is to sequentially expose the tumour to cytotoxic agents after a vascular shutdown induced by anti-angiogenesis therapy. However, two issues complicate this approach. First, the suppression of tumour growth requires long-term administration of angiogenesis inhibitors, whereas chemotherapy is usually given in short treatment cycles[9]. Second, the chronic administration of an anti-angiogenesis agent impairs blood flow inside the tumour microenvironment, precluding the intra-tumoural build-up of a therapeutically effective concentration of the chemotherapeutic agent.

To address this challenge we designed a novel delivery system, termed a nanocell, comprising a nanoscale pegylated-phospholipid block-copolymer envelope coating a nuclear nanoparticle. A chemotherapeutic agent is conjugated to the nanoparticle and an anti-angiogenesis agent is trapped within the lipid envelope. We proposed
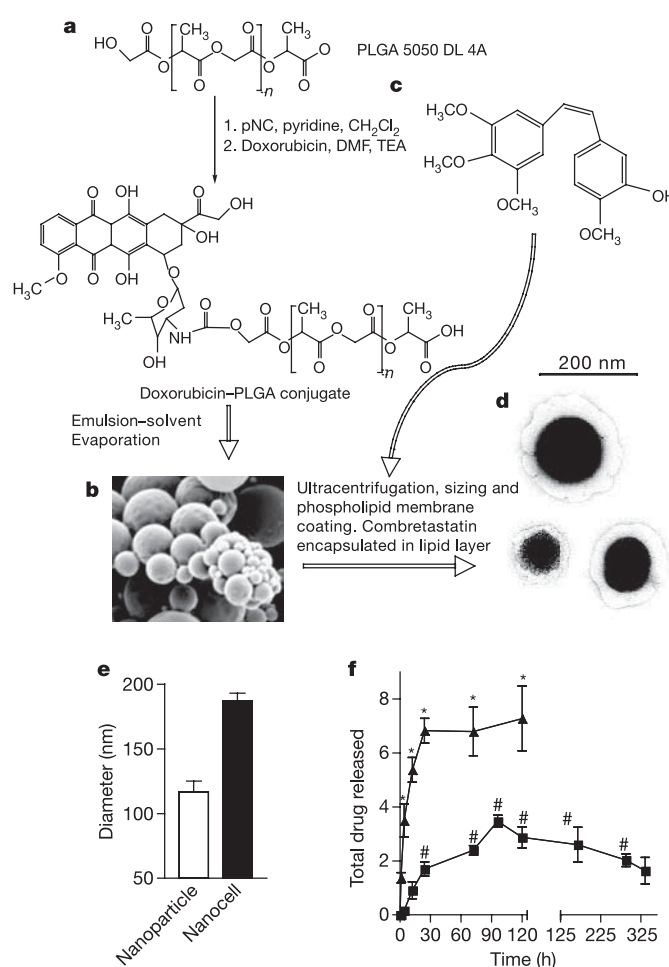
Figure 1 | Synthesis and characterization of a combretastatin–doxorubicin nanocell. a, Diagram of conjugation reactions between doxorubicin and PLGA 5050. DMF, dimethylformamide; pNC, p-nitrophenyl-chloroformate; TEA, triethylamine. **b**, Scanning electron micrograph of heterogeneous nanoparticles. **c**, Combretastatin is encapsulated in the lipid envelope. **d**, Transmission electron micrograph of the cross-section of three nanocells shows the dark nuclear nanoparticle within the phospholipid block-copolymer envelope. **e**, Dynamic light scattering shows that nanoparticles of defined sizes were used for encapsulation within the envelope. **f**, Physicochemical release kinetics shows the temporal release of combretastatin (triangles, scale in $10^2\,\mu g$) and doxorubicin (squares, scale in μg). Results are means ± s.e.m. ($n = 4$). The error is small where not visible. Asterisk, $P < 0.002$; hash, $P < 0.001$.

[1]Biological Engineering Division and [2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA.
*These authors contributed equally to this work.

that the disruption of this envelope inside a tumour would result in a rapid deployment of the anti-angiogenesis agent, leading to vascular collapse and the intra-tumoural trapping of the nanoparticles. The subsequent slow release of the cytotoxic agent from the nanoparticle should then kill the tumour cells.

As a proof of concept, we selected doxorubicin and combretastatin-A4 as the cytotoxic agent and the anti-angiogenesis drug, respectively. Doxorubicin induces apoptosis by intercalating with the DNA, causing scission[10]. Combretastatin causes a rapid vascular shutdown inside a tumour by disrupting the cytoskeletal structures[11].

The nanoparticles were fabricated from the biodegradable and nonbioactive copolymer poly-(lactic-co-glycolic) acid (PLGA)[12]. Doxorubicin was conjugated to PLGA (Fig. 1a) to achieve a slow release profile, distinct from the characteristic 'burst' release associated with nanoparticles. *In vitro* cell culture studies revealed a shift in the concentration–effect curve of the doxorubicin–PLGA conjugate, for which the concentration giving half-maximal response (EC_{50}) was about $10.59\,\mu g$, compared with free doxorubicin ($EC_{50} \approx 5.93\,\mu g$; Fig. 2c). This shift in the concentration–effect curve indicates that the doxorubicin–PLGA oligomers are inactive and have to degrade further into bioactive doxorubicin–PLGA fragments and free drug, which is consistent with earlier reports[13]. Although there was heterogeneity in the sizes of nanoparticles synthesized (Fig. 1b), we isolated a homogeneous population of nanoparticles between 80–120 nm (Fig. 1e).

To synthesize the nanocell, these nanoparticles were nucleated

inside a nanoscale phospholipid block-copolymer envelope composed of 2,000-Da poly-(ethylene glycol) distearoylphosphatidyl-ethanolamine (PEG-DSPE), phosphatidylcholine and cholesterol, in an optimal ratio with combretastatin. The selection of combretastatin, a lipophilic *cis*-stilbene (Fig. 1c), allowed optimal loading by partitioning into the lipid bilayer. Transmission electron microscopy showed the ultrastructure as being similar to a biological cell—a
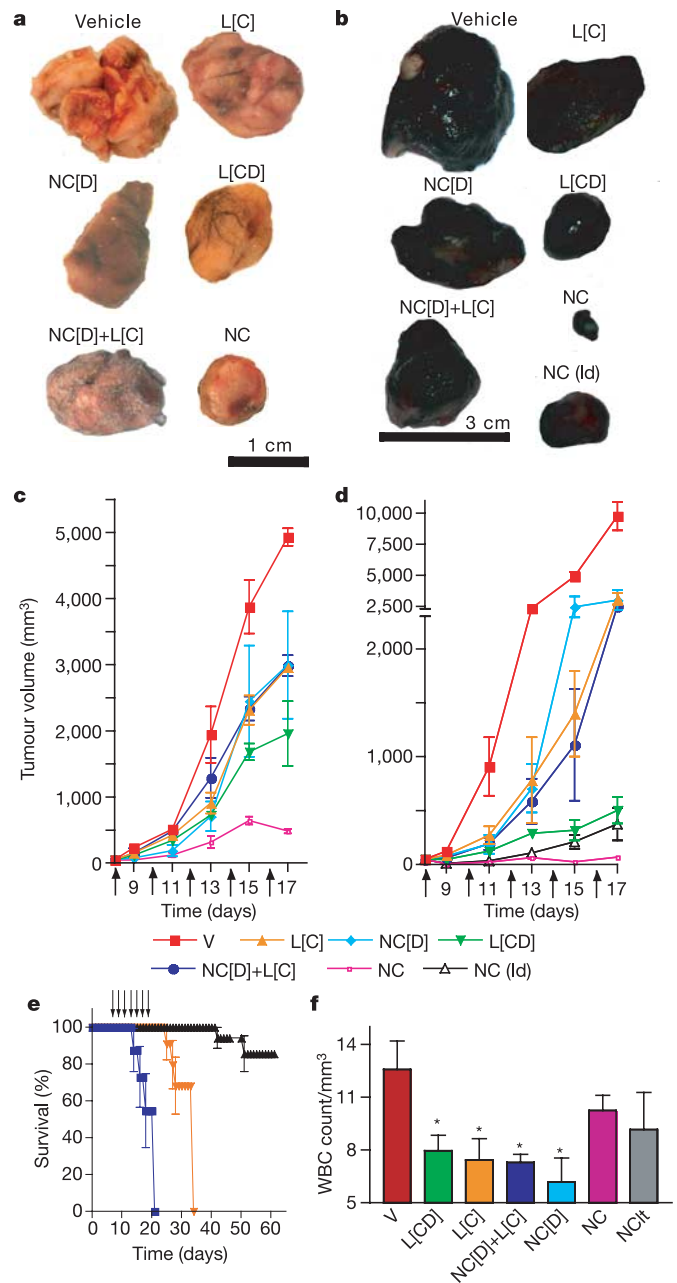
Figure 3 | **Nanocell therapy inhibits B16/F10 melanoma and Lewis lung carcinoma growth. a, b,** Excised Lewis lung carcinoma (**a**) and B16/F10 melanoma (**b**) comparing the effects of NC and the effects of nanocells with NC[D], L[C], co-injection of NC[D] + L[C], L[CD] and a lower dose (ld) of NC. Control groups received saline. **c, d,** Tumour volume in different treatment groups for Lewis lung carcinoma (**c**) and B16/F10 carcinoma (**d**). Results are means ± s.e.m. V, vehicle. **e,** Kaplan–Meier survival graph showing that treatment with NC significantly increases the lifespan: blue squares, vehicle; black triangles, NC; orange triangles, L[CD]. **f,** The effect of different treatments on the white blood cell (WBC) counts. Results are means ± s.e.m. Error is small where hidden. Arrows represent drug administration. *$P < 0.05$ vs. vehicle (V).

Figure 2 | **Bioassay of the nanocell with a GFP-positive melanoma–endothelial cell three-dimensional co-culture system. a,** Micrographs showing the effect of treatments on melanoma (yellow) or endothelium (red). **b,** Stereological quantification of the result. Treatment with nanocells (NC) results in temporal ablation of vasculature followed by destruction of the tumour cells. Liposomal combretastatin ($250\,\mu g\,ml^{-1}$) (L[C]) and doxorubicin-conjugated nanoparticles (ND) ($20\,\mu g\,ml^{-1}$ doxorubicin) results in selective loss of vasculature and tumour respectively. Results are means ± s.e.m. from three independent experiments. V, vehicle. Three asterisks, $P < 0.001$ (analysis of variance with Bonferroni's post-hoc test). **c,** Concentration–effect curve of free doxorubicin (triangles) and PLGA-conjugated doxorubicin (squares) on B16/F10 cells. Results are means ± s.e.m. for two independent experiments.

nuclear core surrounded by a lighter lipid membrane—hence the metaphor 'nanocell' (Fig. 1d). The size of the nanocells ranged between 180 and 200 nm in diameter (Fig. 1e).

To evaluate the nanocell physicochemically, we quantified the release kinetics for each drug. We observed a rapid release of combretastatin, reaching significant levels within 12 h. In contrast, the doxorubicin–PLGA oligomer was found to degrade into smaller doxorubicin–PLGA fragments and free doxorubicin significantly more slowly (Fig. 1f), extending over 15 days. This slow release underlines the shift in the concentration–effect curves, the EC_{50} being calculated at a time before the complete breakdown of the drug–polymer conjugate.

To correlate the temporal-release kinetics with the pathophysiology, we harnessed a tumour-endothelium co-culture bioassay[14]. The co-cultures were exposed to different treatments for defined periods, after which they were evaluated by dual-fluorescence confocal microscopy. As shown in Fig. 2, prolonged (30 h), but not short (12 h), incubation of the co-culture with doxorubicin-conjugated nanoparticles, alone, resulted in a complete ablation of the tumour cells without affecting the endothelial cells. This was consistent with the results obtained with free doxorubicin[14]. In contrast, incubation with combretastatin-encapsulated liposomes resulted in a rapid collapse of the vascular network without affecting the tumour cells. In the nanocell-treated groups, collapse of the vasculature was evident as early as 12 h, at which time point the tumour cells remained unaffected. A complete ablation of the co-culture was achieved by 30 h, which is consistent with the kinetics profile of the drug release rate. Although this release profile may not mimic the *in vivo* state, the close similarity of *ex vitro* release profiles in hypoxic-

tumour cell lysate and PBS, together with the bioassay results, emphasizes that a similar release pattern over time would be functional in the acidic microenvironment of the tumour.

To validate the therapeutic efficacy of this treatment, we randomly sorted mice bearing established (green fluorescent protein (GFP)-positive) B16/F10 melanomas or Lewis lung carcinoma into six groups and treated each group with one of the following: PBS (control); nanocells containing doxorubicin nanoparticles but lacking combretastatin (NC[D]); combretastatin-encapsulated liposomes (L[C]); co-administration of NC[D] + L[C]; nanocells containing both doxorubicin and combretastatin (NC); and a simple liposomal system encapsulating both doxorubicin and combretastatin (L[CD]). A separate melanoma-bearing group received half the dose of NC. The mice injected with PBS formed large tumours by day 17 (the day after the last injection), and consequently were killed. The animals in the other groups were also killed at the same time point to evaluate the effect of the treatments on tumour pathology. In a separate experiment, another group of animals was given seven cycles of treatment with NC to evaluate the effect over a longer period.

As shown in Fig. 3, both the L[C] and NC[D] groups showed tumour inhibition compared with the PBS-treated group. The NC[D] + L[C] treatment group had only a negligible improvement over either of the treatments alone, which could arise from the preclusion of NC[D] after chronic administration of the anti-angiogenic agent. In contrast, L[CD] treatment induced a greater inhibition than NC[D] + L[C], possibly from the simultaneous release of free doxorubicin and combretastatin. A greater exposure to the cytotoxic agent at the acute phases of the therapy, resulting from the normalization of the tortuosity of tumour vessels[15], has
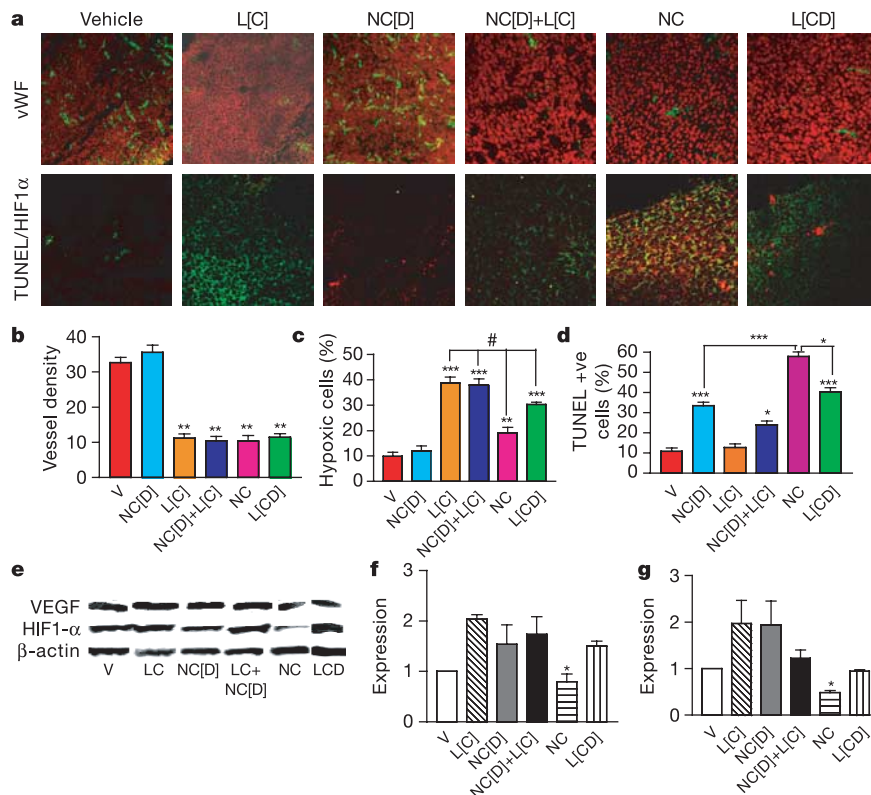
Figure 4 | Effect of nanocell treatment on tumour vasculature and apoptosis. a, Top: cross-section of tumours immunostained for vWF, an endothelial marker. Bottom: tumours TUNEL-labelled for apoptosis with the use of Texas red-labelled nucleotide, and co-stained with an anti-HIF1-α antibody labelled by FITC. **b–d**, Graphs showing tumour vessel density (**b**), hypoxia (**c**) and apoptosis (**d**). Results are means ± s.e.m. ($n = 3$). Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$; all compared with

controls (analysis of variance with Newman–Keul's post-hoc test). **e**, Western blots showing the effect of different treatments on the levels of HIF1-α and VEGF;. **f, g**, Graphs of levels of HIF1-α (**f**) and VEGF (**g**), normalized to β-actin; expression is shown relative to vehicle, normalized to 1. Results are means ± s.e.m. Asterisk, $P < 0.05$ compared with other combretastatin-treated groups.

been reported to underlie such an outcome. Another explanation could be that low-dose (that is, 'metronomic') chemotherapy can also exert an anti-angiogenetic effect[16]. However, the NC-treated groups, when compared directly with the equivalent doses of the NC[D], L[C], NC[D] + L[C] and L[CD] groups, had a distinctly superior outcome. NC treatment-induced inhibition of tumour growth was dose-dependent, although melanoma was more susceptible than lung carcinoma. Furthermore, there was a significant increase in the lifespan of the NC-treated animals (Fig. 3e). These observations bring out two issues. First, it is well known that tumour cell response to chemotherapy is a function of the chemotherapeutic agent. As observed here, the greater susceptibility of melanoma than the Lewis lung carcinoma to this current drug combination is consistent with this view. Second, the effectiveness of a chemotherapeutic agent is likely to be a function of signalling pathways and the responses to extracellular signalling from the heterotypic environment that impinge on the tumour cells. This underscores the need to develop chemotherapeutic approaches while keeping in perspective the different compartments of distinct tumours.

To dissect the mechanism of action, we immunostained tumour cross-sections for an angiogenesis marker, von Willebrand factor (vWF). As shown in Fig. 4a, c, the L[C] group induced a more than 50% loss in vasculature compared with PBS-treated controls. There was no difference in the vessel density between the groups treated with L[C], L[CD], NC and the NC[D] + L[C]. NC[D] had no effect on the tumour vasculature, indicating that 'metronomic' dose anti-angiogenesis was not the underlying cause for the enhanced efficacy observed with nanocells.

We next evaluated the tumours for apoptosis by using TdT-mediated dUTP nick end labelling (TUNEL) staining, staining the same sections simultaneously for HIF1-α expression. As shown in Fig. 4a, d, e, NC[D] treatment induced significant apoptosis in the tumour without any increase in the expression of HIF1-α. L[C] treatment did not induce apoptosis, nor did it potentiate the effect of NC[D] when both were injected together. A possible explanation for this observation could be attributed to the accumulation of HIF1-α, which can upregulate pro-survival growth factors[7]. In contrast, administration of NC significantly increased the percentage of TUNEL-positive cells in the tumour in comparison with NC[D] alone, and also disconnected the expressions of HIF1-α and vascular endothelial growth factor (VEGF) in comparison with other anti-angiogenesis-treated groups (Fig. 4e–g), possibly arising from a focal build-up of doxorubicin inside the tumour, leading to increased apoptosis.

Analysis of the white blood cell count, which is highly susceptible to cytotoxic agents, showed that treatment with NC resulted in the least systemic toxicity (Fig. 3f), consistent with the intra-tumourally restricted release of the cytotoxic agent arising from the vascular shut-down. To confirm the enhanced uptake into the tumour, we fabricated nanocells with fluorescein and monitored the level of the dye in the tumour and other highly vascular organs. The nanocells were detected within 5 h in all the vascularized tissues in equilibrium with the levels in the blood, and preferential accumulation of the nanocells in the tumours was evident at 24 h, with a concomitant fall in the serum level (Supplementary Fig. 2). This selective uptake into the tumour could arise from increased residence time in the circulation as a result of pegylation-induced reduction in immunogenicity[17]. Furthermore, the unique 'leakiness' of the tumour vessels, which, unlike physiological vasculature, have pores that are 400–600 nm in diameter[18], enhances the permeability and retention of the nanocells fabricated in this study. Additionally, the anti-angiogenesis agent-induced decrease in hydrostatic pressure gradient across the vascular wall could explain the increased accumulation at 24 h compared with earlier time points[19]. Although further studies are needed to dissect out these effects, the reduced toxicity and enhanced anti-tumour anti-metastatic effects (Supplementary Fig. 3) emphasizes the advantages of the mechanism-based design of the system.

Several components of the nanocell approach can facilitate future therapy in humans. First, the nuclear arrangement of cytotoxic agent-conjugated nanoparticle and the encapsulation of the anti-angiogenesis agent in the surrounding phospholipid block-copolymer envelope enables temporal targeting of the tumour vasculature, resulting in the intra-tumoural trapping of the nanoparticles. Second, the resultant slow release and focal build-up of the cytotoxic agent within the tumour allows a prolonged exposure and an increase in the apoptotic potential, which can overcome hypoxia-induced reactive resistance. Third, although we show here a selective tumour uptake, we recognize that these vehicles can be further specifically targeted to tumour vasculature by using probes that recognize specific molecular signatures on the vasculature[20]. Thus, the nanocell concept enables a significant advance in cancer therapy over current approaches. This platform technology supports the paradigm shift from a 'reductive' to an 'integrative' approach in cancer therapy[21].

METHODS

Synthesis and characterization of nanocells. Activated PLGA was dissolved in dimethylformamide and reacted with doxorubicin and triethylamine at room temperature (20 °C) under a nitrogen atmosphere (stoichiometric molar ratio of activated PLGA:doxorubicin:triethylamine, 1:1:4)[13]. Nanoparticles were formulated from the PLGA copolymer–doxorubicin conjugates by using an emulsion–solvent evaporation technique. For scanning electron microscopy, dehydrated nanoparticles were coated with gold on a carbon grid. They were analysed using a Jeol electron microscope (magnification × 3,700). Nanoparticle size fractions were recovered by ultracentrifugation at 10,000 g, 25,000 g and 50,000 g, which also removed free doxorubicin from the PLGA–doxorubicin conjugates. The smallest fraction was extruded through a membrane (pore size 100 nm) to obtain nanoparticles for encapsulation within nanocells. Phosphatidylcholine:cholesterol:PEG-DSPE (2:1:0.2 molar ratio) lipid membranes were prepared by dissolving the lipids in chloroform in a round-bottomed flask. Combretastatin A4 was added at a 0.9:1 drug:lipid molar ratio. To fabricate the nanocells, doxorubicin-conjugated nanoparticles were added to the aqueous lipid resuspension buffer, which was extruded through a membrane (pore size 200 nm). Sizing was performed by dynamic light scattering and transmission electron microscopy. For further details see Supplementary Fig. 4.

Cell culture studies. B16/F10 cells were plated at a concentration of 5×10^4 cells per well and left to grow for 24 h in 5% fetal calf serum. The cells were exposed to increasing concentrations of free or nanoparticle-conjugated doxorubicin and incubated for 48 h, after which the cells were counted with the Trypan blue exclusion method. EC_{50} was calculated by curve-fitting.

Release kinetics characterization. Drug-loaded nanocells were suspended in 1 ml of PBS buffer or hypoxic-cell lysate and sealed in a dialysis bag (molecular mass cut-off 10 kDa). The dialysis bag was incubated in 20 ml of PBS buffer at 37 °C with gentle shaking. Aliquots were extracted from the incubation medium at predetermined intervals, and released drug was quantified by reverse-phase HPLC with a C_{18} column, using a linear gradient of acetonitrile and water eluents. For further details see Supplementary Information.

In vitro tumour–endothelium co-culture studies. The co-culture of GFP-positive B16/F10 melanoma with human umbilical vein endothelial cells on a three-dimensional Matrigel matrix was exposed to different treatments. The cells were fixed with 4% paraformaldehyde, incubated with propidium iodide and analysed with a confocal microscope. Quantification was performed with a planimetric point-count method using a 224-intersection point square reticulum. Data were expressed as the ratio of each component to the total area covered by cells. For further details see Supplementary Information.

In vivo tumour studies. GFP-positive BL6/F10 melanoma cells (3×10^5) or Lewis lung carcinoma cells (2.5×10^5) were implanted in male C57/BL6 mice. Treatment was started when the tumours reached 50 mm³ in volume. Each formulation was prepared, quantified and diluted so that 100 μl (intravenous) was equivalent to 50 mg kg^{-1} combretastatin and 500 μg kg^{-1} doxorubicin. The animals were killed at defined time points, and the tumours were excised for histopathology. Simultaneously, 1 ml of blood was drawn through cardiac puncture and analysed for a toxicity profile of the treatment regimens.

Tissue distribution studies. Nanocells were fabricated with fluorescein dye and injected into tumour-bearing mice. The animals were killed at 5, 10 and 24 h after injection. Serum, tumour, liver, lungs and spleen were collected during necropsy, and fluorescein was extracted from these tissues with methanol, detected with a fluorescence plate reader, and normalized to the tissue weight. In another study, nanocells were fabricated with Quantum Dots in the core, and injected intravenously into tumour-bearing mice. The animals were killed at

different time points, and the highly vascular organs were extracted during necropsy. The tissue sections (30 μm thick) were immunostained to delineate the blood vessels. Confocal images were captured at 512 pixels × 512 pixels resolution, with excitation using a 488-nm laser line and emissions at the fluorescein isothiocyanate (FITC)/rhodamine wavelengths. Depth-coding was performed with LSM510 software.

Immunohistology. For analysis of tumour vasculature, permeabilized cryosections of the tumours were immunoprobed for vWF, an endothelial cell marker. For study of apoptosis, TUNEL staining was performed with Texas red-labelled nucleotide in accordance with the manufacturer's instructions (Roche). The same sections were also stained simultaneously for HIF1-α, using appropriate primary and FITC-labelled secondary antibodies. Images were captured with a Leica LSM510 confocal microscope at a 512 pixels × 512 pixels resolution. For quantifying metastatic nodes, the liver and lungs were harvested from the animals during necropsy, fixed in paraformaldehyde and embedded in paraffin. Thin sections were stained with haematoxylin/eosin and imaged with a Zeiss LSM510 confocal microscope with FITC/rhodamine filters. Metastatic nodes were quantified by counting the number of yellow foci in each section. For further details see Supplementary Information.

Western blotting. Tissue samples were lysed in sample buffer and resolved on a 4–12% gradient SDS–polyacrylamide-gel electrophoresis gel. The proteins were transferred to a nitrocellulose membrane, blocked, and probed with the appropriate primary antibodies and secondary horseradish peroxidase-labelled antibodies. Proteins were detected by chemiluminescence and quantified with a Kodak 2000 gel-imaging system.

1. Kerbel, R. S. & Kamen, B. A. The anti-angiogenic basis of metronomic chemotherapy. *Nature Rev. Cancer* **4**, 423–436 (2004).
2. Semenza, G. L. Surviving ischemia: adaptive responses mediated by hypoxia-inducible factor 1. *J. Clin. Invest.* **106**, 809–812 (2000).
3. Tran, J. *et al.* A role for survivin in chemoresistance of endothelial cells mediated by VEGF. *Proc. Natl Acad. Sci. USA* **99**, 4349–4354 (2002).
4. Yu, J. L., Rak, J. W., Coomber, B. L., Hicklin, D. J. & Kerbel, R. S. Effect of p53 status on tumour response to antiangiogenic therapy. *Science* **295**, 1526–1528 (2002).
5. Blagosklonny, M. V. Antiangiogenic therapy and tumour progression. *Cancer Cell* **5**, 13–17 (2004).
6. Pennacchietti, S. *et al.* Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell* **3**, 347–361 (2003).
7. Rofstad, E. K. *et al.* Hypoxia promotes lymph node metastasis in human melanoma xenografts by up-regulating the urokinase-type plasminogen activator receptor. *Cancer Res.* **62**, 1847–1853 (2002).
8. Kieran, M. W., Folkman, J. & Heymach, J. Angiogenesis inhibitors and hypoxia. *Nature Med.* **9**, 1104 (2003).
9. Kerbel, R. & Folkman, J. Clinical translation of angiogenesis inhibitors. *Nature Rev. Cancer* **2**, 727–739 (2002).
10. Chabner, B. A., *et al. Goodman and Gilman's The Pharmacological Basis of Therapeutics* 9th edn (McGraw-Hill, 1996).
11. Tozer, G. M., Kanthou, C., Parkins, C. S. & Hill, S. A. The biology of the combretastatins as tumour vascular targeting agents. *Int. J. Exp. Pathol.* **83**, 21–38 (2002).
12. Miklos, A. G., Lyman, M. D., Freed, L. E. & Langer, R. Wetting of poly(L-lactic acid) and poly(D-lactic-co-glycolic acid) foams for tissue culture. *Biomaterials* **15**, 55–58 (1994).
13. Yoo, H. S., Oh, J. E., Lee, K. H. & Park, T. G. Biodegradable nanoparticles containing doxorubicin-PLGA conjugate for sustained release. *Pharm. Res.* **16**, 1114–1118 (1999).
14. Sengupta, S., Kiziltepe, T. & Sasisekharan, R. A dual-colour fluorescence imaging-based system for the dissection of antiangiogenic and chemotherapeutic activity of molecules. *FASEB J.* **18**, 1565–1567 (2004).
15. Jain, R. K. Normalizing tumour vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. *Nature Med.* **7**, 987–989 (2001).
16. Hanahan, D., Bergers, G. & Bergsland, E. Less is more, regularly: metronomic dosing of cytotoxic drugs can target tumour angiogenesis in mice. *J. Clin. Invest.* **105**, 1045–1047 (2000).
17. Allen, T. M. Ligand-targeted therapeutics in anticancer therapy. *Nature Rev. Drug Discov.* **2**, 750–763 (2002).
18. Yuan, F. *et al.* Vascular permeability in a human tumour xenograft: molecular size dependence and cutoff size. *Cancer Res.* **55**, 3752–3756 (1995).
19. Tong, R. T. *et al.* Vascular normalization by vascular endothelial growth facor receptor-2 blockade induces a pressure gradient across the vasculature and improves drug penetration in tumors. *Cancer Res.* **64**, 3731–3736 (2004).
20. Arap, W. *et al.* Steps toward mapping the human vasculature by phage display. *Nature Med.* **8**, 121–127 (2002).
21. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

BUSINESS

Painkiller in the dock

Still reeling from the loss of its blockbuster pill Vioxx, Merck is now in court to face its first lawsuit over the drug. Meredith Wadman reports.

The attention of the global pharmaceutical industry is fixated this month on a courtroom in Texas. In the little town of Angleton, Merck is defending an allegation that its blockbuster painkiller Vioxx killed Robert Ernst, a 59-year-old personal trainer from nearby Keene.

The case is the first of some 4,000 Vioxx lawsuits filed against Merck to come to trial. The company, based in New Jersey, withdrew the painkiller from the market last September, after one of its clinical trials showed that prolonged use of the drug doubled the risk of heart attacks and strokes.

Ernst died of an arrhythmia in May 2001 after using Vioxx for eight months. Arrhythmias, or irregular heartbeats, range from relatively benign conditions that can be treated by medication to sudden, fatal disturbances of the kind that killed Ernst.

For Merck, "a lot depends on this first case", says Tony Butler, an analyst at Lehman Brothers in New York, who watches the company closely. If the plaintiffs win and secure a big monetary award, it could encourage a huge number of additional lawsuits. A victory for Merck could significantly reduce its financial liability over Vioxx, which analysts have estimated at anywhere between $4 billion and $20 billion. It is fear of this liability — even more than the lost annual revenue of some $2.5 billion — that caused Merck's shares to plunge so sharply when it recalled the drug (see Graph).

Under attack

Lawyers acting for Ernst's widow Carol allege that Merck, which had marketed Vioxx since 1999, knew early on about its dangerous cardiovascular effects and stifled that information in its quest for profit.

Their argument was bolstered early in the trial, on 19 July, when lead attorney Mark Lanier presented as evidence a letter that Merck sent to thousands of doctors during 2001. This said that, in the largest clinical trial of Vioxx, only 0.5% of about 4,000 subjects taking the drug had experienced "cardiovascular events" including heart attacks and strokes. In fact, that rate, as Merck reported it

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Mark Lanier (left) puts the plaintiff's case against Merck in Angleton.

to federal regulators, was 14.6%. Ernst's doctor received the letter in April, 2001, the month before Ernst died.

"You-all didn't tell the doctor the truth about the safety of Vioxx, did you?" Lanier asked Nancy Santanello, Merck's top epidemiologist, who is the company's key witness. "Perhaps it's not very precise," she conceded after prolonged questioning.

But Jonathan Skidmore, a defence attorney with the Washington firm of Fulbright & Jaworski, accused Lanier of clouding the central issue: whether Ernst's death from an arrhythmia was caused by Vioxx. "This is a case about causation and science," said Skidmore. "A sudden death is a tragedy in any family, but the evidence here will show that Vioxx did not cause Mr Ernst's death. There is no credible scientific evidence associating Vioxx with an increased risk of arrhythmia."

Merck's lawyers will also focus on timing.

MERCK'S SHARE PRICE

Ernst took Vioxx for eight months, but in the study that led to the drug's recall, the increased risk of heart attacks and strokes did not appear until patients had been taking the drug for 18 months.

Merck has contended from the outset that it followed Food and Drug Administration (FDA) safety regulations religiously as it developed Vioxx. "Merck acted responsibly," says Kenneth Frazier, the company's general counsel.

The trial, which began on 14 July, has been packed out by dozens of reporters, investors and lawyers involved in other pending Vioxx lawsuits. It is pitting the folksy but formidable Lanier against some of America's most seasoned corporate lawyers, from two powerhouse Washington law firms.

Their respective styles have contrasted vividly during the trial, which is expected to run until late August. Lanier calls the jurors by their first names and projects slides with pithy slogans, such as "Merck-y ethics", as he seeks to build a case that the company knowingly deceived some 20 million patients who took Vioxx. David Kiernan, the surgeon and lawyer who gave Merck's opening arguments on 14 July, read from notes and quoted long excerpts from FDA letters.

Biological actions

But whichever rhetorical method they use, the winner of this case will need to convince the jurors of one central fact: that Vioxx either did or did not contribute to Robert Ernst's death. That may come down to a detailed pathophysiological discussion of how arrhythmias can be provoked by heart attacks, and whether, despite Merck's public statements to the contrary, Vioxx can sometimes cause arrhythmias.

For Merck, which employs more than 62,000 people, including 8,000 researchers, across 150 countries, the stakes could hardly be higher. Wyeth, another drug firm based in New Jersey, has so far paid $21 billion in charges related to lawsuits filed by people who took its diet drug fenfluramine. The drug was pulled from the market in 1997 after being linked to heart-valve damage.

But a prominent Dallas lawyer, Kip Petroff, failed to win substantial settlements for fenfluramine plaintiffs until he took Wyeth to trial a few times in small courtrooms much like the one in Angleton. After that, recalls Joseph Goulden, a Washington writer who will deal with such cases in a forthcoming book, "settlements rained down like hail on a tin roof in an east Texas thunderstorm".
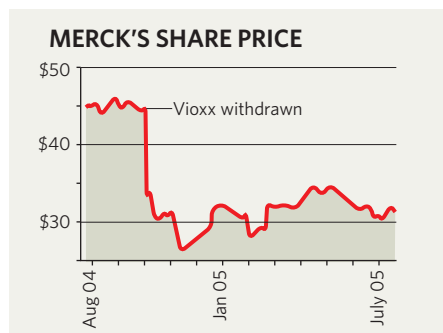
Sanctions should not apply to biomedical research

SIR — I was pleased to read your News Feature "The brains trust of Tehran" (*Nature* **435,** 264–265; 2005), which reminded me, and probably many young Iranian scientists raised after the revolution, of our enthusiasm for scientific research at university, despite access to only the most basic of equipment.

It is disappointing for me to see the effects that the sanctions imposed on Iran by the United States are having on basic biomedical research. In the example cited in your News Feature it took four years for the Iranian lab to secure equipment for neuronal recordings. I understand that the United States would not welcome scientific achievements in fields such as nuclear science or space technology for fear that they could be used to make weapons of mass destruction. But I believe other scientific fields, especially biomedical ones, should be exempt from these sanctions.

Local research on endemic diseases such as haemophilias, thalassaemias and oesophageal carcinoma may lead to new preventive and therapeutic insights beneficial to patients in Iran and worldwide; sanctions should not apply to this kind of work. I do not think it is difficult to differentiate such activities from those used to further geopolitical and governmental goals, and I would like to see the US government reconsider some of its scientific embargos against Iran.

Ahmad Jalili
Division of Immunology, Allergy and Infectious Diseases, Department of Dermatology, Medical University of Vienna, General Hospital–AKH, Währinger Gürtel 18-20, A-1090, Vienna, Austria

Bioterror killed five in US; guns kill 30,000 a year

SIR — I read with concern the call for biological scientists to be more careful in how they communicate their work ("Biologists asked to breed a culture of responsibility in face of terrorism" *Nature* **435,** 860; 2005). Many implements can be used by terrorists to harm people, by far the most common being military-style firearms. Every year in the United States, some 30,000 people are killed by firearms. The only lethal biological attack in the United States was carried out in 2001, using a supply of anthrax that came from a US source and killed five people.

In the United States, every call to restrict access to military-style firearms by civilians is met with fierce opposition from the National Rifle Association (NRA), who proclaim that the Second Amendment prevents any such regulation. When the number of people killed by biological agents gleaned from the scientific literature exceeds the number killed by firearms, perhaps then we should consider restricting that literature in the same way that firearms are restricted. Until then, we should be prepared to uphold the First Amendment and protect the freedom of speech.

To borrow a phrase from the NRA, when can you take away my scientific literature? When you can pry it from my cold dead hand!
David R. Whitlock
40 Nash Street, Watertown, Massachusetts 02472, USA

Ethics debate is what put Newcastle paper in the news

SIR — Your Editorial "Too much, too soon" (*Nature* **435,** 538; 2005) deplored the fact that the abstract for our manuscript submitted to *Reproductive BioMedicine (RBM) Online* became public, before being peer-reviewed, on the same day that a Korean group led by W. S. Hwang announced in a *Science* paper that it had succeeded in producing eleven embryonic stem-cell lines. We would like to clarify our position in this matter.

On 16 May, we submitted a manuscript to *RBM Online* that had previously been offered to both the *British Medical Journal* and *The Lancet*; both journals had declined it after editorial review, claiming insufficient clinical content. At the time of submission to *RBM Online*, we were not aware that the abstract would be made immediately available online. Learning of this the next day, we asked *RBM Online* to remove the abstract to prevent public awareness before publication of the *Science* study. One of us also met Professor Hwang to inform him of our actions and exchange manuscripts. Our paper has now been accepted for publication by *RBM Online* and will appear, in full, in August.

On 16 May, UK journalists also became aware of the forthcoming publication in *Science* and contacted us to ask about our progress with nuclear transfer. We did not comment on the Korean breakthrough in advance of its publication, but felt it natural and necessary to say that we had submitted our work and that it would soon become public. In our subsequent comments, we were very clear that the Koreans had made a major advance, when compared with our own achievement.

We believe the strong interest in our work was not driven primarily by a parochialism that affords prominence to local news, but by the ongoing ethical debate. It is the ethics of human nuclear transfer that keeps embryo research high on the agenda of the UK press and public, not the science. We cannot avoid the resulting close and continuous scrutiny, nor should we.

Alison Murdoch*, Miodrag Stojkovic†
*Newcastle Fertility Centre at Life, Times Square, Newcastle upon Tyne NE1 4EP, UK

†Centre for Stem Cell Biology and Developmental Genetics, University of Newcastle upon Tyne, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK

Misconduct: lower ranks take most of the blame

SIR — The Commentary article by Brian C. Martinson and colleagues, "Scientists behaving badly" (*Nature* **435,** 737–738; 2005), highlights the need for an appropriate response in order to preserve good science practices at the institutional level. Although formal guidelines exist for misconduct investigations in universities, there are no checks or balances, or even public scrutiny, to ensure that the institution behaves appropriately. Ideally, universities should hand over authority to enforce good science guidelines to an independent body of informed, non-partisan reviewers unaffiliated with the institution under investigation.

Graduate students, postdoctoral fellows and junior research associates are too-frequent casualties of self-policing by institutions investigating scientific-misconduct cases. They are often less culpable than their senior colleagues, but naive about the convoluted investigatory process. A December 2004 newsletter from the US Office of Research Integrity (ORI) documents this disparity, revealing that just 15% of full professors were found guilty of fabrication, falsification or plagiarism (FFP) between 1994 and 2003, whereas junior researchers were held accountable 61–77% of the time. Compounding this rank bias, university misconduct inquiries are rarely conducted openly, yet still comply with ORI guidelines. This secrecy makes it difficult to protect the rights of junior whistleblowers.

It is interesting to learn from the survey by Martinson and colleagues that senior investigators seem more ready to admit indiscretions such as data duplication or inaccurate authorship delegation than FFP. It would be interesting to ask scientists in a misconduct survey about their willingness to disclose minor indiscretions and correlate their responses with rank. Like the recent ORI report, this would probably reflect the relative protection that a senior researcher's status provides.
Shawn G. Clouthier
Comprehensive Cancer Center, University of Michigan, 1500 East Medical Center Drive, 6410 CGC, Ann Arbor, Michigan 48109-0942, USA

BOOKS & ARTS

Power and the people

A wide-ranging look at the future of the nuclear power industry.

Nuclear Renaissance: Technologies and Policies for the Future of Nuclear Power
by William Nuttall
Institute of Physics Publishing: 2004.
334 pp. £45, $65

Michael Golay

*Nuclear Renaissanc*e provides a welcome addition to the literature on nuclear power. It is an unusual book, covering the history of nuclear power but also reflecting on recent policy and technological developments — all presented in the clear voice of the author, William Nuttall. The book covers the broad scope of factors important for the future of nuclear power, while avoiding the pitfalls of either cheerleading or simplistic criticism, both of which have characterized much of the nuclear-power literature during recent decades.

Nuttall provides an accurate and concise survey of the important aspects of light- and heavy-water reactors, gas-cooled reactors and liquid metal-cooled reactors. Some of the other technologies he describes, such as actinide transmutation-focused accelerators and reactors, and the DUPIC concept — using spent fuel from light-water reactors in heavy-water reactors — are less likely to play a future role, but are nonetheless interesting. He correctly observes that reprocessing plutonium from spent fuel is unlikely to be economically justified or technologically important for several decades, as are reactor designs that depend on this practice.

Whereas most discussions of nuclear power focus on the hardware — as if it were used in a socioeconomic vacuum — *Nuclear Renaissance* covers the full scope of the nuclear-power scene, from the hardware to national energy policies, schemes for economic deregulation and competition from non-nuclear technologies, and the social factors affecting the acceptability and economics of this energy source.

Chapter 1 alone, which surveys the more important social and policy factors affecting nuclear power in the United States and Britain, is worth the price of the book. It also recognizes that the acceptance of nuclear power has been affected by non-technological factors, such as the cold war, and the failures and frustrations of the Vietnam war — opposition to nuclear power was an effective vehicle for opposition to the establishment that imposed the war on them. Accidents at nuclear power

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A model technology: in the 1950s, the future of nuclear power looked bright.

plants, which have tended to be viewed differently in sociopolitical terms over the years, are another factor. There is also a good discussion of the trends and causes of changes in UK nuclear policy.

Crucially, *Nuclear Renaissance* recognizes the importance of the role of public trust in providing conditions favourable to the success of nuclear power. Compared with most European states, this factor is in short supply in Britain and the United States, and this partly explains why the nuclear power industry is more successful in France and Finland.

However, despite the book's many merits, I have a few quibbles. The unforgiving nature of nuclear power when mistakes occur is not noted in the book, yet this is the primary reason that the technology is so difficult to use successfully. The socially and financially punishing nature of such accidents is also relevant here. These factors create high barriers to the entry of new concepts in nuclear technology, especially in the absence of abundant financial resources for coping with mistakes. Safety regulatory agencies effectively 'lock in'

the technologies that are already licensed, to the exclusion of new entrants. This factor is especially acute in the United States, where efforts to remedy this effective bias are proceeding too slowly to allow the adoption of new technologies.

Nuclear Renaissance provides enthusiastic support for the US Department of Energy's Generation IV programme, which aims to foster technologies for improving the sustainability, safety and economic viability of the next generation of nuclear power plants. However, Nuttall seems too ready to accept the claims and hopes of the initiative's proponents. He is also too kind to the developers of gas-cooled reactors in Britain, France, the United States and Germany, claiming that the failures of these reactors were due to their being ahead of the times, rather than faulty engineering and management.

The book's greatest omission is that it gives no attention to the goal of 'user friendliness' in nuclear technology — in other words, achieving ease of operation, low maintenance burdens and low corrosion rates. Yet these are

some of the greatest problems of the current generation of reactors.

The existence of the Institute of Nuclear Power Operations (INPO) in the United States, and the World Association of Nuclear Operators, which were created following the Three Mile Island and Chernobyl accidents respectively, are not acknowledged, much less discussed. This is despite the consensus that the INPO has been largely responsible for improvements in the performance of the nuclear power industry in the United States over the past 25 years.

The need for nuclear technologies capable of mitigating global warming — for example, with greater power capacity and fuel breeding — is also not recognized as an essential goal.

Technological subtleties such as uncertainties over future economic performance, the trade-offs between passive and active safety features and the role of human error are at the heart of debates over the best paths for technological development, but *Nuclear Renaissance* does not examine these issues in any depth.

It does discuss economic deregulation, but fails to explore the role of Britain's more compact transmission and distribution structure in the greater success of deregulation compared with the United States. Despite this advantage, nuclear power has been less successful in the United Kingdom than other forms of power generation. BNFL, Britain's nationalized nuclear-power company, currently survives on subsidies from the government. This example raises the question of whether nuclear power plants, with their need for cash even when not operating, are suitable for use in economically deregulated markets.

At the same time, *Nuclear Renaissance* is far too kind to the opponents of nuclear power, claiming that they have supported the interim storage of nuclear waste at monitored surface sites where terrestrial disposal — as at Yucca Mountain in the United States — is unavailable. This may be true in Britain, but in the United States, anti-nuclear pressure groups have insisted on perfect and immediate solutions to the problem of nuclear-waste disposal. The US nuclear industry has foolishly accepted this impossible challenge.

Finally, Nuttall asserts that greater public knowledge about nuclear power and debate of the issues will help to win acceptance, but I'm not convinced. He does not recognize the alternative hypothesis that the public's approval of nuclear power may depend more on the industry being the three Bs — beneficial, boring and banal — than on intellectual appeals. This hypothesis seems to explain how the public has come to accept other hazardous and previously feared technologies, such as lifts, steam engines and electricity. These opposing hypotheses for public acceptance are worth exploring: knowing which is closer to the truth may be at least as important as developing better hardware.

The book is essential reading for anyone who is interested in the relationship between energy, society and the environment. The author's observations are thought-provoking, and his knowledge of the current energy scene, particularly in Britain, is deep and subtle. So, despite the few warts noted here, this book is a valuable contribution to the debate about nuclear power. You should read it. ■

Michael Golay is professor of nuclear engineering at the Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

A planet tells its story

Reading the Rocks: The Autobiography of the Earth
by Marcia Bjornerud
Basic Books: 2005. 256 pp. $26, £15.50

Simon Lamb

On a recent flight back to Britain from the United States, the person sitting next to me, after discovering that I was a geologist, asked me if it was true that predictions about global warming were just bad science.

I thought of the scientists around the world who had devoted their careers to unravelling the way our planet works and the impact we humans have had on the environment. Are they bad scientists? And then, with a growing feeling of frustration, I realized how effective the smear campaign against environmental science has been, raising doubts in the minds of so many people, lulling them into a false sense of security just at the time when drastic and urgent action has to be taken if humanity is to save itself from the impending disaster.

We certainly need popular science books like *Reading the Rocks* by Marcia Bjornerud to help science fight back. The book reveals the extraordinary wealth of knowledge that Earth scientists now have about the biosphere and the lithosphere, and the complex interactions between them, that have made this planet the only home for us.

Bjornerud deals with the nature of rocks and the rock record, and what they can tell us about the origins of the Solar System and the subsequent evolution of the Earth's atmos-
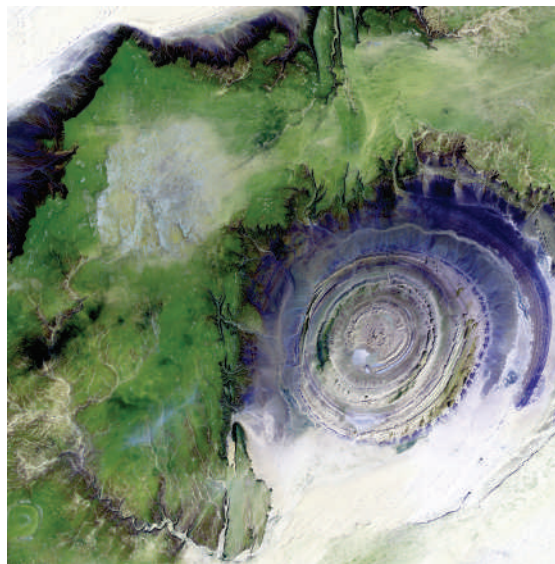
Discovering the past: this Richat structure, 50 kilometres across, provides valuable clues to Earth's history.

LANDSAT/NASA

phere, oceans and continents, as well as its deep interior. Inextricably linked to all this is the story of life itself. This is the context in which we must consider the effects of our actions on the planet.

The book is well written and Bjornerud has a rare talent for explaining scientific ideas clearly with intriguing and helpful analogies, similes and metaphors. One I rather liked was her riposte to Simon Conway Morris's contention that human intelligence is the inevitable endpoint of evolution. This, says Bjornerud, is a bit like declaring that the fork is the ultimate eating utensil without realizing that half the population of the world find chopsticks even better.

Yet despite my enthusiasm for both Bjornerud's aims and subject matter, I was left with a feeling that the book had missed a trick — not because the science was lacking, but because it did not inspire me as much as I felt it should.

I wish Bjornerud had arranged the facts and explanations with more of a sense of an unfolding story about the Earth, the science, and herself and her fellow geologists. When studying the planet's amazing story, the personal quests for understanding, with all the adventures and difficulties along the way, are an overwhelming experience. And these adventures and difficulties create a deeper understanding of the scientific endeavour itself, with its inherent uncertainties.

Nonetheless, Bjornerud tells the fascinating story of how our planet came into being, with its peculiar restless motion of tectonic plates and the extraordinary role of water, not only in this, but in maintaining the delicate environmental balance that allowed life to evolve over geological time. I only hope we don't go and muck it all up in the next blink of our planet's long history. The planet will survive to tell another story, but we might not. ■

Simon Lamb is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

Laying down the laws

When Physics Became King
by Iwan Rhys Morus
University of Chicago Press: 2005. 288 pp.
$60, £42 (hbk); $25, £17.50 (pbk)

Stephen G. Brush

When did physics become king? And what is its kingdom?

According to Iwan Rhys Morus, a lecturer in history at the University of Wales, it was during the nineteenth century that physics became the most powerful and prestigious science in the United Kingdom and Germany.

That does not necessarily mean that other sciences bowed down to physics — the only other sciences he mentions are chemistry and astronomy, both of which had been dominant in the early part of the century. Apparently, chemistry failed to keep up with physics, while astronomy became more and more dependent on physical instruments and theories.

This is a story about the ascendance of physics as viewed by the public and, more importantly, by government and industry. As Morus puts it, physics became "the ultimate authority in nature". It is a fascinating story, well told and mostly based on the latest research by professional historians of science. But the approach is "unashamedly cultural", or rather, social: high culture appears only in a chapter on romantic nature philosophy.

Morus stresses the role of institutions such as universities, journals and public exhibitions, alongside events that depended on the actions of individuals. Readers who don't have some background knowledge of physics and its history may get the impression that objective scientific factors did not have much influence on the success of physics, although that is not the author's position. While acknowledging the importance of mathematics in the development of theoretical physics, he provides few equations and no discussion of any mathematical concepts or techniques; this may please some readers but it fails to convey an understanding of what theoretical physicists such as Hermann von Helmholtz, James Clerk Maxwell and Augustin-Jean Fresnel were actually doing.

Why did physics rather than chemistry become the leading science by 1900? The author's answer is fourfold. First, physics was credited with establishing the law of energy conservation, the most important discovery in nineteenth-century physical science. This law governed all the physical and biological sciences, and was even believed to explain why women should not do scientific research. Some nineteenth-century scientists argued that "only so much energy could be contained in a woman's body. Its proper purpose was to be directed towards childbirth and nurturing."

Second, Morus believes that the mathematical aspects of physics "made laboratory work respectable for the sons of gentlemen", thereby overcoming class prejudice against the manual labour needed to do experimental science. Third, "its connexions with the world of telegraph cables, electric power, and factory engines made it a practical occupation for the sons of trade".

The final reason that Morus gives for the rise of physics is that it gained favourable notice from public demonstrations and exhibits, especially of electrical phenomena, and from applications that benefited industry, government and citizens. For those scientists and others fascinated by "mysterious fluids and forces" at the end of the nineteenth century, he says, the electromagnetic-wave experiments of Heinrich Hertz not only confirmed the remarkable prediction from Maxwell's equations of their existence and properties, but also "demonstrated graphically the physicists' power to manipulate nature, to make things happen as if by magic".

Going back up the causal chain, we learn that the University of Cambridge began to stress advanced mathematics because, at the beginning of the nineteenth century, that was the only subject to be tested by written examinations leading to public honours. Attaining such honours was the best way to enter a career in government service in the expanding British Empire. Success in the mathematical tripos examination was supposed to demonstrate the moral and mental characteristics needed for such service. Preparation for this very competitive examination provided excellent training for research in theoretical physics.

Then, around 1870, despite strong university opposition, a laboratory for experimental work was established. This was only possible because William Cavendish, Duke of Devonshire, who had himself attained high honours in mathematics, was chancellor of the university and paid for the laboratory himself. The Cavendish Laboratory got off to an excellent start with the efforts of Maxwell, its first director, and so the University of Cambridge became a world leader in both theoretical and experimental physics. A similar story, combining social and personal factors, can be told about the rise of physics in Germany.

It is impossible to cover all the important aspects of nineteenth-century physics in a text of fewer than 300 pages. However, a serious omission from *Why Physics Became King* is the establishment of the existence and quantitative properties of atoms. The contributions of Avogadro, Loschmidt, Perrin and others are ignored. Although in the nineteenth century the atom may have been more important to chemistry than to physics, it was thanks to physical theories (especially the kinetic theory of gases) that the speeds, masses and sizes of atoms could be determined. This transformed a philosophical speculation into a legitimate scientific concept and expanded the kingdom of physics into fertile new territory. ∎

Stephen G. Brush is professor of the history of science in the Department of History and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA.

EXHIBITION

Snow business

Wilson 'Snowflake' Bentley had a lifelong obsession with the intricate hexagonal symmetry of snowflakes. He produced more than 5,000 photo-micrographs, compiled in his 1931 landmark book *Snow Crystals*, and in other publications.

An unseasonal exhibition at the New York Academy of Sciences, 'One of a Kind', pairs Bentley's images with works on the same theme by contemporary artists, including Daniele Imperiale-Warner, whose egg tempera on parchment painting, *Untitled*, is shown here.

The exhibition runs until 31 August. **A.P.**
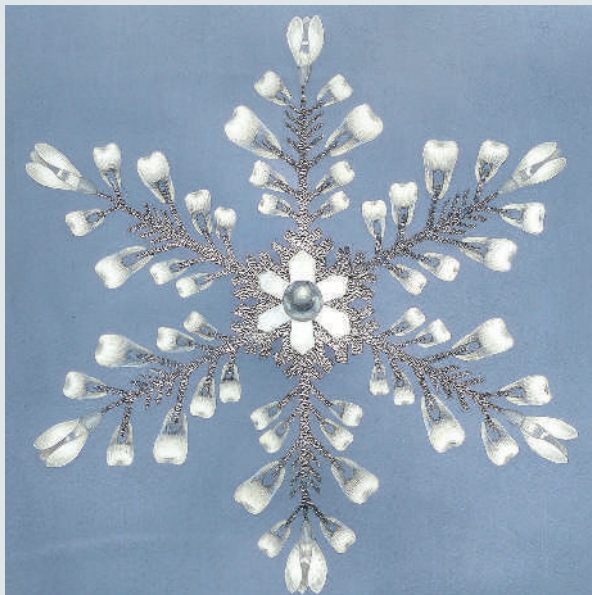
Outsmarted by ants

An elegant orientation solution that is used by ants to get back to their nest eluded even Richard Feynman, suggesting that social insects could help to solve many of our engineering problems.

Francis Ratnieks

Richard Feynman, the Nobel prizewinning physicist who died in 1988, was smart. *Omni* magazine once declared him "The smartest man in the world". But clever as he was, Feynman was outwitted by an ant, or rather by a colony of ants.

How could ants outsmart one of the twentieth century's leading theoretical physicists? When the physicist plays ants at their own game — trying to understand a problem fundamental to colony survival that ants have been working on, by means of natural selection, for millions of years — ants can come out ahead.

Feynman was well known for his curiosity and practical approach. During the space shuttle Challenger enquiry he famously dunked pieces of rubber o-ring from the shuttle's booster rockets into iced water to demonstrate that the material lost its elasticity at low temperatures, and so failed during a cold-weather launch. Many years before this, while a PhD student at Princeton University, the same qualities had led him to think about ants — ants which had entered his imagination by inviting themselves into his room and even raiding his larder.

At Princeton, Feynman "got curious as to how they [ants] found things. How do they know where to go? Can they tell each other where food is, like bees can? Do they have any sense of geometry?"

All these questions are at the heart of ant foraging biology. Natural selection will make foragers good at going to where the food is. It will also favour ants that help nestmates forage more effectively, for example, by laying chemical trails to guide each other to the food. And a sense of geometry will be favoured, provided there is some workable mechanism, if it boosts foraging efficiency.

Feynman's experiments with sugar baits, which involved ferrying individual ants from place to place on small pieces of paper, told him that ants help each other and that they can learn where to go to get food. But his studies on geometry gave conflicting results. He found that the Princeton ants apparently could not tell whether they were walking the wrong way along a trail, but he later found that Brazilian leafcutter ants could. He suggested that Brazilian ants did this using a series of chemicals that polarized the trail. Thus, a trail marked A-B-space-A-B-space is polarized because it reads differently backwards. But a trail with a single chemical A-space-A-space reads the same in both directions. Further experiments at his home in California left the geometry question unresolved: "The ants look as if they have a good sense of geometry. Yet the experiments that I did to try to demonstrate their sense of geometry didn't work".

There are probably 20,000 ant species and they do not all use the same navigational methods or have equal navigational

Small but clever: pharaoh ants are natural geometers.

abilities. Many can reorient on trails by using external cues, including landmarks and the position of the sun. Leafcutter ants are even thought to use the Earth's magnetic field. But recent research has shown that one common ant, the pharaoh's ant, *Monomorium pharaonis*, does have a sense of geometry, and other species probably do as well.

A pharaoh's ant colony forms a foraging-trail network leading from the nest entrance into the surrounding environment. These trails form Y-shaped branches with an internal angle of approximately 60 degrees as they lead away from the entrance. Ants walking the wrong way along a trail are unable to reorient at a trail bifurcation if the angle is 120 degrees. But if the angle is less, then they can. Angles less than 120 degrees give the 'Y' bifurcation a nest–environment polarity, whereas at 120 degrees there is only symmetry. The ability to reorient is maximized at the natural bifurcation angle of 60 degrees.

It is not surprising that Feynman's hypothesized A-B-space polarity mechanism is not used: it is impractical. With hundreds or thousands of ants walking along a trail it would be virtually impossible not to blur the all-important positions of the two different chemical markers and the space marker as ants reinforced the trail by applying more pheromone. What is perhaps more surprising is the unexpected and elegant solution that pharaoh's ants actually use.

Are there any general lessons to take from this? One is that Feynman's research strategy of using whatever ants were around was not ideal for doing biology. But then biologists are probably more used to diversity than are physicists.

The main lesson, however, is one that we have yet to grasp: that we can learn from ants. Natural selection has made insect societies good at solving a problem that is simple to state but hard to solve — to send foragers to where the food is. Because social insects have been solving this complex dynamic problem for millions of years, they have probably evolved some simple and elegant solutions. We should care about these solutions because human life depends more and more on engineering systems that must solve similar problems to function efficiently — electronic messaging, grid computing, transmitting electricity and traffic regulation to name a few. One obvious lesson we might learn is how to make our systems more reliable and robust. If there is one thing that natural selection should be good at, it is eliminating solutions that are not robust. The colony or organism that 'crashes' will soon be a dead one.

If Feynman were alive today he would surely be smart enough to realize that it is no disgrace to be outsmarted by ants. But are we smart enough to learn from them? ■

Professor Francis Ratnieks is in the Laboratory of Apiculture and Social Insects, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, and at the Institute for Advanced Study, Wallostrasse 19, Berlin 14193, Germany.

FURTHER READING
Feynman, R. P. *Don't You Have Time To Think?* (ed. Feynman, M.) (Allen Lane, 2005).
Feynman, R. P. *Surely you're joking, Mr. Feynman!* (Norton, New York, 1985).
Jackson, D. E., Holcombe, M., Ratnieks, F. L. W. *Nature* **432,** 907–909 (2004).

ESSAY

T. WENSELEERS

NEWS & VIEWS

EARTH SCIENCES

Ghosts from within

William F. McDonough

The first detection of geoneutrinos from beneath our feet is a landmark result. It will allow better estimation of the abundances and distributions of radioactive elements in the Earth, and of the Earth's overall heat budget.

The decay of unstable isotopes of chemical elements within the Earth produces heat that contributes to its overall energy output — a fact recognized shortly after Henri Becquerel first discovered radioactivity in 1896. More than 100 years on, Araki *et al.* (page 499 of this issue)[1] report the first measurement of antineutrinos produced by radioactive β^--decay at the heart of the Earth. The results obtained from these so-called geoneutrinos are consistent with geochemical and geophysical models of the planet, and provide a new way of determining where the unstable isotopes — radionuclides — are stored inside the Earth, and in what concentrations.

Antineutrinos, like their counterpart neutrinos, come in three varieties, each named after the charged particle with which they are paired: electron, muon and tau. Electron antineutrinos are produced in β^--decays of an atomic nucleus that occur, for example, when potassium (^{40}K) is transformed to the calcium isotope ^{40}Ca, and also in the decay series of uranium (U) and thorium (Th). Neutrinos and antineutrinos are ghostly particles — they have no charge and almost no mass, and pass through most matter without interacting with it at all. Detecting them is thus extremely difficult.

The KamLAND (Kamioka Liquid-scintillator Anti-Neutrino Detector) apparatus was purpose-built to catch a glimpse of these elusive particles (see Fig. 5 on page 502). The detector is situated in the centre of the largest Japanese island, Honshu, in a mine one kilometre below the summit of Mt Ikenoyama, to reduce the effects of cosmic rays formed from particles other than

antineutrinos. Antineutrinos are occasionally captured by protons in KamLAND's 1-kilotonne, 13-metre-diameter scintillation detector (pictured above) in a process known as inverse β-decay. This produces a neutron, which combines with a proton to form a deuteron and produces a characteristic γ-ray ('scintillation light') with an energy of 2.2 MeV. The light that this reaction produces is detected as an electrical signal by an array of photomultiplier devices surrounding the detector.

In 2003, KamLAND played a fundamental role in documenting the phenomenon known as antineutrino oscillation[2], in which the varieties of antineutrino are thought to change spontaneously into one another. That result demonstrated that antineutrinos have a mass (albeit a small one), and reinforced the discovery of oscillation among neutrino types from the Sun by the Sudbury Neutrino Observatory in Canada[3]. The oscillating antineutrinos detected by KamLAND were produced in nuclear reactors on average some 180 kilometres away. Now the detector has seen antineutrinos from an even more distant source — geoneutrinos, key to understanding where the energy output of the Earth comes from.

The total power dissipated by the Earth's interior is estimated to be between 30 and 44 terawatts (1 TW is 10^{12} watts)[4,5]. The spread in values stems from differences between global models of heat flow at the Earth's surface on the one hand, and estimates of heat dissipation at mid-ocean ridges, corrected for the effects of hydrothermal circulation in the oceans themselves, on the other. Several compositional models for the Earth find that the

amount of K, U and Th in the planet contributes only around 19 TW of power[6,7] to the total. These observations result in a Urey ratio (an assessment of the amount of heat produced by radioactive decay to total heat flow on the surface of a planet) of 0.4–0.6. The remaining heat must come from other potential contributors, such as core segregation, inner-core crystallization, accretion energy or extinct radionuclides — for example, the gravitational energy gained by metal accumulating at the centre of the Earth, which is converted to thermal energy, and the energy added by impacts during the Earth's initial growth.

Alternative models suggest that there is also K in the Earth's core[8,9], and predict a higher Urey ratio. These models are, however, accompanied by geochemical consequences that limit their acceptability (see ref. 4 for a review). The heat flux across the core–mantle boundary and the nature of heat sources in the core are also subject to considerable speculation[10]. The KamLAND results[1] show an upper limit (at the 99% confidence level) of radiogenic heat power from Th and U of 60 TW, and a central value of 16 TW that is consistent with model predictions.

The KamLAND results were not straightforward to obtain, and are not simple to interpret. Various 'pollutants' must be removed from the energy spectrum of the antineutrinos to achieve a pure signal: of a total of 152 events that were potentially from geoneutrinos, only 20–25 were considered true candidate geoneutrinos. The remaining, 'background' antineutrinos came from nearby nuclear power reactors (more than 50% of the total signal)

and radioactive contamination in the detector (around 28%). Moreover, geoneutrinos produced from K decays are not — yet — detected at KamLAND, because their energies are below the threshold of 1.8 MeV required to trigger the existing detector system.

The data reported by Araki et al.[1] are the results from their first experiment, which comprised just over two years of counting. Future observations at KamLAND, and at the Borexino detector under the Gran Sasso mountain in central Italy, which begins operation in 2006, will generate more data and provide greater sensitivity in testing the nature and sources of geoneutrinos. A crucial advance will be to confirm that the geoneutrino heat flux moving radially outwards from the Earth is directly proportional to the radiogenic heat flux. This will, however, require an exact knowledge of the abundance and distribution of K, Th and U in the Earth.

To this end, a first detailed assessment has been made[11] of the predicted geoneutrino flux relative to the distribution of radioactive elements in the regional crust and underlying mantle near KamLAND, and throughout the Earth's interior. Further in the future, combining angle-integrated geoneutrino fluxes at different detector sites[12] with element distribution maps will enable us to construct geoneutrino tomographic maps of the Earth that will tell us more about the planet-wide distribution of K, Th and U. Proposed sites for future (anti)neutrino detectors must therefore be sure to include areas beneath both continental regions rich in K, Th and U and oceanic regions where the three radionuclides are depleted.

The pioneering results from KamLAND presented by Araki et al.[1], along with data from future work, will provide a fundamental constraint for the Earth's U and Th budget (and, it is to be hoped, shortly for that of K), and define the fractional contribution of radioactive heating to the total energy budget. Later this year, particle physicists and Earth scientists will gather to discuss these exciting and common areas of research at a meeting on Hawaii[13]. ■

William F. McDonough is in the Department of Geology, University of Maryland, College Park, Maryland 20742, USA.
e-mail: mcdonoug@geol.umd.edu

1. Araki, T. et al. Nature 436, 499–503 (2005).
2. Eguchi, K. et al. Phys. Rev. Lett. 90, 021802 (2003).
3. Ahmad, Q. R. et al. Phys. Rev. Lett. 89, 011301 (2002).
4. Pollack, H. N., Hurter, S. J. & Johnson, J. R. Rev. Geophys. 31, 267–280 (1993).
5. Hofmeister, A. M. & Criss, R. E. Tectonophysics 395, 159–177 (2005).
6. McDonough, W. F. in Treatise on Geochemistry Vol. 2 (ed. Carlson, R. W.) 547–568 (Elsevier, Oxford, 2003).
7. Palme, H. & O'Neill, H. St C. in Treatise on Geochemistry Vol. 2 (ed. Carlson, R. W.) 1–38 (Elsevier, Oxford, 2003).
8. Rama Murthy, V., van Westrenen, W. & Fei, Y. Nature 423, 163–165 (2003).
9. Lee, K. K. M. & Jeanloz, R. Geophys. Res. Lett. 30, 2212 (2003).
10. Labrosse, S. Phys. Earth Planet. Inter. 140, 127–143 (2003).
11. Fiorentini, G., Lissia, M., Mantovani, F. & Vannucci, R. preprint at www.arxiv.org/hep-ph/0501111 (2005).
12. Field, B. D. & Hochmuth, K. A. preprint at www.arxiv.org/hep-ph/0406001 (2004).
13. www.phys.hawaii.edu/~sdye/hnsc.html

468

CANCER

One step at a time

David Mooney

Traditional chemotherapy kills tumour cells directly; some newer drugs work instead by cutting the tumour's blood supply. An innovative approach combines these strategies sequentially to pack a double whammy.

In 1971, Judah Folkman proposed that the progression of cancer might be halted by preventing tumours from recruiting new blood vessels (a process called angiogenesis) to provide them with oxygen and nutrients. Last year, this theory bore fruit with the approval by the US Food and Drug Administration of the first anti-angiogenic cancer treatment, Avastin (also known as bevacizumab)[1]. Sengupta and colleagues (page 568 of this issue)[2] advance this concept by designing a drug-delivery vehicle that sequentially releases an anti-angiogenic drug and a traditional chemotherapeutic drug at high concentrations specifically into a tumour. They report that their strategy can slow tumour growth in mice more than can either drug alone or the two drugs delivered at the same time.

Traditional chemotherapeutic agents kill all rapidly growing cells in the body — both cancer cells and other cells that divide quickly (for example, blood, hair and cells lining the intestine). This leads to the distressing side effects of chemotherapy, and limits the practical dose and frequency of application of the drugs. One tactic to avoid these effects is to target the drug specifically to the tumour, and approaches being tested include the incorporation of drugs into materials or complexes that can either be placed in, or directed to, tumours[3].

A second issue, however, is that some tumours develop resistance to a particular drug, so efforts to identify targets that are not prone to developing resistance continue. Endothelial cells, which line blood vessels, may provide an attractive target, as they are thought to be genetically more stable than cancer cells and so less likely to develop mutations that might promote resistance. A number of drugs that kill endothelial cells or prevent their growth are proving effective in phase III clinical trials for treating colon, kidney and lung cancer, and gastrointestinal stromal tumours[1,4–6]. These drugs can be useful alone, but they are commonly combined with traditional chemotherapy to prevent blood-vessel growth while also killing cancerous cells.

Simultaneous delivery of chemotherapeutic and anti-angiogenic drugs is clearly beneficial, but because chemotherapy is blood-borne, shutting down the tumour's blood supply with anti-angiogenic drugs may decrease the delivery of drugs designed to kill the tumour cells. Sengupta et al.[2] hypothesized that a more effective strategy would be to use a delivery vehicle that became concentrated in tumours before the vasculature shut down, and allowed the staged release of the two drugs. More specifically, the delivery of the anti-angiogenic factor could lead to a collapse of the vascular network and imprison the vehicle — still bearing its second payload of chemotherapeutic drug — in the tumour. The subsequent release of the latter drug within the tumour would kill the cancer cells.

The authors exploited the fact that the blood vessels of tumours are 'leaky'[7], so tumour tissue can take up larger particles than can normal tissues, promoting selectivity. They created composite vehicle particles of 80–120 nm, consisting of a solid biodegradable polymer core surrounded by a lipid membrane (Fig. 1). The anti-angiogenic drug combretastatin was dissolved in the lipid layer, from which it rapidly escaped. This drug attacks the internal skeleton of cells, and quickly disrupts blood vessels. The chemotherapeutic drug doxorubicin was bound chemically to the inner core of the particle, and so was released more slowly as the bond holding the drug to the polymer broke down. Doxorubicin is a common chemotherapeutic agent, and its structure consists of chemical groups that are amenable to attachment to polymers.

Sengupta et al. examined the effects of the drugs on two types of tumour in mice, and showed that, unsurprisingly, either drug alone slowed tumour growth, and that when the drugs were delivered simultaneously there was an additive effect. Strikingly, however, the staged release of the two drugs using the new delivery vehicle improved the outcome still further — survival time increased from approximately 30 days when the drugs were delivered simultaneously to more than 60 days when they were released sequentially. The delivery vehicles tended to accumulate in the tumours, rather than in other body tissues, and the drugs they transported killed both endothelial and cancer cells.

The effect of the sequential delivery of these two drugs on tumour growth is dramatic, but we cannot assume a quick translation of these results to therapy for humans. The biological differences between mice and humans prevent direct comparison between the systems, and it will also be important to extend these studies to longer time periods. Moreover, it has been speculated that anti-angiogenic drugs may actually promote the spread of tumours to other tissues, owing to a complex feedback loop, although there is no evidence of this in
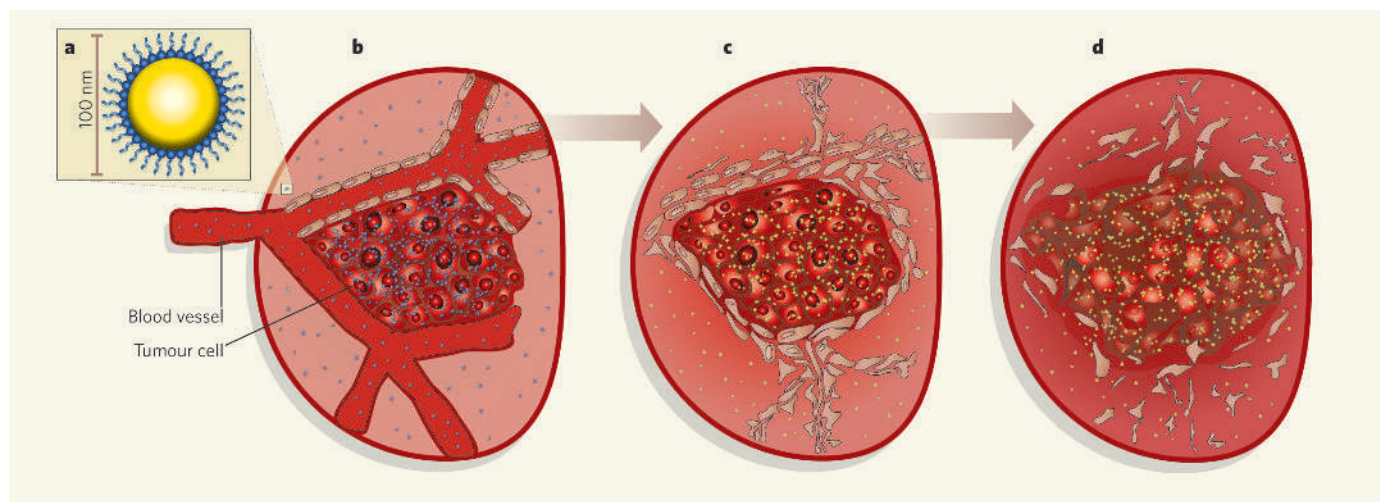
Figure 1 | Step-by-step in fighting cancer. The delivery system of Sengupta *et al.*[2] causes the sequential loss of blood vessels and the death of tumour cells. **a**, Nanometre-scale particles have an outer lipid layer (blue) and an inner core (yellow). **b**, Once injected into the bloodstream, the particle is selectively taken up into tumour tissues, where the lipid layer rapidly releases a drug that kills endothelial cells and disrupts blood vessels. **c**, The inner core gradually releases a chemotherapeutic drug to destroy the cancer cells (**d**).

humans[8]. It is promising, in this regard, that Sengupta and colleagues' system produced no increase in the expression of a factor (HIF-1α) that can link the low oxygen levels resulting from reduced blood flow with potential resistance to drug therapy and tumour invasiveness. Finally, in contrast to combretastatin, many anti-angiogenic drugs require prolonged tissue exposure to shut down the vasculature, and so may not be amenable to the particular approach described by Sengupta and colleagues.

The general concept of timing the availability of drugs aimed at specific stages or targets in cancer is widely applicable, however, and is consistent with similar efforts to promote blood-vessel formation in diseases involving insufficient blood flow[9]. Appropriate design of drugs will allow targeting of cancer cells or other specific cell types[10], and the delivery device described by Sengupta *et al.* could readily be modified for this. It may also be necessary to target multiple aspects of angiogenesis, either by using several drugs or by using a drug that interferes with several pathways (for example, MAPK inhibitors)[11], to prevent tumours from switching on alternative angiogenesis pathways. Ultimately, combining the development of advanced drug-delivery systems with the identification of early markers of cancer may allow early and highly effective intervention, and help to accomplish the US National Cancer Institute's stated goal of eliminating the suffering and death from cancer by 2015. ■

David Mooney is in the Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.
e-mail: mooneyd@deas.harvard.edu

1. Hurwitz, H. *et al. N. Engl. J. Med.* **350,** 2335–2342 (2004).
2. Sengupta, S. *et al. Nature* **436,** 568–572 (2005).
3. Moses, M. A. *et al. Cancer Cell* **4,** 337–341 (2003).
4. Escudier, B. *et al. Annu. Meet. Am. Soc. Clin. Oncol.* abstr. LBA4510; www.asco.org/ac/1,1003,_12-002636-00_18-0034-00_19-0032211,00.asp (2005).
5. Sandler, A. B. *et al. Annu. Meet. Am. Soc. Clin. Oncol.* abstr. LBA4; www.asco.org/ac/1,1003,_12-002643-00_18-0034-00_19-0033325,00.asp (2005).
6. Demetri, G. D. *et al. Annu. Meet. Am. Soc. Clin. Oncol.* abstr. 4500; www.asco.org/ac/1,1003,_12-002643-00_18-0034-00_19-0034169,00.asp (2005).
7. Jain, R. K. *Adv. Drug Deliv. Rev.* **46,** 149–168 (2001).
8. Blagosklonny, M. V. *Cancer Cell* **5,** 13–17 (2004).
9. Richardson, T. P., Peters, M. C., Ennett, A. & Mooney, D. J. *Nature Biotechnol.* **19,** 1029–1034 (2001).
10. Nori, A. & Kopecek, J. *Adv. Drug Deliv. Rev.* **57,** 609–636 (2005).
11. Kim, D. W., Lu, B. & Hallahan, D. E. *Curr. Opin. Invest. Drugs* **5,** 597–604 (2004).

CARBON CYCLE

The age of the Amazon's breath

Peter A. Raymond

The inorganic carbon carried in rivers of the Amazon basin seems to originate largely from the decomposition of young plant material — a finding that improves our understanding of the role of rivers in the carbon cycle.

Increases in atmospheric carbon dioxide from the burning of fossil fuels have unknown effects on the global climate and economy. Scientists aim to understand more about these effects by studying the mechanisms that control the exchange of carbon between land, the atmosphere and the oceans. Processes that remove CO_2 from the atmosphere, where it could cause global warming, and move it into long-term storage on land or in the oceans, are of particular interest in this context. Because they connect land and sea, rivers are a vital link in these processes (Fig. 1, overleaf). On page 538 of this issue, Mayorga *et al.*[1] provide insights into how this river linkage works for the world's largest river system — the Amazon.

There are two main forms of carbon: organic (such as the biomass within a tree) and inorganic (CO_2 in the atmosphere, for example). These forms are intimately coupled through photosynthesis in plants, which creates organic from inorganic carbon, and decomposition, which returns plant-produced carbon to its inorganic form. In rivers, organic and inorganic

carbon exist in approximately equal proportions, and originate mainly when rainfall hits continental surfaces and either dissolves carbon, or carries it to rivers in particulate form.

A single river can drain a landscape that has a wide array of plant species, land uses, soils and climatic zones. This complexity has made it difficult to pin down exactly where most river carbon originates, how long it existed on land before being carried to a river, and how reactive it might be once in a river and, later, in the coastal ocean. Most of the carbon in rivers ultimately comes from atmospheric CO_2 and therefore represents mobile 'greenhouse carbon' that either cycles back to the atmosphere and contributes to global warming, or enters a storage compartment that is not in contact with the atmosphere (coastal sediments, for example; Fig. 1).

The Amazon basin is a central player in the global carbon balance because it stores large amounts of carbon in biomass above ground, and this carbon is being returned to the atmosphere by slash-and-burn agriculture[2]. But the
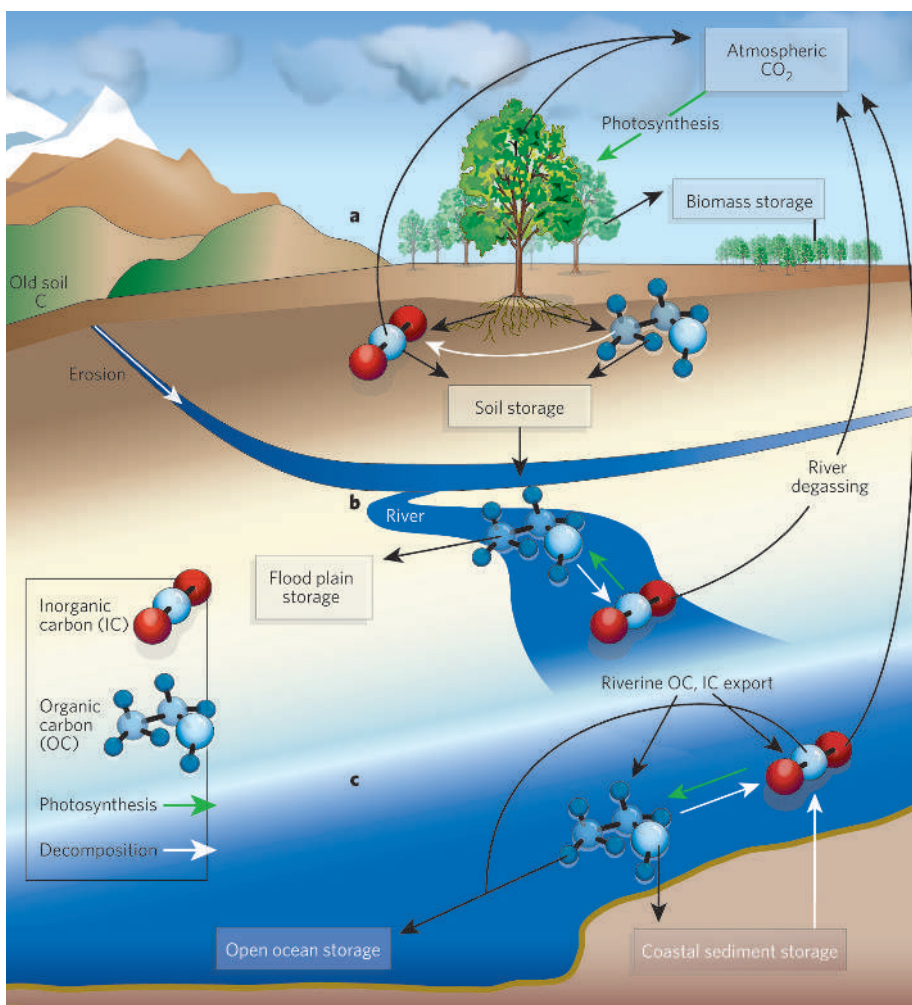
Figure 1 | Rivers and the carbon cycle. a, Photosynthesis in land plants fixes atmospheric CO_2 (inorganic carbon) as organic carbon, which is either stored as plant biomass or in soil, or is decomposed back to CO_2 through plant and soil respiration. This CO_2 can return to the atmosphere or enter rivers; alternatively, it can react with soil minerals to form inorganic dissolved carbonates that remain stored in soils or are exported to rivers. **b,** The transformations of organic to inorganic carbon through decomposition and photosynthesis continue in rivers; here, CO_2 will re-exchange with the atmosphere ('degassing'), or be converted to dissolved carbonates. These carbonates do not exchange with the atmosphere and are mainly exported to the coastal ocean. Organic carbon is also exported to the ocean, or stored in flood plains. **c,** In the coastal ocean, photosynthesis, decomposition and re-exchanging of CO_2 with the atmosphere still continue. Solid organic carbon (such as soil particles and phytoplankton cells) is buried in coastal sediments, where it is stored or decomposes to inorganic carbon and diffuses back into coastal waters. Dissolved inorganic and organic carbon are also exported to the open ocean, and possibly deep-ocean waters, where they are stored for many centuries.

remoteness of the Amazon and the rapid changes in land use in the region have made it difficult to determine the exact size of carbon fluxes within the basin[2]. It was only recently shown[3] that a significant amount of terrestrial organic carbon in the Amazon basin decomposes to CO_2 and is cycled back to the atmosphere through the basin's network of rivers, which 'breathe' CO_2 out in a process known as degassing. Before this, rivers had not been incorporated into the Amazon's carbon budget at all. The finding was significant, as the degassing of CO_2 from rivers occurs in many different regions of the globe[4]. The importance of understanding where this carbon comes from, and how old it is, is therefore not limited to the tropics.

Mayorga *et al.*[1] argue convincingly that the

CO_2 added back to the atmosphere from the Amazonian rivers by degassing is balanced by the decomposition of organic carbon from recent plant growth. This implies that the inorganic carbon does not originate from large-scale decomposition of organic carbon previously held in long-term storage on land. The CO_2-degassing from the tropics thus simply represents the cyclical movement of the gas from the atmosphere, through land and rivers and then back to the atmosphere, and does not represent an additional input of greenhouse gas.

According to Mayorga *et al.*, this cyclical movement seems to happen, at least in the tropics, on a timescale of less than five years. They come to this conclusion by analysing the ratios of the two carbon isotopes, ^{14}C and ^{13}C, contained in both inorganic and organic

carbon at different locations in the Amazon basin. The amount of ^{14}C, a radioactive isotope with a half-life of 5,730 years, can be used to determine the age of a sample (the familiar technique of carbon dating) and also provides information on the source of the carbon. The proportion of the stable carbon isotope ^{13}C supplies more clues to the origin of the carbon[5].

Although Mayorga and colleagues find that most of the organic carbon, particularly in the lowland rivers of the Amazon basin where much of the decomposition to CO_2 and degassing occurs, is indeed young, they also show that the Andean headwaters of the Amazon carry suspended particulate organic carbon that is many centuries to thousands of years old. This result is consistent with other studies on the ages of organic carbon in rivers[6–8] that also pinpoint the origin of old organic carbon in mountain headwaters as mechanically eroded old soil carbon and organic matter, known as kerogen, that is found in sedimentary rocks. Interestingly, the source of old carbon in temperate rivers is still unknown[9].

In the Mayorga study[1], the isotopic signal of the older headwater organic carbon disappears farther down the river network, the carbon presumably becoming buried or decomposed. Although the mountainous regions of the Amazon headwaters are losing soil carbon to erosion, little of this material seems to reach the coastal ocean. The small fraction of aged particulate carbon that does escape the Amazon is then reworked considerably in coastal-shelf sediments[10,11], leaving a very small amount for export and burial in the open ocean.

Mayorga and colleagues' contribution[1] provides a conceptual framework of the role of streams and rivers in tropical carbon budgets on which future studies can build. Research should now focus on how land use and climate change have altered the timescales of the connections between the atmosphere, land and water reported by Mayorga *et al.*. While we continue to burn fossil fuels, and atmospheric CO_2 concentrations continue to increase, scientists will be able to return to this benchmark study to understand how the Amazon basin continues to respond to environmental change. ■

Peter A. Raymond is in the School of Forestry and Environmental Studies, Yale University, 205 Prospect Street, New Haven, Connecticut 06511, USA.
e-mail: peter.raymond@yale.edu

1. Mayorga, E. *et al. Nature* **436,** 538–541 (2005).
2. Houghton, R. A., Lawrence, K. T., Hackler, J. L. & Brown, S. *Glob. Change Biol.* **7,** 731 (2001).
3. Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M. & Hess, L. L. *Nature* **416,** 617–620 (2002).
4. Cole, J. J. & Caraco, N. F. *Mar. Freshwat. Res.* **52,** 101–110 (2001).
5. Raymond, P. A. & Bauer, J. E. *Org. Geochem.* **32,** 469–485 (2001).
6. Blair, N. E. *et al. Geochim. Cosmochim. Acta* **67,** 63–73 (2003).
7. Kao, S. & Liu, K. *Limnol. Oceanogr.* **41,** 1749–1757 (1996).
8. Masiello, C. A. & Druffel, E. R. M. *Glob. Biogeochem. Cycles* **15,** 407–416 (2001).
9. Raymond, P. A. *et al. Mar. Chem.* **92,** 353–366 (2004).
10. Blair, N. E. *et al. Mar. Chem.* **92,** 141 (2004).
11. Druffel, E. R. M., Bauer, J. E. & Griffin, S. *Geochem. Geophys. Geosyst.* **6,** Q03009 (2005).

COGNITIVE NEUROSCIENCE

Decision amid uncertainty

Jonathan D. Cohen and Gary Aston-Jones

Choosing whether to stick to a belief or to abandon it in the face of uncertainty is central to human behaviour. Modelling implicates brain chemicals called neuromodulators in adjudicating this essential decision.

Why did you choose to read this article? Perhaps you are a neuroscientist eager to refine your knowledge. Or perhaps you are keen to broaden your horizons outside your current discipline. These motivations reflect a fundamental trade-off in how we invest our time and effort: individuals must continually decide whether it would be better to pursue known sources of reward, or whether there is more to be gained by searching for new strategies or opportunities. In reinforcement learning, this dilemma is referred to as the trade-off between exploitation and exploration. There is growing evidence that the mechanisms used to resolve this trade-off are directly regulated by neuromodulators[1–3]. Yu and Dayan[4], writing in *Neuron*, extend this work by using simple principles from Bayesian probability theory to derive a sophisticated model of how neuromodulatory systems are central to the trade-off decision.

As their name suggests, neuromodulators such as dopamine, acetylcholine and noradrenaline seem to modify the effects of neurotransmitters — the molecules that allow communication between neurons. Neuromodulatory systems are implicated in almost every mental function, including attention, learning and emotion, and they are disturbed in major neurological and psychiatric disorders, from Alzheimer's disease and post-traumatic stress disorder to depression and schizophrenia.

The conventional view of neuromodulators has been that they have broad, nonspecific functions, such as signalling reward (dopamine) and regulating arousal (noradrenaline). But a current renaissance in the field is showing that neuromodulators have more-specific functions in learning and decision-making. For example, dopamine neurons are implicated in signalling errors in reward prediction, a role that is central to reinforcement learning[1,5]. Moreover, noradrenaline may be key in facilitating the responses to decision-making processes[6,7], and in regulating the balance between exploitation and exploration[3,8].

To examine the role of neuromodulators in such processes, Yu and Dayan[4] exposed subjects to a task involving a set of cues — differently coloured arrows pointing to the left or right — one of which points to where a target will subsequently appear. The participant must respond as quickly as possible to the target; if they work out what the predictive cue is, they tend to do better. In a typical set-up, the predictive cue stays the same for a number of trials, but then changes without warning.

The crux of the task is that there are two forms of uncertainty associated with the cues. Subjects must work out which cue predicts where the target will appear, as well as how reliably it does so. If the predictive cue were 100% accurate, the task would be trivial: most people would quickly discover which arrow reliably points to the target, and notice as soon as this changed. However, the cue is usually set to be only partly predictive (for example, 80% of the time). And herein lies the intrigue — and generality — of the problem. Suppose you have come to believe that a particular cue predicts the target, but in the last few trials it has failed to do so. How do you know whether this is because the cue is not a perfect predictor of the target (like most cues in the world), or because the relevant cue has changed? More generally, how do we decide whether to stand by our beliefs, even as we recognize their fallibility, or to abandon them in search of better ones?

Yu and Dayan cast this dilemma in terms of a distinction between expected uncertainty (in their task, the less-than-perfect reliability of a cue) and unexpected uncertainty (a surreptitious switch in the relevant cue). They propose that information about these forms of uncertainty is coded in the brain by different neuromodulatory systems — with acetylcholine reflecting the degree of expected uncertainty, and noradrenaline gauging unexpected uncertainty.

Yu and Dayan develop a model of how acetylcholine and noradrenaline levels encode uncertainty, and how their interaction determines whether we should abide by or abandon an existing belief. Their analysis implies that if the optimal strategy was computed on each trial, the process would be so demanding as to be biologically unfeasible. However, they demonstrate that an alternative probability algorithm that is biologically plausible can approximate the optimal strategy.

Their model allows Yu and Dayan to make detailed quantitative — and sometimes counterintuitive — predictions about neuromodulatory function and its influence on behaviour. For example, according to their theory the degree of unexpected uncertainty that causes you to abandon a belief should depend on the level of expected uncertainty; that is, if you know that the selected cue is not reliable, you will have a higher tolerance for its failure to predict the target. This makes interesting predictions about how disturbances of acetylcholine and noradrenaline levels will affect behaviour. Confirmation of these predictions, in turn, is likely to provide deeper insight into the patterns of behavioural deficits observed in clinical disorders involving disturbances

---

MATERIALS SCIENCE

Sticky business

Geckos are small tropical lizards known for their ability to attach themselves to walls and ceilings. Their sticking power has been attributed to the millions of little hairs on their feet: each hair exerts a tiny force on the surface, which, added together, enables geckos to hang off surfaces at any angle — a phenomenon that has been exploited to produce adhesive 'gecko tape'.

But what happens to water droplets on a surface that has the potential sticking qualities of a gecko's foot but that is also extremely hydrophobic? M. Jin and colleagues produced such a superhydrophobic, adhesive surface by packing polystyrene nanotubes into an array at a density of 6 million per square millimetre (*Adv. Mater.* doi:10.1002/adma.200401726). The water contact angle — a measure of surface wettability — for the nanotube array was 162°, compared with 95° for a smooth polystyrene surface.

Using a microelectromechanical balance, Jin and colleagues

found that ten times more force was required to remove water from the nanotube surface than from a normal superhydrophobic surface. They attribute this increase to the same mechanism — cumulative van der Waals forces — that governs gecko adhesion. They also discovered that water droplets weighing up to 8 milligrams stuck to the nanotube surface regardless of the angle at which the surface was held (see images), and that the adhesive strength of the surface increased with increasing density of the nanotubes.

Water droplets on the superhydrophobic structured surface could be transferred to a hydrophilic surface without any loss of fluid or introduction of contamination — properties that the authors hope might be useful for practical fluid manipulation.

Rosamund Daw

WILEY

of attention, decision-making and learning.

It is perhaps in this regard that Yu and Dayan's work is most noteworthy. Research on the role of neuromodulatory disturbances in mental disorders has tended to focus on simple hypotheses concerning static excesses or deficits of activity in individual neuromodulatory systems, with little consideration of interactions between systems. A more profound understanding of these dynamics, and their relationship to cognition and behaviour, is crucial if we are to understand how disruption of these systems contributes to the clinical symptoms associated with neurological and psychiatric disorders, and ultimately, how to design effective treatments.

A future problem will be to bring this theory into contact with ones that have framed the question more specifically in terms of the trade-off between exploration and exploitation, and the maximization of reward (for instance, theories about reinforcement learning from neuroscience and utility maximization from economics). The theory also needs to address the high temporal specificity of neuromodulatory systems, which can show rapid phasic responses following task-relevant events (within 100–200 ms), and which may have an immediate impact on task performance[2,6,9–11]. Finally, an understanding of the biophysical and circuit mechanisms by which acetylcholine and noradrenaline interact to produce the proposed functions is needed to advance this view of neuromodulators in brain function.

Yu and Dayan's work is an impressive contribution to the evidence that neuromodulatory function is more specific than previously thought. We can say with some certainty that their theory represents a particularly promising direction of research, and that their paper is a highly rewarding read. ∎

Jonathan D. Cohen is in the Department of Psychology and the Center for the Study of Brain, Mind and Behavior, Green Hall, Princeton University, Princeton, New Jersey 08544-1010, USA. Gary Aston-Jones is in the Laboratory of Neuromodulation and Behavior, Department of Psychiatry, University of Pennsylvania School of Medicine, 705 Stellar Chance, Philadelphia, Pennsylvania 19104-6100, USA.
e-mail: gaj@mail.med.upenn.edu

1. Schultz, W., Dayan, P. & Montague, P. R. Science 275, 1593–1599 (1997).
2. Hasselmo, M. E. & McGaughy, J. Prog. Brain Res. 145, 207–231 (2004).
3. Aston-Jones, G. & Cohen, J. D. Annu. Rev. Neurosci. 28, 403–450 (2005).
4. Yu, A. J. & Dayan, P. Neuron 46, 681–692 (2005).
5. Montague, P. R., Dayan, P. & Sejnowski, T. J. J. Neurosci. 16, 1936–1947 (1996).
6. Clayton, E. C., Rajkowski, J., Cohen, J. D. & Aston-Jones, G. J. Neurosci. 24, 9914–9920 (2004).
7. Brown, E. T. et al. Int. J. Bifurcat. Chaos 15, 803–826 (2005).
8. Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J. & Aston-Jones, G. Science 283, 549–554 (1999).
9. Hasselmo, M. E. Behav. Brain Res. 67, 1–27 (1995).
10. Aston-Jones, G., Rajkowski, J., Kubiak, P. & Alexinsky, T. J. Neurosci. 14, 4467–4480 (1994).
11. Hasselmo, M. E. Neuron 46, 526–528 (2005).

QUANTUM OPTICS

Crystal-clear images

Claude Fabre

Two experiments that use nonlinear crystals to control the spatial distribution of photons in optical images bring the field of quantum imaging closer to maturity. Quantum information processing could ultimately benefit.

God, it seems, plays dice. Randomness lies at the root of many quantum effects: in the arrival times of the photons on a detector, for example; or, if the intensity is such that the photons cannot be distinguished individually, in the temporal fluctuations of the current generated in the detector by those photons. In the past two decades, physicists have found ways to tame this 'quantum noise', and to master, at least in some instances, the temporal distribution of photons.

Two papers now show that it is also possible to control photons' spatial distribution. Writing in *Physical Review Letters*, Mosset and colleagues[1] provide the first experimental demonstration of an optical device that amplifies an image without increasing the spatial distribution of where the photons hit. In a contribution to the same journal last year, Jedrkiewicz et al.[2] used a similar device to produce two images with photons that are identically distributed in space.

Because of the quantum nature of light, the amplification of an optical signal is not a simple multiplication of the input signal by the gain. For very large gains, the signal-to-noise ratio at the output is a factor of two smaller than that at the input[3]. Fortunately, a special class of amplifiers that is not doomed to degrade the signal being amplified does exist. The 'degenerate parametric amplifier' is one of these (Fig. 1a). This amplifier consists of a crystal that responds in a nonlinear manner when it is subjected to an intense 'pump' beam, the electric field E of which varies cosinusoidally with time t as $E = E_0\cos(\pi ft)$. (E_0 is the maximum amplitude of the wave, and f is the frequency of its oscillation.) An optical input signal at a frequency $f/2$ can be amplified while propagating in this crystal without changing the signal-to-noise ratio — that is, without degrading the quality of the information channel[3]. The drawback of this type of amplifier compared with a more conventional one is that it is 'phase-sensitive': only cosinusoidally varying fields of the form $E = E_0\cos(\pi ft)$ are amplified, whereas sinusoidally varying fields $E = E_0\sin(\pi ft)$ are attenuated.

The first experimental demonstration of this 'noiseless amplification' was made in 1993 for single-channel signals (those that have a single frequency and a fixed transverse shape), such as are used in optical telecommunications[4], and in 1999 for signals of an arbitrary transverse shape — images[5].

The novel aspect of the work of Mosset
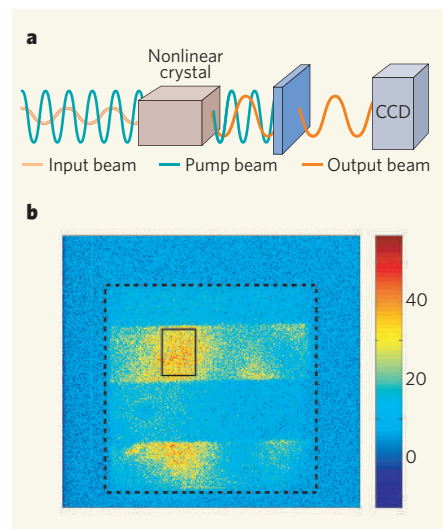
Figure 1 | Parametric amplification of light.
a, Principle of a parametric optical amplifier. The nonlinear crystal allows energy transfer between the two optical beams that propagate through it: the intense pump beam (green) oscillating at frequency f and the weak input beam (orange) oscillating at frequency $f/2$. When the input beam has a well-defined phase relation with the pump beam, it is amplified when it propagates in the crystal. This amplification process is 'noiseless' — no extra quantum fluctuations are added to the input signal. The filter behind the crystal blocks out the pump beam, and the output is recorded on a charge-coupled device, or CCD. **b**, An amplified image recorded on the CCD when the input image is a single rectangular slit of constant intensity (lighter area at the centre). Strong spatial fluctuations between pixels are observed in the output image, related to the randomness of the impact positions of the photons making up the image. By measuring the spatial intensity averages in the rectangular region inside the solid black line (3,266 pixels), Mosset and colleagues[1] show that the amplification process does not degrade the signal-to-noise ratio of the image. The dashed square shows the projection of edges of the crystal.

et al.[1] is its emphasis on spatial, rather than temporal, quantum noise. Spatial quantum fluctuations appear only in images, not in single-channel signals. They are related to the degree of randomness that exists in the distribution of photons composing a beam of light carrying an image; this distribution is measured in the plane transverse to the direction of propagation. In essence, the authors show that images are not degraded by additional spatial quantum fluctuations when they are amplified by a degenerate parametric amplifier.

COGNITIVE NEUROSCIENCE

Decision amid uncertainty

Jonathan D. Cohen and Gary Aston-Jones

Choosing whether to stick to a belief or to abandon it in the face of uncertainty is central to human behaviour. Modelling implicates brain chemicals called neuromodulators in adjudicating this essential decision.

Why did you choose to read this article? Perhaps you are a neuroscientist eager to refine your knowledge. Or perhaps you are keen to broaden your horizons outside your current discipline. These motivations reflect a fundamental trade-off in how we invest our time and effort: individuals must continually decide whether it would be better to pursue known sources of reward, or whether there is more to be gained by searching for new strategies or opportunities. In reinforcement learning, this dilemma is referred to as the trade-off between exploitation and exploration. There is growing evidence that the mechanisms used to resolve this trade-off are directly regulated by neuromodulators[1–3]. Yu and Dayan[4], writing in *Neuron*, extend this work by using simple principles from Bayesian probability theory to derive a sophisticated model of how neuromodulatory systems are central to the trade-off decision.

As their name suggests, neuromodulators such as dopamine, acetylcholine and noradrenaline seem to modify the effects of neurotransmitters — the molecules that allow communication between neurons. Neuromodulatory systems are implicated in almost every mental function, including attention, learning and emotion, and they are disturbed in major neurological and psychiatric disorders, from Alzheimer's disease and post-traumatic stress disorder to depression and schizophrenia.

The conventional view of neuromodulators has been that they have broad, nonspecific functions, such as signalling reward (dopamine) and regulating arousal (noradrenaline). But a current renaissance in the field is showing that neuromodulators have more-specific functions in learning and decision-making. For example, dopamine neurons are implicated in signalling errors in reward prediction, a role that is central to reinforcement learning[1,5]. Moreover, noradrenaline may be key in facilitating the responses to decision-making processes[6,7], and in regulating the balance between exploitation and exploration[3,8].

To examine the role of neuromodulators in such processes, Yu and Dayan[4] exposed subjects to a task involving a set of cues — differently coloured arrows pointing to the left or right — one of which points to where a target will subsequently appear. The participant must respond as quickly as possible to the target; if they work out what the predictive cue is, they tend to do better. In a typical set-up, the predictive cue stays the same for a number of trials, but then changes without warning.

The crux of the task is that there are two forms of uncertainty associated with the cues. Subjects must work out which cue predicts where the target will appear, as well as how reliably it does so. If the predictive cue were 100% accurate, the task would be trivial: most people would quickly discover which arrow reliably points to the target, and notice as soon as this changed. However, the cue is usually set to be only partly predictive (for example, 80% of the time). And herein lies the intrigue — and generality — of the problem. Suppose you have come to believe that a particular cue predicts the target, but in the last few trials it has failed to do so. How do you know whether this is because the cue is not a perfect predictor of the target (like most cues in the world), or because the relevant cue has changed? More generally, how do we decide whether to stand by our beliefs, even as we recognize their fallibility, or to abandon them in search of better ones?

Yu and Dayan cast this dilemma in terms of a distinction between expected uncertainty (in their task, the less-than-perfect reliability of a cue) and unexpected uncertainty (a surreptitious switch in the relevant cue). They propose that information about these forms of uncertainty is coded in the brain by different neuromodulatory systems — with acetylcholine reflecting the degree of expected uncertainty, and noradrenaline gauging unexpected uncertainty.

Yu and Dayan develop a model of how acetylcholine and noradrenaline levels encode uncertainty, and how their interaction determines whether we should abide by or abandon an existing belief. Their analysis implies that if the optimal strategy was computed on each trial, the process would be so demanding as to be biologically unfeasible. However, they demonstrate that an alternative probability algorithm that is biologically plausible can approximate the optimal strategy.

Their model allows Yu and Dayan to make detailed quantitative — and sometimes counterintuitive — predictions about neuromodulatory function and its influence on behaviour. For example, according to their theory the degree of unexpected uncertainty that causes you to abandon a belief should depend on the level of expected uncertainty; that is, if you know that the selected cue is not reliable, you will have a higher tolerance for its failure to predict the target. This makes interesting predictions about how disturbances of acetylcholine and noradrenaline levels will affect behaviour. Confirmation of these predictions, in turn, is likely to provide deeper insight into the patterns of behavioural deficits observed in clinical disorders involving disturbances

---

MATERIALS SCIENCE

Sticky business

Geckos are small tropical lizards known for their ability to attach themselves to walls and ceilings. Their sticking power has been attributed to the millions of little hairs on their feet: each hair exerts a tiny force on the surface, which, added together, enables geckos to hang off surfaces at any angle — a phenomenon that has been exploited to produce adhesive 'gecko tape'.

But what happens to water droplets on a surface that has the potential sticking qualities of a gecko's foot but that is also extremely hydrophobic? M. Jin and colleagues produced such a superhydrophobic, adhesive surface by packing polystyrene nanotubes into an array at a density of 6 million per square millimetre (*Adv. Mater.* doi:10.1002/adma.200401726). The water contact angle — a measure of surface wettability — for the nanotube array was 162°, compared with 95° for a smooth polystyrene surface.

Using a microelectromechanical balance, Jin and colleagues

found that ten times more force was required to remove water from the nanotube surface than from a normal superhydrophobic surface. They attribute this increase to the same mechanism — cumulative van der Waals forces — that governs gecko adhesion. They also discovered that water droplets weighing up to 8 milligrams stuck to the nanotube surface regardless of the angle at which the surface was held (see images), and that the adhesive strength of the surface increased with increasing density of the nanotubes.

Water droplets on the superhydrophobic structured surface could be transferred to a hydrophilic surface without any loss of fluid or introduction of contamination — properties that the authors hope might be useful for practical fluid manipulation.

Rosamund Daw

WILEY

Mosset and colleagues used an intense pulsed laser as the pump beam for the non-linear crystal of the amplifier, and took photographs of the amplified image during a single pulse. The output image, a slit of rectangular shape (Fig. 1b), exhibits strong spatial intensity fluctuations from one pixel to the next. To quantify the importance of this spatial noise, the authors took averages of it over the pixels composing the recorded image. They measured the signal-to-noise ratio in the image with and without amplification, and found that it was not degraded when the parametric amplifier was operated in the phase-sensitive regime; that is, when the field varied cosinusoidally.

This seemingly simple experiment posed many experimental challenges. Quantifying pixel-to-pixel fluctuations reliably at the quantum level, for example, requires delicate calibration and control of the spatial homogeneity of the detection system. Furthermore, the size of the pixels on which the spatial fluctuations are measured must be carefully adjusted.

Jedrkiewicz et al.[2] had already used a non-linear crystal to produce identical twin images by taking advantage of a process known as parametric down-conversion. This effect, in which a single incoming photon of the pump beam is split by a crystal into two photons whose combined energy and momentum is equal to that of the original photon, has featured in many spectacular quantum-optical experiments performed in the low-intensity regime, where photons can still be counted individually[6]. Jedrkiewicz and colleagues used a very intense pulsed laser as a pump, so that the twin photons generated could not be counted individually on each pixel of the detector. They showed that, even though the two generated images exhibited strong pixel-to-pixel quantum fluctuations, these fluctuations were identical in the two images. Their device thus produces twin images instead of twin photons.

The experiments of Mosset et al.[1] and Jedrkiewicz et al.[2] are notable achievements in the rapidly developing domain of quantum imaging[7], which deals with the spatial ordering of photons, even in instances where a macroscopic, uncountable number of photons is produced. Other experiments[8], using different techniques, have shown that photons inside a beam can be ordered four by four in the transverse plane, and that this kind of light can be used to improve the accuracy of beam positioning on the nanoscale.

Such studies open the way to the observation of a phenomenon known as spatial quantum entanglement[9,10], which extends to images the entangled states that were introduced in a famous paper by Einstein, Podolsky and Rosen[11]. This specific quantum feature, in which two images form a single quantum object — even when they are very far away from each other — is a prerequisite for implementing quantum information protocols not only in time-varying signals, but also in an optical image considered as a quantum object in its own right. The pay-off, some time in the future, could be the development of highly parallel quantum-information processors. ■

Claude Fabre is in the Laboratoire Kastler-Brossel, Ecole Normale Supérieure and Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris, France.
e-mail: fabre@spectro.jussieu.fr

1. Mosset, A., Devaux, F. & Lantz, E. *Phys. Rev. Lett.* **94,** 223603 (2005).
2. Jedrkiewicz, O. *et al. Phys. Rev. Lett.* **93,** 243601 (2004).
3. Caves, C. M. *Phys. Rev. D* **26,** 1817–1839 (1982).
4. Levenson, J. A., Abram, I., Rivera, T. & Grangier, P. *J. Opt. Soc. Am. B* **10,** 2233–2238 (1993).
5. Choi, S.-K., Vasilyev, M. & Kumar, P. *Phys. Rev. Lett.* **83,** 1938–1941 (1999).
6. Mair, A. *et al. Nature* **412,** 313–316 (2001).
7. Fabre, C. & Treps, N. in *Lectures on Quantum Information* (eds Bruss, D. & Leuchs, G.) (Wiley-VCH, Berlin, in the press).
8. Treps, N., Grosse, N., Fabre, C., Bachor, H. & Lam, P. K. *Science* **301,** 940–943 (2003).
9. Gatti, A., Brambilla, E. & Lugiato, L. A. *Phys. Rev. Lett.* **90,** 133603 (2003).
10. Hsu, M. T. L. *et al. Phys. Rev. A* **72,** 013802 (2005).
11. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47,** 777 (1935).

NEUROSCIENCE

Genomics reaches the synapse

Cori Bargmann

A genomic survey uses innovative genetics to make neurons susceptible to RNA-mediated gene inactivation. The results implicate many genes in communication at the synapse between neurons and muscle.

RNA interference (RNAi) is a spectacularly useful technique for selectively reducing gene activity — and thereby garnering clues about gene function[1]. In this issue, a collaborative group of geneticists, genomicists and neuroscientists report the first large-scale RNAi screen in neuroscience (Sieburth *et al.*, page 510)[2]. They identify more than 100 novel genes that have specific functions in the transmission of signals across the junction between neurons and muscle cells.

The nematode worm *Caenorhabditis elegans* was the subject of the first genome-wide RNAi screens, aimed at finding the genes involved in development[3,4]. But neurons in *C. elegans* have proved inexplicably resistant to

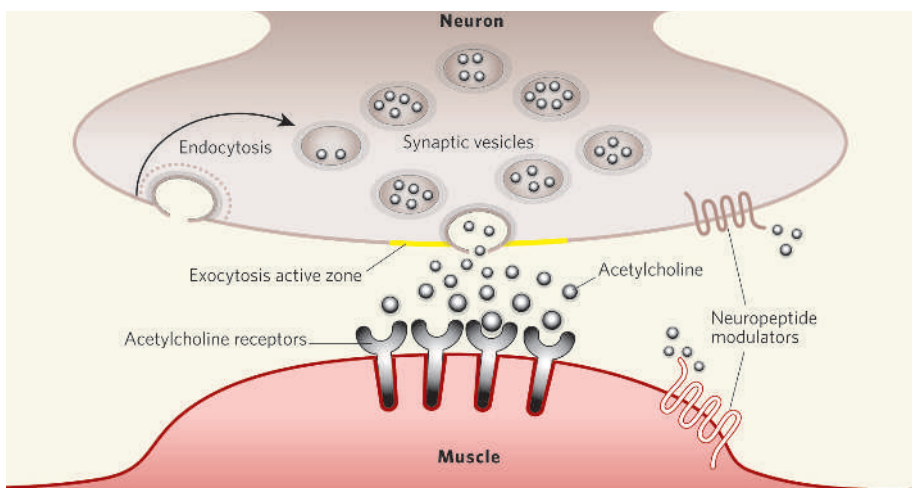**Figure 1 | Signal transmission at the neuromuscular junction (the synapse).** The neurotransmitter acetylcholine is stored within neurons in bubble-like 'vesicles'. When the neuron is activated, the neurotransmitter is released into the synapse at active zones by a process called exocytosis. Acetylcholine binds to receptors on the muscle cell to stimulate muscle contraction. The strength of this signal can be modified by the action of neuropeptide modulators. Synaptic vesicles are re-formed by endocytosis at peri-active zones (near the release site). Sieburth et al.[2] have implicated the products of 185 genes in various synaptic functions, including exocytosis, endocytosis, formation of the active and peri-active zones, vesicle transport and neuropeptide modulation at the synapse.

the technique. The Ruvkun lab[5,6] has taken the geneticist's gambit to solve this problem: if you want something but have no idea how to get it, just find the right mutant. Their clever two-step screen to look for such mutants yielded a strain, designated eri-1;lin-15B, in which neuronal RNAi is possible.

The first gene of this combination, eri-1, encodes a ribonuclease enzyme that may suppress RNAi by degrading small interfering RNAs (siRNAs)[5] — the intermediates in the RNAi process that mark out specific gene products for destruction. The second gene, lin-15B, is the C. elegans equivalent of the mammalian retinoblastoma tumour-suppressor gene (Rb), involved in gene regulation. In this issue, Ruvkun and colleagues (page 593)[6] report that in lin-15B mutants, neurons and other mature cells are partly transformed into immature germline cells. Germline cells are responsive to RNAi, and this transformation also sensitizes neurons to RNAi. The eri-1;lin-15B strain provided the key element that allowed Sieburth et al.[2] to find neural genes using RNAi.

Synaptic transmission is the crucial process by which neurons communicate with their neuronal or non-neuronal targets (Fig. 1). Acetylcholine is the neurotransmitter at the neuromuscular synapse of C. elegans, as it is in vertebrates. The neuron releases acetylcholine from bubble-like vesicles that burst into the synaptic space. The neurotransmitter binds to receptors on the target cell to pass the signal across the synapse, and any excess neurotransmitter is broken down by enzymes called acetylcholinesterases, so that the synapse is primed for the next signal.

In an early application of chemical genetics, Nguyen et al.[7] found that drugs that inhibited acetylcholinesterases kill normal worms, but

are less poisonous to worms with reduced neuromuscular transmission. The researchers screened a series of mutant worms for resistance to the inhibitors, and found many proteins that are involved in the cycle of synaptic-vesicle formation and breakdown (for example, UNC-13 and UNC-18). But much clearly remained to be discovered. In their work, Sieburth et al.[2] applied RNAi together with this highly selective drug screen in the eri-1;lin-15B strain and in a second eri-1;lin-15B strain sensitized by another mutation in neurotransmission. Even with these sensitized and specific assays, the experiments were difficult, so they screened only 2,072 of the roughly 19,500 C. elegans genes, cherry-picking those with relatives in other species or possessing particular structural domains that hinted at neural functions. This sub-screen yielded 185 genes, 132 of which had not previously been implicated in synaptic transmission in any animal.

The next hurdle was to characterize this huge set of 185 genes further. Following an approach first used on genes involved in fat metabolism[8], the authors analysed 60 RNAi candidates in the presence of various drugs and in different mutant-gene combinations. Genes with related functions should behave similarly in these assays. By clustering the patterns of defects seen in new RNAi genes with the defects caused by known genes, the authors assigned some of the novel genes to various steps of synaptic neurotransmission (Fig. 1).

But they didn't stop there. They examined the subcellular distribution of 100 of the proteins encoded by the genes by fusing them with green fluorescent protein (GFP). This work was facilitated by the 'ORFeome' project, led by one of the authors, Marc Vidal, in which full-length clones of all C. elegans genes are being isolated. Of the 100 proteins, 26 were

located in presynaptic regions, some in association with synaptic vesicles, others associated with the plasma membrane at or near the sites of synaptic release. Ten proteins seemed to be in postsynaptic regions. Given that so many of the proteins reside in or near the synapse, and that many had associations with known synaptic components such as the kinesin UNC-104 (KIF-1A), this confirmed that the screens yielded exactly the desired molecules.

For many of the candidate genes from the screens, mutants were available in other gene collections created by classical genetic methods or from gene-knockout screens. When these traditional mutants were examined, 28 out of 34 mutants had characteristics suggesting functions at the synapse. Finally, the authors examined synapses in live animals using tagged synaptic-vesicle proteins to subdivide the mutants into even more specialized functional groups.

Most of the genes identified have close relatives in other animals, and detailed mechanistic studies should elucidate their exact molecular functions. For the present, the framework provided by Sieburth et al. is an impressive beginning. The work raises the standard for future experiments in the field: it's not just a gene list from a screen, but a thorough and thoughtful combination of genetics, molecular biology and data analysis.

RNAi is subject to false-negative results from inefficiencies in the system, and to false-positive results from effects on unintended targets, but Sieburth et al. were successful largely because they used highly selective screens in sensitized genetic backgrounds. Moreover, they could validate their methods using existing mutations (19 of 21 known cholinesterase-resistance genes were confirmed by at least one of their screens), and were able to verify their results by reference to classical genetic mutations — tools that are not available in every system.

The considerable efforts under way to define the enzymes involved in RNAi should make it an even more powerful and reliable tool. It has the potential to bring genetics to an exciting array of unsolved questions in neuroscience. Armed with the data from genome projects, we can look forward to bold and imaginative RNAi screens in previously intractable, but highly desirable, experimental systems — perhaps songbirds, honeybees, cichlid fish or even the primate brain. ■

Cori Bargmann is at the Howard Hughes Medical Institute, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.
e-mail: cori@rockefeller.edu

1. Fire, A. Nature **391,** 806–811 (1998).
2. Sieburth, D. et al. Nature **436,** 510–517 (2005).
3. Fraser, A. G. et al. Nature **408,** 325–330 (2000).
4. Gönczy, P. et al. Nature **408,** 331–336 (2000).
5. Kennedy, S., Wang, D. & Ruvkun, G. Nature **427,** 645–649 (2004).
6. Wang, D. et al. Nature **436,** 593–597 (2005).
7. Nguyen, M. et al. Genetics **140,** 527–535 (1995).
8. Ashrafi, K. et al. Nature **421,** 268–272 (2003).

BRIEF COMMUNICATIONS

'Fingerprinting' documents and packaging

Unique surface imperfections serve as an easily identifiable feature in the fight against fraud.

We have found that almost all paper documents, plastic cards and product packaging contain a unique physical identity code formed from microscopic imperfections in the surface. This covert 'fingerprint' is intrinsic and virtually impossible to modify controllably. It can be rapidly read using a low-cost portable laser scanner. Most forms of document and branded-product fraud could be rendered obsolete by use of this code.

Naturally occurring randomness in the physical properties of an attached tag or token is a means of ascribing a unique identifier to documents and objects[1–4]. We have investigated the possibility of using the intrinsic roughness present on all non-reflective surfaces as a source of physical randomness. This has the potential to provide strong, in-built, hidden security for a wide range of paper, plastic or cardboard objects.

Making use of the optical phenomenon of laser speckle[5], we examined the fine structure of different surfaces using the diffuse scattering of a focused laser. Speckle has been used for determining, among other things, surface roughness[6], small deformations of metal and wet paper[7], shear moduli of paper[8] and for visualizing blood flow *in vivo*[9].

Figure 1a shows the results of scanning a focused laser beam across a sheet of standard white paper and continuously recording the reflected intensity from four different angles

Figure 1 | Comparing document fingerprints. a, The fluctuation in diffuse intensity at one photodetector as a focused laser beam is scanned across a 40-mm length of white paper. **b,** Digital cross-correlation between two scans from different sheets of paper. **c,** Digital cross-correlation between two scans from the same sheet of paper.

by using four photodetectors. (See supplementary information for methods.) Statistical analysis indicates that there are pseudo-random fluctuations that have a minimum wavelength of 70 μm. The fluctuations from the mean intensity are digitized into ones and zeros to form the fingerprint code for the object.

Figure 1b shows the digital cross-correlation between this scan and a similar scan from a different sheet of paper from the pack. The absence of any strong peak indicates that the scans are independent of each other. By contrast, the digital cross-correlation between the original scan and a subsequent scan from the same sheet of paper taken three days later, with normal handling of the paper in between, shows a strong peak close to zero positional shift (Fig. 1c); this indicates that the scans are largely identical. Similar results were obtained from matt-finish plastic cards (such as credit cards), identity cards and coated paperboard packaging (as used to pack pharmaceuticals and cosmetics, for example).

Recognition was good even after the object had been roughly handled. For paper, this included screwing it into a tight ball, followed by smoothing to leave a badly creased surface; submerging it in cold water for 5 min, followed by natural drying; baking it in air at 180 °C for 30 min to scorch the surface; scribbling heavily over the scanned area with a ball-point pen and a thick black marker pen; or scrubbing the surface with an abrasive cleaning pad. Misplacements of the object on the scanner by up to ±1 mm and 2° were judged acceptable.

The amplitude of the cross-correlation peak can be used to determine the probability of two objects sharing indistinguishable fingerprints. For the paper studied here, the probability was less than 10^{-72} (see supplementary information). Smoother surfaces, such as matt-finish plastic cards and coated paperboard, typically give probabilities of less than 10^{-20}. The speckle signal therefore serves as a virtually unique fingerprint for the object. Each fingerprint requires about 200–500 bytes of storage space.

Most existing security validation schemes rely on a proprietary manufacturing process that would be difficult for a fraudster to reproduce (for example, holograms or security inks). Our findings open the way to a new approach to authentication and tracking — even the inventors would not be able to carry out a physical attack on this fingerprint as there is no
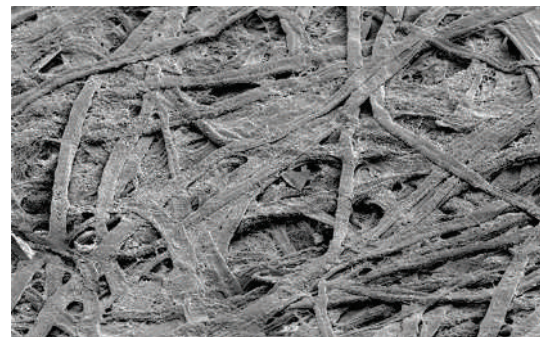
Duck and weave: the microscopic patterns on the surface of paper will survive a thorough soaking.

known manufacturing process for copying surface imperfections at the required level of precision. There is no need to modify the protected item in any way through the addition of tags, chips or inks, so protection is covert, low-cost, simple to integrate into the manufacturing process, and immune to attacks directed against the security feature itself.

James D. R. Buchanan*, Russell P. Cowburn*, Ana-Vanessa Jausovec*, Dorothée Petit*, Peter Seem*, Gang Xiong*, Del Atkinson†, Kate Fenton†, Dan A. Allwood‡, Matthew T. Bryan‡

*Blackett Physics Laboratory, Imperial College London, London SW7 2BW, UK
e-mail: r.cowburn@imperial.ac.uk
†Durham University Physics Department, Rochester Building, Durham DH1 3LE, UK
‡Department of Engineering Materials, Sir Robert Hadfield Building, University of Sheffield, Sheffield S1 3JD, UK

1. Smith, J. R. & Sutherland, A. V. *Proc. AutoID'99, Workshop on Automatic Identification Advanced Technologies* 79–83 (IEEE, New York, 1999).
2. van Renesse, R. in *European Convention on Security and Detection* 45–49 (Institution of Electrical Engineers, London, 1995).
3. Indeck, R. S. & Glavinas, E. *IEEE Trans. Magn.* **29,** 4095–4097 (1993).
4. Pappu, R., Recht, B., Taylor, J. & Gershenfeld, N. *Science* **297,** 2026–2030 (2002).
5. Goodman, J. W. *Topics in Applied Physics: Laser Speckle and Related Phenomena* (ed. Dainty, J. C.) vol. 9 (Springer, Berlin, 1984).
6. Leonard, L. C. & Toal, V. *Opt. Lasers Eng.* **30,** 433–440 (1998).
7. Sjodahl, M. & Larsson, L. *Opt. Lasers Eng.* **42,** 193–201 (2004).
8. Seo, Y. B *J. Pulp Paper Sci.* **25,** 321–325 (1999).
9. Briers, J. D., Richards, G. & He, X. W. *J. Biomed. Opt.* **4,** 164–175 (1999).
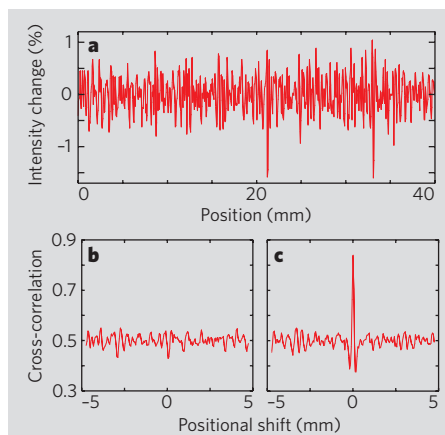
CLIMATE

How unusual is today's solar activity?

Arising from: S. K. Solanki, I. G. Usoskin, B. Kromer, M. Schüssler & J. Beer *Nature* **431**, 1084–1087 (2004)

To put global warming into context requires knowledge about past changes in solar activity and the role of the Sun in climate change. Solanki *et al.*[1] propose that solar activity during recent decades was exceptionally high compared with that over the preceding 8,000 years. However, our extended analysis of the radiocarbon record reveals several periods during past centuries in which the strength of the magnetic field in the solar wind was similar to, or even higher than, that of today.

Solanki *et al.* combine radiocarbon (^{14}C) data, visually observed sunspot numbers and models to extend the historical sunspot record over the Holocene. They exclude the most recent 100 years of the ^{14}C record, which are influenced by ^{14}C-depleted fossil-fuel emissions and atomic-bomb tests conducted since AD 1950. We extend the analysis of the radiocarbon record to AD 1950, which allows us to link the ^{14}C-based solar reconstruction to instrumental measurements of solar magnetic modulation that cover the past 68 years[2,3].

The Sun influences the production rate of ^{14}C in the Earth's atmosphere by modulating the galactic cosmic-ray flux through its magnetic field. Increased magnetic field in the solar wind causes a stronger deflection of galactic cosmic rays and lower radionuclide production rates in the atmosphere and vice versa. However, the atmospheric ^{14}C concentration also depends nonlinearly on the geomagnetic field intensity and the global carbon cycle. These factors and their uncertainties need to be carefully included in the reconstruction of solar activity.

We reconstructed ^{14}C production rates from radiocarbon records[4,5] (Fig. 1a) using two box-diffusion type carbon-cycle models, the box-diffusion[6] and the Bern model[7], and a three-dimensional ocean model coupled to a four-box land biosphere. The ^{14}C production record was transformed[8,9] into a record of the solar modulation parameter that describes the solar influence on galactic cosmic-ray deflection by normalizing it to neutron-monitor and ionization-chamber data covering the recent decades[2]. Alternatively, balloon-borne estimates of galactic cosmic-ray deflection[3] were used instead of the ionization-chamber data. Like Solanki *et al.*[1], we assumed that natural variations in the carbon cycle were small during the past millennium, which is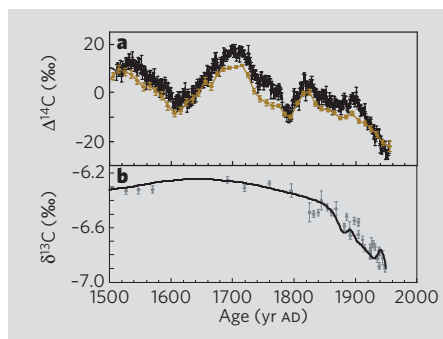 consistent with ice-core CO_2 and $^{13}CO_2$ data and models[10]. We considered uncertainties in ^{14}C data, fossil CO_2 emissions, the geomagnetic field and model parameters. Results are similar across the range of models and simulations. Calculations using the Northern[4] or Southern[5] Hemisphere ^{14}C record, or both, reveal that differences between the ^{14}C records do not affect our main conclusions.

It is standard practice to model the observed dilution of the atmospheric isotope ratios caused by the addition of isotopically depleted carbon from fossil and land-use sources. Emissions from fossil sources are prescribed, emissions from land-use sources are inferred from the atmospheric carbon budget, and the two-way exchange fluxes between reservoirs are simulated. The dilution of ^{14}C is governed by the same processes that affect ^{13}C. The good agreement between modelled ^{13}C and ice-core data (Fig. 1b) supports the reconstructed rate of ^{14}C production.

The comparison between the physical quantities, the ^{14}C-production rate and the solar-modulation parameter, and the visually based sunspot record reveals similarities and striking differences (Fig. 2). The 11-year solar cycle is distinct in all records. On the other hand, solar magnetic modulation was higher or comparable to today during the late eighteenth (and twelfth) century and around AD 1600, whereas sunspot numbers were highest over the recent decades. Sunspot numbers fell to zero during the Maunder Minimum (AD 1650–1700), whereas ^{14}C production and solar modulation continued to vary. The balloon-borne measurements[3] imply lower values for the solar modulation parameter than the ionization-chamber data[2]. Those data seem preferable because they agree better with the solar-modulation changes inferred by Solanki *et al.*[1] than do the balloon-borne measurements[3]. If

we include the Northern and Southern Hemisphere ^{14}C records (three-dimensional model) and normalize the data to the balloon-borne record, we obtain very similar levels of solar modulation (black curve in Fig. 2b).

In any case, and irrespective of the data set applied, the recent solar activity is not exceptionally high (Fig. 2). The ^{14}C results are broadly consistent with earlier reconstructions based on ^{10}Be data from the South Pole, which show that production rates around AD 1780 and in the twelfth century were comparable to those observed today[11]. We conclude that
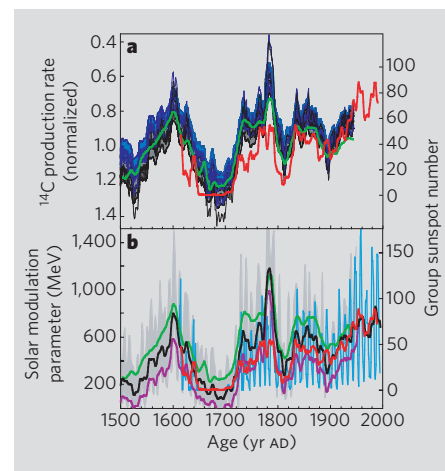
Figure 2 | Radiocarbon production rate and solar modulation parameter compared with the group sunspot number. a, ^{14}C-production rates (blue) were calculated from the Northern Hemisphere ^{14}C record with two carbon-cycle models[6,7]; fossil-fuel emissions ($\pm 10\%$) and air–sea CO_2 exchange rate ($\pm 30\%$) were varied systematically in sensitivity calculations (various lines). The shadowed area shows the 1σ errors resulting from 100 Monte Carlo simulations. We included Northern Hemisphere[4] and Southern Hemisphere[5] ^{14}C data in a three-dimensional ocean model calculation (green line). The black curves show the ^{14}C-production rate corrected for the geomagnetic field intensity[8,9]. The red curve shows the group sunspot number[15]. All records are normalized and show 11-yr averages. **b,** The solar modulation parameter was calculated[8] from the ^{14}C-production rates and from ionization-chamber and neutron-monitor data[2]. The grey band shows results from annual ^{14}C data, including the error range based on the Monte Carlo calculations and uncertainties in the geomagnetic data[9]. The black line shows 11-yr averages (best estimate). Unfiltered (blue) and 11-yr-averaged (red) sunspot numbers[15] are shown for comparison. The green line shows the three-dimensional model results. The purple line depicts results using an alternative neutron-monitor record[3] that indicates a stronger trend from 1937 to 1950 than the ionization-chamber data[2].

Figure 1 | Measured atmospheric $^{14}CO_2$:$^{12}CO_2$ and $^{13}CO_2$:$^{12}CO_2$ ratios and model results for $\delta^{13}C$ from AD 1500 to 1950. $\Delta^{14}C$ and $\delta^{13}C$ denote the per mil deviations from a standard. **a,** Tropospheric radiocarbon from tree-ring measurements from the Northern Hemisphere[4] (black) and Southern Hemisphere[5] (brown). Error bars show one standard deviation. **b,** The Law Dome (Antarctica) $\delta^{13}C$ record[14] (dots) compared with the output of the Bern model (black line). Error bars show one standard deviation.

the link between the visually based sunspot numbers and solar-modulation parameter is neither straightforward nor yet understood, and also that solar modulation must have reached or exceeded today's magnitudes three times during the past millennium.

Uncertainties in low-frequency changes increase when reconstructions are extended from the past few centuries to the past millennia. The low-frequency Holocene ^{14}C variations can largely be explained by changes in the geomagnetic field as they lie within the errors of the archaeomagnetic data set used for correction[12]. Some palaeomagnetic records[13] indicate higher geomagnetic intensities around 7000 BC, which indicate that solar activity could have been lower during this period than is suggested by Solanki et al.[1].

What do our results mean for climate change? It is speculative to translate solar magnetic modulation quantitatively into irradiance because we do not have a clear mechanistic understanding or evidence from data. Still, records of solar magnetic modulation proxies are often used as direct indicators of solar irradiance in climate and carbon-cycle model calculations (see ref. 10, for example). The reconstruction by Solanki et al. implies generally less solar forcing during the past millennium than in the second part of the twentieth century, whereas our reconstruction indicates that solar activity around AD 1150 and 1600 and in the late eighteenth century was probably comparable to the recent satellite-based observations. In any case, as noted by Solanki et al., solar activity reconstructions tell us that only a minor fraction of the recent global warming can be explained by the variable Sun.

**Raimund Muscheler*, Fortunat Joos†,
Simon A. Müller†, Ian Snowball‡**
*National Center for Atmospheric Research, Climate and Global Dynamics Division, Paleoclimatology, Boulder, Colorado 80305-3000, USA
e-mail: raimund@ucar.edu
†Climate and Environmental Physics, Physics Institute, University of Bern, 3012 Bern, Switzerland
‡GeoBiosphere Science Centre, Quaternary Sciences, Lund University, 22362 Lund, Sweden

1. Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M. & Beer, J. *Nature* **431,** 1084–1087 (2004).
2. Beer, J. *Space Sci. Rev.* **93,** 107–119 (2000).
3. McCracken, K. G. & Heikkila, B. in *Proc. 28th Int. Cosmic Ray Conf. 2003, Tsukuba, Japan* 4117–4120 (Univ. Acad. Press, Tokyo, 2003).
4. Stuiver, M., Reimer, P. J. & Braziunas, T. F. *Radiocarbon* **40,** 1127–1151 (1998).
5. McCormac, F. G. et al. *Radiocarbon* **44,** 641–651 (2002).
6. Siegenthaler, U. *J. Geophys. Res.* **88,** 3599–3608 (1983).
7. Joos, F. et al. *Tellus B* **48,** 397–417 (1996).
8. Masarik, J. & Beer, J. *J. Geophys. Res.* **104,** 12099–12111 (1999).
9. Yang, S., Odah, H. & Shaw, J. *Geophys. J. Int.* **140,** 158–162 (2000).
10. Gerber, S. et al. *Clim. Dynam.* **20,** 281–299 (2003).
11. Bard, E., Raisbeck, G. M., Yiou, F. & Jouzel, J. *Tellus B* **52,** 985–992 (2000).
12. Muscheler, R., Beer, J., Kubik, P. W. & Synal, H.-A. *Quat. Sci. Rev.* **24,** 1849–1860 (2005).
13. Snowball, I. & Sandgren, P. *Earth Planet. Sci. Lett.* **227,** 361–376 (2004).
14. Francey, R. J. et al. *Tellus B* **51,** 170–193 (1999).
15. Hoyt, D. V. & Schatten, K. H. *Solar Phys.* **179,** 189–219 (1998).

CLIMATE

Solanki et al. *reply*

Reply to: R. Muscheler et al. doi:10.1038/nature04045 (2005)

Muscheler et al.[1] claim that the solar activity affecting cosmic rays was much higher in the past than we deduced[2] from ^{14}C measurements. However, this claim is based on a problematic normalization and is in conflict with independent results, such as the [44]Ti activity in meteorites and the [10]Be concentration in ice cores.

Our results[2] are based on ^{14}C-production rates, Q, before AD 1900, which largely avoids the uncertainties arising from the extensive fossil-fuel-burning signal[3] commonly called the Suess effect; however, Muscheler et al.[1] determine the relative ^{14}C-production rate (normalized to a mean value of unity) up to AD 1950. Their values were then scaled by a constant factor, determined such that the inferred cosmic-ray modulation strength, Φ, matches the values determined from ionization-chamber data measured after AD 1937. Uncertainties in the correction for the Suess effect thus directly translate into errors in Φ. Muscheler et al. assert that the dilution of ^{13}C is governed by the same processes that affect ^{14}C, but this is an oversimplification. Although both isotopes are affected by fossil-fuel emissions, ^{13}C is, in addition, influenced by land-use changes. Further model parameters are thus available for adjustment when reproducing ^{13}C, so that this isotope cannot be used as an independent check on the ^{14}C reconstructions.

The calibration procedure for Φ seems problematic because ionization chambers have uncontrollable drifts[4]. Moreover, the combined data record[5] of the Cheltenham ionization chamber (AD 1937–53) and neutron monitors (since AD 1953), on which Muscheler et al. base their analysis, represents not the real cosmic-ray intensity but rather its detrended and normalized variation[6]. Direct balloon-borne measurements show that the cosmic-ray intensity before AD 1950 had a strong declining trend. As a result of scaling the ^{14}C-production rate on the basis of these inappropriate data, Muscheler et al. infer too low an average value of Q and, accordingly, too high a value of Φ. Because of the nonlinearity of the relationship between both quantities, this leads to particularly significant effects for small values of Q and results in a strong amplification of the associated large Φ values. The use of a more appropriate data set[4] leads Muscheler et al. to results that are largely consistent with our reconstruction, except for a short period around AD 1780 (purple curve in their Fig. 2b). However, they instead use their problematic original scaling ('best estimate', black curve).

By contrast, our model[2] consistently reproduces the values of Φ determined from modern cosmic-ray measurements without any scaling or parameter adjustment. Comparing the values of Φ determined from the ^{14}C-production rate before AD 1900 and the values computed from the group sunspot number up to the present[7], we find that both curves match each other before AD 1900 (see supplementary information in Solanki et al.[2]) and, at the same time, that the latter Φ agrees very well with the values derived from neutron-monitor and balloon data[8], in contrast to the claim by Muscheler et al.[1].

Furthermore, their large values of Φ contradict the integrated cosmic-ray flux measured by the abundance of [44]Ti (half-life of about 60 years) in meteorites[9,10] that have fallen since AD 1766. The [44]Ti activity in meteorites is completely independent of transport effects and redistribution in the Earth's atmosphere, so it provides direct measurements of past cosmic-ray flux. The 'best estimate' of Muscheler et al. yields a [44]Ti activity that is systematically too low, whereas our reconstruction fits the measurements well.

The abnormally high modulation parameter around AD 1780 obtained by Muscheler et al. is also not reflected in results obtained for [10]Be. South Pole data[11] from around AD 1780 show about 55% of the Maunder minimum level, whereas the value of $\Phi = 1,200$ MeV proposed by Muscheler et al.[1] would imply a much stronger reduction, to about 30% (ref. 12). Similarly, [10]Be data from Greenland do not show a prominent dip at around AD 1780. Neither do other proxies (such as sunspots[13], aurorae[14] and polar nitrates[15]) indicate particularly strong solar activity around AD 1780.

We conclude that by basing their normalization procedure on inappropriate data, Muscheler et al. have heavily overestimated the solar modulation parameter before AD 1950, which was further exaggerated by the nonlinear relation between Q and Φ.

**S. K. Solanki*, I. G. Usoskin†, B. Kromer‡,
M. Schüssler*, J. Beer§**
*Max-Planck-Institut für Sonnensystemforschung,

37191 Katlenburg-Lindau, Germany
e-mail: solanki@mps.mpg.de
†Sodankylä Geophysical Observatory (Oulu unit),
University of Oulu, 90014 Oulu, Finland
‡Heidelberger Akademie der Wissenschaften,
Institut für Umweltphysik, 69120 Heidelberg,
Germany
§Department of Surface Waters,
Eidgenössische Anstalt für Wasser-
versorgung Abwasserreinigung und
Gewässerschutz (EAWAG), 8600 Dübendorf,
Switzerland

1. Muscheler, R., Joos, F., Müller, S. A. & Snowball, I. *Nature* **436,** doi:10.1038/nature04045 (2005).
2. Solanki, S. K., Usoskin, I. G., Kromer B., Schüssler, M. & Beer, J. *Nature* **431**, 1084–1087 (2004).
3. Suess, H. E. *Science* **122,** 415–417 (1955).
4. McCracken, K. G. & Heikkila, B. in *Proc. 28th Int. Cosmic Ray Conf. 2003, Tsukuba, Japan* 4117–4120 (Univ. Acad. Press, Tokyo, 2003).
5. Beer, J. *Space Sci. Rev.* **93**, 107–119 (2000).
6. Beer, J., Raisbeck, G. M. & Yiou, F. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Matthews, M. S.) 343–359 (Univ. of Arizona Press, Tucson, 1991).
7. Usoskin, I. G., Mursula, K., Solanki, S. K., Schüssler, M. & Kovaltsov, G. A. *J. Geophys. Res.* **107**, doi:10.1029/2002JA009343 (2002).
8. Masarik, J. & Beer, J. *J. J. Geophys. Res.* **104**, 12099–12111 (1999).
9. Bonino, G., Castagnoli, G. C., Bhandari, N. & Taricco, C. *Science* **270**, 1648–1650 (1995).
10. Cini Castagnoli, G., Cane, D., Taricco, C. & Bhandari, N. in *Proc. 28th Int. Cosmic Ray Conf. 2003, Tsukuba, Japan* 4045–4048 (Univ. Acad. Press, Tokyo, 2003).
11. Bard, E., Raisbek, G. M., Yiou, F. & Jouzel, J. *Earth Planet. Sci. Lett.* **150**, 453–462 (1997).
12. Webber, W. R. & Higbie, P. R. *J. Geophys. Res.* **108**, doi:10.1029/2003JA009863 (2003).
13. Hoyt, D. V. & Schatten, K. H. *Solar Phys.* **179**, 189–219 (1998).
14. Feynman, J. & Silverman, S. M. *J. Geophys. Res.* **85**, 2991–2997 (1980).
15. McCracken, K. G., Dreschhoff, G. A. M., Smart, D. F. & Shea, M. A. *J. Geophys. Res.* **106**, 21599–21609 (2001).

INSECT DEFENCES
Taste alteration and endoparasites

Taste sensation and food selection by animals can change adaptively in response to experience, for example to redress specific nutrient deficiencies[1]. We show here, in two species of caterpillar, that infection by lethal parasites alters the taste of specific phytochemicals for the larvae. Given that these compounds are toxic to the parasites and are found in plants eaten by the caterpillars, their changed taste may encourage parasitized caterpillars to increase consumption of plants that provide a biochemical defence against the invaders.

Caterpillars of the tiger moths (Lepidoptera Arctiidae) *Grammia geneura* and *Estigmene acrea* are dietary generalists that occupy a grassland and savanna habitat in southern Arizona. Both are susceptible to several species of parasite (insect parasitoids)[2] that typically wound their host mortally. The caterpillars defend themselves by sequestering pyrrolizidine alkaloids, which are toxic to a range of predators and parasitoids[3,4], from favoured food plants[5] such as *Senecio longilobus*.

The caterpillars can detect pyrrolizidine alkaloids even at extremely low concentration by virtue of specialist receptor cells, which stimulate feeding behaviour[6,7]. *Grammia geneura* also sequesters iridoid glycosides, which induce rapid firing of sucrose-receptor cells and promote feeding on the iridoid-containing plant *Plantago insularis*[8]. This plant is known to confer protection against predators[9,10].

To quantify these gustatory-cell responses and determine whether they are linked to parasitism, we collected *G. geneura* and *E. acrea* from the field and tested how the response of their gustatory cells to these plant compounds and to selected feeding deterrents was affected by the presence of parasites (for methods, see supplementary information).

We found that the taste cells of parasitized *G. geneura* caterpillars increased their firing rates in response to the pyrrolizidine alkaloid

seneciphylline *N*-oxide and the iridoid catalpol, but reduced their firing rate in response to the deterrent caffeine, as compared with unparasitized caterpillars; there were no differences in the response to a sucrose control (Fig. 1a) (one-way analyses of variance (ANOVAs): seneciphylline *N*-oxide, $F_{44,1} = 19.448$, $P < 0.0001$; catalpol, $F_{44,1} = 8.1$, $P = 0.007$; caffeine, $F_{44,1} = 4.398$, $P = 0.0424$; sucrose, $F_{44,1} = 0.0045$, $P = 0.947$).

Of 20 *E. acrea* caterpillars tested from Gardner Canyon, 12 were parasitized, and of 57 tested from Box Canyon, 46 were parasitized. Parasitized larvae showed a greater response to the pyrrolizidine alkaloid (two-way ANOVA: seneciphylline *N*-oxide, $F_{75,1} = 11.745$, $P < 0.0001$; population, $F_{75,1} = 0.0354$, $P = 0.851$; *N*-oxide × population, $F_{75,1} = 0.868$, $P = 0.41$) (Fig. 1b). However, the responses to the deterrent protocatechuic acid were reduced (two-way ANOVA: deterrent, $F_{75,1} = 14.62$, $P = 0.003$; population, $F_{75,1} = 1.29$, $P = 0.259$; deterrent × population, $F_{75,1} = 0.359$, $P = 0.551$), and responses to sucrose were inconsistently different (two-way ANOVA: sucrose, $F_{75,1} = 4.438$, $P = 0.039$; population, $F_{75,1} = 0.651$, $P = 0.422$; sucrose × population, $F_{75,1} = 5.296$, $P = 0.024$) (Fig. 1b).

Previous work has indicated that increased gustatory responsiveness in caterpillars to pyrrolizidine alkaloids and iridoid glycosides may cause them to increase their consumption of host plants containing these compounds[10–12]. The reduced responsiveness to deterrent compounds demonstrated here may increase the likelihood that the caterpillar will feed on normally unpalatable plants, which could also provide chemical defence. Other species of generalist arctiid caterpillars are known to sequester opportunistically a wide range of toxins from plants, including ones foreign to their natural habitats[13].

The opposite directions of sensory change for feeding stimulants and deterrents, and the absence of a consistent change in response to the phagostimulatory nutrient sucrose in relation to parasitism, dispel the possibility that we were recording a generalized change in sensory neurons. In *G. geneura*, the sucrose sensitivity lies on the same gustatory neuron as iridoid sensitivity[8], indicating that the change does not affect the whole cell, but rather the particular receptor protein or its second messengers.

Elizabeth A. Bernays*, Michael S. Singer†
*Department of Entomology and Arizona Research Laboratories Division of Neurobiology, University of Arizona, Tucson, Arizona 85721-0036, USA
e-mail: bernays@comcast.net
†Department of Biology, Wesleyan University, Middletown, Connecticut 06459, USA

Estigmene acrea gets a taste for toxins in plants to foil parasites.
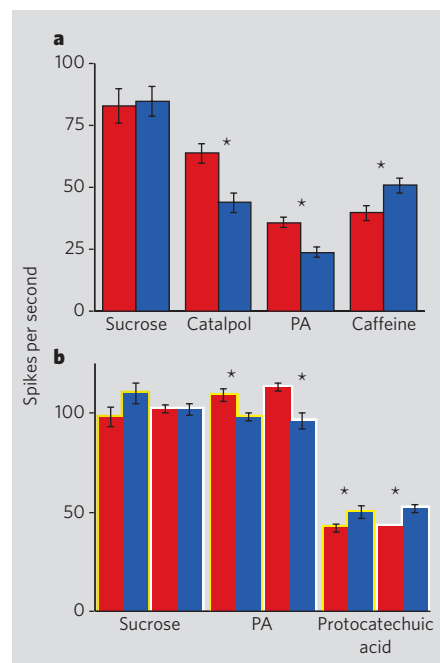
Figure 1 | **Gustatory responses of parasitized and unparasitized *Grammia geneura* and *Estigmene acrea* caterpillars.** **a,** Depolarizations occurring in the first second of stimulation are shown for: a phagostimulant-specific gustatory cell from *G. geneura*, in response to a sucrose control and to the iridoid catalpol; a gustatory cell specific for the pyrrolizidine alkaloid (PA) seneciphylline *N*-oxide; and a gustatory cell specific for the deterrent caffeine; mean and s.e.m. are shown; asterisk indicates significant difference of $P < 0.01$. Parasitized larvae (red bars), $n = 23$; unparasitized larvae (blue bars), $n = 21$. **b,** As in **a**, but in *E. acrea* and using protocatechuic acid as deterrent. Gardner Canyon population (yellow outline): parasitized larvae, $n = 12$; unparasitized larvae, $n = 8$; Box Canyon population (white outline): parasitized larvae, $n = 46$; unparasitized larvae, $n = 11$.

1. Simpson, S. J., James, S., Simmonds, M. S. J. & Blaney, W. M. *Appetite* **17,** 141–154 (1991).
2. Stireman, J. O. & Singer, M. S. *Ecology* **84,** 296–310 (2003).
3. Nishida, R. *Annu. Rev. Entomol.* **47,** 57–92 (2002).
4. Bezzerides, A. *et al. Proc. Natl Acad. Sci. USA* **101,** 9029–9032 (2004).
5. Hartmann, T. *et al. J. Chem. Ecol.* **30,** 229–254 (2004).
6. Bernays, E. A., Chapman, R. F. & Hartmann, T. *J. Comp. Physiol.* **188,** 715–723 (2002).
7. Bernays, E. A., Chapman, R. F. & Hartmann, T. *Physiol. Entomol.* **27,** 312–321 (2002).
8. Bernays, E. A., Chapman, R. F. & Singer, M. S. *J. Comp. Physiol.* **186,** 13–19 (2000).
9. Bowers, M. D. in *Caterpillars: Ecological and Evolutionary Constraints on Foraging* (eds Stamp, N. E. & Casey, T. M.) 331–371 (Chapman & Hall, New York, 1993).
10. Singer, M. S. & Stireman J. O. *Oikos* **100,** 554–562 (2003).
11. Singer, M. S., Carriere, Y., Theuring, C. & Hartmann, T. *Am. Nat.* **164,** 423–429 (2004).
12. Singer, M. S., Rodrigues, D., Stireman, J. O. & Carriere, Y. *Ecology* **85,** 2747–2753 (2004).
13. Rothschild, M. *Biol. J. Linn. Soc.* **12,** 305–326 (1979).

BRIEF COMMUNICATIONS ARISING online
▸ **www.nature.com/bca** see *Nature* contents.

nature
outlook

INDIA

nature outlook INDIA

Reaching for the top

REUTERS/CORBIS

The classic image of India that most people can conjure has cows, beggars, small children and sari-clad women all jostling for space on crowded streets. That image still reflects reality — but with palpable differences.

Along some of those streets now are gleaming, modern buildings where men and women churn out medicines for poor countries. Many children are being immunized with affordable vaccines produced by India's own biotechnology industry. And if the country continues to prosper as it has for the past decade, there soon may not be many beggars left.

Since 1991, when India discarded its socialist past and instituted broad reforms, its economy has been growing rapidly. By 2032, India's economy could be larger than those of all but the United States and China, according to an estimate by the investment banking firm Goldman Sachs.

In the following pages, we look at what effect these changes have had on India's life sciences. Indian biotechnology companies have been remarkably successful, but they have made most of their money copying patented drugs. To sustain growth, they will have to become more innovative. The same is true of basic-research institutes, which have only recently begun to be globally competitive.

While developed nations are seeing their populations shrink, half of India's growing population is under the age of 25. The government must improve India's ailing universities to educate and train these young people and stem the haemorrhage of its brightest students to the United States and Europe. It must also untangle the still-rampant red tape that stifles creativity and set up ethics panels to monitor clinical research.

Most important, it must find a way to protect its people from grave health crises. India has among the highest rates of HIV/AIDS, diabetes and tuberculosis, which threaten to sap the country's resources and derail its progress.

This is a critical juncture for India. The opportunities, whether in partnership with multinational pharmaceutical companies or as an outsourcing centre for clinical trials, are many. Seized judiciously, they can allow India to build a scientific and technological future to be proud of.

Apoorva Mandavilli, Senior News Editor,
Nature Medicine.

p496

features

p488

Then and now

Inder Verma celebrates the recent success of India's biotechnology industry, applauds the increased investment and looks to the future.

"Your job is to collect urine in the men's bathroom and bring it to the laboratory at the end of the day." These were the instructions my thesis adviser gave me when I began my graduate career in 1966. Others would then purify factors from the urine that might function as a male contraceptive in monkeys. He pointed to a 20-litre glass bottle with a broken rim and a small glass funnel. I was expected to obey his orders — and I did. (My only intellectual contribution was to switch to a larger funnel, which made it easier for both the donors and for me.)

A few months later I was told that monkeys were no longer available and the study was being shelved. From now on I would work with a plant researcher in another lab. At no point was I consulted during these arrangements. I was merely a graduate student expected to do as my supervisor commanded. In the years since, the situation has improved, and the overbearing attitude of many advisers is becoming a thing of the past. Young scientists in India are more independent and increasingly resentful of the autocratic system.

For many decades, Indian universities have churned out biologists, many of whom move to the United States or Europe. Those that remain in India are usually destined to publish in local or low-impact scientific journals. Life-science research in India has been constrained by insufficient funds and equipment coupled with a lack of will to compete with the rest of

Inder Verma believes that young scientists must be able to challenge the status quo. As a graduate student (inset), Verma felt that he had to obey all his thesis adviser's commands.

the world. Funding agencies such as the Council of Scientific and Industrial Research have always emphasized translational research but, because the success of such programmes rests on a strong foundation of basic science, the results were often dismal. Indian medical schools, lacking resources for research, have failed to train scientists in translational research. Not surprisingly, India has not produced a single original drug sold in the world market. All this, I hope, is about to change.

I've been to India every year for the past 35 years, have visited many institutions, met many colleagues, students and young investigators and participated in countless meetings. In that time, I've witnessed the growth of Indian biosciences with concern and pride.

Recently, I've been delighted by a run of excellent papers published in top-tier journals by scientists in India. And I've been impressed by the desire of the biotech and pharmaceutical industry to undertake novel challenges.

The first prime minister of India, Jawaharlal Nehru, believed that science is the way out of poverty. The scientific opportunities in India may well prove him to be a true visionary, but there is much work to be done first.

Novel ideas

Biotechnology companies are among the most profitable in the country. Today, Indian companies provide 22% of the world's generic drugs — copies of brand-name drugs. India also manufactures a significant proportion of vaccines made for the developing world.

Science has always been important to the Indian government. It's first prime minister, Jawaharlal Nehru, greatly admired Albert Einstein. Current leader Manmohan Singh has made science a priority.

Within India, vaccines against hepatitis B are manufactured by Indian companies and sold for less than 30 cents per dose. True, most of India's biotech and pharmaceutical industry is based on duplicating existing products, but that still requires considerable sophistication and know-how. But with the passage of a new law on 23 March this year forbidding companies from making copycat drugs (see page 480), Indian companies will be forced to be more innovative.

But how do we encourage such innovation? In the mid-1980s, the Indian government established the Department of Biotechnology (DBT) and invited several researchers —including me — to serve on its overseas advisory committee. We encouraged the DBT to furnish state-of-the-art laboratories, provide increased funds for research, reduce the bureaucracy required for importing research materials such as restriction enzymes, initiate projects that promote the sharing of expensive equipment, offer visiting fellowships and build manpower to support the biotech industry.

Since the DBT's inception, its budget has grown from a meagre US$15 million to more than US$125 million a year. That may not be much compared with the US$27 billion budget of the US National Institutes of Health, but it's a good start.

Increased investment
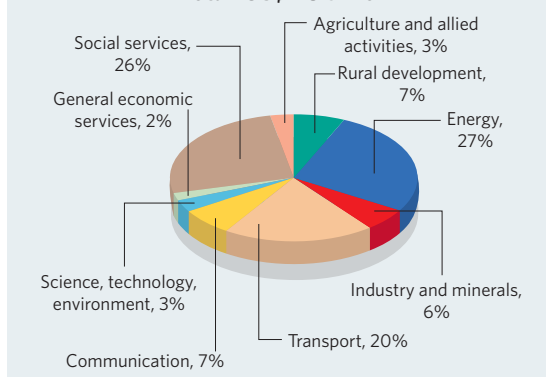In addition to the DBT, there are at least three major funding sources for life scientists within India, so finding funds for a good project is no longer a hurdle. Indian scientists have been able to win significant funding from international sources such as the US National Institutes of Health, the Bill & Melinda Gates Foundation, the European Union and the Wellcome Trust (see page 489).

Despite this, India will not succeed unless it encourages innovation and rewards excellence. Most Indian universities still operate under a feudal system, which stifles creativity. Science is best carried out in an irreverent environment, where the status quo is challenged, often at the risk of offending superiors. But the Indian scientific enterprise frowns on questioning authority and rewards obedience. Senior scientists are too often selected by seniority and rank, rather than their ability and achievements.

In common with many scientists at institutes in neighbouring China, Indian researchers once hired are there to stay. I do not know of any case in which someone was denied tenure or did not have their contract renewed because of low productivity. This

INDIAN GOVERNMENT BUDGET, 2005–06
Total: US$118 billion

Social services, 26%
General economic services, 2%
Science, technology, environment, 3%
Communication, 7%
Transport, 20%
Industry and minerals, 6%
Energy, 27%
Rural development, 7%
Agriculture and allied activities, 3%

problem is compounded by the lack of lateral mobility in India. Most scientists who train abroad return to India for family reasons and are loath to live away from their home town. They would sooner go back abroad than swap cities or states within India.

But science thrives when there is a nucleus of scientists striving for excellence. The Indian Institute of Science in Bangalore and the National Institute of Immunology in New Delhi have achieved this. Scientists at these institutes, which are independent of universities, are publishing in international journals and are being invited to prestigious national and international meetings. To be competitive on a global scale, India needs to nurture such centres rather than worry about equitable distribution of the country's resources.

These centres can also encourage collaborations across cities and disciplines. Modern biology requires the expertise of scientists from many different fields. Unfortunately, scientific departments at Indian institutions and universities have traditionally operated as islands. Experts in microbiology, biophysics and biochemistry might all work within a few feet of each other and yet hardly interact, much less collaborate. The government should provide incentives for interdisciplinary research. It should also encourage academics to forge alliances with industry, market their inventions and set up technology-transfer offices.

Government officials are always coming up with catchy slogans, such as 'IT today, BT tomorrow', but often they do not follow through on those ambitious plans. India is still far behind the United States, Europe and Japan. Although India has the advantage that its citizens know English, it still lags behind other Asian countries such as China and South Korea.

India sometimes gives the impression that it will never compete properly in the global arena. But life scientists will be able to

if they can learn from the success of the IT sector.

India's position in the global IT industry is a source of pride and confidence. Graduates of the Indian Institutes of Technology (IIT) have started numerous companies worldwide, and their innovative spirit should inspire biologists to be more adventurous.

Last year, *60 Minutes*, a popular investigative television programme in the United States, featured the IITs. When the interviewer asked N. R. Naryana Murthy, founder and chairman of the multibillion-dollar company Infosys, what his son would do if he did not pass the arduous entrance examination for the IITs, Murthy replied, without skipping a beat: "Well, we have back-ups like Cornell, MIT and Stanford."

I don't usually support starting new institutions, but an Indian Institute of Integrative Biology, much like the successful IITs, is worth considering. Alternatively, the IITs could broaden their curriculum to include degrees in biology and biotechnology.

Promoting science
The time is ripe for life sciences to blossom in India. Every day, newspapers carry headlines reporting Indian successes in information technology, tales of rich Indian biotech tycoons and highlighting the enormous purchasing power of the growing middle class. In my experience, the Indian government, regardless of which party was in power, has always been highly supportive of science. But the current left-leaning administration is particularly so (see graphic, above). For example, the government recently gave the Indian Institute of Science a one-off grant of about US$25.3 million. Prime minister Manmohan Singh has also endorsed the creation of a US$250-million independent agency to support basic research.

India is not a poor country — indeed, it is rich in natural and intellectual resources — but it has many poor people. More than 700 million people, nearly 70% of the population, live in rural areas but contribute only 20% of the GDP. Until this disparity changes, vast sectors of the Indian population will never see the benefits of biotechnology or modern medicine. It's very clear what India needs to do to become a world player in the life sciences, but unless the government, researchers and the industry work together to put the solutions into practice, all their best laid plans will not succeed. ■

Inder Verma is American Cancer Society Professor of Molecular Biology at the Salk Institute, La Jolla, San Diego

Biotech boom

India's thriving biotechnology industry is threatened by a change in the law. Will the current high levels of investment be enough to secure its future? **K. S. Jayaraman** finds out.

Every year, on the presumed birthday of Krishna, the blue-skinned Hindu god, New Delhi's Mathura Road is choked with traffic. Buses, cars and bullock carts packed with devotees advance toward Vrindavan, where Krishna is believed to have been born 5,000 years ago.

This August, pilgrims to Krishna's legendary birthplace will pass an imposing three-story structure of sandstone and glass. Inside, at The Centre for Genomic Applications (appropriately shortened to TCGA, the letters of the genetic alphabet), white-coated scientists are devoted to a god of a different kind. On the centre's first birthday in May 2005, scientists acquired a new Hewlett-Packard supercomputer, the fastest in India, worth a whopping US$2.2 million. Capable of four trillion operations a second, the computer places the centre among the ranks of international institutions such as the Wellcome Trust Sanger Institute in Britain and the Institute for Molecular Science in Japan.

TCGA is the youngest of several biotechnology laboratories that have sprung up in India in recent years. It is a shared venture between the Chatterjee Group, a Kolkata-based industrial house, and two government agencies — the Council of Scientific and Industrial Research (CSIR) and the Department of Science and Technology — which shared the US$ 5.7-million cost of construction.

The centre, which aims to provide world-class facilities for genomics and proteomics, is the first to be built by a public–private partnership. It also symbolizes the current revolution in India's biotechnology industry.

With fewer than 300 registered companies, the biotechnology sector is small but is gaining in global stature. According to the World Health Organization, India is the fourth largest producer of pharmaceuticals, and 66.7% of its exports go to developing countries. For example, the Pune-based Serum Institute of India, once a small manufacturer of tetanus toxin, now makes 80% of the world's measles and DTP (diphtheria, tetanus and pertussis) vaccines (see 'A shot of success', opposite).

Other companies that have made a global impact include the Mumbai-based company Cipla, which rewrote the rules of the international AIDS drug market when, in 2001, it introduced inexpensive generic antiretroviral drugs. Strand Genomics in Bangalore has formed a partnership with Japan's MediBIC to develop informatics solutions for pharmaceutical companies in Japan. And Avesthagen, also based in Bangalore, has joined up with France's bioMérieux to develop diagnostic instruments for medical and industrial applications.

Profits soar

In 2003–04, Indian biotechnology companies together had a revenue of more than US$700 million (see graphic, below). This year, they have surpassed US$1 billion. "The biotech-

LOCATION OF TOP 20 BIOTECHNOLOGY COMPANIES

- 1 company (17) — New Delhi
- 2 companies (3, 15)
- 7 companies (5, 8, 9, 10, 11, 16, 19)
- Bulandshahr
- Calcutta
- Mumbai
- Pune
- Hyderabad
- Attur
- 2 companies (2, 4)
- 4 companies (12, 13, 18, 20)
- Bangalore
- 1 company (7)
- 3 companies (1, 6, 14)
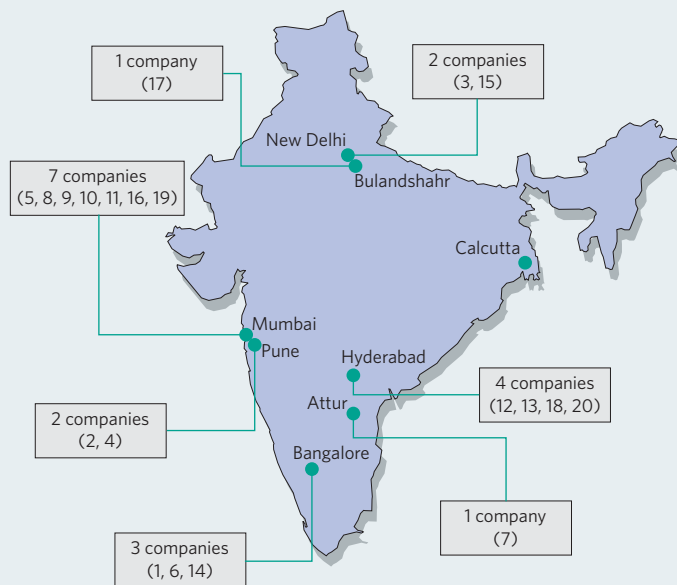
INDIA'S TOP 20 BIOTECHNOLOGY COMPANIES

Company name	Revenue 2003–04 (US$million)	Revenue 2004–05 (US$million)	Growth (percentage)
1. Biocon	115	148	29
2. Serum Institute of India	126	130	3
3. Panacea Biotec	34	50	46
4. Venkateshwara Hatcheries	20	43	114
5. Mahyco Monsanto	12	38	208
6. Novo Nordisk	25	31	23
7. Rasi Seeds	—	20	—
8. Aventis Pharma	17	19	15
9. Bharat Serums & Vaccines	18	19	2
10. Chiron Behring Vaccines	15	18	16
11. GlaxoSmithKline	14	18	28
12. Indian Immunologicals	13	17	28
13. Shantha Biotechnics	9	16	75
14. Novozymes	12	16	30
15. Eli Lilly & Company	15	16	1
16. Wockhardt	9	15	68
17. Bharat Immunologicals & Biological	5	12	147
18. Bharat Biotech	8	9	14
19. Advanced Biochemicals	7	9	31
20. Biological E	9	8	–6

Data from the ABLE-*Spectrum* Industry survey
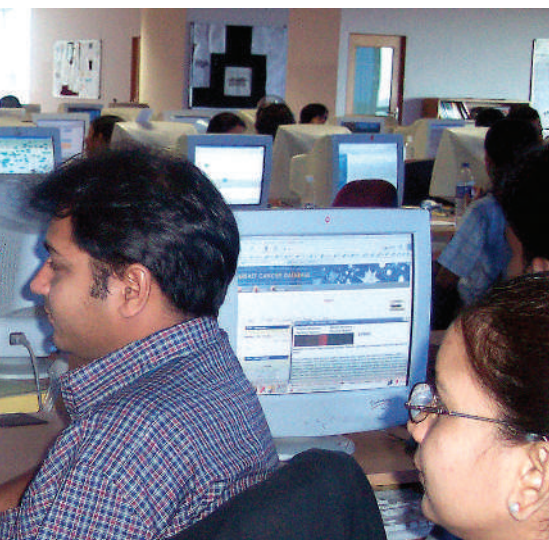
World class: researchers from the Institute of Bioinformatics in Bangalore managed to characterize the human X chromosome.

nology sector is witnessing an impressive 40% annual growth," says Kiran Mazumdar Shaw, chief executive of Biocon, Bangalore. "The profile of Indian biotech companies is undergoing a change and they are becoming international." Biocon, the most profitable biotechnology company in India, this year posted revenues of US$150 million, a 30% increase over the previous year.

Shaw, sometimes called India's 'biotech queen', and several other company chiefs are quickly becoming household names in India. Five of them won this year's Padma awards, the nation's highest honour for civilians. "Never in the history of these coveted awards has a small section of the industry grabbed the entire limelight," says N. Suresh, editor of *BioSpectrum* magazine, published in Bangalore.

The biggest boost to the biotechnology industry has come from the government itself. "Biotech is the government's priority area," says science minister Kapil Sibal. Less than a year after Sibal took office, the Department of Biotechnology (DBT) released an ambitious plan to create a biotechnology industry that would generate US$5 billion in revenues per year and create one million jobs by 2010.

As part of its strategy, the DBT is planning to make it easier for foreign-owned companies to set up in India. Foreign investors have in the past had to knock on the doors of several different government agencies. But the DBT's new plan is to set up a single independent authority to replace the committees at different ministries.

The DBT has also subsidized the construction of three biotechnology parks, including Genome Valley (see 'Land for rent', overleaf) and aims to help finance at least ten such parks by 2010. Together with the Ministry of Information Technology, the DBT plans to build the country's first biotech/IT park. This

A SHOT OF SUCCESS

On a sunny Friday morning in January, the students of the Shiva Shivani School in Hyderabad were ushered, one by one, into a room where a doctor and nurse stood with syringes.

The school's 470 students, ranging in age from 5 to 12 years, all received a shot of Shanvac-B, a hepatitis B vaccine produced by a local company, Shantha Biotechnics. At just Rs18 per dose (equivalent to US 40 cents), the vaccine is so affordable that scenes like this are commonplace.

Such widespread vaccinations were unthinkable early in 1997 when GlaxoSmithKline (GSK) sold its product at Rs500 per dose. But later that year, Shantha's hepatitis B vaccine became India's first recombinant product to be made indigenously in the healthcare sector. Today this vaccine is available in 52 countries and makes up 40% of UNICEF's supply.

Varaprasad Reddy says he launched the company, named after his mother Shantha, after he heard multinational companies disparage India's dependence on them for its vaccine needs. "India can position itself as an affordable base for producing vaccines for the entire developing world," he says.

Shantha's success inspired four other Indian companies to bring out their own hepatitis B

vaccine. As a result, GSK's share of the vaccine in the Indian market has plummeted to 10% from a virtual monopoly (see graph, below). The Serum Institute, based in Pune, is the world's largest manufacturer of the diphtheria, tetanus and pertussis vaccine, which now protects three out of ten of the world's children, according to figures released by UNICEF. The Mumbai-based Haffkine Institute supplies 75% of India's oral polio vaccine, and Indian Immunologicals in Hyderabad, already the world's largest manufacturer of vaccines for foot-and-mouth disease, plans to make human vaccines against rabies, hepatitis B and measles.

"Not many people are aware that the United States is one of the largest importers of measles vaccine from India," says Nirmal Kumar Ganguly, director-general of the Indian Council of Medical Research.

But Indian companies are only beginning to venture into original territory.

"It is true we have not made any new vaccine on our own,"

says Krishna Ella, managing director of Bharat Biotech.

To encourage innovation the Department of Biotechnology is supporting the development of 11 original vaccines. Four of these, for rotavirus, cholera, Japanese encephalitis and malaria, will enter clinical trials this year.

Outside of India, the number of companies producing vaccines has dwindled to 4 in 2005 from 26 in 1967. This reduction presents an opportunity for India and other developing countries, says Ganguly.

The Global Alliance for Vaccine and Immunization, launched by the Bill & Melinda Gates Foundation in 2000 is purchasing vaccines from Indian companies at low prices. In return, these companies are assured a long-term contract to guarantee revenue. The alliance has awarded contracts to the Serum Institute, Shantha Biotechnics and Panacea Biotec to make and test heat-stable vaccines.

K. S. Jayaraman

NATIVE COMPANIES DOMINATE INDIA'S PHARMACEUTICALS MARKET

LAND FOR RENT

K. S. JAYARAMAN

Beyond the gateway of the Biotech Park near Hyderabad the land lies empty. Companies are reluctant to build there until the infrastructure is complete.

"Welcome to Genome Valley". The giant green-and-white sign stretches across the road on the way from the Hyderabad airport to this bustling south Indian city. But don't blame the taxi driver if he meets your queries with a blank look — almost no one knows where the valley is.

Not too long ago, Hyderabad was dubbed 'Cyberabad', reflecting its strength in information technology. These days, it's aiming for pastures new: the biotechnology sector. For now though, the 'genome valley' exists only in the minds of those who conceived it eight years ago.

The name refers to a 600-square-km area bound by an imaginary line that links nearly every institution in the city that has anything even remotely to do with science.

"I actually suggested that the whole of Hyderabad should be declared Genome Valley," says Bhim Sen Bajaj, chairman of the southern chapter of the All India Biotech Association. "They did not listen to me."

In reality, Genome Valley is the Knowledge Park, 40 km from the city on 80 hectares of government land and set up by ICICI, a premier financial institution. There is no public transport to the area and, at one point, cars have to snake through muddy roads, occasionally interrupted by cows and hens.

The park offers modern modular laboratory units of approximately 300 square metres to rent. In the past six years, 13 companies have rented spaces, and an additional 7 have booked space.

A second area of the valley is the Biotech Park, which is intended for companies to use for manufacturing units. This park, which opened in 2001, has an impressive arched entrance to its 126 hectares — but it is almost empty. Officials say 19 companies have purchased plots and 7 have begun construction, but the promised infrastructure for these companies is not yet in place. The water pipe was laid only a few months ago.

To make matters worse, builders are selling the land at about US$112,000 per hectare, nearly ten times the government's original price.

"I had no choice," says S. P. Vasireddi, chairman and managing director of Vimta Labs, who is building a US$10-million facility in a 4.5-hectare plot. "Land of this size is not available inside the city."

For the moment, the valley has one undisputed king: Bharat Biotech, built by US-educated microbiologist Krishna Ella and his wife Suchitra. "Whenever a government guest wants to visit the genome valley, he or she is brought to our company," says Seema Kumar, general manager of public relations for the company.

Bharat Biotech, which earned US$10 million in 2004, produces vaccines, but the most used product at the company's headquarters may not be one that it manufactures itself. Instead it's the anti-snake venom stocked in the first-aid kit. "The place is crawling with snakes," says the company's communication manager U. V. L. Ananda.

To help employees with the three-hour daily commute, Bharat Biotech operates shuttle buses to the city. In the meantime, there is no hospital, bank or police station nearby. ICICI bank is noted for its cash machines, but it has yet to set one up in its own Knowledge Park.

Bharat Biotech buys water from private tankers and generates its own electricity to operate its vaccine unit because the unit in nearby Turkapally village crashes at the slightest environmental disturbance, such as rain or lightning. The company has gleaming state-of-the-art technology, but no broadband Internet connectivity and, until recently, workers there could not use their mobile phones. Says one employee: "We do cutting edge science in such a primitive setting."

K. S. Jayaraman

is expected to attract bioinformatics contracts from around the world and foster innovative companies.

"We not only want to build on the existing platform but expand the base to create global leadership in biotechnology," says Maharaj K. Bhan, secretary of the DBT. "This will require larger investments."

The science ministry has already announced a 50% increase in its budget over the past year for drug-discovery research and called for proposals from the industry.

To encourage small businesses, the DBT gives out grants of Rs5 million (US$115,000) for proof-of-concept research and low-interest loans for subsequent product development and commercialization. Any money that companies spend on maintaining patents is also exempt from income tax.

Patent crackdown

One reason for this increased investment is that from 1 January 2005, a new patent law, which brings India in line with World Trade Organization (WTO) rules, came into effect. Indian companies will have to honour international patents and stop producing unlicensed generic drugs, a major source of their revenue over the past 30 years (see graphic, above).

The pharmaceutical industry's research and development spending has shot up from Rs2 billion in 2000 to Rs8 billion in 2004, says D. G. Shah, secretary-general of the Indian Pharmaceutical Alliance. Chennai-based Orchid, which earned US$82 million in 2004, is building a US$15-million plant. Biocon is adding a US$170-million research lab and manufacturing plant to its assets in

GROWTH IN DRUG PRODUCTION IN INDIA
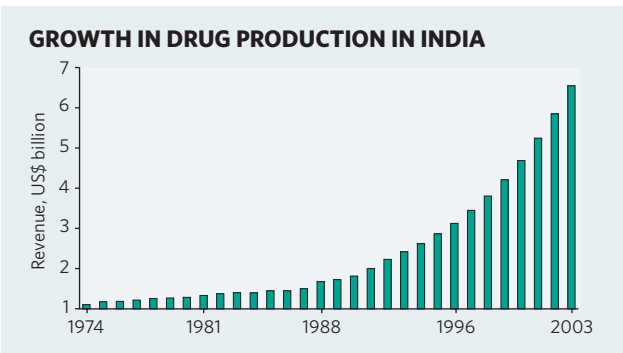
Bangalore. And Wockhardt, a large pharmaceutical company, last year built a US$50-million complex — India's largest — in Aurangabad. Nicholas Piramal India, which opened its US$25-million research facility in September 2004, says it already has eight original products, including potential treatments for cancer and diabetes, in preclinical tests. Other generic producers such as Dr Reddy's Laboratories in Hyderabad and Delhi-based Ranbaxy Laboratories are moving along this innovation route even as they continue to scout for patent-expired generics in the US market.

Still, "the odds are not in India's favour in the innovation driven global drug markets", says Prasanta Ghosh, a consultant to Cadila Pharmaceuticals in Ahmedabad and former adviser to the DBT. "We are nowhere compared with China, South Korea, Cuba or Brazil," he says. "Our research has not gone into product development."

Local companies will also have to compete on Indian soil with powerful multinational companies, others warn. Companies such as Merck and Bristol-Myers Squibb left India years ago because cheap copycat drugs cut into their profits. But with the WTO rules enforced to protect their patents, those companies will return, says Yusuf Hamied, Cipla's chairman. "Multinationals will invade India in force and wipe us out in five years," he warns.

Companies face other obstacles too. "Biology is extremely technology-driven and yet we do not make any instruments in this country," says Syed Hasnain, director of the Centre for DNA Fingerprinting and Diagnostics in Hyderabad. "Except for flasks and syringes, every instrument is imported. We do not trust the local centrifuges, so we import even these." Most institutes also have limited access to good animal-research facilities.

Another problem is the shortage of people trained in the latest techniques. India's role in information technology is supported by the massive number of software experts in cities such as Bangalore, but such a pool of biotechnologists is not available.

When the Bangalore-based Institute of Bioinformatics (IOB) needs someone who knows the computer programming language Oracle, it can find a good candidate within a week. That's because there are so many good programmers in the Industrial Technology Park, where the institute is based.

Akhilesh Pandey, a professor at the Johns Hopkins University, founded the IOB in 2002. In three years, it has published more than 40 papers, including several in high-impact journals. In April 2005, the IOB published its annotation of the X chromosome (H. C. Harsha *et al. Nature Genet.* **37,** 331–332; 2005). The study, carried out in direct competition with Britain's Wellcome Sanger Institute, marks the first time that any institute other than a genome-sequencing centre has characterized an entire human chromosome.

Training biotechnologists

With a booming biotechnology business and a demonstrated prowess in information technology, there should be a dozen such bioinformatics success stories in India. But even the big Indian pharmaceutical companies have been slow to adopt bioinformatics.

Although the country grants nearly 300,000 degrees and diplomas in biotechnology, bioinformatics and the biological sciences each year, companies struggle to find skilled staff. "Most of them get their degrees without seeing a biotech lab," says Krishna Ella, managing director of Bharat Biotech International in Hyderabad. Most of those qualified also leave for greener pastures: up to 90% of those who finish their PhDs at the Indian Institute of Science go abroad. India needs to find ways to stem that massive brain drain, says Ella.

In the meantime, companies are wooing scientists to India from around the globe. According to estimates from the Indian Pharmaceutical Alliance, 10% of the new recruits at senior levels are expatriate Indians or foreigners.

Geetha Vani Rayasam, an IISc graduate, joined the Ranbaxy Research Labs in November 2004 after spending several years in Europe and the United States. Rayasam says many more like her would return to India if a few essential things were fixed. "There are several things that need to be improved, like a more professional approach, less bureaucracy and providing better salaries," she says. "But the day is not far off when India might be leading the way in drug discovery." ■

K. S. Jayaraman is *Nature's* **India correspondent.**

A. MANDAVILLI

Science minister Kabil Sibal is backing a plan to create a thriving biotechnology sector.

Vaccines on trial

One of the largest ever vaccine studies is under way in Kolkata. **Paroma Basu** uncovers the benefits, and difficulties, of inoculating 60,000 people against cholera and typhoid fever.

A rhyming couplet written in Bengali invites Kolkata's residents to receive free vaccinations.

Famed for its poets and filmmakers, disgruntled Marxists and sleepy government clerks, Kolkata, a city of about 14 million people in northeastern India, may seem an unlikely setting for one of the largest clinical trials in the world.

In the poorest areas of this city, residents live in homes jammed together along winding sewage-littered pathways and rely on shared toilets and drinking water. Typhoid fever and cholera are endemic in India, and are chronic problems in Kolkata. Cholera, in particular, has a tenacious grip on the state of West Bengal, often called the 'homeland of cholera'.

The vaccine industry has been reluctant to commit resources to the development of vaccines for some of the world's poorest people. But a grant of US$40 million from the Bill & Melinda Gates Foundation is helping to introduce affordable vaccines to cities such as Kolkata. The money has funded the five-year Diseases of the Most Impoverished (DOMI) Program, an initiative of the South Korea-based International Vaccine Institute (IVI).

DOMI is studying the social, economic and clinical effects of introducing vaccines into six other countries: Pakistan, Bangladesh, China, Indonesia, Vietnam and Thailand. Since 2000, it has launched two cholera studies, six projects investigating typhoid fever and another six exploring shigellosis, a common bacterial disease. The work is unprecedented, says John Clemens, the IVI's director, not only because of the project's size but also because it focuses on a demographic long ignored by drug companies, lawmakers and public-health officials.

In a unique research effort, 60,000 Kolkata slum-dwellers will participate this summer in phase III trials of an oral cholera vaccine. Last November, researchers injected the same population with a vaccine against typhoid fever.

The typhoid fever vaccine was developed at the US National Institutes of Health, and was donated by GlaxoSmithKline. The cholera vaccine is modelled on a widely available oral vaccine, Dukoral, which was developed in Sweden during the 1970s.

But at several dollars a dose, Dukoral is too expensive for most developing countries. Vietnamese scientists licensed the technology in the early 1990s and formulated a cheaper version at about 20 cents per dose.

With the IVI and Kolkata-based National Institute of Cholera and Enteric Diseases (NICED), Shantha Biotechnics, an Indian biotechnology company, is developing a local version of the Vietnamese vaccine and has procured regulatory clearances for the Kolkata trials. It has also secured the right to market the vaccine if it is approved, says Raman Rao, head of clinical research for Shantha.

Meanwhile, the roadblocks encountered during these trials in Kolkata are an example of the difficulties of carrying out such a programme, from political and religious tensions and bureaucratic delays to mistruths spreading like wildfire among the largely illiterate trial participants.

The project got off to a smooth start in 2002, with Kolkata epidemiologists performing a year-long surveillance study before identifying two slums that were particularly hard-hit by cholera and typhoid fever. But the next step was a bureaucratic nightmare, says Dipika Sur, deputy director of epidemiology at the NICED. The institute had to get an endless list of clearances from, among others, the national health ministry committee, local councillors, ethics and human rights groups, Hindu priests, Muslim imams and community thugs.

But the clearances were easy compared with gaining the confidence of the study participants. During the typhoid vaccine trials, for instance, rumours spread that scientists were injecting cancer cells into people. Others believed they were being sterilized. "There was mass panic," Sur recalls.

About 65% of the targeted study group eventually gave their informed consent and received the typhoid jab. Sur thinks a big reason for this level of success was her tactic of employing 250 slum-dwellers to serve as the 'face' of the study, working as community health workers, field supervisors and sample collectors.

The strategy paid off largely because of staggeringly high unemployment levels in the slums. Montu Chandra Das, who was previously unemployed, now earns roughly US$60 per month as a health worker. "It feels good to help others," he says.

Das goes door-to-door sending patients with persistent symptoms of diarrhoea or fever to one of seven 'health outposts', where patients receive free blood tests and medicines if diagnosed with cholera or typhoid.

"I like the way that the health workers come to check on us all the time," says Kamala Das, a 35-year old domestic helper who is taking part in the study. "Now we can see a doctor and be treated right away," adds another study member, Gopal Balmiki, a car driver who shares a room with 11 family members. "We know now that it's really important to keep the bathroom clean and drink unpolluted water."

Paroma Basu is a freelance writer based in Madison, Wisconsin.

India's drug tests

Drug companies are converging on India to conduct low-cost clinical trials. But is it ready to become the outsourcing centre for the world? **T. V. Padma** investigates.

The average cost of a clinical trial in the United States is US$180 million. The average cost in India is US$100 million. No surprise, then, that multinational companies are flocking to India to launch their trials. "It's a lot cheaper to do things in India," says Sameer Deb, general manager of government affairs at Glaxo-SmithKline's office in Mumbai.

Multinational companies are not the only ones to benefit financially from these studies — India does too. The consultancy firm McKinsey estimates that US and European pharmaceutical companies will spend US$1.5 billion per year on clinical trials in India by 2010.

India has several advantages as a host for such trials. Its biggest asset is probably the size of its population at more than 1 billion. In addition, Indians are increasingly suffering from the same illnesses as Americans and Europeans — diseases for which companies are desperate to find cures. For instance, at least 70 million Indians suffer from heart disease and 35 million have diabetes. It also has the edge over most developing countries because of its sophisticated hospitals and because many of its medical personnel speak English.

But India does not have the robust infrastructure needed to sustain this expansion. It does not even have a central database listing all the trials currently under way, despite each one having to be cleared by the Drug Controller General of India.

In addition, contract research organizations (CROs) are struggling to recruit enough participants to its trials, despite India's enormous population. An even greater obstacle is the lack of trained staff. Trials of this size will need an estimated 3,000 investigators, 600 medical institutions and 9,000 other professionals — numbers India cannot produce at present. Some CROs, in partnership with educational institutes and trial sponsors, are addressing this by setting up specialized institutes and centres of clinical excellence.

The bureaucracy-ridden Indian system is no small barrier, either. Officially, the government has limited the time it takes to accredit trials to 90 days for phase I, 45 days for phase II and 45 days for phase III trials. But "we are yet to see such speedy approvals in reality", notes Dhananjay Bakhle, director of regulatory affairs at Aventis Pharma, based in Mumbai.

The thorniest issue by far may be the ethics of conducting these trials. In one instance, scientists at Johns Hopkins University in Baltimore, Maryland, and the Regional Cancer Centre in Thiruvananthapuram, India, tested two experimental anticancer molecules on 27 oral cancer patients from 1999 to 2000 (see *Nature* **412,** 466; 2001). But the researchers tested the drugs without the required federal or university approvals and without adequate preliminary tests in animals (see *Nature* **414,** 835; 2001).

Following this incident, the Indian government ordered a full review of the ethics and safety of all trials. In 2000, the Indian Council of Medical Research (ICMR), the country's premier agency for biomedical research, issued guidelines for research on humans. The following year, the health ministry released its policy on good clinical practice.

But scandals continued to surface. In 2003, private clinics across India used a generic version of the anticancer drug letrozole to treat more than 435 women with fertility problems. This trial did not have clearance from the health ministry, and the women involved did not know that the drug was not approved for this use. The manufacturer, Mumbai-based Sun Pharmaceuticals, denies ordering this trial.

Government agencies concede that the regulations need to be tightened. "We are looking into strengthening institutional ethics committees and are conducting training workshops for them," says Vasantha Muthuswamy, deputy director-general of the ICMR.

Tightening the regulations will protect the trial participants, but it may be possible to go further. "Participants should benefit from any drugs that result from the trials," says Prasanta Ghosh, president of the biotechnology division of Cadila Pharmaceuticals, based in Ahmedabad. In which case, expansion of clinical trials in India could benefit everyone. ■

T. V. Padma is a freelance writer based in New Delhi.

India's enormous population makes it an attractive destination for clinical drug trials.

C. STOWERS/PANOS

Ayurveda

Science and business are racing to tap the 3,000-year-old system of medicine for new drugs, says **T. V. Padma**.

Ayurvedic herbs (left) — numbered at more than 10,000 by one source — are made into remedies that are sold across India (above).

I n November 1987, ethnobotanist Palpu Pushpangadan was trekking through the tropical forests in the southern Indian state of Kerala. Pushpangadan and his colleague were struggling to keep up, but their guides, men from the local Kani tribe, kept popping brownish-black fruit about the size of a cardamom pod into their mouths, and walking briskly ahead. Curious, the scientists ate the fruit and felt a surge of energy.

Locally known as *arogyapacha*, meaning 'evergreen health', the fruit, *Trichopus zeylanicus*, is noted in ancient Indian texts as a source of health and vigour. With the tribe's help, Pushpangadan and his colleagues studied the plant extracts, isolated the active ingredients and developed an energy-boosting drug, Jeevani, combining *T. zeylanicus* with two other herbs mentioned in the texts.

The team patented the drug and granted a manufacturing licence to the Indian company Arya Vaidya Pharmacy. Sales now garner nearly US$230,000 a year. Under an agreement signed in 1995, the Kani got a 50% share of the US$50,000 licence fee, and have earned nearly US$5,000 a year in royalties.

The Jeevani story is one often told in India as a success both in scientifically validating traditional knowledge and in setting up a model that benefits the local population.

Updating traditional remedies

Today, several laboratories and private companies are racing to tap the ancient Indian system of medicine, Ayurveda, for new drugs. The government's Council of Scientific and Industrial Research (CSIR) and the health ministry are jointly digitizing Ayurvedic knowledge and translating the information into English, Spanish, German, French and Japanese to forestall contentious patent applications.

"India can benefit enormously if it can build a golden triangle between traditional medicine, modern medicine and modern science," says Raghunath Mashelkar, CSIR director-general. The CSIR has funded three Ayurveda-based projects seeking drug candidates for arthritis, type-2 diabetes and liver disease.

The Ayurvedic system of medicine is nearly 3,000 years old and prescribes remedies for a range of problems, from diarrhoea to contraception. The Sanskrit classics *Atharvaveda*, *Charak Samhita* and *Sushrut Samhita* together describe more than 700 medicinal herbs, cataloguing everything from their taste, appearance and digestive effects to safety, efficacy, dosage and benefits.

For example, the texts refer to neem (*Azadirachta indica*) as an antidiabetic, *Withania somnifera* or the Indian ginseng as an antitumour agent, and curcumin — the active ingredient in the spice turmeric — as an anti-infective, anti-inflammatory and antidiabetic. Scientists are following intriguing leads that curcumin might have antimalarial activity and inhibit the replication of some viruses.

"The Ayurvedic database in classical texts can be used for bioprospecting and new drug discovery," says Bhushan Patwardhan, coordinator of the CSIR's arthritis Ayurveda project.

But that task is easier said than done. Some species have become extinct. Time and environmental factors, as well as pesticide and heavy-metal contamination of soils and waters, may also have wrought subtle, but important, alterations in the plant products.

In some cases, there is simply not enough raw material to meet market demand. For example, the Central Drug Research Institute in Lucknow developed an anticholesterol drug, guggul, from the resin of the tree *Commiphora mukul*. The institute patented the drug in Europe and the United States, and in 1987 licensed it to a Mumbai-based company. The drug brought in about US$80,000 in sales per year, but in 1999, the company was forced to stop making it because the slow-growing tree failed to produce fast enough. The institute has since developed a synthetic version of the drug, which is set for phase III clinical trials in June.

Because plants, in adapting to local conditions, produce the substances at different stages of growth, the herbs must be cultivated and collected in a certain way. The original texts described the specific location, time of year and developmental phase of the plant, and this knowledge was passed down orally. "The best collection practices have eroded with the passage of time," says Pushpangadan.

With the herbs in hand, scientists must correctly identify the active ingredient, standardize the manufacturing process, and conduct clinical trials to test their effectiveness.

Data mountain

There is no centralized database of the herbs under investigation, but one database of tribal traditional knowledge at the National Botanical Research Institute in Lucknow catalogues more than 10,000 plant species. Of these, fewer than 200 have so far been investigated.

Ayurveda-based drug discovery uses 'reverse pharmacology', in which drug candidates are first identified based on large-scale use in the population, then validated in clinical trials. Experts say this approach can cut the time for drug discovery from 12 years to 5 years or less, and for a fraction of the usual cost.

Unlike Jeevani, most herb-derived drugs are widely used across India and cannot be traced back to a particular community. In general, there are few guidelines to ensure that the local population benefits from these drugs, notes Darshan Shankar, president of the Foundation for Revitalisation of Local Health Traditions in Bangalore.

At the moment, even the Kani agreement has problems. The licence agreement expired last year and has not yet been renewed. ■

T. V. Padma is a freelance writer based in New Delhi.

Breathing life into biology

Mriganka Sur says that life sciences will prosper in India once research and teaching reconnect.

When I was a teenager, I thought I had only two choices — to be a doctor or an engineer. Growing up in Allahabad in the northern state of Uttar Pradesh, students like me were so frequently asked to choose between these two careers that most of us never considered any other options. Becoming a biology researcher did not figure on the list.

There is no question that to capitalize on the life-sciences and biotechnology revolution, India must build excellence in life-science training. In most countries, this is the task of universities. Unfortunately, the Indian university system is in serious decline — acknowledged to be so both within India and elsewhere — and unequal to this task.

As it turns out, I eventually became a neuroscientist, driven mainly by my interest in understanding how the brain works. I never had a university-level course in biology — there were simply none to be taken at the Indian Institute of Technology (IIT) in Kanpur, where I studied electrical engineering as an undergraduate in the early 1970s. But in a sign of the increasing importance of the life sciences, the IITs are beginning to embrace biology. When I went back to give a lecture at the IIT Kanpur a few years ago, I was asked to advise the institute on its newly created Department of Biological Sciences and Bioengineering and a new programme in cognitive sciences.

At the IIT Kanpur, I attended classes taught by excellent teachers and had a few hands-on projects. But the IITs are largely an exception. Most Indian universities are ill-equipped to tackle the complex, interdisciplinary nature of modern biology. Faculty members at both undergraduate teaching colleges and universities offering advanced degrees are largely concerned with teaching, and tend to focus more on theory than on experimental science. But as most scientists can attest, research and teaching are inseparable components of a modern science education.

Many Indian universities don't have the equipment or the faculty members and staff to give students a solid grounding in techniques and instruments. At some biology departments, even introductory procedures such as DNA extraction are merely described.

This unfortunate division between research and teaching runs deep. The Indian government decided more than 50 years ago to create focused research institutes. The scheme set up distinct priorities: universities would focus on teaching, and the institutes would concentrate on research. That policy led to a handful of excellent scientific institutes but also impoverished university-based research. There is no shortage of funds for university research or teaching laboratories, but a heavy teaching load, an overly bureaucratic system of appointments and promotions, and the lack of infrastructure have all made it difficult to recruit capable researchers.

The institutes, meanwhile, have good research projects, but little teaching. Yet good researchers are also often the better teachers.

Still, many institutes have begun training students by evolving graduate-level courses for small classes. Unfortunately, the top students who train at these institutes often choose to leave India and complete their PhDs abroad — at least in part because few of the centres can match the breadth of education at a first-rate university in the United States or Europe.

What might be done? There is little question that the government should actively support university-based research. Universities need help in upgrading their research infrastructure and laboratories, and in recruiting scientists with dedicated research space and healthy start-up packages. The government could contribute to these costs.

Perhaps federal research grants can include infrastructure costs.

There are signs that the government is responding to these concerns. In March 2005, the Science Advisory Council recommended setting up a National Science and Engineering Research Foundation, on the lines of the US National Science Foundation, to support research in various disciplines. Universities are expected to be important beneficiaries.

As funding for these universities increases, there should be accompanying changes in culture. Research and teaching should be valued as mandatory components of faculty appointments and promotions. Universities should also institute periodic review of the faculty members; this approach is also beginning to find favour in China, a country beset by similar problems.

The particular structure of Indian science also suggests solutions. For instance, funding agencies might consider establishing long-term research faculty positions within universities. New research centres could be closely allied with or even located on university campuses. The researchers would be required to teach in the university, and university students would have access to research labs.

One model for effectively integrating research and teaching already exists within the IITs. These institutes continue to attract the best Indian undergraduates in engineering and the physical sciences, and give them a world-class education. Their success has positioned India as a key player in the IT industry. One small step towards boosting the life sciences may be to encourage the IITs to expand their biological sciences curriculum. But that cannot be the only solution, if only because a wider transformation is needed.

Innovation in universities is part of a broader theme in how a society educates not only its élite but all of its citizens. A hard look at education in the life sciences is particularly urgent for a country such as India, which has both a strong need for development and the ambition to match it. ◼

Mriganka Sur is head of the Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology.

Indian universities focus on theory rather than experiments, says Mriganka Sur.

E. QUINN

Coming home

Reagents may be slow to arrive, but the freedom to explore his own research interests more than compensates. **Satyajit Mayor** is thriving in India.

Happy times: Satyajit Mayor enjoys his membrane-biology research at the National Centre for Biological Sciences in Bangalore.

"It can never be the same as doing science in America." This was the near-unanimous response from my peers when, in late 1995, I announced my decision to move back to India from the United States. By then, I had spent nearly 11 years in America, first as a graduate student at Rockefeller University and then as a postdoctoral fellow at Columbia University.

Like most expatriates who long to return home, I had planned for many years to return to India — but I kept putting it off. I worried that my career in India would become marked by obscurantist science, an ossified research set-up and cash-strapped institutes; that I would be surrounded by mediocre science and scientists; and that I would have to wait for weeks for supplies that can be in summoned in New York within a day.

Happily, I was wrong on most counts.

And I'm not the only one. Like me, many other researchers are choosing to return home,

and they're leading productive lives at the many small-sized institutions that are committed to competing in the international arena.

The National Centre for Biological Sciences (NCBS) in Bangalore recruited me in 1995, and I have remained here ever since. The centre was set up in 1992 as an arm of the Mumbai-based Tata Institute of Fundamental Research. The founding director, noted *Drosophila* geneticist Obaid Siddiqi, wanted to establish an institute where a diverse group of people with distinct approaches to biology could do world-class research. I knew from the beginning that this would be a good place.
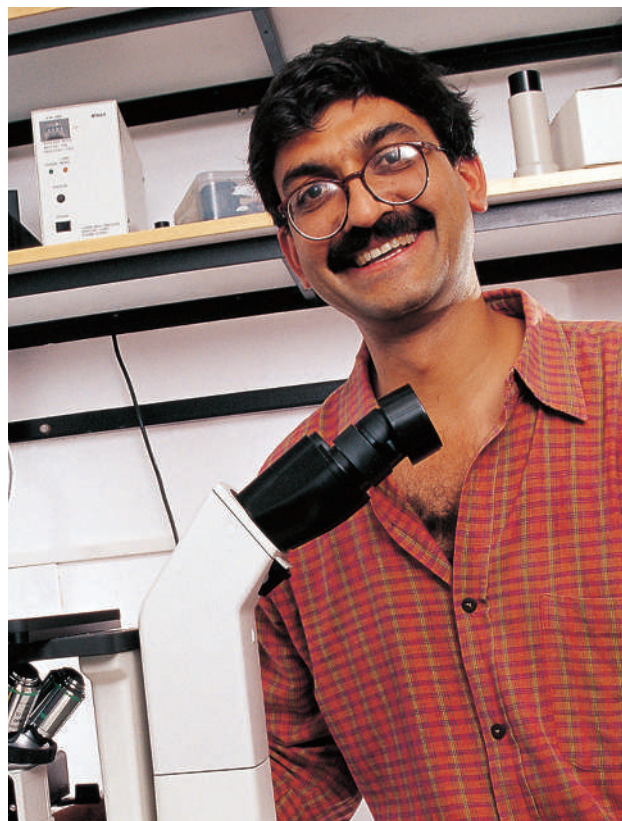
Since its launch, the institute has moved into a much bigger campus and hopes to grow from its current 15 research groups to a full capacity of around 40 teams. Together, NCBS researchers published 60 papers in 2004, including 18 in high-impact journals such as *Nature* and *Cell* (see table, left).

The institute's current director, K. Vijay-Raghavan, is generous to a fault, and accommodates the most outrageous of whims of his colleagues in the interest of furthering science. Most laboratories at the institute are engaged in research on different biological scales — from understanding the structure of molecules to the evolution of behaviour. My own lab aims to unravel the molecular mechanisms of endocytosis, the process by which cells internalize viruses or molecules, using a multidisciplinary approach that combines biology, physics and chemistry.

Being a researcher in India is not as difficult as I had expected. In some ways, my transition from being a privileged graduate student at Rockefeller University to a post-doctoral fellow at Columbia University was more of a culture shock than my move to the NCBS, where the scope of my research is limited only by my imagination. More difficult

was the adjustment to life in an Indian city after living in New York. I was also not looking forward to confronting the depressing political reality in India at the time. (Happily, things have changed for the better, thanks to India's democratic tradition.) But exciting scientific interactions at the NCBS and elsewhere in India have been more than adequate compensation for these travails.

Beating bureaucracy

When I joined the NCBS, my area of research, cell biology, was not yet the focus of much study there: protein structure and developmental neurobiology were more in vogue. I had trained almost entirely outside India so I had very different expectations when I arrived, but I was able to shape my environment as much as it shaped me. For example, my colleagues and I have helped to create an administrative structure that is both efficient and responsive to our needs. This is something most scientists in developed countries might take for granted, but revising the old bureaucracy-ridden model is an important step forward for Indian science.

Upon arrival, my priority was to set up my laboratory, and the NCBS seemed to share this emphasis, something that is rare at most Indian institutes. Still, some of my fears were legitimate. When I was at Columbia University, for instance, I could simply pick up the phone and have reagents instantly delivered to my bench. Here, it can take up to four weeks depending on where the manufacturer is located.

PUBLICATION RECORD OF THE NATIONAL CENTRE FOR BIOLOGICAL SCIENCES

Year	Total no. of publications	Publications in high-impact journals
1997	24	*Nature Struct. Biol.* (1)
1998	23	*Nature* (1)
1999	26	
2000	28	*Nature* (1), *Development* (1), *J. Mol. Biol.* (3), *Curr. Biol.* (1)
2001	28	*Nature Struct. Biol.* (1), *Dev. Cell* (1), *Virology* (1), *EMBO J.* (2), *J. Mol. Biol.* (2)
2002	35	*Science* (1), *J. Mol. Biol.* (1), *Development* (1), *Dev. Cell* (1), *J. Neurosci.* (2), *Cell* (1,
2003	45	*Virology* (1), *EMBO J.* (1), *J. Neurosci.* (1), *J. Cell Biol.* (1), *Immunology* (1), *J. Cell Sci.* (1)
2004	60	*Science* (1), *Mol. Cell. Biol.* (2), *Virology* (2), *J. Neurosci.* (2), *J. Cell Sci.* (1), *Development* (2), *J. Mol. Biol.* (2), *Cell* (1), *Neuron* (2), *Immunology* (2)
2005	20	(up to June 2005) *Nature Struct. Mol. Biol.* (1), *J. Virology* (1)

M.SUDHAKARAN

Another formidable challenge is convincing customs officials that my high-resolution cameras need not be exposed to the elements while numerous public holidays pass by. Things have changed for the better as suppliers have set up shop in or near Bangalore, but there is much room for improvement.

For the majority of Indian researchers, publication in prestigious international journals is as much a marker of success as it is elsewhere. But scientists here have not been able to fully meet this goal. Most of my colleagues in India believe that this is, partly, the result of a bias, conscious or unconscious, against Indian science. The critical mass of biologists — except perhaps in protein chemistry (see page 494) — in India is too small to have any visibility in international publishing. The lack of name recognition may make it easier to dismiss exceptional and controversial work from unfamiliar institutions.

Manual work

At many of the biology-based institutions in India, almost all the work is done by graduate students, and they can take up to four years to become productive. This is the result of an educational system that prevents students from getting their hands wet, partly because the system does not value experimental training and partly because of a genuine lack of resources. One reason for this is a bias against manual work among the élite classes who populate the universities. Fortunately, this seems to be changing (see page 487).

Although the graduate-student pool generates a group of committed researchers, the lack of a vibrant postdoctoral research culture is the sign of an unsophisticated research community. Some institutions in India are beginning to recruit postdocs from around the world, albeit slowly. But even among Indian students, going abroad to train is still the most attractive option. The handsome salaries offered by research positions in Europe and the United States are no small factor in their decision. More attractive salaries would go a long way to retaining researchers in India.

At 1.1% of the gross domestic product, India's budget for science is woefully inadequate. But the government has pledged to increase this to 2% by 2007. Already, funding opportunities for biological sciences in India are at an all-time high. Several agencies fund projects and any well-structured proposal has a high chance of getting funded.

In addition, researchers have access to international agencies. The arrival of the Wellcome Trust in India six years ago wooed many young talented researchers back to India (see 'Wellcome funding', below). Such international agencies have helped bridge the gap between the salaries in India and more developed countries.

Wellcome Trust fellows are required to provide frequent research reports and attend meetings to evaluate progress, which ensures that they stay focused. A major flaw in India's system is that the funding agencies seem to expect little intellectual return on their investment, which can breed complacency.

One of the biggest attractions as a scientist in India today is the freedom to take on unusual and innovative projects without much interference. Because the pool of active researchers is small, it fosters truly multidisciplinary collaborative links. I have been able to forge meaningful collaborations with a condensed-matter physicist, a fruitfly geneticist interested in neuronal conduction and a physiologist of the immune system.

India's biological research enterprise requires a broad and visibly multidisciplinary scientific base. Institutions such as the NCBS can help accomplish this goal. Another ten similar institutions could completely transform the nature and quality of biological sciences in India. If these institutes then form productive alliances with universities, they could tap into India's vast student pool. At the very least, this could boost the woefully underfunded and often stifling university departments. The much touted biotech revolution in India may even acquire well-trained scientific personnel and a semblance of reality.

In an encouraging development, the government in February this year announced a US$250-million National Science and Engineering Research Fund to fund basic research in India. The goal is to establish two integrated universities to teach and train a new breed of basic scientist, the most important resource for a productive scientific future. ■

Satyajit (Jitu) Mayor is an associate professor at the National Centre for Biological Sciences, Bangalore.

WELLCOME FUNDING

Many of Indian biology's rising stars, such as Satyajit Mayor (see main text), are recipients of the Wellcome Trust International Senior Research scheme.

These fellowships, awarded for five years and renewable for another five, are highly coveted because of their size — bigger than the budget of the science departments in some Indian universities — and because they can be used flexibly. In 2004, 14 scientists from India were interviewed for the grants; all won.

Unencumbered by the pressure to raise their own money, many of the fellows are doing very well. Utpal Bhadra, a fellow at the Centre for Cell and Molecular Biology (CCMB) in Hyderabad, has co-authored six papers in high-impact journals, including *Science*, *Nature Genetics* and *Molecular Cell*, since his return to India in 2002. His book on RNA interference, written with his wife, a 2005 fellow, is to be published by the Cold Spring Harbor Laboratory Press. Sanjeev Galande, a fellow at the National Centre for Cell Sciences, has also published six papers in prestigious journals in the past four years.

Rajan Sankarnarayanan, another fellow at the CCMB, used his Rs33 million (US$800,000) award to set up a world-class crystallography lab to study protein structure with equipment that would otherwise have been out of his reach.

"This kind of fellowship has helped many young scientists, including myself, to return after postdoctoral training abroad and establish their own laboratories in the home country," says Sankarnarayanan.

Altogether, 37 Indian biomedical researchers have benefited from the grant. The trust has also supported researchers in South Africa since 1992 and in Central Europe since 2002. These countries shared 15% of the funding; the rest was earmarked for scientists from Britain. But from 2005, all applicants will have to compete with UK scientists.

K. S. Jayaraman

Rajan Sankarnarayanan's fellowship money helped to pay for his state-of-the-art structural biology laboratory.

P. ANIL KUMAR

Rivalry and red tape

Researchers are suffering as a result of the conflicts between funding agencies. **T. V. Padma** uncovers plans to heal the rift.

Awaiting approval: Pradeep Seth waited for more than two years for the go-ahead from the government for phase I trials of his HIV vaccine.

Pradeep Seth has given up. For nearly two years, Seth, a microbiologist at the prestigious All India Institute of Medical Sciences (AIIMS) in New Delhi, has been awaiting word on his HIV vaccine candidate from the health ministry.

Seth's vaccine research was funded by the Department of Biotechnology (DBT), part of the science ministry. He published his work in peer-reviewed journals and received positive feedback from several evaluation committees. In March 2003, the DBT gave him the green light to form a partnership with an Indian company and manufacture the vaccine for phase I human trials.

Everything went according to plan. A committee at the AIIMS helped Seth select an industrial partner and write a licence agreement. The institute's legal and financial departments approved the agreement. Then, in December 2003, an over-cautious AIIMS director sought additional approval for the agreement from the health ministry. Seth has been waiting for that approval ever since.

The quality of the potential vaccine does not seem to be the problem. The healthy ministry formed a committee to review the vaccine in December 2003. But before it made a decision, India headed for national polls, and a new government took charge in May 2004. "It was

back to square one," says Seth. A new committee then recommended in November 2004 that trials on the vaccine should proceed without further delay. Then the health secretary, who had to okay the licence agreement, retired. "Again back to square one," says Seth. Finally, a third committee reviewed the vaccine in February 2005. This committee, led by Nirmal Kumar Ganguly, director-general of the Indian Council of Medical Research (ICMR) — the research arm of the health ministry — recommended on 29 March 2005 that the trials should proceed. By then Seth had almost retired.

Lack of coordination

Seth's story is well-known in Indian academic circles. Some say the story is a tragicomic case of classic Indian bureaucracy. But others say the events are no accident. Seth's vaccine did not fare well because of a deep-seated rivalry between the DBT, which funded his research, and the ICMR, they say. "Most DBT-funded projects in the AIIMS have met a similar fate," says Seth.

There are several branches of government that fund biomedical research, but money comes primarily from the health ministry, which finances research through the ICMR, and the science ministry, which oversees the DBT, the Department of Science and Technology and the Council of Scientific and Industrial Research (see graph).

Each agency has — or is at least supposed to have — a well defined role. For example, the Department of Science and Technology looks after basic research; the DBT focuses on the development of biotechnology products and processes; and the

ICMR funds projects in epidemiology and clinical research.

In theory, the health and science ministries would coordinate research. "In practice, however, overlaps do take place," says Martanda Sankaran Valiathan, former president of the Indian National Science Academy.

These glitches occur despite representatives from funding agencies sitting on each other's committees and asking applicants to note whether they receive funds from elsewhere.
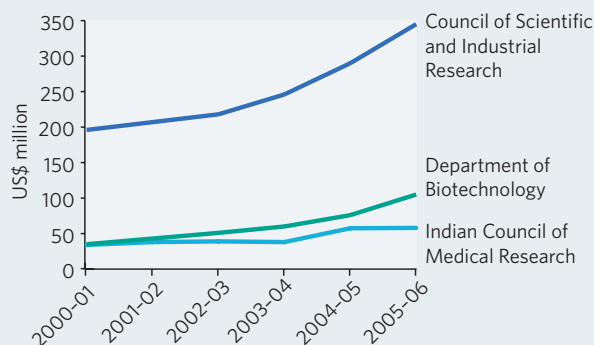
Some projects can be neatly classified as basic or clinical, but increasingly there are many projects that are 'applied'. Stem-cell research falls squarely into this hazy overlap.

No single government agency in India has been assigned responsibility to licence or monitor stem-cell work. As a result, the DBT and the ICMR each issued guidelines — the DBT in 2001 and the ICMR in 2002 — without consulting each other. The result was a plethora of public- and private-funded institutes that joined the stem-cell bandwagon.

Researchers at Nutech Mediworld, a private clinic in Delhi, say they have used stem cells to successfully treat paralysis in 24 patients since May 2003; their work is as yet unpublished. P. Venugopal, a surgeon at the All India Institute of Medical Sciences in Delhi has used bone-marrow derived stem cells to repair damaged heart muscles (see *Nature* **434,** 259; 2005). And the L. V. Prasad Eye Institute in Hyderabad has since 2001 used limbal stem cells to repair damaged corneas in 260 patients (V. S. Sangwan *et al. Biosci. Rep.* **23,** 169–174; 2005).

The competition between the DBT and the ICMR in particular has been the focus of much discussion among researchers. "Greater collaboration between agencies is a critical issue," says secretary of the DBT Maharaj Kishan Bhan. "Our research efforts are getting fragmented because of this lack of coordina-

OVERALL BUDGETS OF KEY FUNDING AGENCIES

Council of Scientific and Industrial Research

Department of Biotechnology

Indian Council of Medical Research

tion." If the different departments worked together, they could avoid duplication and save enormous resources, Bhan says. "Right now all agencies work subcritically."

Bhan, who has led the DBT since 2004, and Ganguly both say they are taking steps to put an end to the famous rivalry. After much discussion, the two agencies have agreed to meet in June 2005 to combine the two sets of guidelines. They are also jointly drafting a national stem-cell initiative and plan to work together on HIV vaccines and microbicides.

A waiting game

Another area of researchers' lives that could be improved is that of paperwork. For example, to get a grant approved, Indian scientists have to be prepared to wait for more than a year while their application is processed, and months more to get the money after approval. The funds go to the institution and not to the investigator, so additional delays occur during the transfer of money to the researcher.

Researchers have to submit nearly 30 copies of their proposal to funding agencies such as the ICMR. So cumbersome is the red tape that in January 2005, at the opening of the science congress in Ahmedabad, prime minister Manmohan Singh lamented the "tyranny of bureaucracy" in Indian agencies. He went on to ask whether India has "allowed bureaucratic systems and patron–client relationships to stifle creativity" and "scare away" young researchers.

"Most of the red tape is due to the fact that there are no serious consequences for incom-petence," says Sandip Basu, director of National Institute of Immunology in Delhi. "It is not easy to take severe disciplinary action, and all jobs are until retirement," he says.

Some delays are entirely avoidable, Basu adds. These are the ones agency chiefs are committed to cutting down. Some researchers are reporting that recently submitted proposals to the Department of Science and Technology and the Council of Scientific and Industrial Research cleared in six months. What's more, the process required only two hard copies: the rest was handled through e-mail.

But this may have all come too late for Seth whose last working day was 30 April. "I am through with it," Seth says. ■

T. V. Padma is a freelance writer based in New Delhi.

A TOUGH JOURNEY

To outsiders, Vijaylakshmi Ravindranath's story might seem a resounding success. In 2000, she was appointed director of the nascent National Brain Research Centre in Manesar, on the outskirts of New Delhi.

But her achievement did not come easily. When Ravindranath was a postdoc at the US National Institutes of Health, her husband, also a scientist, remained in India. She initially took her two-year-old son with her, but, unable to arrange for his care when she was in the lab, she sent him to be looked after by her parents. "I had to be away during two critical years of my son's childhood," she recalls.

Twenty years later, she was once again faced with a difficult choice. The offer to head the new centre was a rare opportunity, but it meant a five-year separation from her husband and son. "Had my husband and son not cooperated, it would have been impossible," she says.

Ravindranath is one of the lucky few. Unable to juggle professional and domestic demands, and sometimes coping with nepotism and sexual harassment, most Indian women scientists give up early in their careers. According to a 2004 report of the Indian National Science Academy (INSA), only

Vijaylakshmi Ravindranath is one of the very few high-achieving women scientists in India. She is the director of the National Brain Research Centre, near New Delhi, but she could not have reached this position without the support of her husband and son.

nine of 398 Shanti Swarup Bhatnagar awards, India's highest honour for science, went to women. Within the INSA, women have won only 14 of the academy's 502 awards and medals, and no woman has ever led the academy.

Like their counterparts worldwide, Indian women scientists fare better in biology than in engineering or physics. In 2000, women accounted for 32% of medical students, compared with 16% in engineering. But, as elsewhere, there is a steep drop-off in women at graduate level. Many of those who continue opt for a career in medicine or teaching, rather than research, says Manju Sharma, former secretary of the Department of Biotechnology

and co-chair of an international advisory panel on women scientists.

Apart from the usual factors, such as marriage and lack of adequate childcare, Indian women also have to endure the country's staunch patriarchal culture. Many complain that male colleagues do not treat them as peers and often assign them to reception committees or to choosing menus for conferences.

Women also face considerable bias in the interview process, says Sharma, who has served on the interview committees of several national institutes. "I have noticed a tendency to ignore the excellent career records of women scientists when it comes to selecting top positions," she says.

The government is beginning to take action. In 2003, the Department of Science and Technology initiated a scheme to offer fellowships to women whose careers were interrupted by their husbands' geographical moves. More than half of the 2,000 applicants were from life sciences.

Women scientists particularly need assistance when they are struggling to balance their early career with a growing family, says Ravindranath. "If the woman can tide over the period of marriage and child rearing, a critical mass of women scientists can develop in biology," she says. "Today, my son and husband are among the proudest and happiest about my success."

T. V. Padma

Among the best

India's life-sciences institutes are rewriting the rule books for research. **K. S. Jayaraman** finds that they are focusing on higher standards and enterprise.

Just two years ago, the tree-lined roads on the campus of the Indian Institute of Science (IISc) in Bangalore were a joggers' paradise. The only sounds were those of rustling leaves and chirping birds.

These days, blaring horns and speeding Toyotas are more the scene, occasionally forcing terrified pedestrians into the wedge-shaped drains on the sides of the roads. "Everyone here is in hurry," says S. Vijaya, a professor of microbiology at the institute.

This flurry of activity is a result of a radical culture change at the IISc, widely considered India's premier scientific institute. Fourteen companies — including Belgium's Janssen Pharmaceutica and New Jersey-based Novozymes Biologicals — now rent laboratory space on campus. Researchers at the IISc have acted as consultants for 161 companies and completed more than 100 projects in the past ten years. And in the past three, five professors have launched their own companies.

The Indian Department of Biotechnology is encouraging this trend by allowing faculty members to hold positions in industry and academia simultaneously. It has earmarked 30% of its US$105-million budget for 2005–06 to such partnerships.

A similar scene is being played out all over the country. With nearly 450 faculty members in disciplines ranging from biochemistry to metallurgy, the IISc is the largest and the most established Indian institute. But a handful of smaller life-sciences institutes that have sprung up in the past decade are also rewriting the rules of Indian research. "We have succeeded in setting higher standards for ourselves and achieving them," says G. C. Mishra, director of the National Centre for Cell Sciences (NCCS) in Pune.

Not everyone is happy with this change. For many Indian researchers, making money from science is an uncomfortable concept. Hindu mythology even holds that Lakshmi, the

STRENGTH IN NUMBERS

India's huge and diverse population is an asset to researchers studying complex genetic diseases.

The massive size of India's population has always been seen as its biggest weakness, but geneticists are turning this burden into a boon. India could be the ideal place to study complex diseases.

"Given the enormous genetic diversity, India has a strategic advantage," says Samir Brahmachari, director of the Institute of Genomics and Integrative Biology in Delhi.

With roughly one-sixth of the world's population, India is home to 4,693 communities, 532 tribes, 325 functional languages (representing four major linguistic groups) and 25 distinct scripts. Because members of many of these communities tend to marry within their group, they have evolved over time into distinctive gene pools.

To map genes for complex diseases, researchers have often turned to geographically isolated groups such as those in Iceland or Estonia. But Indian ethnic groups are genetically more

Leading the way: the Indian Institute of Science in Bangalore is the most productive of India's life-science institutes.

goddess of wealth, and Saraswati, the goddess of learning, cannot coexist.

"Half of us would be happy teaching and publishing papers," says P. N. Rangarajan, associate professor of biochemistry at the IISc. But the other half are making the most of this new opportunity.

Rangarajan is firmly in the latter group: this year he won an award for transferring the DNA vaccine he developed against rabies into a commercial setting. His vaccine can be stored at room temperature and so is ideal for tropical countries. The IISc licensed the technology to Hyderabad-based company Indian Immunologicals.

But this achievement came at a price for Rangarajan. His promotion was delayed by two years because the number of papers he published declined during the time he spent with his industrial partner.

Measuring success

This attitude is slowly changing, but many administrators still insist that academic pursuit should be the faculty's first priority. "Our basic objectives continue to be research and teaching," says G. Mehta, director of the IISc. "If the faculty finds time to hold hands with industry, that is welcome."

Some researchers feel that this attitude is a little short-sighted. "We should have a different yardstick for assessment," says Govindarajan

homogeneous than the Icelandic population, says Partha Majumdar, head of the human genetics unit at the Indian Statistical Institute (ISI) in Kolkata, who compared published genetic data of Icelandic populations with those of ethnic Indian groups.

In India, it is possible to study genetic variation both within and between communities, Brahmachari adds. He is coordinating a five-year US$5.8-million project that includes researchers from his institute and five others under the Council of Scientific and Industrial Research. Launched in 2003, the Indian Genome Variation consortium aims to map genetic markers for diseases such as diabetes, cancer and heart disease. The ISI will analyse the data from the project and help choose sample populations for the studies.

In 2003, the International HapMap Project, which aims to find genes associated with human disease and response to pharmaceuticals, asked Majumdar for samples from some Indian ethnic groups. But the ISI does not have the technology to be a proper partner in this project, and Majumdar says he declined because he did not want only to supply samples.

The Indian project's goal is to collect blood samples from 15,000 individuals drawn from different communities across the country, select about 1,000 genes relevant to complex diseases common in India and assess any differences in drug response. The scientists plan to study candidate genes reported by other groups as well as any new ones that crop up.

The team plans to record DNA sequence variations that occur when a single nucleotide in the genome sequence is altered. These are known as single nucleotide polymorphisms or SNPs. The group also hopes to note repeats of nucleotides and compare their frequency across different groups. Brahmachari's team has already collected samples from more than 40 populations, but because the report is unpublished, he declined to reveal further details.

Other genomics projects are also under way in increasingly well-equipped labs at the Institute of Genomics and Integrative Biology, the Centre for Cellular and Molecular Biology (CCMB), the Centre for DNA Fingerprinting and Diagnostics in Hyderabad, the Indian Institute of Science in Bangalore and the University of Delhi.

In November 2004, India agreed to take part in a pan-Asian initiative to study genetic diversity and similarities in Asia. The ten-nation study, an initiative of the Pacific chapter of the Human Genome Organisation, is slated to begin later this year. Researchers at the Hyderabad-based CCMB are also studying disease susceptibility in various castes and tribes. In collaboration with the Anthropological Survey of India and 15 universities, the researchers have collected more than 9,000 DNA samples from 130 groups.

From these samples, scientists have already identified genetic diseases that are particularly prevalent among certain groups, says Kumarasamy Thangaraj, a researcher at the CCMB.

For example, the Onges tribe has dwindled to only 98 people from 700 in 1858. Genetic analysis of 46 members of the tribe has revealed 23 mutations in the mitochondrial genome and deletions of some Y-chromosome genes needed for sperm production.

"It is astonishing that in several populations, the load due to genetic diseases is so high it is leading to their extinction," says Thangaraj.
T. V. Padma

STRONG BONDS

Basic research in the life sciences has not been India's strong suit. But in one field — protein science — Indian researchers have made their mark.

India's love affair with proteins began more than 50 years ago, when G. N. Ramachandran, a physics professor at the University of Madras, and his colleague, G. Kartha, used X-ray data analysis to unveil the triple-helical structure of collagen, one of the most abundant proteins in mammals.

The X-ray machine was primitive, and Ramachandran performed painstaking manual calculations to work out collagen's structure. He also came up with a formula, later dubbed the Ramachandran plot, to predict the conformation of proteins. During his career, he trained several biophysicists who worked with him first in Madras, now called Chennai, and then at the Indian Institute of Science (IISc) in Bangalore.

Around the same time, teaching and research in biochemistry gained ground in India, and scientists from both disciplines began to collaborate. "India could thus

Highly respected: Indian scientist G. N. Ramachandran (first row, centre) made a significant contribution to the field of protein science. Here, he is shown sitting next to his wife (first row, second from left). Ramachandran is talking animatedly with S. Ramasesham (first row, fourth from left), another eminent crystallographer. G. Kartha, the co-discoverer with Ramachandran of the triple-helical structure of collagen, is seated in the fourth row (far left).

produce groups of people and enough institutional expertise to work on both protein physics and protein chemistry, which proved to be an advantage," says Padmanabhan Balaram, a protein physicist and chairman of biological sciences at the IISc. In the 1980s, the field received a further boost when government agencies began providing funds to buy equipment for selected centres.

Work on proteins also suits the research culture. Indian scientists prefer projects that rely more on individual creativity and competence because they fear large groups will detract from individual recognition, says Govindarajan Padmanaban, scientist emeritus at the department of biochemistry at the IISc. "Indian scientists on the whole do not integrate in large groups," he says.

Because protein research rarely requires large interdisciplinary collaborations, it has remained a popular choice among Indian scientists.

Countries around the world may be able to benefit from India's expertise. "With rising costs and the closure of some chemistry departments in the United States and Europe," Balaram says, "India could become a destination for outsourced chemistry research and development, including theoretical work such as the study of protein dynamics and computer modelling for drug design."

T. V. Padma

Padmanaban, a biochemist and former director of the IISc, who has discovered a potential drug target for malaria. "Publication is not everything," he adds.

The rules are a bit looser at the government laboratories, such as those under the Council of Scientific and Industrial Research (CSIR). In 2000, the CSIR launched a programme to promote partnerships with industry.

For instance, since 2001 Mumbai-based Nicholas Piramal India has been paying the CSIR's Institute of Genomics and Integrative Biology (IGIB) in Delhi US$460,000 each year as a 'knowledge' fee. The arrangement allows the company to develop any leads the institute's researchers find. The alliance has already resulted in two joint patents. A potential drug against tuberculosis that has shown promise in animal studies is the outcome of a similar alliance between Mumbai-based Lupin Laboratories and four CSIR institutes.

To promote technology transfer, the IISc has launched an 'incubation' centre, where faculty members help students to develop commercial applications. The institute also supports scientists who launch biotechnology companies by granting them 'entrepreneurial leave', making it a popular partner for multinational companies. General Motors has invested US$1.4 million and Boeing US$500,000 in research at the IISc.

The IISc is careful to accept only projects that involve innovative research. The Society for Innovation and Development, the commercial arm of the institute, has turned down about US$20 million in projects because they

COMPARISON OF FOUR HIGH-PROFILE INDIAN LIFE-SCIENCE RESEARCH INSTITUTES IN 2004–05

	No. of faculty members	No. of publications	No. of patents	No. of faculty members educated abroad	Percentage of faculty members trained abroad
Indian Institute of Science, Bangalore	450	1,367	25	129	98
National Centre for Cell Sciences, Pune	24	37	6	16	96
Centre for Cellular and Molecular Biology, Hyderabad	88	111	15	20	80
National Centre for Biological Sciences, Bangalore	21	49	Nil	14	100

E. SUBRAMANIAN

lacked research content, says S. Mohan, the society's chief executive. "Even then, the demand from companies is so great we cannot cope," he says.

Some partnerships with industry are providing much-needed training for young scientists. For example, Hyderabad-based Bharat Biotech has trained nearly 500 interns in biotechnology. In addition, 12 students of Jawaharlal Nehru Technological University in Hyderabad are doing their graduate work at the company.

Critical mass

With more than 1,300 papers published in peer-reviewed journals last year, the IISc is the most productive of the Indian institutes (see table, left), and continues to enjoy the government's favour. In February 2005, the government gave the institute a massive grant of US$24 million.

But in the past ten years, a handful of other institutes have begun to make their mark. The National Centre for Biological Sciences in Bangalore, an offshoot of the famous Tata Institute of Fundamental Research in Mumbai, produced India's first paper to be published in the high-profile journal *Cell* just four years after it moved to its own building in Bangalore. The institute's faculty members were recruited almost entirely from the United States, and they work under a US-style contract system: scientists are hired for five years and their performance is periodically reviewed by the institute's research council. One contract has already been terminated — the first time a publicly funded scientist in India has been sacked for underperformance.

But the model is unlikely to catch on. "Indian scientists want secure jobs," says Rangarajan. "If there was a poll, 80% would opt for security in preference to contract jobs at higher salary."

As biotech companies continue to proliferate, academic centres will find it difficult to hold on to scientists unless they raise salaries by an order of magnitude, warns C. N. R. Rao, former IISc director and the prime minister's science adviser. Rao says three of his students joined General Electric's research labs in Bangalore at a salary of US$1,400 per month, twice what a professor at the IISc can expect to earn.

Some institutes offer attractions to compensate for the lower salaries. "What they need is freedom and a conducive atmosphere," says Lalji Singh, director of the Centre for Cellular and Molecular Biology in Hyderabad.

NCCS director Mishra says his institute's high productivity — 35 papers were published in high-impact journals in 2004 — helps to attract good scientists. The NCCS has grown from 13 faculty members in 1995 to 25 in

PRODUCTIVITY OF NATIONS BASED ON THEIR PUBLICATION RECORD

Country	Country classification	No. of publications (1997–2001)	GDP per capita	Publications per GDP per capita	Publications per GDP per capita per year
India	IDC	77,201	487	158.52	31.7
China	IDC	115,339	989	116.62	23.32
United States	G8, OECD	1,265,808	36,006	35.16	7.0
Brazil	IDC	43,971	2,593	16.96	3.39
Germany	G8, OECD	318,286	24,051	13.23	2.64
United Kingdom	G8, OECD	342,535	26,445	12.95	2.59
Japan	G8, OECD	336,858	31,407	10.73	2.15
South Africa	IDC	18,123	2,299	7.88	1.58
Canada	G8, OECD	166,216	22,777	7.30	1.46
Italy	G8, OECD	147,023	20,528	7.16	1.43

PRODUCTIVITY OF NATIONS BASED ON THEIR CITATION RECORD

Country	Country classification	No. of citations (1997–2001)	GDP per capita	Citations per GDP per capita	Citations per GDP per capita per year
India	IDC	188,481	487	387.02	77.40
China	IDC	341,519	989	345.32	69.06
United States	G8, OECD	10,850,549	36,006	301.35	67.27
United Kingdom	G8, OECD	2,500,035	26,445	94.54	18.91
Germany	G8, OECD	2,199,617	24,051	91.46	18.29
Brazil	IDC	155,357	2,593	59.91	11.98
Japan	G8, OECD	1,852,271	31,407	58.98	11.79
Canada	G8, OECD	1,164,450	22,777	51.12	10.22
Italy	G8, OECD	964,164	20,528	46.97	9.39
Australia	OECD	623,636	20,822	29.95	5.99

Source: D. A. King *Nature* **430,** 311–316 (2004). IDC, internationally developing country; OECD, Organisation for Economic Co-operation and Development.

2005 (see table, opposite); all except one are Indians who trained abroad. In the early years, Mishra says, he had to go to the United States to recruit scientists, but now he gets about one overseas application each week.

Lost in translation

India's life-science institutes are improving at a rapid rate (see 'Strength in numbers', previous page, and 'Strong bonds', opposite). Already, Indian scientists are more productive per dollar invested in their research than their counterparts worldwide (see tables, above). But overall, India lacks a critical mass of competent scientists in basic and clinical research, notes Sandip Basu, director of the National Institute of Immunology in New Delhi. As a result, there is limited collegial interaction and unreliable peer review. "The resulting isolationist, nepotistically inclined atmosphere does no good to the cause of fostering high-quality research in the country," he says.

The country's medical schools lack the infrastructure and the faculty staff to train researchers. A quarter of India's 229 medical colleges have never published research reports in indexed journals; most clinical research

comes from just three medical institutes.

"We have a fair amount of curiosity- and individual-driven research," says Maharaj Kishan Bhan, secretary of the Department of Biotechnology. But without clinical researchers to translate advances into treatments, "the bulk of Indian research in basic science is lost," he says. "The whole chain does not function in an orderly fashion."

Bhan says what's needed is an institute for translational research where basic-science researchers and physician researchers can collaborate on finding solutions to India's public health problems. It would focus on developing simple, low-cost technologies that can be used in settings where the resources are poor. "We need to build one or two such new technology-focused institutes," he adds.

In the United States, the biotechnology sector accelerated rapidly when scientists moved from academia to industry, patenting their discoveries. "The opportunities that are unfolding in India are breathtaking," says Ragunath A. Mashelkar, director-general of the CSIR. ■

K. S. Jayaraman is *Nature*'s India correspondent.

The coming epidemic

A staggering 5.1 million people are estimated to be HIV positive in India. **Apoorva Mandavilli** finds a country on the brink of a crisis.

Our son was born in June 1998. He was healthy but after eight months, he had diarrhoea and fever all the time. He was in the hospital many times. The sixth time, they diagnosed him with AIDS. That was when we found out that my wife and I have HIV. Our son died 12 July 2002. I also started to get sick. I didn't take my medicines regularly; they didn't tell me not to do that. Now my health has become worse. I haven't worked in six months. We'll be paupers. I don't know what we're going to do.

— Suresh, air-conditioning technician

Suresh is sitting in a small, dark room at an AIDS clinic in the southern Indian city of Chennai. This city is where the first cases of HIV in India were discovered in 1986 after a police sweep of sex workers. Nearly 20 years later, there are an estimated 5.1 million cases in India, a number second only to that in South Africa. The difference is that in India the epidemic has not yet peaked. According to the CIA, the number of cases in India could top 20 million by 2010.

The people hit by AIDS in India are not the ones you would expect. About 86% of cases are a result of sexual transmission, most of it heterosexual. Intravenous drug use accounts for only 2.4% of infections, except in Nagaland and Manipur, two states where HIV infection is highly prevalent. The earliest cases appeared in high-risk groups: among the country's 3 million sex workers or the 5 million truckers who haul the virus up and down the highways. But increasingly, the people who are flooding hospitals are men like Suresh, and their wives and children.

Like many Indian men, Suresh visited sex workers before he got married. He is in his thirties, but looks at least 20 years older. He is thin, with large sores on his lips. His wife Sumithra Devi was once near death, with a T-cell count of 40; a person is said to have AIDS once their T-cell count drops below 200. Afraid to tell anyone they have AIDS — their family believes their son died of tuberculosis — they travel nine hours from Kancheepuram to Chennai to visit YRG Care, a non-profit clinic.

The clinic enrolled Sumithra first in a clinical trial and then in a Global Fund to Fight AIDS, Malaria and Tuberculosis programme that provides antiretroviral drugs. She looks healthy now, and her T-cell count has shot up to 1,000. But Suresh, who is resistant to the drugs that the programmes provide, is desperate for newer treatments — and the money to buy them. "It's too late for me," he says.

AIDS is a big problem in India but we have been able to keep it in place so that it doesn't become a bigger problem. We are not going the South Africa way.

— Nirmal Kumar Ganguly, director-general, Indian Council of Medical Research

In May this year, the Indian government announced that during 2004, there were just 28,000 new cases of HIV infection, down 95% from the previous year — proof, it said, that its strategies are working. But even in the United States there are an estimated 40,000 new cases each year, says Kevin Frost, director of TREAT Asia, a network of HIV/AIDS clinics and institutions. "So do you believe that Indian number?" he asks. "Of course not, that would be ridiculous."

S. Y. Quraishi, director-general of the National AIDS Control Organization (NACO) says that the figures were collected in exactly the same way as the previous year. "Tell me, what can we do? This is the best possible estimate," he says.

Compared with South Africa, where the prevalence of HIV is 21.5%, India's prevalence is just 0.91%, Quraishi notes. That is less than the 1% generally considered to be the tipping point, beyond which the epidemic will escalate out of control. It took just seven years for South Africa's prevalence to rise from 1% to 20%. "Considering that our prevalence has been kept to less than 1%, that shows that our strategies have been on the right track," Quraishi says. "It's not by sleeping that we're not South Africa. Obviously we did something."
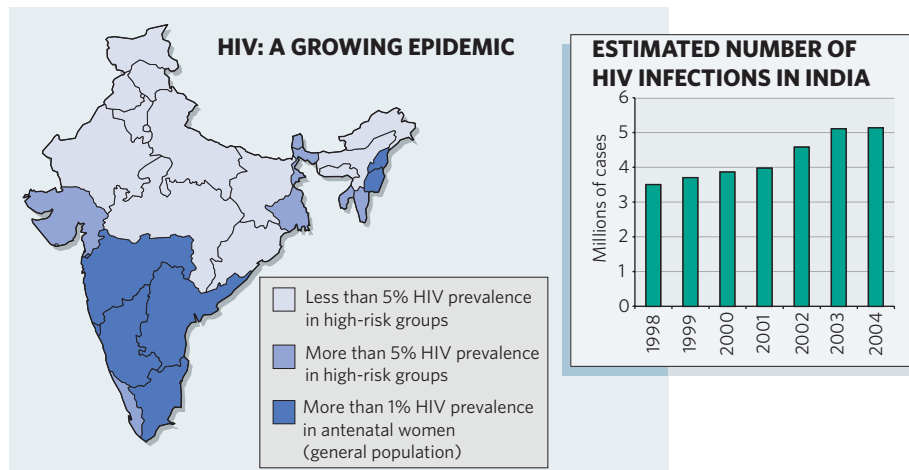
But the statistics are sobering (see graphic, opposite). Six states in India already have a prevalence greater than 1%, and in 2004 the number of high-prevalence districts jumped from 49 to 116. Even in the low-prevalence or 'highly vulnerable' states, as they are now dubbed, there are pockets where more than 4% of adults are infected. "The epidemic is progressing to the tipping point, there's no doubt about that," says Ashok Alexander, director of Avahan, the Bill & Melinda Gates Foundation's AIDS initiative in India.

Some experts say that the new administration, which in April 2004 replaced the previous right-leaning government, could yet turn things around. NACO's strategy is to reach high-risk groups — particularly sex workers and truck drivers — and scale up programmes that have proved successful. Over the next five years, it plans to expand its 670 voluntary testing sites to 24,000 and increase the number of people on antiretroviral therapy from an estimated 8,000 to 188,000 (see 'Seeking care', opposite).

But in the world's most populous democracy, good intentions don't go very far. Despite doubling its AIDS budget since 2003–04, India still only spends about 29 US cents per person on AIDS. In contrast, Uganda, credited with curbing its AIDS crisis, dedicates US$1.85 per person. Already, India accounts for more than 13% of the world's AIDS burden but with its prevalence so close to the tipping point, the stakes are high. "Overall," says Frost, "the Indian government's response has been pathetic and not what it needs to be for an epidemic of this size."

Out in the field: in a bid a quash India's escalating HIV problem (below) health workers are running outreach programmes to teach people about AIDS.

HIV: A GROWING EPIDEMIC

- ☐ Less than 5% HIV prevalence in high-risk groups
- More than 5% HIV prevalence in high-risk groups
- More than 1% HIV prevalence in antenatal women (general population)

ESTIMATED NUMBER OF HIV INFECTIONS IN INDIA

Millions of cases (y-axis 0–6); years 1998–2004

I saw people with AIDS die. That's when I started getting scared. I have been teaching other sex workers about AIDS, telling them to use condoms. Now clients also want to put on condoms. They are available at medical shops. But I know where to get them for free.
— *Padma, sex worker*

In Mumbai, the bustling metropolis on India's west coast, there are nearly 100,000 sex workers. Volunteers from non-governmental organizations (NGOs) walk through the streets of Kamathipura, the red-light district, handing out condoms and leaflets.

But in the southern states, where prevalence is highest, there are no red-light districts to canvass (see 'The epicentre', overleaf).

Instead, sex workers — many of whom are married women — work out of their homes, hotels or on the highways, earning themselves the nickname 'highway queens'. Some towns, such as Peddapuram and Amalapuram in the state of Andhra Pradesh, are "like one big red-light district", says Ashok Babu, a programme manager for the AIDS Prevention and Control Project in the neighbouring state of Tamil Nadu. "For US$230, you can stay there for one month and be treated like a *mappillai*," he says, meaning a son-in-law, who is treated like a prince by the girl's family. In Andhra Pradesh, HIV prevalence jumped from 1.25% in 2003 to 2.25% in 2004.

Experience has shown that the best way to reach these unconventional sex workers is through their peers. At the Community Health Education Society, an NGO in Chennai, 'peer educators' such as Padma explain to the women how to use a condom, the difference between HIV and AIDS, and why they are more at risk of getting infected. But the message is cleverly delivered through games, songs or as a cost–benefit analysis of, for instance, the amount of money squandered on abortions and treatments for sexually transmitted diseases. "When you talk to them just about AIDS, they get bored," says Pinagapany Manorama, a physician who runs the NGO.

In some ways, sex workers are now better informed, and more empowered, than

SEEKING CARE

When you're trying to manage an AIDS epidemic and you have limited resources, preventing infection is the logical priority. But where does that leave those who are already infected?

Treating people with AIDS is not easy. At the very least, it requires trained medical staff and the resources to make sure patients take the drugs on time. Nobody knows that better than the doctors at Tambaram Hospital.

Built in 1928 as a sanatorium for patients with tuberculosis, the government centre is 45 km outside Chennai and has more AIDS patients than any other Indian hospital. There are often more than 900 inpatients for its 776 beds, so some have to sleep on the floor. Every hallway is flooded with patients who look skeletal, with shrunken limbs and sallow skin. Outside the wards, hairy black pigs roam beneath drying laundry, accompanied by the rancid smell of sewage.

The hospital was one of eight government centres that together were meant to roll out antiretroviral drugs (ARVs) to 100,000 people over five years. In the first year, which began April 2004, it treated fewer than 1,000.

"From the outside, you may think it is a low number, but for people working here, there are a lot of problems," says S. Rajasekar, the hospital's deputy superintendent. Despite repeated requests, he says, the centre has the same resources it did in 1993, when it had just two HIV-positive patients. In 2004, it saw 14,991 new patients and had 140,000 hospital visits from HIV-positive patients. "With just 25 doctors," says Rajasekar. "Amazing, right?"

By June 2005, government centres, including Tambaram Hospital, had doled out ARVs to 8,000 people. In the same time, since April 2004, small private and non-profit clinics reached an estimated 30,000 sufferers. But these clinics are in a constant struggle for survival.

One such centre is the Naz Foundation's orphanage in New Delhi. Of the 24 children there — ranging in age from 19 months to 17 years — 10 are on ARVs. Despite one child dying two years ago, only the oldest one knows that she is HIV positive. To spare the children from stigma, their status has also been kept secret from their teachers and neighbours.

One child's monthly supply of ARVs can cost about Rs900 (US$20). The home was funded by the Gere Foundation until March 2005, but since then money has come almost entirely from small, private donations. "Care is something no traditional donor wants to fund," says the centre's director, Anjali Gopalan. "They see it as a black hole, as one donor told me. There's no return on the dollar."

Scrambling to treat their patients, doctors at some clinics use medicines that are past their expiry date; others bring free drugs they are given in the United States or elsewhere. Staff at the YRG Care Clinic in Chennai last year began asking people to donate just $10 each. "It's always beg, borrow, steal, donations, fundraise. That's how we get funds for care," says Suniti Solomon, who runs the YRG. "We cannot save the millions out there. The government has to do that." **A.M.**

THE EPICENTRE

In Bangalore, the bustling capital of the southern state of Karnataka, people drive shiny new cars, work in gleaming new buildings and carry mobile phones. Just a few hundred kilometres north, the residents are poorer than many in sub-Saharan Africa. In Bangalore, dubbed India's Silicon Valley, the land is green, but here it is brown and dry. This is the home of one of India's two AIDS hotspots.

"Someone's got to wake up to the fact that there are two Karnatakas," says Ashok Alexander, director of Avahan, the Bill & Melinda Gates Foundation's AIDS programme in India.

Driven by poverty and unemployment, women from northern towns such as Bijapur and Belgaum travel across the border to work as prostitutes in the richer cities in the adjoining

Maharashtra state. When they return home, they bring the money they've earned — and HIV. On a map (right), the districts along this 'AIDS corridor' are immediately visible. "This is the epicentre of the whole Indian AIDS epidemic," Alexander says.

He has discussed his theory with several people in the government. "We entirely agree with him because those pockets are surely high-prevalence," says S. Y. Quraishi, director-general of the National AIDS Control Organization.

Yet until Avahan — meaning 'call to action' in Sanskrit — was launched in April 2003, Karnataka was largely ignored by AIDS groups. Most donor agencies 'adopted' other high-prevalence states: the US Agency for International Development

SOURCE: AVAHAN, 2004

HIV prevalence in pregnant women
- Less than 1%
- 1–2.9%
- 3% or more
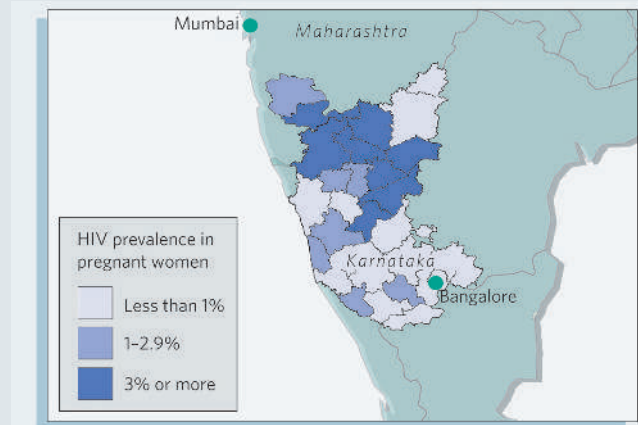
earmarks its money for Tamil Nadu and Maharashtra, for instance, and Britain's Department for International Development champions states such as Andhra Pradesh.

Avahan has already spent more than US$17 million working with sex workers, truck drivers and

drug users. It is also backing research on migration patterns in the corridor, large-scale surveys to measure behavioural changes and mathematical models that chart the epidemic's course. "We're going to be here as long as it takes to make an impact," says Alexander.   **A.M.**

housewives. In much of India, arranged marriages are still common, notes Rochelle D'Souza Yepthomi, information officer at YRG Care. Half of Indian women marry by the age of 18, when most are still virgins, and the majority remain faithful. "Their only risk is their husband. It's scary," she says.

My brother died of AIDS. When he was sick, the doctors kept telling him he had tuberculosis. When they said he had AIDS, they didn't explain. They said there was no treatment, that he would die. We took him back to the village. Nobody there knew anything about AIDS. I found out about it in the city. I told him not to sleep with his wife, to not eat from the same plate. I didn't see that anywhere, that you can't eat from the same plate. He died one and half years ago. There were 1,400 people at the funeral, he was that kind of man.
  — Mohan Singh, tourist-car driver

In the past year, two movies from Bollywood, India's prolific film industry, have revolved around HIV and AIDS, and the popular television show *Indian Idol* featured an AIDS-related song. Film and cricket stars have been featured in adverts sponsored by the Heroes Project, an AIDS awareness scheme, launched last year by the actor Richard Gere. Since November 2003, the non-profit organization Population Services International has also run a series of adverts, on television and billboards, featuring the fictitious character Balbir Pasha in Mumbai and Pulli Raja in Chennai. But in a country where more than 70% of the population lives in villages, the reach of the adverts is limited.

There are 638,000 villages in India, many of them with bad roads, no running water and

no medical care. The best estimates are that 59% of those infected with HIV live in rural areas, but because there is limited surveillance there, the real number is likely to be higher. Compounded by ignorance and illiteracy, myriad myths about AIDS have taken root: that eating chicken and brinjal (aubergine) is harmful, for example, or that the virus can be transmitted by mosquito bites.

Some largely rural states such as Uttar Pradesh and Bihar have all the ingredients for a massive epidemic — illiteracy, poverty, migrant labour and sex traffic — but have so far received little attention either from the government or from donor agencies. The model of success for those states is perhaps Tamil Nadu, where the epidemic spawned dozens of NGOs that conduct education programmes and provide counselling and support. As a result, it is the only state where prevalence has dropped — from 1.13% in 2001 to 0.63% at the last count.

Before marriage I used to go to those women. Now I don't. I just drink and go to sleep. I know I shouldn't go to women, that's all I know. My conscience tells me I'm healthy. I can't have any such thing as AIDS.
  — Sekar, truck driver

It's 10 p.m. on a warm night in February at a truck stop on National Highway 45, about 45 kilometres outside Chennai. Every day, more than 300 trucks stop here on their way south. At this hour, there are about 20 trucks parked, and drivers are eating dinner and preparing to sleep. At one end of the truck stop is a tent, plastered with AIDS posters, where some are watching an AIDS-themed show. In front, Anthony Selvaraj uses a penis model to show a driver how to use a condom, then convinces

the man to visit the makeshift clinic next to the tent.

"The word HIV we keep quiet. If you start with HIV, they'll run away," Selvaraj explains. "That's my last chance to talk to him so I should prepare to interact with him in such a way that he remembers."

These truckers go home only once every few months. On the road, they have multiple sexual partners and often abuse drugs and alcohol. Many believe that if they don't have sex they will build up *garmi*, a Hindi word meaning heat, and will go blind.

In recent years, they've become accustomed to being chased by volunteers with flipcharts and penis models in their pockets. Selvaraj works with the HOPE Foundation, which has counsellors and doctors at several stops along a 60-km stretch on the highway. But the message has not entirely got through. Believing that AIDS can only be transmitted by women, many truckers now have sex with 'cleaners', the young boys who travel with them. Homosexuality is illegal in India, so the government's campaigns mostly neglect men who have sex with men.

These truckers have ample access to condoms, but many use the condoms to plug pipes in the trucks. According to a report by India's comptroller and auditor-general, 75% of the 1.5 billion condoms made each year are used as sealants for leaky roofs, as lubricants in weaving gold-embroidered saris, or are mixed with concrete and tar to pave roads.

"We're going for 100% condom promotion," says health minister Anbumani Ramadoss. "Our job is giving them awareness. After that, if they don't use it, what can we do?" ■
Apoorva Mandavilli is senior news editor of *Nature Medicine*.

ARTICLES

Experimental investigation of geologically produced antineutrinos with KamLAND

T. Araki[1], S. Enomoto[1], K. Furuno[1], Y. Gando[1], K. Ichimura[1], H. Ikeda[1], K. Inoue[1], Y. Kishimoto[1], M. Koga[1], Y. Koseki[1], T. Maeda[1], T. Mitsui[1], M. Motoki[1], K. Nakajima[1], H. Ogawa[1], M. Ogawa[1], K. Owada[1], J.-S. Ricol[1], I. Shimizu[1], J. Shirai[1], F. Suekane[1], A. Suzuki[1], K. Tada[1], S. Takeuchi[1], K. Tamae[1], Y. Tsuda[1], H. Watanabe[1], J. Busenitz[2], T. Classen[2], Z. Djurcic[2], G. Keefer[2], D. Leonard[2], A. Piepke[2], E. Yakushev[2], B. E. Berger[3], Y. D. Chan[3], M. P. Decowski[3], D. A. Dwyer[3], S. J. Freedman[3], B. K. Fujikawa[3], J. Goldman[3], F. Gray[3], K. M. Heeger[3], L. Hsu[3], K. T. Lesko[3], K.-B. Luk[3], H. Murayama[3], T. O'Donnell[3], A. W. P. Poon[3], H. M. Steiner[3], L. A. Winslow[3], C. Mauger[4], R. D. McKeown[4], P. Vogel[4], C. E. Lane[5], T. Miletic[5], G. Guillian[6], J. G. Learned[6], J. Maricic[6], S. Matsuno[6], S. Pakvasa[6], G. A. Horton-Smith[7], S. Dazeley[8], S. Hatakeyama[8], A. Rojas[8], R. Svoboda[8], B. D. Dieterle[9], J. Detwiler[10], G. Gratta[10], K. Ishii[10], N. Tolich[10], Y. Uchida[10], M. Batygov[11], W. Bugg[11], Y. Efremenko[11], Y. Kamyshkov[11], A. Kozlov[11], Y. Nakamura[11], H. J. Karwowski[12], D. M. Markoff[12], K. Nakamura[12], R. M. Rohm[12], W. Tornow[12], R. Wendell[12], M.-J. Chen[13], Y.-F. Wang[13] & F. Piquemal[14]

The detection of electron antineutrinos produced by natural radioactivity in the Earth could yield important geophysical information. The Kamioka liquid scintillator antineutrino detector (KamLAND) has the sensitivity to detect electron antineutrinos produced by the decay of ^{238}U and ^{232}Th within the Earth. Earth composition models suggest that the radiogenic power from these isotope decays is 16 TW, approximately half of the total measured heat dissipation rate from the Earth. Here we present results from a search for geoneutrinos with KamLAND. Assuming a Th/U mass concentration ratio of 3.9, the 90 per cent confidence interval for the total number of geoneutrinos detected is 4.5 to 54.2. This result is consistent with the central value of 19 predicted by geophysical models. Although our present data have limited statistical power, they nevertheless provide by direct means an upper limit (60 TW) for the radiogenic power of U and Th in the Earth, a quantity that is currently poorly constrained.

The Kamioka liquid scintillator antineutrino detector (KamLAND) has demonstrated neutrino oscillation using electron antineutrinos ($\overline{\nu}_e$s) with energies of a few MeV from nuclear reactors[1,2]. Additionally, KamLAND is the first detector sensitive enough to measure $\overline{\nu}_e$s produced in the Earth from the ^{238}U and ^{232}Th decay chains. Using $\overline{\nu}_e$s to study processes inside the Earth was first suggested by Eder[3] and Marx[4], and has been reviewed a number of times[5–10]. As $\overline{\nu}_e$s produced from the ^{238}U and ^{232}Th decay chains have exceedingly small interaction cross-sections, they propagate undisturbed in the Earth's interior, and their measurement near the Earth's surface can be used to gain information on their sources. Although the detection of $\overline{\nu}_e$s from ^{40}K decay would also be of great interest in geophysics, with possible applications in the interpretation of geo-magnetism, their energies are too low to be detected with KamLAND. The antineutrino flux above our detection threshold from other long-lived isotopes is expected to be negligible.

The radiogenic Earth

The total power dissipated from the Earth (heat flow) has been measured with thermal techniques[11] to be 44.2 ± 1.0 TW. Despite this small quoted error, a more recent evaluation[12] of the same data (assuming much lower hydrothermal heat flow near mid-ocean ridges) has led to a lower figure of 31 ± 1 TW. On the basis of studies of chondritic meteorites[13] the calculated radiogenic power is thought to be 19 TW, 84% of which is produced by ^{238}U and ^{232}Th decay. Some models of mantle convection suggest that radiogenic power is a larger fraction of the total power[14,15].

^{238}U and ^{232}Th decay via a series of well-established α and β^- processes[16] terminating in the stable isotopes ^{206}Pb and ^{208}Pb, respectively. Each β^- decay produces a daughter nucleus, an electron and a $\overline{\nu}_e$. The $\overline{\nu}_e$ energy distribution is well established[17], and includes a correction for the electromagnetic interaction between the electron and the charge distribution of the daughter nucleus. Figure 1 shows the expected $\overline{\nu}_e$ distribution, $dn(E_\nu)/dE_\nu$, as a function of $\overline{\nu}_e$ energy, E_ν, for the ^{238}U and ^{232}Th decay chains.

Ignoring the negligible neutrino absorption, the expected $\overline{\nu}_e$ flux at a position **r** for each isotope is given by:

$$\frac{d\phi(E_\nu, \mathbf{r})}{dE_\nu} = A \frac{dn(E_\nu)}{dE_\nu} \int_{V_\oplus} d^3\mathbf{r}' \frac{a(\mathbf{r}')\rho(\mathbf{r}')P(E_\nu, |\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|^2} \quad (1)$$

[1]Research Center for Neutrino Science, Tohoku University, Sendai 980-8578, Japan. [2]Department of Physics and Astronomy, University of Alabama, Tuscaloosa, Alabama 35487, USA. [3]Physics Department, University of California at Berkeley and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]W. K. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125, USA. [5]Physics Department, Drexel University, Philadelphia, Pennsylvania 19104, USA. [6]Department of Physics and Astronomy, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA. [7]Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA. [8]Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA. [9]Physics Department, University of New Mexico, Albuquerque, New Mexico 87131, USA. [10]Physics Department, Stanford University, Stanford, California 94305, USA. [11]Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996, USA. [12]Physics Department, Duke University, Durham, North Carolina 27008, USA, and Physics Department, North Carolina State, Raleigh, North Carolina 27695, USA, and Physics Department, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [13]Institute of High Energy Physics, Beijing 100039, China. [14]CEN Bordeaux-Gradignan, IN2P3-CNRS and University Bordeaux I, F-33175 Gradignan Cedex, France.

where A is the decay rate per unit mass, the integral is over the volume of the Earth, $a(\mathbf{r}')$ is the isotope mass per unit rock mass, $\rho(\mathbf{r}')$ is the rock density, and $P(E_\nu, |\mathbf{r} - \mathbf{r}'|)$ is the $\bar{\nu}_e$ 'survival' probability after travelling a distance $|\mathbf{r} - \mathbf{r}'|$. This probability derives from the now accepted phenomenon of neutrino oscillation, and can be written, for two neutrino flavours as[18]

$$P(E_\nu, L) \cong 1 - \sin^2 2\theta_{12} \sin^2 \left(\frac{1.27\Delta m_{12}^2[\text{eV}^2]L[\text{m}]}{E_\nu[\text{MeV}]} \right) \qquad (2)$$

where $L = |\mathbf{r} - \mathbf{r}'|$. The neutrino oscillation parameters $\Delta m_{12}^2 = 7.9^{+0.6}_{-0.5} \times 10^{-5}\,\text{eV}^2$ and $\sin^2 2\theta_{12} = 0.82 \pm 0.07$ are also determined with KamLAND[2] using reactor $\bar{\nu}_e$s with energies above those of geoneutrinos, combined with solar neutrino experiments[19]. Corrections from three flavour neutrino oscillation ($<5\%$) and "matter effects"[20] ($\sim 1\%$) are ignored. For typical geoneutrino energies, the approximation $P(E_\nu, |L|) = 1 - 0.5\sin^2 2\theta_{12}$ only affects the accuracy of the integral in equation (1) at 1% owing to the distributed $\bar{\nu}_e$ production points. This approximation, used in this paper, neglects energy spectrum distortions.

Geoneutrino detection

KamLAND is located in the Kamioka mine, 1,000 m below the summit of Mt Ikenoyama, Gifu prefecture, Japan (36° 25′ 36″ N, 137° 18′ 43″ E). It detects electron antineutrinos in ~ 1 kton of liquid scintillator via neutron inverse β-decay,

$$\bar{\nu}_e + p \rightarrow e^+ + n \qquad (3)$$

which has a well-established cross-section[21] as a function of E_ν. Scintillation light from the e^+, 'prompt event', gives an estimate of the incident $\bar{\nu}_e$ energy, $E_\nu \approx E_{e^+} + 0.8\,\text{MeV}$ (neglecting the small neutron recoil), where E_{e^+} is the kinetic energy of the positron plus the electron–positron annihilation energy. With a mean time of $\sim 200\,\mu\text{s}$, the neutron is captured by a proton, producing a deuteron and a 2.2 MeV γ-ray. The detection of scintillation light from this 2.2 MeV γ-ray is referred to as the 'delayed event'. The spatial and temporal coincidences between the prompt and delayed events provide a powerful tool for reducing backgrounds, which generally limit the sensitivity in low energy neutrino studies.

A reference model[22] is constructed using seismic data to divide the Earth into continental crust, oceanic crust, mantle, core and sediment. Some of these regions are further sub-divided, with each sub-region having different U and Th concentrations. This model assumes that U and Th are absent from the core. The expected geoneutrino flux at KamLAND, including a suppression factor of 0.59 due to neutrino oscillations, is $2.34 \times 10^6\,\text{cm}^{-2}\,\text{s}^{-1}$ and
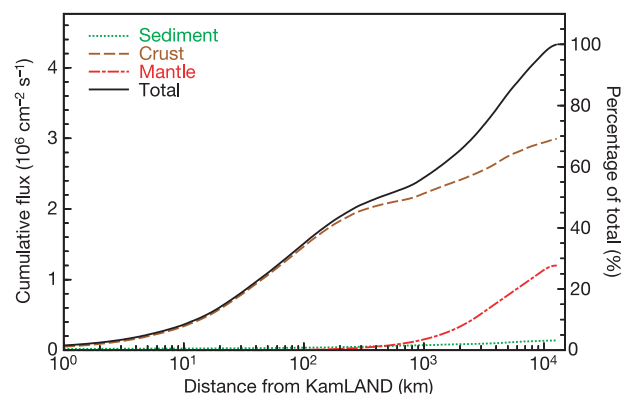
$1.98 \times 10^6\,\text{cm}^{-2}\,\text{s}^{-1}$ from the ^{238}U and ^{232}Th decay chains, respectively. Including the detection cross-section, the number of geoneutrinos expected at KamLAND from ^{238}U and ^{232}Th decay is $3.85 \times 10^{-31}\,\bar{\nu}_e$ per target proton per year, 79% of which is due to ^{238}U. Figure 2 shows that a large fraction of the expected geoneutrino flux originates in the area surrounding KamLAND. The effect of local geology was studied extensively in the context of the reference model[22] and was found to produce less than a 10% error on the total expected flux.

The data presented here are based on a total detector live-time of 749.1 ± 0.5 d after basic cuts to ensure the reliability of the data. The number of target protons is estimated at $(3.46 \pm 0.17) \times 10^{31}$ on the basis of target proton density and a spherical fiducial scintillator volume with 5 m radius, resulting in a total exposure of $(7.09 \pm 0.35) \times 10^{31}$ target proton years. The overall efficiency for detecting geoneutrino candidates with energies between 1.7 and 3.4 MeV in the fiducial volume is estimated to be 0.687 ± 0.007. The energy range reaches below the inverse β-decay threshold owing to the detector energy resolution.

Backgrounds for geoneutrino candidates are dominated by $\bar{\nu}_e$s from nuclear reactors in the vicinity of the detector, and by α-particle induced neutron backgrounds due to radioactive contamination within the detector. Reactor $\bar{\nu}_e$s reach substantially higher energies, as shown in Fig. 3. Therefore, the oscillation parameters in ref. 2 were determined by analysing $\bar{\nu}_e$s with energies greater than 3.4 MeV, where there is no signal from the geoneutrinos. Using these parameters, the number of nuclear reactor $\bar{\nu}_e$ background events used by the 'rate only' analysis discussed below is determined to be 80.4 ± 7.2.

The α-particle-induced neutron background is due to the ^{13}C$(\alpha,n)^{16}$O reaction where the α-particle is produced in ^{210}Po decay with a kinetic energy of 5.3 MeV. The ^{210}Po is produced by the decay of ^{210}Pb, which has a half-life of 22 yr. The ^{210}Pb resulted from the decay of ^{222}Rn contamination, and is distributed throughout the detector. The neutrons in the ^{13}C$(\alpha,n)^{16}$O reaction are produced with kinetic energy up to 7.3 MeV. Owing to scintillation light quenching for high ionization density, only about one-third of this energy is converted into 'visible' energy as the neutrons thermalize. The thermal neutrons are captured by protons with a mean capture time of $\sim 200\,\mu\text{s}$, producing a delayed signal identical to that from neutron inverse β-decay. The number of ^{13}C nuclei in the fiducial volume is determined from the measured ^{13}C/^{12}C ratio in the KamLAND scintillator. On the basis of the ^{13}C$(\alpha,n)^{16}$O reaction cross-section[23], the α-particle energy loss in the scintillator[24], and the number of ^{210}Po decays, the total number of neutrons produced is expected to be 93 ± 22. This error is dominated by estimated 20% and 14% uncertainties in the total ^{13}C$(\alpha,n)^{16}$O reaction cross-section

Figure 1 | The expected ^{238}U, ^{232}Th and ^{40}K decay chain electron antineutrino energy distributions. KamLAND can only detect electron antineutrinos to the right of the vertical dotted black line; hence it is insensitive to ^{40}K electron antineutrinos.

Figure 2 | The expected total ^{238}U and ^{232}Th geoneutrino flux within a given distance from KamLAND[22]. Approximately 25% and 50% of the total flux originates within 50 km and 500 km of KamLAND, respectively. The line representing the crust includes both the continental and the almost negligible oceanic contribution.
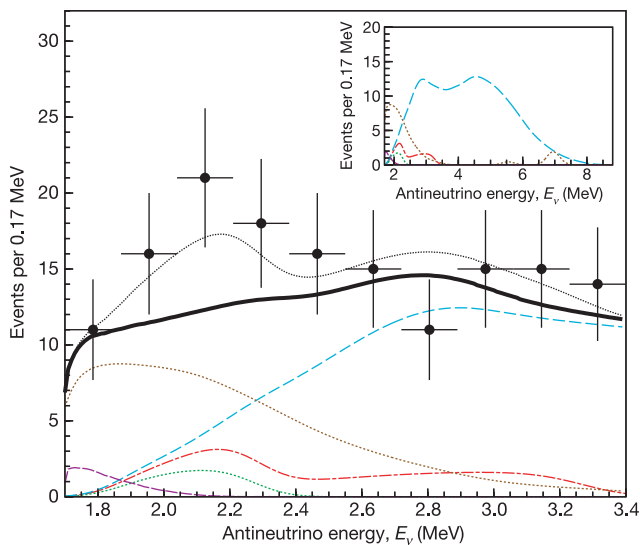
and the number of ^{210}Po decays, respectively. The neutron energy distribution is calculated using the measured neutron angular distributions in the centre of mass frame[25,26]. Including the efficiency for passing the $\bar{\nu}_e$ candidate cuts, the number of (α,n) background events is estimated to be 42 ± 11.

There is a small contribution to the background from random coincidences, $\bar{\nu}_e$s from the β^- decay of long lived nuclear reactor fission products, and radioactive isotopes produced by cosmic rays. Using an out-of-time coincidence cut from 10 ms to 20 s, the random coincidence background is estimated to be 2.38 ± 0.01 events. Using the expected $\bar{\nu}_e$ energy spectrum[27] for long lived nuclear reactor fission products, the corresponding background is estimated to be 1.9 ± 0.2 events. The most significant background due to radioactive isotopes produced by cosmic rays is from the β^- decay $^9\text{Li} \rightarrow 2\alpha + n + e^- + \bar{\nu}_e$, which has a neutron in the final state. On the basis of events correlated with cosmic rays, the estimated number of background events caused by radioactive ^{9}Li is 0.30 ± 0.05. Other backgrounds considered and found to be negligible include spontaneous fission, neutron emitters and correlated decays in the radioactive background decay chains, fast neutrons from cosmic ray interactions, (γ,n) reactions and solar ν_e induced break-up of ^{2}H. The total background is estimated to be 127 ± 13 events (1σ error).

The total number of observed $\bar{\nu}_e$ candidates is 152, with their energy distribution shown in Fig. 3. Including the geoneutrino detection systematic errors, parts of which are correlated with the background estimation errors, a 'rate only' analysis gives 25^{+19}_{-18} geoneutrino candidates from the ^{238}U and ^{232}Th decay chains. Dividing by the detection efficiency, live-time, and number of target protons, the total geoneutrino detected rate obtained is $5.1^{+3.9}_{-3.6} \times 10^{-31}$ $\bar{\nu}_e$ per target proton per year.

We also perform an un-binned maximum likelihood analysis of the $\bar{\nu}_e$ energy spectrum between 1.7 and 3.4 MeV, using the known shape of the signal and background spectra. As the neutrino oscillation parameters do not significantly affect the expected shape of the geoneutrino signal, the un-oscillated shape is assumed. However, the

oscillation parameters are included in the reactor background shape. Figure 4a shows the confidence intervals for the number of observed ^{238}U and ^{232}Th geoneutrinos. Based on a study of chondritic meteorites[28], the Th/U mass ratio in the Earth is believed to be between 3.7 and 4.1, and is known better than either absolute concentration. Assuming a Th/U mass ratio of 3.9, we estimate the 90% confidence interval for the total number of ^{238}U and ^{232}Th geoneutrino candidates to be 4.5 to 54.2, as shown in Fig. 4b. The central value of 28.0 is consistent with the 'rate only' analysis. At this point, the value of the fit parameters are $\Delta m^2_{12} = 7.8 \times 10^{-5}$ eV^2, $\sin^2 2\theta_{12} = 0.82$, $p_\alpha = 1.0$, and $q_\alpha = 1.0$, where these last two parameters are defined in the Methods section. The 99% confidence upper limit obtained on the total detected ^{238}U and ^{232}Th geoneutrino rate is 1.45×10^{-30} $\bar{\nu}_e$ per target proton per year, corresponding to a flux at KamLAND of 1.62×10^7 cm^{-2} s^{-1}. On the basis of our reference model, this corresponds to an upper limit on the radiogenic power from ^{238}U and ^{232}Th decay of 60 TW.

As a cross-check, an independent analysis[29] has been performed using a partial data set, including detection efficiency, of 2.6×10^{31} target proton years. In this analysis, the ^{13}C$(\alpha,n)^{16}$O background was

Figure 3 | $\bar{\nu}_e$ energy spectra in KamLAND. Main panel, experimental points together with the total expectation (thin dotted black line). Also shown are the total expected spectrum excluding the geoneutrino signal (thick solid black line), the expected signals from ^{238}U (dot-dashed red line) and ^{232}Th (dotted green line) geoneutrinos, and the backgrounds due to reactor $\bar{\nu}_e$ (dashed light blue line), ^{13}C$(\alpha,n)^{16}$O reactions (dotted brown line), and random coincidences (dashed purple line). Inset, expected spectra extended to higher energy. The geoneutrino spectra are calculated from our reference model, which assumes 16 TW radiogenic power from ^{238}U and ^{232}Th. The error bars represent ± 1 standard deviation intervals.

Figure 4 | Confidence intervals for the number of geoneutrinos detected. Panel a shows the 68.3% confidence level (CL; red), 95.4% CL (green) and 99.7% CL (blue) contours for detected ^{238}U and ^{232}Th geoneutrinos. The small shaded area represents the prediction from the geophysical model. The vertical dashed line represents the value of $(N_U - N_{Th})/(N_U + N_{Th})$ assuming the mass ratio, Th/U = 3.9, derived from chondritic meteorites, and accounting for the ^{238}U and ^{232}Th decay rates and the $\bar{\nu}_e$ detection efficiencies in KamLAND. The dot represents our best fit point, favouring 3 ^{238}U geoneutrinos and 18 ^{232}Th geoneutrinos. Panel b shows $\Delta\chi^2$ as a function of the total number of ^{238}U and ^{232}Th geoneutrino candidates, fixing the normalized difference to the chondritic meteorites constraint. The grey band gives the value of $N_U + N_{Th}$ predicted by the geophysical model.
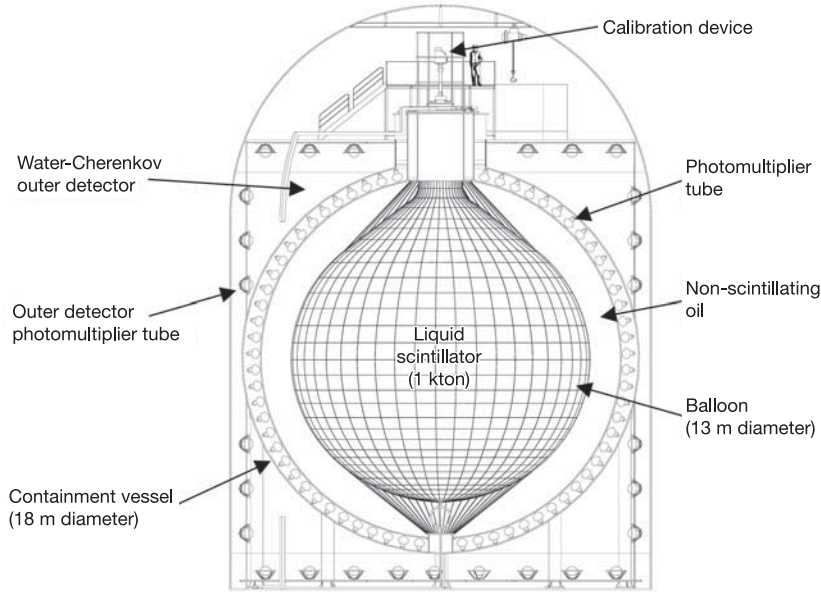
Figure 5 | Schematic diagram of the KamLAND detector.

verified using the minute differences in the time structures of scintillation light from different particle species. Scintillation light in the prompt part of $\overline{\nu}_e$ events is caused by positrons, whereas scintillation light in the prompt part of $^{13}C(\alpha,n)^{16}O$ background events is caused by neutron thermalization. This alternative analysis produced a slightly larger geoneutrino signal, which is consistent with the results presented here.

Discussion and future prospects

In conclusion, we have performed the first experimental study of antineutrinos from the Earth's interior using KamLAND. The present measurement is consistent with current geophysical models, and constrains the $\overline{\nu}_e$ emission from U and Th in the planet to be less than 1.45×10^{-30} $\overline{\nu}_e$ per target proton per year at 99% confidence limits, corresponding to a flux of $1.62 \times 10^7\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. There is currently a programme underway to reduce the ^{210}Pb content of the detector. This should help to reduce the substantial systematic error due to the $^{13}C(\alpha,n)^{16}O$ background. Further background reduction will require a new detector location, far away from nuclear reactors. The reported investigation of geoneutrinos should pave the way to future and more accurate measurements, which may provide a new window for the exploration of the Earth.

METHODS

As shown in Fig. 5, KamLAND[1] consists of 1 kton of ultrapure liquid scintillator contained in a transparent nylon/EVOH (ethylene vinyl alcohol copolymer) composite film balloon suspended in non-scintillating oil. Charged particles deposit their kinetic energy in the scintillator; some of this energy is converted into scintillation light. The scintillation light is then detected by an array of 1,325 17-inch-diameter photomultiplier tubes (PMTs) and 554 20-inch-diameter PMTs mounted on the inner surface of an 18-m-diameter spherical stainless-steel containment vessel. A 3.2-kton water-Cherenkov detector with 225 20-inch-diameter PMTs surrounds the containment sphere. This outer detector tags cosmic-ray muons and absorbs γ-rays and neutrons from the surrounding rock.

The arrival times of photons at the PMTs allow us to determine the location of particle interactions inside the detector, and the amount of detected light after correcting for spatial variation of the detector response allows us to determine the particle's energy. The event location and energy determination is calibrated with γ-ray sources deployed vertically down the centre of the detector. To be classified as a $\overline{\nu}_e$ candidate, the time coincidence between the prompt and delayed events (ΔT) is required to satisfy $0.5\,\mu s < \Delta T < 500\,\mu s$. The position of the prompt ($\mathbf{r}_p$) and delayed ($\mathbf{r}_d$) events with respect to the centre of the detector are required to satisfy $|\mathbf{r}_p| < 5\,\mathrm{m}$, $|\mathbf{r}_d| < 5\,\mathrm{m}$ and $|\mathbf{r}_p - \mathbf{r}_d| < 1.0\,\mathrm{m}$. The energy of

the electron antineutrino is required to satisfy $1.7\,\mathrm{MeV} < E_\nu < 3.4\,\mathrm{MeV}$ and the energy of the delayed event (E_d) is required to satisfy $1.8\,\mathrm{MeV} < E_d < 2.6\,\mathrm{MeV}$.

Given N_U and N_{Th} geoneutrinos detected from the ^{238}U and ^{232}Th decay chains, the expected energy distribution of the candidates is

$$\frac{d\overline{N}(E_\nu)}{dE_\nu} = N_U \frac{dP_U(E_\nu)}{dE_\nu} + N_{Th} \frac{dP_{Th}(E_\nu)}{dE_\nu} + \frac{dN_r\left(E_\nu; \Delta m_{12}^2, \sin^2 2\theta_{12}\right)}{dE_\nu}$$
$$+ p_\alpha \frac{dN_\alpha(E_\nu/q_\alpha)}{dE_\nu} + \sum_k \frac{dN_k(E_\nu)}{dE_\nu} \quad (4)$$

where $dP_U(E_\nu)/dE_\nu$ and $dP_{Th}(E_\nu)/dE_\nu$ are the normalized expected geoneutrino spectra from ^{238}U and ^{232}Th decay chains. The third term on the right hand side of equation (4) is the energy spectrum of the expected $\overline{\nu}_e$ reactor background, which is a function of the neutrino oscillation parameters Δm_{12}^2 and $\sin^2 2\theta_{12}$. $dN_\alpha(E_\nu/q_\alpha)/dE_\nu$ is the energy spectrum of the expected $^{13}C(\alpha,n)^{16}O$ background with energy and rate scaling factors q_α and p_α, respectively. The sum is over the other known backgrounds where $dN_k(E_\nu)/dE_\nu$ is the expected energy spectrum of the background. All expected spectra include energy smearing due to the detector energy resolution. Integrating equation (4) between 1.7 and 3.4 MeV gives the total number of expected candidates, $\overline{N}$.

The number of geoneutrinos from the ^{238}U and ^{232}Th decay chains is determined from an unbinned maximum likelihood fit. The log likelihood is defined by

$$\log L = -\frac{(N - \overline{N})^2}{2\left(\overline{N} + \sigma_{\overline{N}}^2\right)} + \sum_{i=1}^N \log\left(\frac{1}{\overline{N}} \frac{d\overline{N}(E_i)}{dE_\nu}\right) - \frac{(p_\alpha - 1)^2}{2\sigma_p^2} - \frac{(q_\alpha - 1)^2}{2\sigma_p^2}$$
$$- \frac{\chi^2\left(\Delta m_{12}^2, \sin^2 2\theta_{12}\right)}{2} \quad (5)$$

where N is the total number of observed candidates and $\sigma_{\overline{N}}$ is the error on $\overline{N}$. E_i is the energy of the ith $\overline{\nu}_e$ candidate. $\sigma_p = 0.24$ and $\sigma_q = 0.1$ are the fractional errors on p_α and q_α, respectively. The term $\chi^2(\Delta m^2, \sin^2 2\theta)$ provides a constraint on the neutrino oscillation parameters from the KamLAND reactor measurements and the solar neutrino results[30]. $\log L$ is maximized at different values of N_U and N_{Th} by varying Δm_{12}^2, $\sin^2 2\theta_{12}$, p_α and q_α. The best fit point for N_U and N_{Th} corresponds to the maximum $\log L$. A $\Delta\chi^2$ is defined by

$$\Delta\chi^2 = 2(\log L_{max} - \log L) \quad (6)$$

where $\log L_{max}$ is the $\log L$ at the best fit point. The confidence intervals are calculated from this $\Delta\chi^2$.

1. Eguchi, K. *et al.* First results from KamLAND: Evidence for reactor antineutrino disappearance. *Phys. Rev. Lett.* **90**, 021802 (2003).

2. Araki, T. *et al.* Measurement of neutrino oscillation with KamLAND: Evidence of spectral distortion. *Phys. Rev. Lett.* **94**, 081801 (2005).
3. Eder, G. Terrestrial neutrinos. *Nucl. Phys.* **78**, 657–662 (1966).
4. Marx, G. Geophysics by neutrinos. *Czech. J. Phys. B* **19**, 1471–1479 (1969).
5. Avilez, C., Marx, G. & Fuentes, B. Earth as a source of antineutrinos. *Phys. Rev. D* **23**, 1116–1117 (1981).
6. Krauss, L. M., Glashow, S. L. & Schramm, D. N. Antineutrino astronomy and geophysics. *Nature* **310**, 191–198 (1984).
7. Kobayashi, M. & Fukao, Y. The Earth as an antineutrino star. *Geophys. Res. Lett.* **18**, 633–636 (1991).
8. Raghavan, R. S. *et al.* Measuring the global radioactivity in the Earth by multidetector antineutrino spectroscopy. *Phys. Rev. Lett.* **80**, 635–638 (1998).
9. Rothschild, C. G., Chen, M. C. & Calaprice, F. P. Antineutrino geophysics with liquid scintillator detectors. *Geophys. Res. Lett.* **25**, 1083–1086 (1998).
10. Mantovani, F., Carmignani, L., Fiorentini, G. & Lissia, M. Antineutrinos from Earth: A reference model and its uncertainties. *Phys. Rev. D* **69**, 013001 (2004).
11. Pollack, H. N., Hurter, S. J. & Johnson, J. R. Heat flow from the Earth's interior: analysis of the global data set. *Rev. Geophys.* **31**, 267–280 (1993).
12. Hofmeister, A. M. & Criss, R. E. Earth's heat flux revised and linked to chemistry. *Tectonophysics* **395**, 159–177 (2005).
13. McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
14. Jackson, M. J. & Pollack, H. N. On the sensitivity of parameterized convection to the rate of decay of internal heat sources. *J. Geophys. Res.* **89**, 10103–10108 (1984).
15. Richter, F. M. Regionalized models for the thermal evolution of the Earth. *Earth Planet. Sci. Lett.* **68**, 471–484 (1984).
16. Firestone, R. B. *Table of Isotopes* 8th edn (John Wiley, New York, 1996).
17. Behrens, H. & Jänecke, J. *Landolt-Börnstein - Group I, Elementary Particles, Nuclei and Atoms* Vol. 4 (Springer, Berlin, 1969).
18. McKeown, R. D. & Vogel, P. Neutrino masses and oscillations: triumphs and challenges. *Phys. Rep.* **394**, 315–356 (2004).
19. Ahmed, S. N. *et al.* Measurement of the total active ^{8}B solar neutrino flux at the Sudbury Neutrino Observatory with enhanced neutral current sensitivity. *Phys. Rev. Lett.* **92**, 181301 (2004).
20. Wolfenstein, L. Neutrino oscillations in matter. *Phys. Rev. D* **17**, 2369–2374 (1978).
21. Vogel, P. & Beacom, J. F. Angular distribution of neutron inverse beta decay, $\bar{\nu}_e + p \rightarrow e^+ + n$. *Phys. Rev. D* **60**, 053003 (1999).
22. Enomoto, S. *Neutrino Geophysics and Observation of Geo-neutrinos at KamLAND.* Thesis, Tohoku Univ. (2005); available at ⟨http://www.awa.tohoku.ac.jp/KamLAND/publications/Sanshiro_thesis.pdf⟩.
23. JENDL Japanese Evaluated Nuclear Data Library. ⟨http://wwwndc.tokai.jaeri.go.jp/jendl/jendl.html⟩ (2004).
24. Apostolakis, J. Geant—Detector description and simulation tool. ⟨http://wwwasd.web.cern.ch/wwwasd/geant/index.html⟩ (2003).
25. Walton, R. B., Clement, J. D. & Borlei, F. Interaction of neutrons with oxygen and a study of the C$^{13}(\alpha,n)$O^{16} reaction. *Phys. Rev.* **107**, 1065–1075 (1957).
26. Kerr, G. W., Morris, J. M. & Risser, J. R. Energy levels of ^{17}O from ^{13}C$(\alpha,\alpha_0)^{13}$C and ^{13}C$(\alpha,n)^{16}$O. *Nucl. Phys. A* **110**, 637–656 (1968).
27. Kopeikin, V. I. *et al.* Inverse beta decay in a nonequilibrium antineutrino flux from a nuclear reactor. *Phys. Atom. Nuclei* **64**, 849–854 (2001).
28. Rocholl, A. & Jochum, K. P. Th, U and other trace elements in carbonaceous chondrites: Implications for the terrestrial and solar-system Th/U ratios. *Earth Planet. Sci. Lett.* **117**, 265–278 (1993).
29. Tolich, N. *Experimental Study of Terrestrial Electron Anti-neutrinos with KamLAND* Thesis, Stanford Univ. (2005); available at ⟨http://www.awa.tohoku.ac.jp/KamLAND/publications/Nikolai_thesis.pdf⟩.
30. KamLAND collaboration. Data release accompanying the 2nd KamLAND reactor result. ⟨http://www.awa.tohoku.ac.jp/KamLAND/datarelease/2ndresult.html⟩ (2005).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.E. (sanshiro@awa.tohoku.ac.jp) or N.T. (nrtolich@lbl.gov).

ARTICLES

Evidence for magmatic evolution and diversity on Mars from infrared observations

P. R. Christensen[1], H. Y. McSween Jr[2], J. L. Bandfield[1], S. W. Ruff[1], A. D. Rogers[1], V. E. Hamilton[3], N. Gorelick[1], M. B. Wyatt[1], B. M. Jakosky[4], H. H. Kieffer[5], M. C. Malin[6] & J. E. Moersch[2]

Compositional mapping of Mars at the 100-metre scale with the Mars Odyssey Thermal Emission Imaging System (THEMIS) has revealed a wide diversity of igneous materials. Volcanic evolution produced compositions from low-silica basalts to high-silica dacite in the Syrtis Major caldera. The existence of dacite demonstrates that highly evolved lavas have been produced, at least locally, by magma evolution through fractional crystallization. Olivine basalts are observed on crater floors and in layers exposed in canyon walls up to 4.5 km beneath the surface. This vertical distribution suggests that olivine-rich lavas were emplaced at various times throughout the formation of the upper crust, with their growing inventory suggesting that such ultramafic (picritic) basalts may be relatively common. Quartz-bearing granitoid rocks have also been discovered, demonstrating that extreme differentiation has occurred. These observations show that the martian crust, while dominated by basalt, contains a diversity of igneous materials whose range in composition from picritic basalts to granitoids rivals that found on the Earth.

High-resolution imaging of the martian surface has allowed the construction of morphologic maps showing considerable diversity at local and regional scales. However, petrology has not kept pace and so virtually no information exists about mineralogic or geochemical variations at comparable spatial scales. Remote sensing data from the Mars Global Surveyor Thermal Emission Spectrometer (TES) and visible/near-infrared (IR) spectroscopy from spacecraft-based and Earth-based telescopes have revealed mineralogic variations between regional-scale units[1–4], but martian igneous and sedimentary processes almost certainly resulted in much more diversity than is presently recognized. Volcanic units are probably composed of materials formed by different degrees of partial melting and modified by fractional crystallization and interaction with crustal compositions during magma ascent and eruption. Similarly, sedimentary units presumably experienced mineral fractionations during transport and deposition, and erosion, tectonic activity and impact may have excavated different subsurface materials at local scales.

THEMIS provides IR spectroscopic data in eight surface-sensing wavelength bands from 6.8 to 13.6 µm at sufficiently high spatial resolution (~100 m per pixel) to identify and map local geologic units having different mineralogies[5]. Here we use a highly effective strategy for compositional mapping that combines THEMIS and TES data, selecting TES pixels within mapped THEMIS units for detailed spectroscopic analysis, to extend the mineralogic mapping achieved with TES[1,2,6–8] to spatial scales relevant to many geologic processes.

Magma evolution in the Nili Patera volcano

Nili Patera forms the northernmost caldera of a caldera complex 400 by 200 km at the summit of the ~1,300-km-diameter Syrtis Major

volcanic construct of Hesperian (~2–3 billion years old) age[9] (Fig. 1). Nili Patera is ~50 km in diameter and lies ~1.8 km below the western caldera rim on the basis of MOLA topography[10]. The floor has been domed by ~300 m following the initial emplacement of the floor units, producing a N–S-trending fracture system, with subsequent flows that cross these grabens (Fig. 1a).

Approximately a third of the floor materials (~900 km^2) have unusually high night-time temperatures (208 to 214 K; Fig. 1b) that indicate the presence of bedrock or highly consolidated materials[5]. Barchan dunes, which form through sand transport across a hard surface where sand is limited, are seen on the western portion of these bedrock surfaces (Figs 1 and 2). These dunes formed in a southwest-directed wind regime, suggesting that wind-driven sand may be the erosive agent that has scoured the upwind surface to bedrock. The entire caldera floor, including the dunes, has night-time temperatures $> 192 \pm 1$ K, corresponding to thermal inertias $>400\,\mathrm{J\,m^{-2}\,s^{-1/2}\,K^{-1}}$ (refs 11, 12) and unconsolidated particle diameters $>\sim 1$ mm (ref. 13), indicating that all of the floor materials are coarse-grained. The low TES albedo (<0.12; ref. 7) and low spectrally derived dust cover index[14] confirm that these are dust-free surfaces.

Compositional mapping with THEMIS multi-spectral and TES hyperspectral[15] IR imaging shows distinct variations across the caldera floor and the surrounding flanks of the volcano (Figs 3 and 4). The two primary floor units are a 'magenta' unit (unit A) that occurs in an area ~12 by 12 km on the central portion of the caldera floor with several nearby outliers, and a 'blue' unit (unit B) that covers much of the eastern and western portions of the caldera floor (Fig. 4). Several 5–20-km-sized flow units that are spectrally similar to unit A occur 50–60 km to the south along an arcuate trend (Fig. 3).

Unit A is lobate with steep margins, stands ~80 m above the floor,

[1]Department of Geological Sciences, Arizona State University Tempe, Arizona 85287, USA. [2]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, Tennessee 37996, USA. [3]Institute of Geophysics and Planetology, University of Hawaii, Honolulu, Hawaii 96822, USA. [4]University of Colorado, Boulder, Colorado, USA. [5]US Geological Survey, Emeritus, Carson City, Nevada 89703, USA. [6]Malin Space Science Systems, San Diego, California 92191, USA.
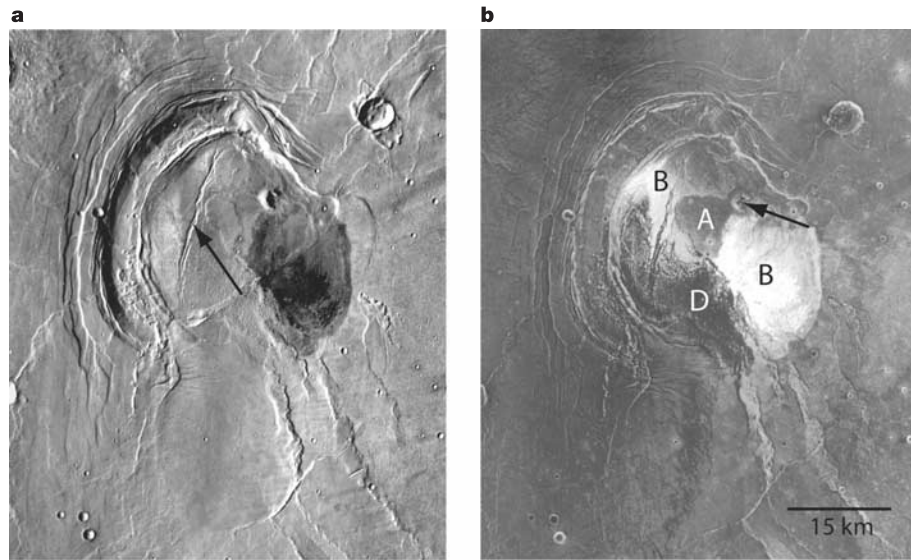
Figure 1 | **Daytime and night-time mosaics of THEMIS thermal images of the summit region of Syrtis Major showing the Nili Patera caldera and surrounding lava plains.** **a**, Daytime mosaic. The caldera is delineated on the western side by concentric ring fractures. The cool (dark) region of the southeast caldera floor is the bedrock surface. A N–S-trending graben system (arrow) was produced by doming after floor emplacement. **b**, Night-time mosaic. The bedrock surfaces of unit B are bright (warm) at night with temperatures >210 K. The colder (dark) lobate unit A (A) is clearly delineated from the surrounding caldera floor by its uniformly lower temperatures and is associated with a construct (arrow). Dunes (D) cover much of the southeastern portions of the caldera floor. Each mosaic is centred near 9° N, 67° E. North is towards the top in all figures.
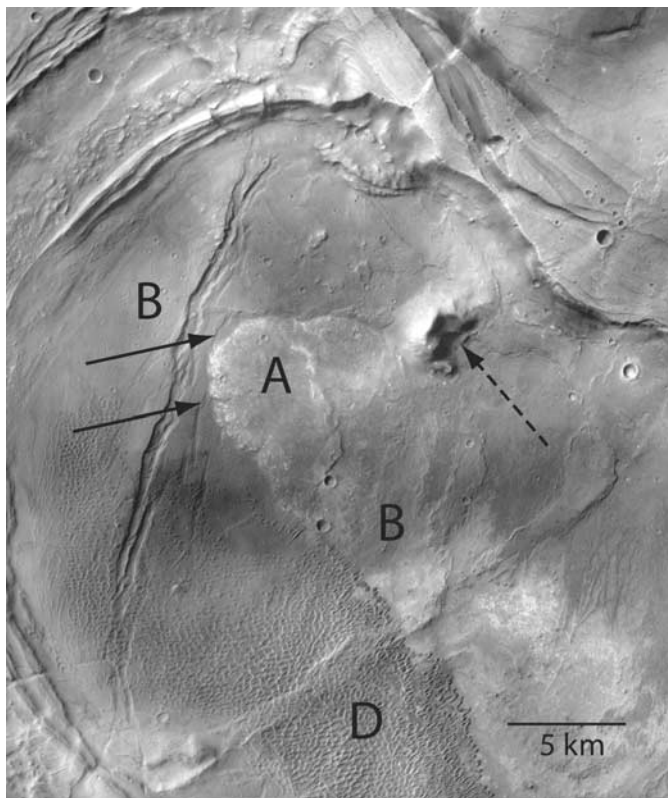
Figure 2 | **The floor of the Nili Patera caldera seen in a mosaic of THEMIS visual images.** Unit A is an ~80-m-thick lobate flow feature that is stratigraphically above older floor units, as evidenced by the superposition relationship of unit A on sections of the underlying N–S trending graben system (arrows) that cut unit B. Flow unit A appears to originate from a 300-m-high construct (dashed arrow) also identified in Fig. 1. Barchan dunes (D) cover the southern and southwestern caldera floor.
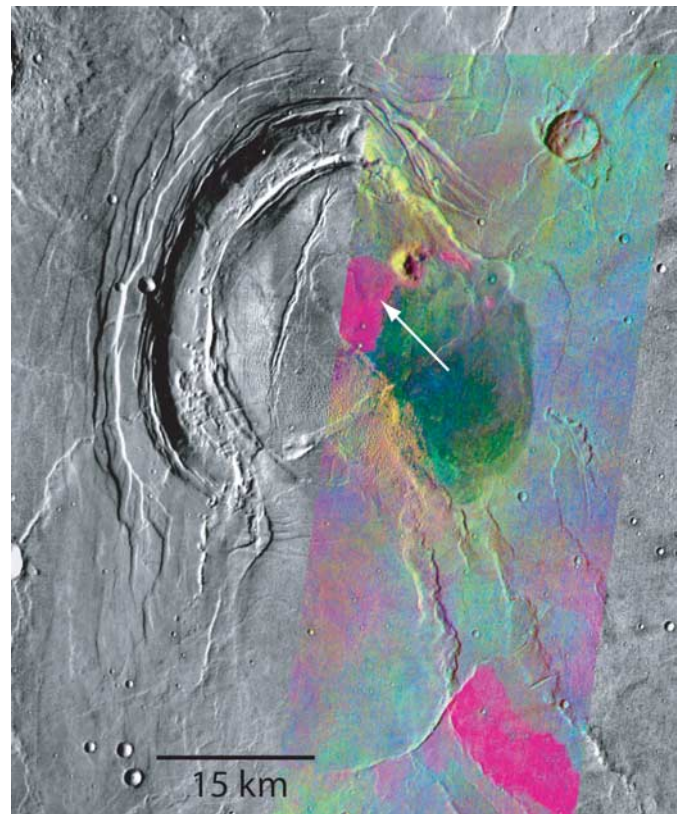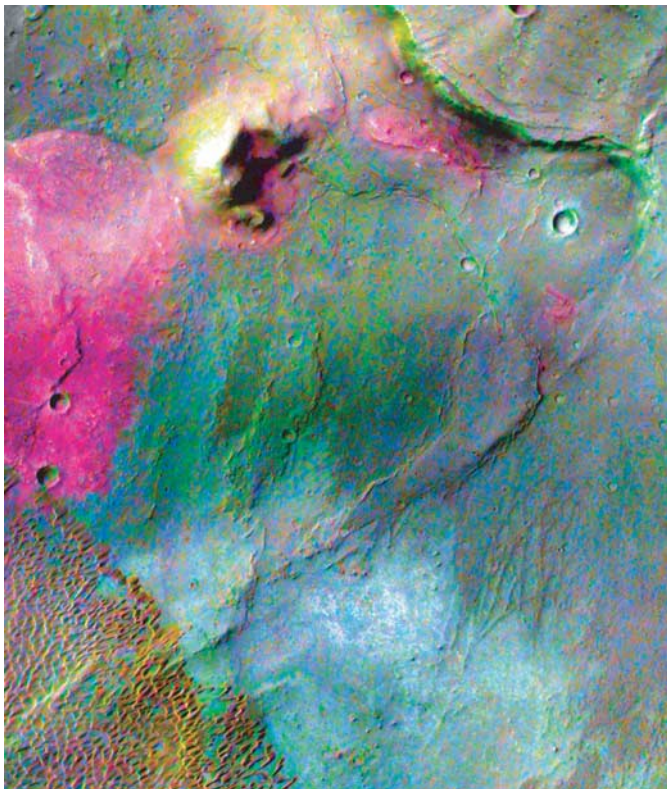

Figure 3 | **Compositional unit map of Nili Patera and the surrounding lava plains derived from a THEMIS multispectral IR image.** THEMIS bands 9, 7 and 5 were processed using a decorrelation stretch to enhance the spectral differences[49], and are shown superimposed of the daytime IR mosaic for geologic context. Unit A is bright magenta and occurs in a lobate flow unit on the caldera floor and several small outliers to the west. Flows spectrally similar to unit A also occur outside the caldera in units that appear to emanate from a positive-relief linear feature. These flows also differ significantly in composition from their surroundings.

505

and has a surface morphology that is distinct from the surrounding floor materials (Fig. 2). It surrounds a relatively steep-sided (~10°), ~300-m-high cone, and post-dates the unit-B floor materials based on its superposition across one of the grabens formed in unit B (Fig. 2). Units A and B have similarly relatively high crater densities, indicating similarly old ages. Night-time temperatures are uniform within units A and B, but differ significantly between them, with unit B consistently 8–9 K warmer than unit A (Fig. 1b), indicating a significant difference in the average particle size or rock abundance between these units[5,11].

The mineral compositions and abundances of units A and B were determined using a linear least-squares deconvolution[16,17] that has been applied extensively to the analysis of TES, Miniature TES (Mini-TES)[18,19] and laboratory spectra[1,7,8,20–23]. Unit B is composed of 35% plagioclase, 25% high-Ca clinopyroxene, 10% orthopyroxene, 10% sulphate, and minor glass, olivine, and carbonate components (Fig. 5a). Unit A is 30% plagioclase, 30% high-Si glass, and 15% clinopyroxene, with lesser sulphate and minor orthopyroxene and carbonate components (Fig. 5b).

Unit B is a basalt that is similar in composition to the typical basalt (surface type I) mapped using orbital TES data throughout much of the ancient cratered terrain[1,2]. Basalts of similar composition have been identified at the Opportunity landing site in the Meridiani plain[19,24].

The composition of unit A is much more silicic than the adjacent basalt, with high-Si glass, plagioclase and clinopyroxene. The spectral bands near 1,075 and 475 cm^{-1} are well matched by a SiO_2-K_2O glass (Fig. 5b). Illite clay and some zeolites, which can develop from the

aqueous alteration of high-Si glass, provide spectral alternatives to silica glass, although neither matches unit A as well as glass (Fig. 6), so we conclude that glass is a major component in unit A.

The derived SiO_2 abundance[21,25,26] of unit A is 60–63 wt%, as compared to values of 56 to 59 wt% for TES surface type II[26]. Given the probable contribution of the surrounding basaltic terrain to the deconvolved TES spectra, this derived SiO_2 abundance is a lower limit. Unit A thus has the highest silica abundance, and the most-evolved rock composition, of any volcanic unit yet found on Mars, with a composition corresponding to the low-silica end of the dacite composition field. A high-viscosity glassy dacite composition is consistent with unit A's occurrence as a relatively short, thick flow, its association with a steep-sided volcanic cone, and its distinct morphology.

The sulphate abundance derived for both units is slightly higher than the background abundance found for much of Mars[7]. The short-wavelength edge of the IR absorption at ~1,200 cm^{-1} is shifted to shorter wavelengths than typical for either type-I or -II material. This shift is characteristic of the Ca and Mg sulphates observed by Mini-TES at Meridiani Planum[19], and suggests that sulphates may
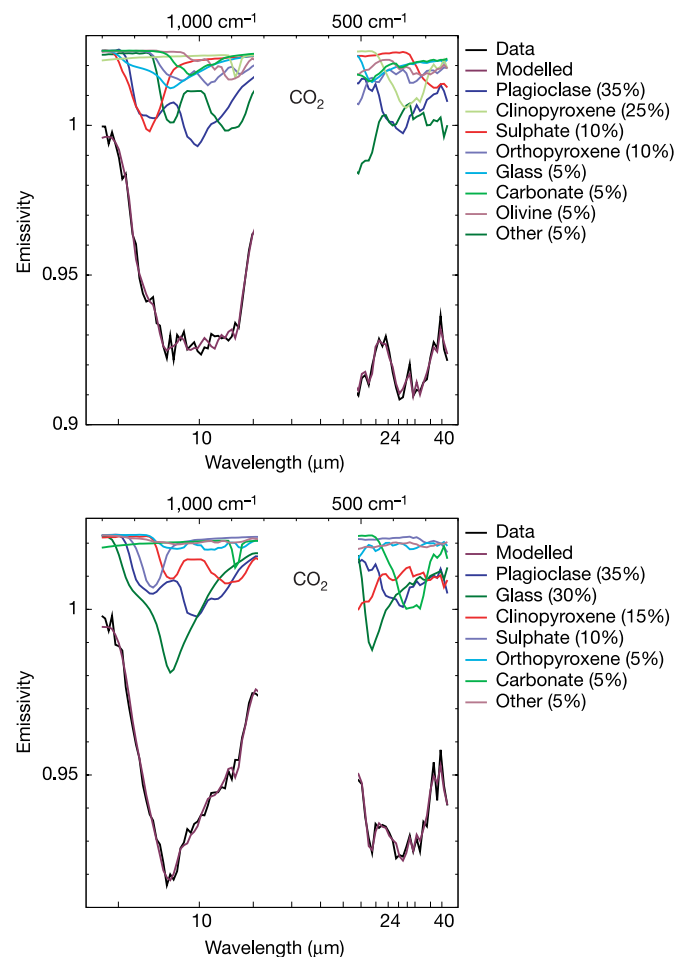

Figure 4 | **The compositional units of the Nili Patera floor.** A portion of the multispectral THEMIS IR image (Fig. 3) is superimposed on an 18 m per pixel THEMIS visual image. The colours are derived from the decorrelation stretch of IR bands 9, 7 and 5. This superposition illustrates the correlation between the composition and morphologic properties of the Nili Patera floor materials. The lobate flow (magenta; unit A) and associated cone are have a unique composition from the lavas that comprise most of the caldera floor (unit B; blue). Small outliers of unit-A material are visible east of the cone. Image width is 16 km.

Figure 5 | **Mineral composition and abundance of Nili Patera floor units derived from deconvolution of TES spectra. a**, Unit B has the mineral composition and abundance of basalt. Mineral abundances derived from the deconvolution model have been rounded to the nearest 5%. Minor components with <2.5% abundance have been grouped into the category 'Other'. The endmember spectra have been scaled on the basis of their derived abundance to indicate their relative contribution to the measured spectrum. Wavelengths from ~12 to 19 μm are blocked by absorption of atmospheric CO_2. **b**, The mineralogy of unit A, including a high abundance of Si-K glass (30%) and plagioclase (30%), indicates that this flow unit is dacite in composition.
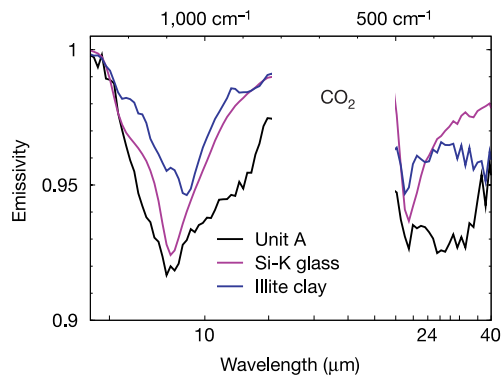
Figure 6 | Comparison of the spectra of Si-K glass and illite clay with the spectrum of unit A. Glass provides a better match than either clay or zeolites to the spectral bands from ∼900 to 1,050 cm^{-1} and 460 to 500 cm^{-1}.

be present within this region, possibly associated with volcanic emanations or fumaroles.

THEMIS eight-point spectra closely match the TES spectra and were deconvolved to provide a quantitative measure of the dacite abundance at 100-m scales (Fig. 7). Unit A is a coherent dacite unit that includes the cone, suggesting that the flow originated from the same source vent as the cone. The spectrally distinct flows to the south are also dacites that were probably erupted from a separate source (Fig. 7). Both the THEMIS and TES spectra suggest that the dunes are composed primarily of basalt, probably eroded from the upwind bedrock exposure. Along their western margin the dunes show subtle colour and spectral variations that suggest an influx of dacitic sands.

We interpret unit B to be basaltic lavas that covered the floor of the caldera following its collapse. Unit A is a later, high-silica, glassy volcanic unit associated with, and probably emanating from, a volcanic cone. Fractional crystallization of magma within a chamber commonly results in variations in the silica content of the erupted lavas over time. It appears that magmatic evolution has occurred within Syrtis Major, producing low-silica basalts followed by higher-silica glassy dacites in cones and flows. Orbital TES data have identified two surface components (TES surface type I and II; ref. 2) whose fundamental difference is silica content. The high-silica component in type-II material, which extends across much of the northern plains and throughout the southern highlands, has been attributed to primary volcanic glass in a basaltic andesite[2]. Alternative interpretations include high-silica, poorly crystalline materials or other weathering/alteration products[27–31], in part based on the apparent necessity for a wet mantle to produce large volumes of andesitic material[26]. The close proximity and similar geologic ages of units A and B argue against compositional differences based on differential weathering, whereas the geologic relationships support
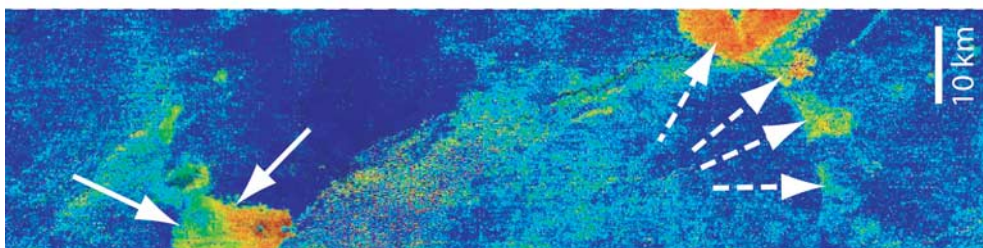
differences in primary volcanic composition. We therefore conclude that, at least locally, high-silica lavas have been produced on Mars, and could be an important component of the globally occurring TES type-II material.

Olivine-rich basalt

The occurrence of olivine is important because: (1) it indicates silica undersaturation (a critical parameter in recognizing primitive, mantle-derived magmas); and (2) it is readily weathered under common aqueous conditions and thus provides an indicator of environmental conditions in the time following rock emplacement.

Olivine-rich rocks have been identified from orbit at regional scales in several locations[23,32]. THEMIS data show that these occurrences are common at localized (hundreds of metres) scales[5,33–35]. An example of an olivine-rich unit occurs on the floor of a 60-km-diameter crater in Aurorae Planum immediately south of western Ganges Chasma[33,36] (Fig. 8). This unit is mapped as a distinct spectral unit using THEMIS, and positively identified as olivine-rich basalt using TES spectra. It is 10–12 K warmer than the surrounding terrain in night-time THEMIS images, with a thermal inertia >800 J m^{-2} s$^{-1/2}$ K^{-1}, indicating the presence of significant rocky material. This unit has been incised 15 to 20 m by channels (Fig. 8), providing additional evidence that it is a competent layer. The olivine basalt forms the uppermost layer in a crater that is superimposed on the top of the thick sequence of layered rocks that comprise Aurorae Planum. Its position, composition, and morphology suggest that it was emplaced as a lava flow at a late stage in plains formation.

A similar olivine basalt layer is exposed 4.5 km below the surface, 600 km to the east, in the walls of Ganges Chasma. This ∼100-m-thick unit crops out over an area 30 by 100 km in size[5]. The composition of this layer is basalt with >∼15% olivine with an Mg/(Mg + Fe) ratio of 0.60 to 0.70 (Fo60-70)[5]; the unit exposed below it on the floor of Ganges Chasma is determined from TES data to be basaltic.

The similarity in the derived composition of the Ganges and Aurorae Planum units suggests that they formed by similar processes, most probably eruption as lava flows. The occurrence of the Ganges unit at a significantly lower stratigraphic level than the Aurorae Planum unit suggests that olivine-rich basalts were erupted episodically at significantly different times throughout the formation of the martian plains.

Picritic basalts have also been observed *in situ* by the Spirit rover in the Gusev crater using IR, Mossbauer, and alpha particle X-ray spectra[18,37–39]. Similar examples of olivine-rich basalts have been identified using both TES data and THEMIS data in Ares Valles[35], Nili Fossae[34], and elsewhere[33,36]. This growing inventory of olivine basalts suggests that mafic basalts may be a relatively common variant of martian basalts. They probably represent less-evolved lavas or contain higher abundances of cumulates or mantle xenoliths than typical basalts.

Figure 7 | Dacite abundance derived from the deconvolution of a THEMIS eight-band multi-spectra IR image of Nili Patera. Red colour represents 80% dacite abundance; blue is 0% abundance. Green and yellow colours also represent high (>50%) dacite abundance, but with the shallower spectral contrast that is probably due to smaller particle size. In addition to the flow and cone mapped as unit A (arrows), several large dacite flows occur to the south (dashed arrows). North is towards '9 o'clock'.
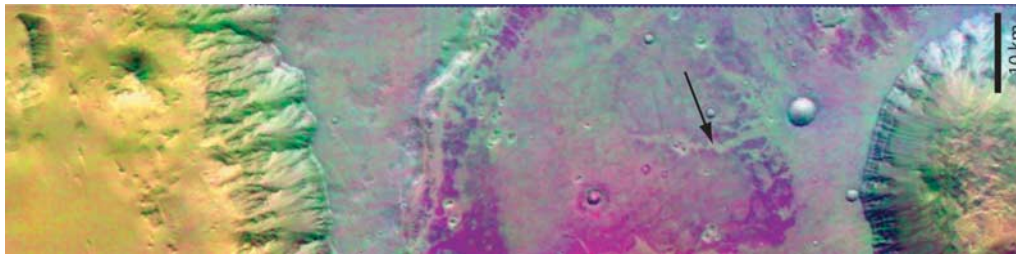
Figure 8 | THEMIS compositional map of a portion of a crater in Aurora Planum. The purple regions are outcrops of olivine-rich basalt. This basalt unit has been incised by channels (arrow), indicating that it is a competent rock layer. THEMIS image I01513001 centred near 308° E; 8° S. Image width is 32 km; north is towards '9 o'clock'.
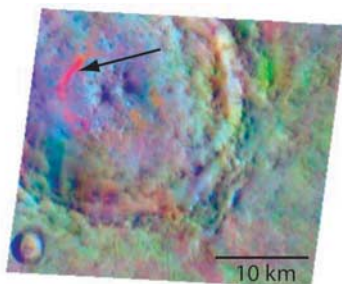
Quartz-bearing granitoid rocks

Quartz has not been recognized on Mars at regional or global scales. Its presence is important for recognizing igneous rocks that have experienced extreme fractionation. Unlike olivine, quartz persists during sedimentary weathering and transport and, if ever formed, should persist in primary or sedimentary rock units.

THEMIS observations have revealed the presence of quartz-bearing rocks exposed in crater central uplifts in two 30-km-diameter craters on the northern flank of the Syrtis Major volcanic construct[40] (Fig. 9). The corresponding TES spectra match laboratory spectra of granitoid rocks composed primarily of quartz and plagioclase[40]. The association of these outcrops with central peaks suggests that these rocks were uplifted from depths of several kilometres by the impact process[40]. The two craters in which these rocks are found are 95 km apart, but no rocks of similar composition have been found on the planet, suggesting that they are part of an unusual regional rock unit that is postulated to be a single granitoid pluton[40].

All of the remote, *in situ*, and martian meteorite analyses to date indicate that Mars is dominated by ultramafic to possibly intermediate volcanic compositions[1,2,4,7,18,19,32,37,41,42]. While fractional crystallization of magmas can produce highly differentiated, high-silica melts[43], the volumes required to produce the Syrtis Major pluton would produce enormous quantities of mafic cumulates[40], which are not observed in any central peaks. A more likely process is partial melting of thick sequences of pre-existing basaltic rocks at relatively low pressures to form high-silica trondhjemites and tonalites[40,43,44].

Implications for crustal formation and evolution

Multi-spectral IR imaging at 100-m scales has produced petrologic maps that are, for the first time, at scales appropriate for studying local geologic processes. At these scales we have identified rock units in basaltic terrains that have olivine abundances of >20%, which on Earth would be classified as picrites. These rocks have been identified in eroded canyon walls and crater floors of ancient terrains in several areas, suggesting that they may be relatively common throughout the ancient martian crust but remain unexposed in most locations. Local eruptive sequences are observed in the Syrtis Major volcano that appear to have evolved in composition from basaltic to dacitic, demonstrating that fractional crystallization occurs in martian magma chambers. This diversity in composition is the result of igneous processes analogous to those found in large volcanic complexes on Earth. Rare occurrences of exposed quartz-bearing granitoid rocks have been discovered on Mars, suggesting the formation of highly evolved magmas. Taken together, these observations show that the martian crust, while dominated by basalt, has produced a diversity of igneous materials whose range in composition from ultramafic to highly silicic rivals that found on the Earth.

METHODS

The spectral deconvolution of TES spectra was done using averages of 5–10 atmospherically corrected TES spectra that were isolated from each unit[1,8,45,46]. The TES has collected over 200 million IR spectra of the martian surface with 3×8 km spatial resolution[7]. These spectra have been used extensively to map global and regional lithologic units[1,2,7,8,30,32,36,40,47]. Each spectrum collected from units A and B was analysed using the same 36-component endmember library used for TES global mapping[8]. This library was constructed from laboratory minerals and TES surface dust spectra[48], and contained representative suites of pyroxene, feldspar, olivine, sulphate, carbonate, and oxide minerals. Spectral effects due to small particle sizes ($<\sim 60\,\mu m$; ref. 17) can be ignored owing to the coarse-grained materials that are shown to comprise both units A and B, based on the night-time temperature and derived thermal inertia. Uncertainties for the mineral abundances derived from the deconvolution are estimated to be ±5%, based on laboratory analyses and *in situ* validation of spectra acquired by the Mini-TES on the MER rovers[1,8,18–20]. Recent *in situ* measurements from the Mössbauer[38], Mini-TES[18], and Alpha Particle X-ray spectrometers[39] on the Mars Exploration Rovers have determined mineral compositions that have verified the TES determinations from orbit[18,37], providing higher confidence in interpretations made elsewhere on the planet.

Deconvolution of the THEMIS spectra was done in a similar manner to the TES analysis, but used the atmospherically corrected TES spectra of units A (dacite) and B (basalt). These spectra were convolved to the THEMIS spectral bandpasses, and used as endmembers in the deconvolution of the THEMIS multi-spectral IR image. This method provides a quantitative measure of abundance of dacite and basalt. The dacite abundance derived using this method is shown in Fig. 7.

Figure 9 | The occurrence of quartz-bearing granitoid rocks mapped using THEMIS multi-spectral IR imaging. This compositional unit map of a 30-km-diameter unnamed crater in Syrtis Major shows the occurrence of quartz-bearing granitoid rocks (red colour; arrow) within the central peak[40]. THEMIS bands 6, 5 and 4 were processed using a decorrelation stretch to enhance the spectral differences[49].

1. Christensen, P. R., Bandfield, J. L., Smith, M. D., Hamilton, V. E. & Clark, R. N. Identification of a basaltic component on the Martian surface from Thermal Emission Spectrometer data. *J. Geophys. Res.* **105**, 9609–9622 (2000).
2. Bandfield, J. L., Hamilton, V. E. & Christensen, P. R. A global view of Martian volcanic compositions. *Science* **287**, 1626–1630 (2000).
3. Soderblom, L. A. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 557–593 (Univ. Arizona Press, Tucson, 1992).
4. Mustard, J. F. & Sunshine, J. M. Seeing through the dust: Martian crustal heterogeneity and links to the SNC meteorites. *Science* **267**, 1623–1626 (1995).
5. Christensen, P. R. *et al.* Morphology and composition of the surface of Mars: Mars Odyssey THEMIS results. *Science* **300**, 2056–2061 (2003).

6. Christensen, P. R. *et al.* Detection of crystalline hematite mineralization on Mars by the Thermal Emission Spectrometer: Evidence for near-surface water. *J. Geophys. Res.* **105,** 9623–9642 (2000).

7. Christensen, P. R. *et al.* The Mars Global Surveyor Thermal Emission Spectrometer experiment: Investigation description and surface science results. *J. Geophys. Res.* **106,** 23823–23871 (2001).

8. Bandfield, J. L. Global mineral distributions on Mars. *J. Geophys. Res.* **107,** doi:10.1029/2001JE001510 (2002).

9. Schaber, G. G. Syrtis Major: A low-relief volcanic shield. *J. Geophys. Res.* **87,** 9852–9866 (1982).

10. Smith, D. E. *et al.* Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.* **106,** 23689–23722 (2001).

11. Kieffer, H. H. *et al.* Thermal and albedo mapping of Mars during the Viking primary mission. *J. Geophys. Res.* **82,** 4249–4292 (1977).

12. Mellon, M. T., Jakosky, B. M., Kieffer, H. H. & Christensen, P. R. High resolution thermal inertia mapping from the Mars Global Surveyor Thermal Emission Spectrometer. *Icarus* **148,** 437–455 (2000).

13. Presley, M. A. & Christensen, P. R. Thermal conductivity measurements of particulate materials, Part II: Results. *J. Geophys. Res.* **102,** 6551–6566 (1997).

14. Ruff, S. W. & Christensen, P. R. Bright and dark regions on Mars: Particle size and mineralogical characteristics based on Thermal Emission Spectrometer data. *J. Geophys. Res.* **107,** doi:10.1029/2001JE001580 (2002).

15. Ruff, S. W. & Hamilton, V. E. Mineralogical anomalies in Mars Nili Patera caldera observed with Thermal Emission Spectrometer data. *Lunar Planet. Sc.* **XXXII,** abstr. 2186 [CD-ROM] (2001).

16. Adams, J. B., Smith, M. O. & Johnson, P. E. Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site. *J. Geophys. Res.* **91,** 8098–8112 (1986).

17. Ramsey, M. S. & Christensen, P. R. Mineral abundance determination: Quantitative deconvolution of thermal emission spectra. *J. Geophys. Res.* **103,** 577–596 (1998).

18. Christensen, P. R. *et al.* Initial results from the Miniature Thermal Emission Spectrometer experiment at the Spirit landing site at Gusev crater. *Science* **305,** 837–842 (2004).

19. Christensen, P. R. *et al.* Mineralogy at Meridiani Planum from the Mini-TES experiment on the Opportunity rover. *Science* **306,** 1733–1739 (2004).

20. Feely, K. C. & Christensen, P. R. Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks. *J. Geophys. Res.* **104,** 24195–24210 (1999).

21. Wyatt, M. B., Hamilton, V. E., McSween, H. Y. Jr, Christensen, P. R. & Taylor, L. A. Analysis of terrestrial and martian volcanic compositions using thermal emission spectroscopy: I. Determination of mineralogy, chemistry, and classification strategies. *J. Geophys. Res.* **106,** 14711–14732 (2001).

22. Hamilton, V. E., Wyatt, M. B., McSween, H. Y. Jr & Christensen, P. R. Analysis of terrestrial and martian volcanic compositions using thermal emission spectroscopy: II. Application to martian surface spectra from MGS TES *J. Geophys. Res.* **106,** 14733–14747 (2001).

23. Hamilton, V. E., Christensen, P. R., McSween, H. Y. Jr & Bandfield, J. L. Searching for the source regions of martian meteorites using MGS TES: Integrating martian meteorites into the global distribution of igneous materials on Mars. *Meteorit. Planet. Sci.* **38,** 871–885 (2003).

24. Rieder, R. *et al.* Chemistry of rocks and soils at Meridiani Planum from the Alpha Particle X-ray Spectrometer. *Science* **306,** 1746–1749 (2004).

25. Hamilton, V. E. & Christensen, P. R. Determining the modal mineralogy of mafic and ultramafic igneous rocks using thermal emission spectroscopy. *J. Geophys. Res.* **105,** 9717–9734 (2000).

26. McSween, H. Y. Jr, Grove, T. L. & Wyatt, W. B. Constraints on the composition and petrogenesis of the martian crust. *J. Geophys. Res.* **108,** doi:10.1029/2003JE002175 (2003).

27. Wyatt, M. B. & McSween, H. Y. Jr Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars. *Nature* **417,** 263–266 (2002).

28. Wyatt, M. B., McSween, H. Y. Jr, Tanaka, K. L. & Head, J. W. III Global geologic context for rock types and surface alteration on Mars. *Geology* **32,** 645–648 (2004).

29. Kraft, M. D., Michalski, J. R. & Sharp, T. G. High-silica rock coatings: TES surface-type 2 and weathering in Mars. *Lunar Planet. Sci.* **XXXV,** abstr. 1936 [CD-ROM] (2004).

30. Ruff, S. W. Spectral evidence for zeolite in the dust on Mars. *Icarus* **168,** 131–143 (2004).

31. Michalski, J. R., Kraft, M. D., Sharp, T. G., Williams, L. B. & Christensen, P. R. Mineralogical constraints on the high-silica Martian surface component observed by TES: Clay-rich mineralogy does not sufficiently explain the Acidalia Planitia-type spectra. *Icarus* (in the press).

32. Hoefen, T. *et al.* Discovery of olivine in the Nili Fossae region of Mars. *Science* **302,** 627–630 (2003).

33. Hamilton, V. E. & Christensen, P. R. Detailed mineralogical analyses of Martian meteorite-like terrains using MGS TES and Odyssey THEMIS data. *Lunar Planet. Sci.* **XXXIV,** abstr. 1982 [CD-ROM] (2003).

34. Hamilton, V. E. & Christensen, P. R. Evidence for extensive olivine-rich bedrock on Mars. *Geology* (in the press).

35. Rogers, A. D., Christensen, P. R. & Bandfield, J. L. Compositional heterogeneity of the ancient martian crust: Analysis of Ares Vallis bedrock the THEMIS and TES data. *J. Geophys. Res.* (submitted).

36. Hamilton, V. E., Christensen, P. R. & McSween, H. Y. Jr Global constraints on the source regions of martian meteorites from MGS TES data. *Meteorit. Planet. Sci.* **37,** 59 (2003).

37. McSween, H. Y. Jr *et al.* Basaltic rocks analyzed by the Spirit rover in Gusev crater. *Science* **305,** 842–845 (2004).

38. Morris, R. V. *et al.* Mineralogy at Gusev crater from the Mössbauer spectrometer on the Spirit rover. *Science* **305,** 833–836 (2004).

39. Gellert, R. *et al.* Chemistry of rocks and soils at Gusev crater from the Alpha Particle X-ray Spectrometer. *Science* **305,** 829–832 (2004).

40. Bandfield, J. L., Hamilton, V. E., Christensen, P. R. & McSween, H. Y. Jr Identification of quartzofeldspathic materials on Mars. *J. Geophys. Res.* **109,** doi:10.1029/2004JE002290 (2004).

41. McSween, H. Y. Jr What have we learned about Mars from SNC meteorites. *Meteoritics* **29,** 757–779 (1994).

42. McSween, H. Y. Jr The rocks of Mars, from far and near. *Meteorit. Planet. Sci.* **37,** 7–25 (2002).

43. Clark, D. B. *Granitoid Rocks* (Chapman and Hall, New York, 1992).

44. Barker, F. in *Trondhjemites, Dacites, and Related Rocks* (ed. Barker, F.) 1–12 (Elsevier, New York, 1979).

45. Smith, M. D., Bandfield, J. L. & Christensen, P. R. Separation of atmospheric and surface spectral features in Mars Global Surveyor Thermal Emission Spectrometer (TES) spectra: Models and atmospheric properties. *J. Geophys. Res.* **105,** 9589–9608 (2000).

46. Bandfield, J. L. & Smith, M. D. Multiple emission angle surface-atmosphere separations of Thermal Emission Spectrometer data. *Icarus* **161,** 47–65 (2003).

47. Christensen, P. R., Morris, R. V., Lane, M. D., Bandfield, J. L. & Malin, M. C. Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. *J. Geophys. Res.* **106,** 23873–23885 (2001).

48. Christensen, P. R. *et al.* A thermal emission spectral library of rock forming minerals. *J. Geophys. Res.* **105,** 9735–9738 (2000).

49. Gillespie, A. R. Spectral mixture analysis of multispectral thermal infrared images. *Remote Sens. Environ.* **42,** 137–145 (1992).

ARTICLES

Systematic analysis of genes required for synapse structure and function

Derek Sieburth[1]*, QueeLim Ch'ng[1]*, Michael Dybbs[1], Masoud Tavazoie[1]†, Scott Kennedy[2], Duo Wang[1], Denis Dupuy[3], Jean-François Rual[3], David E. Hill[3], Marc Vidal[3,4], Gary Ruvkun[1,4] & Joshua M. Kaplan[1,4]

Chemical synapses are complex structures that mediate rapid intercellular signalling in the nervous system. Proteomic studies suggest that several hundred proteins will be found at synaptic specializations. Here we describe a systematic screen to identify genes required for the function or development of *Caenorhabditis elegans* neuromuscular junctions. A total of 185 genes were identified in an RNA interference screen for decreased acetylcholine secretion; 132 of these genes had not previously been implicated in synaptic transmission. Functional profiles for these genes were determined by comparing secretion defects observed after RNA interference under a variety of conditions. Hierarchical clustering identified groups of functionally related genes, including those involved in the synaptic vesicle cycle, neuropeptide signalling and responsiveness to phorbol esters. Twenty-four genes encoded proteins that were localized to presynaptic specializations. Loss-of-function mutations in 12 genes caused defects in presynaptic structure.

Large-scale double-stranded (ds)RNA interference (RNAi) screens in *C. elegans* are possible because expression of endogenous genes can be specifically knocked down by feeding nematodes bacteria expressing dsRNAs from the corresponding genes[1]. Systematic genome-wide RNAi screens using this technique have been described for several phenotypes[2,3]; however, RNAi has had limited success in the study of neuronal genes due to the insensitivity of *C. elegans* neurons to RNAi[2]. Recently, mutations were described that enhance the effectiveness of RNAi[4,5]. Here we use an enhanced RNAi strain[55] to identify genes required for synaptic transmission.

RNAi screen for aldicarb resistance

Steady-state acetylcholine (ACh) secretion in living nematodes can be assayed by measuring resistance to the acetylcholine esterase inhibitor aldicarb. Aldicarb treatment causes accumulation of ACh at neuromuscular junctions (NMJs), leading to acute paralysis and subsequent death. Mutations that decrease ACh secretion confer resistance to aldicarb-induced paralysis and lethality. Screens for aldicarb resistance have identified many genes required for synaptic transmission[6,7]. We reasoned that new genes could be identified in a systematic RNAi screen for aldicarb resistance. A total of 2,072 genes (approximately 10% of the genome) were selected for this screen based on their availability in the RNAi feeding library[2], and on the presence of domains that predict involvement in signal transduction, synaptic localization, cytoskeletal regulation or membrane trafficking (Supplementary Fig. S1 and Supplementary Table S1).

To maximize the number of genes identified, we simultaneously screened for aldicarb resistance in two genotypes: the *eri-1; lin-15B* double mutant (details and lines available on request) and an *eri-1; dgk-1 lin-15B* triple mutant, hereafter referred to as the Ric (resistant to inhibitors of cholinesterase) and Dgk screens, respectively. Mutants lacking the DGK-1 diacylglycerol (DAG) kinase are hypersensitive to aldicarb, suggestive of increased ACh secretion[8]. We

reasoned that *dgk-1* DAG kinase mutants might be more sensitive to modest reductions of ACh secretion produced by RNAi treatments. To identify genes that would otherwise be essential for viability, RNAi treatment was initiated after embryonic and early larval development was completed, and aldicarb resistance of the resulting adults was assayed (Fig. 1a).

On the basis of the false positive rates for empty vector controls (10.9% and 3.2% in Ric and Dgk screens, respectively), we identified 185 genes (8.9% of genes screened) for which inactivation caused significantly more aldicarb resistance than controls ($P \leq 0.0001$) (Fig. 1b; see also Supplementary Table S2). Of these, 132 genes had not been previously implicated in synaptic transmission in *C. elegans*, *Drosophila*, or mammals (Supplementary Table S2). Eighty-eight per cent (163) of the positive genes have mouse or human homologues and 63% (117) have putative mouse or human orthologues (as defined by best reciprocal BLAST score) (Supplementary Table S3).

To validate our screens, we compared aldicarb resistance caused by RNAi and loss-of-function alleles. Of 21 genes previously shown to mutate to an aldicarb resistance phenotype, 19 (90%) were aldicarb resistant by RNAi (Fig. 1c; see also Supplementary Table S4). Loss-of-function alleles were available for 44 positive genes, excluding positive controls. We analysed the aldicarb responsiveness for all homozygous viable alleles ($n = 34$) and found that 28 (82%) were aldicarb resistant (Supplementary Fig. 2 and Supplementary Table S5). Thus, these screens effectively identified many genes required for aldicarb responsiveness.

Expression patterns for 72 of the 185 positive genes have been annotated in the WormBase database (http://www.wormbase.org). In addition, we analysed the expression of 16 positive genes using transcriptional reporter constructs (detailed below; see also Supplementary Fig. S4 and Supplementary Table S7). Of these 88 genes, 78% were expressed in neurons and 38% were expressed in muscles (Supplementary Table S2).

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Department of Pharmacology, University of Wisconsin, Madison, Wisconsin 53706, USA. [3]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present address: College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA.
*These authors contributed equally to this work.

Functional profiling of aldicarb resistance genes

Hierarchical clustering of phenotypic data has been used to predict groups of genes with related functions[9,10]. We reasoned that genes that function in similar pathways or processes should produce similar patterns of aldicarb resistance after RNAi across a variety of conditions. We used a panel of mutants and drug treatments that increase ACh release and tested whether RNAi of each gene could produce aldicarb resistance (Fig. 2). In addition to *dgk-1* (which encodes the protein DAG kinase; hereafter, homologous mammalian proteins are given in parentheses after each occurrence of the corresponding *C. elegans* gene or protein) mutants, we used mutations in three other genes that stimulate ACh release: *goa-1* (G$_{o\alpha}$), *eat-16* (regulator of G protein signalling (RGS)) and *unc-43* (CaMKII)[11]. We also treated animals with phorbol esters (for example, PMA), a DAG analogue that enhances neurotransmitter secretion in a variety of systems, including the *C. elegans* NMJ[12]. Finally, we mimicked the direct activation of ACh receptors using the nicotinic agonist levamisole.

Functional profiling by RNAi differs from traditional double mutant analysis in one important respect: RNAi typically does not eliminate gene activity whereas null mutations are generally used in double mutant studies. To maximize the severity of the secretion defects produced, for this analysis we used two-generation RNAi treatments (see Methods for details). Consequently, profiling was restricted to 60 genes for which two-generation RNAi treatments caused significant aldicarb resistance ($P < 0.05$) but did not cause lethality (Fig. 2; see also Supplementary Table S2). Even under these conditions, RNAi is unlikely to produce null phenotypes. For example, *unc-13* and *ccb-1* mutate to lethality, yet two-generation RNAi did not cause lethality. For each gene, we report whether two-generation RNAi produced significant aldicarb resistance ($P < 0.05$) for each condition (Fig. 2). Unlike traditional epistasis analysis, a positive response in a functional profile does not imply that RNAi treatment completely abolished the stimulatory effects of the drug or mutation on ACh signalling. Using available loss-of-function alleles, we confirmed the functional profiles predicted by RNAi for three conditions (Fig. 2; see also Supplementary Table S5). Corresponding patterns of aldicarb resistance were observed for 85% ($n = 55$) of

genes predicted to be required for aldicarb responsiveness, 80% ($n = 10$) of genes predicted to confer aldicarb resistance in *dgk-1* double mutants, and 83% ($n = 12$) of genes predicted to confer resistance to phorbol esters. These results suggest that functional profiles obtained by RNAi provide a useful approximation of the results that would be seen with double mutants.

The patterns of drug resistance observed for each gene were organized by hierarchical clustering. We identified seven clusters (A–G, Fig. 2), which were detected using multiple clustering strategies (see Methods for details). The arrangement of clusters relative to each other in the dendrogram varied between clustering algorithms; therefore, the relative positioning of clusters within the tree is less meaningful than the membership in individual clusters. These clusters identify groups of genes that appear to have related functions, as detailed below.

Synaptic vesicle cycle. Nine genes analysed have pivotal roles in the synaptic vesicle cycle (Fig. 2, red boxes). Six of these core synaptic components were present in clusters B and C, including *unc-17* (vesicular ACh transporter (VAChT)), *snb-1* (synaptobrevin), *ric-4* (SNAP25), *rab-3*, *unc-10* (RIM1) and *unc-18*. The remaining four genes in clusters B and C had no previously described function in synaptic transmission: D2030.9 (WD40 repeat protein), C31H2.1 (RabGAP), F46G11.3 (cyclin-G-associated kinase (GAK)) and C34C6.5 (sphingosine kinase). We propose that the proteins encoded by these genes are involved in synaptic vesicle exocytosis or endocytosis. Consistent with this idea, F46G11.3 (GAK) is targeted to presynaptic nerve terminals (detailed below) and the mammalian GAK protein co-purifies with clathrin-coated vesicles and stimulates uncoating of vesicles *in vitro*, suggesting a possible function in endocytosis[13].

Neuropeptide signalling. Eleven positive genes are involved in neuropeptide signalling (Supplementary Table S2), indicating an important role for neuropeptides in regulating ACh release at the NMJ. The *egl-3* gene encodes a proprotein convertase most similar to vertebrate PC2, and *egl-21* encodes carboxypeptidase E (CPE), both of which mediate processing of proneuropeptides[14,15]. We previously showed that *egl-3* (PC2) and *egl-21* (CPE) mutations cause aldicarb resistance[15]; however, the neuropeptides and receptors involved
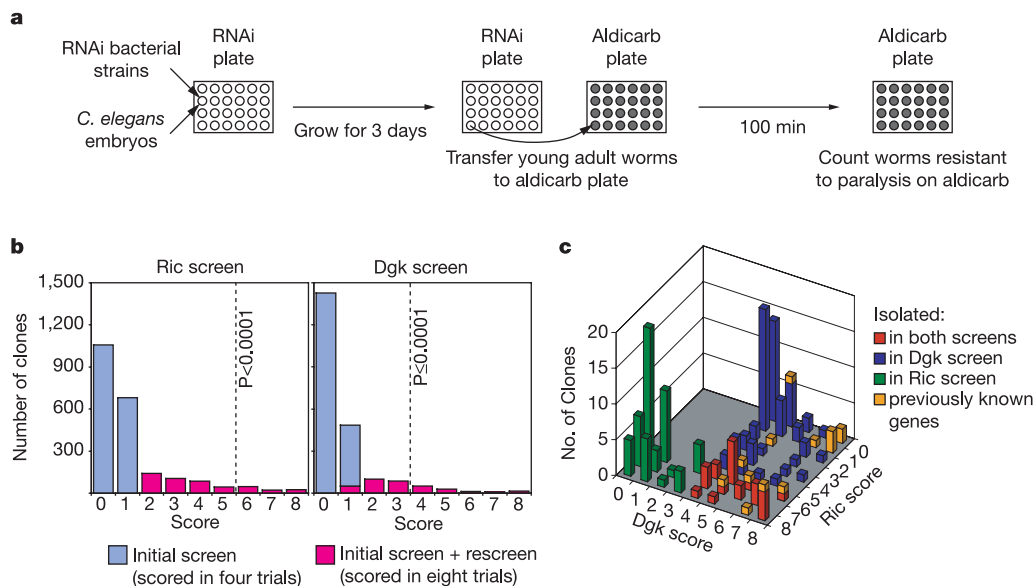
Figure 1 | Summary of RNAi screens. a, *eri-1; lin-15B* (Ric screen) and *eri-1; dgk-1 lin-15B* embryos (Dgk screen) were fed bacteria expressing dsRNA. After 3 days, young adults were scored for resistance to aldicarb. **b**, Histogram of clone scores from the Ric (left) and the Dgk screens (right). After one round of quadruplicate screening, genes scoring positive in one or more replicates underwent a second round of quadruplicate screening. On the basis of the false positive rate (that is, empty vector controls), clones scoring greater than five out of eight from the Ric screen and greater than three out of eight from the Dgk screen scored above threshold and were considered positive clones ($P < 0.0001$). **c**, Positive clones from each screen (green and blue), overlap between the two screens (red), and scores of the previously known Ric genes (orange) are shown. See also Supplementary Table S4.

in this effect were unknown. Two insulin-like growth factors (IGFs; *ins-22* and *ins-31*), two FMRFamide related peptides (FaRPs) (*flp-1* and *nlp-12*), and a predicted neuropeptide receptor (*fshr-1* (FSHR)) were identified in our screen. A downstream target of insulin and mTOR, *daf-15* (Raptor[16]), was also identified. These results suggest that insulin regulation of *C. elegans* diapause, longevity and fat storage[17] might be attributable to changes in neuronal signalling. The *unc-31* (CAPS) gene was also identified. CAPS is proposed to promote secretion of dense core vesicles (DCVs)[18].

Four neuropeptide signalling genes were present in clusters E (*unc-31* (CAPS) and *sbt-1* (7B2)) and F (*nlp-12* (FaRP) and *egl-3* (PC2)) (Fig. 2). The rat 7B2 protein is a chaperonin required for the production of active PC2 (ref. 19). The *nlp-12* gene encodes three FaRPs that contain predicted dibasic cleavage sites[20]. Thus, PC2, a PC2 chaperone and a predicted PC2 substrate, had very similar functional profiles. Several other predicted neuropeptide signalling genes were not found in these clusters, including *egl-21* (CPE). We previously showed that *egl-3*; *egl-21* double mutants have a more

severe phenotype than either single mutant, suggesting that these enzymes process distinct classes of proneuropeptides[15]. Thus, it is not surprising that these two genes had distinct functional profiles.

Phorbol ester response genes. Endogenous DAG or exogenous phorbol esters stimulate acetylcholine release in part by recruiting UNC-13 to synapses where it is thought to promote priming of synaptic vesicles[8,12,21]. Twenty-three positive genes conferred partial resistance to phorbol-ester-stimulated ACh release after RNAi treatment (Fig. 2). One prominent class of phorbol response genes was a subset of the neuropeptide signalling genes, including *egl-3* (PC2), *sbt-1* (7B2) and *unc-31* (CAPS), all confirmed by analysis of loss-of-function alleles (Supplementary Fig. S2). One attractive model to explain these results is that phorbol esters stimulate secretion of neuropeptides, which in turn stimulate ACh secretion at NMJs. Consistent with this idea, phorbol esters are often found to promote secretion of DCVs in several neurosecretory cell types[22].

Other genes required for the phorbol response included *pkc-1* (protein kinase Cη/ε), *sos-1* (RasGEF), *let-60* (Ras) and *osm-9*
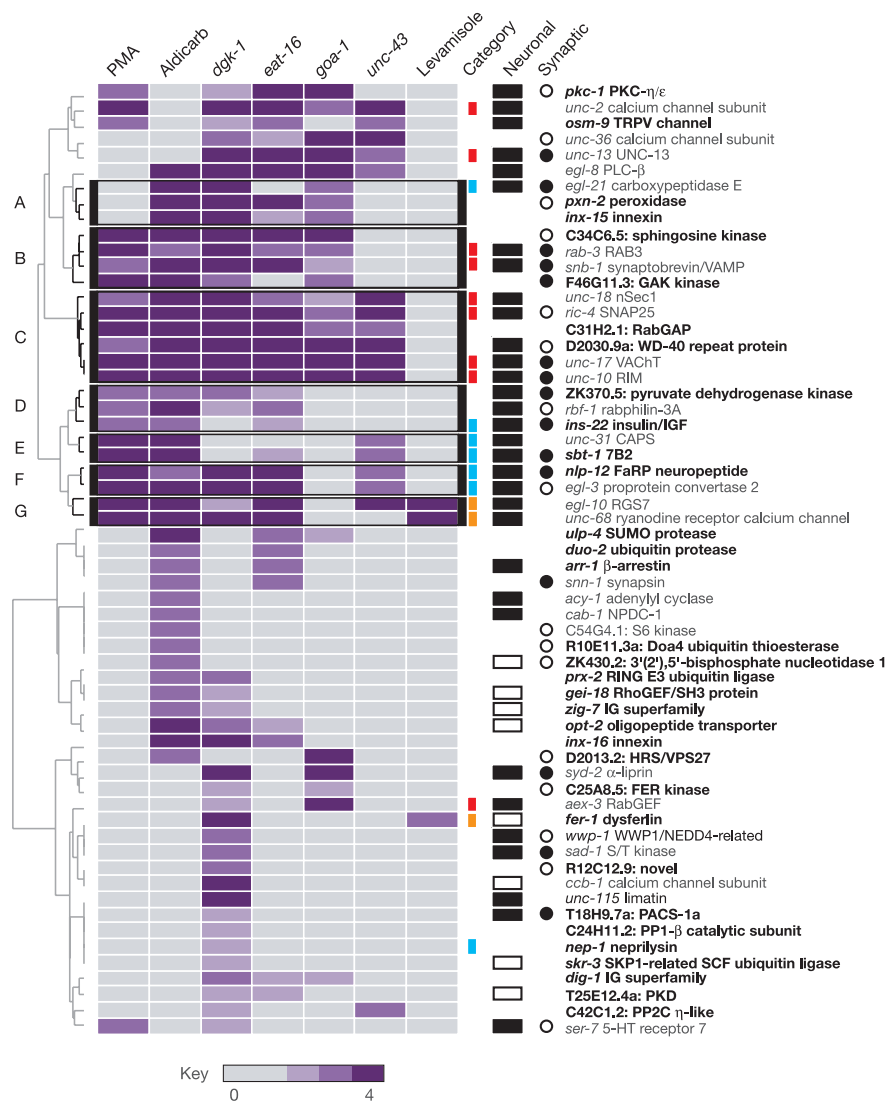
Figure 2 | Functional profiling of 60 aldicarb-resistance genes. Columns represent different genetic backgrounds or drug treatments. Rows represent knockdown of different genes (all scored in quadruplicate). Shading indicates the number of positive trials. Scores not meeting the threshold for significance ($P < 0.05$) were set to zero. Robust gene clusters (A–G), which remained together under multiple clustering algorithms, are indicated by black boxes and black trees in the dendrogram on the left. Annotated functions for genes (category column) are indicated as follows: core exocytosis machinery, red; neuropeptide signalling, blue; muscle, orange.

For the neuronal column, filled squares indicate genes expressed in neurons; open squares indicate genes not expressed in neurons (see Supplementary Tables S2 and S7 and Supplementary Fig. S4 for expression patterns). Blanks indicate that gene expression is unknown. For the synaptic column, filled circles indicate presynaptic localization; open squares indicate non-synaptic protein (see Fig. 3, Table 1, Supplementary Fig. S3 and Supplementary Table S6 for protein localization details). Novel genes are highlighted in bold; previously known synaptic genes are indicated in grey.
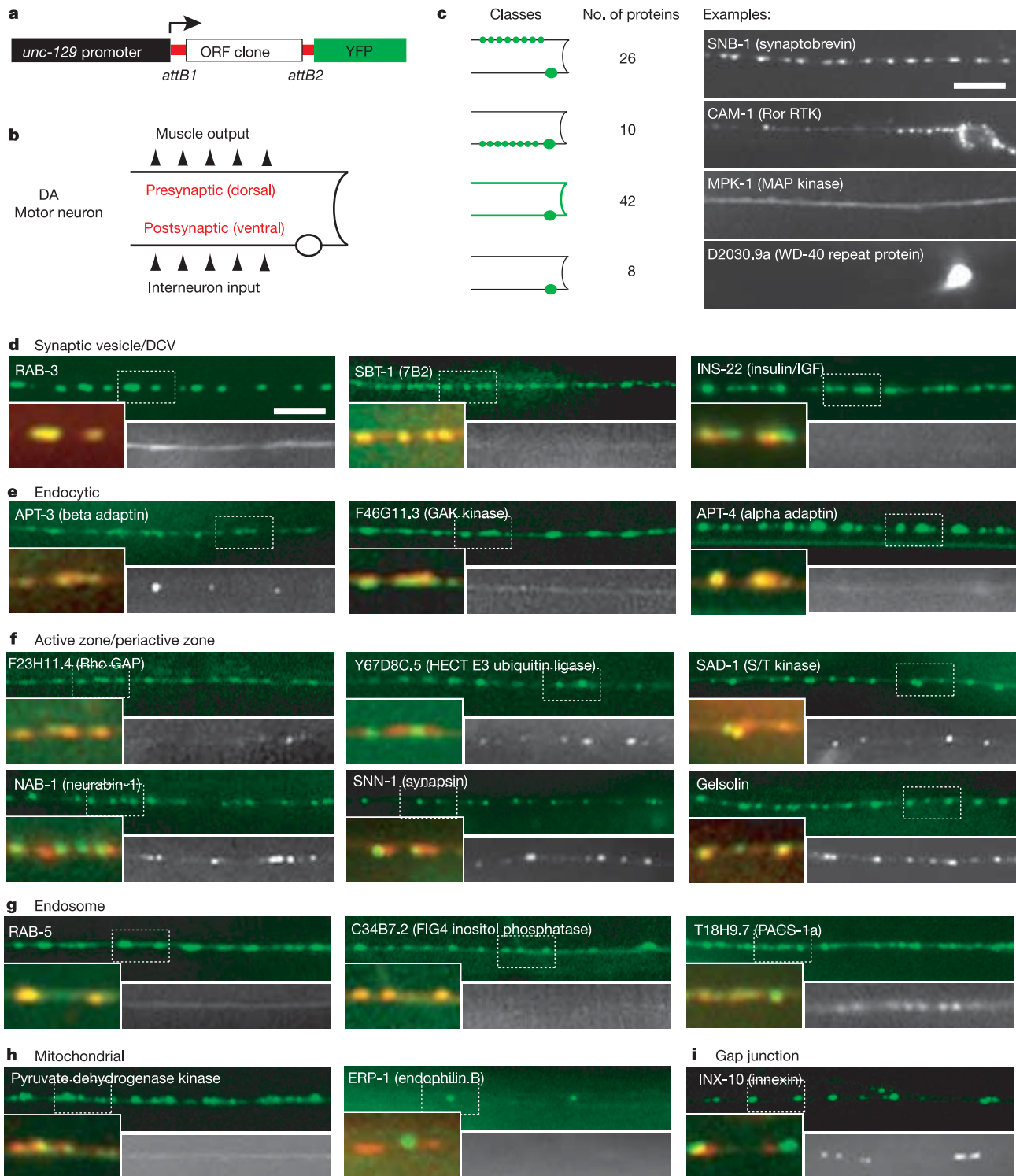
Figure 3 | Subcellular localization of proteins in motor neurons. a, Full-length cDNA clones were fused to YFP using the Gateway cloning system, and expressed in DA cholinergic motor neurons with the *unc-129* promoter. *attB1* and *attB2* recombination signals flank each open reading frame. **b**, DA motor neurons have cell bodies in the ventral cord, form neuromuscular junctions in the dorsal cord and receive synaptic inputs in the ventral cord. **c**, Summary of subcellular localization of 86 YFP fusions. Examples of localization classes from top to bottom are: dorsal cord punctate; ventral cord punctate; diffuse in nerve cords; cell body/nucleus. The number of fusion proteins in each class is indicated to the left of examples. Expression was not detected for 14 YFP fusion proteins. **d–i**, Representative images of YFP-tagged proteins localized in the dorsal cord. Localization of indicated YFP fusion protein (top) and co-localization with RFP::SNB-1 (bottom left, × 2 magnification) in the indicated region (dashed rectangle) is shown. Dorsal cord images of fusion proteins in *unc-104* (KIF1A) mutants are shown (bottom right). Scale bars, 10 μm. All genes listed are expressed in neurons except for the following: gelsolin (expression was detectable only in body wall muscle), SNN-1 and RAB-5 (both undetermined but synaptic in other systems).

(transient receptor potential V (TRPV)) (Fig. 2; see also Supplementary Fig. S2 and Supplementary Table S5). Thus, multiple molecular pathways apparently mediate the phorbol ester response in *C. elegans*. Previous studies suggested that Munc13 is the only phorbol ester target involved in neurotransmission in cultured rat hippocampal neurons[21]. This discrepancy probably reflects differences between these two preparations. Here, we measure the effects of phorbol esters on intact animals. Consequently, phorbol esters could be acting at any level of the sensory or motor circuitry. Consistent with this idea, expression of the *osm-9* (TRPV) gene is limited to sensory neurons, where it has a role in transducing several sensory responses[23]. These results support the idea that changes in sensory neuron activity contribute to the phorbol response. Several recent reports suggest that phorbol ester treatment stimulates the activity of rat TRPV channels[24].

Subcellular localization of positive proteins

To determine whether our screen was enriched for genes that encode synaptic proteins, we examined the subcellular localization of 100 proteins identified here (Supplementary Table S6). We tagged a set of full-length complementary DNA clones from the *C. elegans* ORFeome project[25] with yellow fluorescent protein (YFP) (Fig. 3a). We expressed the tagged proteins in DA motor neurons, which form NMJs in the dorsal nerve cord and receive synaptic inputs in the ventral cord[26] (Fig. 3b). Proteins localized to the dorsal cord were considered candidate axonal or pre-synaptic proteins, whereas those localized to the ventral cord were considered candidate dendritic or post-synaptic proteins. Of 100 proteins analysed, ten were localized in a punctate pattern in the ventral cord (candidate dendritic proteins) and 26 were punctate in the dorsal cord (candidate axonal

proteins) (Fig. 3c and Table 1). The candidate dendritic proteins included the CAM-1 (ROR1) receptor tyrosine kinase (Fig. 3c), which had a characteristic pattern of puncta that were proximal to the DA cell bodies, consistent with the somatodendritic localization of mouse ROR1 in hippocampal neurons[27].

To determine whether the axonal proteins are targeted to presynaptic specializations, we compared spatial fluorescence patterns to the synaptic vesicle protein synaptobrevin tagged with monomeric red fluorescent protein (RFP::SNB-1). Of 26 axonal proteins, 24 were co-localized with RFP::SNB-1, suggesting that they are concentrated at presynaptic nerve terminals (Fig. 3d–h). Although YFP fusion proteins are frequently used to study subcellular localization, it remains possible that some of the localization patterns reported here do not accurately reflect the localization of the corresponding endogenously expressed proteins. For example, some of the proteins analysed here may not normally be expressed in neurons and consequently could not be considered synaptic proteins. To address this point, we examined the expression of 20 predicted presynaptic genes (Supplementary Fig. S4 and Supplementary Table S7) and found that 19 are expressed in neurons (Table 1). These results support our prediction that the corresponding endogenously expressed proteins are found at presynaptic elements. We found several classes of presynaptic proteins, as detailed below.

unc-104 **(KIF1A) cargo.** For each candidate presynaptic protein, we determined whether its distribution was dependent upon the anterograde motor protein KIF1A (encoded by *unc-104*). We expected four classes of KIF1A cargo: synaptic vesicle[28], DCV[15,29], endocytic[30] and mitochondrial proteins[31]. Our results are largely consistent with these precedents (Fig. 3; see also Supplementary Table S6). In addition to the predicted classes of KIF1A cargo, a novel protein

Table 1 | Thirty positive genes that encode synaptic proteins or that affect synaptic structure

Locus	Gene	Description	Orthologues*			Novel†		Neuronal‡	*unc-104* (KIF1A)-dependent§	Synaptic‖	Mutant	SNB-1::GFP imaging¶
			H.s.	M.m.	D.m.	C.e.	Other					
fshr-1	C50H2.1	GPCR	■	■	■	Yes	Yes	■			Ric	▲ Punctal fluorescence
tap-1	C44H4.5	TAB1	■	■	■	Yes	Yes	■			Ric	▲ Inter-punctal fluorescence
arr-1	F53H8.2	β-Arrestin	■	■	■	Yes	Yes	■			Ric	▲ Inter-punctal fluorescence
wwp-1	Y65B4BR.4a	WWP1/NEDD4-related	■	■	■	Yes	Yes	■		○	Ric	▲ Inter-punctal fluorescence
osm-9	B0212.5	TRPV channel	▧	▧	■	Yes	Yes	■			Ric	▼ Punctal fluorescence
cam-1	C01G6.8	Ror RTK	■	■	▧	Yes	Yes	■		○	Ric	▼ Punctal fluorescence
	F46G11.3	GAK kinase	■	■	▧	Yes	Yes	ND	Yes	● S	NA	
nlp-12	M01D7.5	FaRP neuropeptide				Yes	Yes	■	Yes	● S	Ric	
sbt-1	T03D8.3	7B2	■	■		Yes	Yes	■	Yes	● P	Ric	
	C29F5.7	Globin-like				Yes	Yes	■	Yes	● S	NA	
	ZK370.5	Pyruvate dehydrogenase kinase	■	■		Yes	Yes	■	Yes	● P	NA	
	F13A2.4	Novel				Yes	Yes	■	Yes	● S	NA	
	F33D11.1	PDZ-LIM	▧	▧	▧	Yes	Yes	ND	No	● P	NA	
	F23H11.4	RhoGAP	■	■	■	Yes	Yes	■	Yes	● S	NA	
	C34B7.2	FIG4 inositol phosphatase	■	■	■	Yes	Yes	■	Yes	● S	NA	
	F23C8.6	Chmp-1/VPS46	▧	■	■	Yes	Yes	■	No	● S	NA	
	T18H9.7a	PACS-1a	■	■	■	Yes	Yes	■	No	● S	NA	
	Y67D8C.5	HECT E3 ubiquitin ligase	■	■	■	Yes	Yes	■	No	● P	NA	
unc-31	ZK897.1	CAPS				Yes		■			Ric	▲ Punctal fluorescence
let-60	ZK792.6	Ras	■	■	■	Yes		■			Ric	▼ Punctal fluorescence
nab-1	C43E11.6a	Neurabin				Yes		■	No	● P	Ric	▼ Punctal density
prk-2	F45H7.4	S/T kinase	■	■	▧	Yes		ND			Ric	No DA9 fluorescence
ins-22	M04D8.2	Insulin/IGF				Yes		■	Yes	● P	NA	
ins-31	T10D4.4	Insulin/IGF				Yes		ND	Yes	● P	NA	
apt-3	Y71H2B.10	β-Adaptin	■	■	■	Yes		■	No	● P	NA	
rab-5	F26H9.6	RAB5				Yes		ND	Yes	● S	NA	
apt-4	T20B5.1	α-Adaptin				Yes		■	Yes	● S	NA	
snn-1	Y38C1BA.2	Synapsin				Yes		ND	No	● P	NA	
egl-3	C51E3.7a	Proprotein convertase 2	■	■	■		Yes	■		○	Ric	▼ Punctal fluorescence
sad-1	F15A2.6	S/T kinase	■	■	■	Yes		■	No	● S	Dgk	▼ Punctal fluorescence

NA, no mutant available. ND, not determined.

*Orthologues (best reciprocal BLAST hits) are indicated by black shading and homologues by light grey shading. H.s., M.m. and D.m. refer to human, mouse and *Drosophila*, respectively.

†'Yes' indicates genes with no previously characterized roles in neurotransmission in *C. elegans* (C.e.) or in other organisms (Other).

‡Genes expressed in neurons are indicated by black shading.

§*unc-104* (KIF1A) dependence was scored as 'Yes' if the punctate pattern became diffuse in *unc-104* mutants, and was scored as 'No' if any puncta remained in *unc-104* mutants.

‖Symbols and letters indicate the following protein localization: filled circles, YFP fusion protein is punctate in the dorsal cord; open circles, YFP fusion protein is diffuse in the dorsal cord; S, synaptic; P, peri-synaptic.

¶Arrowheads indicate the direction of changes that are significantly different from wild type (*P* < 0.05, Student's *t*-test; see also Fig. 4).

(F13A2.4), a globin-related protein (C29F5.7) and a RhoGAP protein (F23H11.4) also fell into this class of presynaptic proteins (Fig. 3; see also Supplementary Fig. S3).

Cytoskeletal and active zone proteins. In contrast to many vesicular proteins, anterograde transport of active zone proteins (for example, UNC-13, UNC-10 (RIM1) and SYD-2 (α-liprin)) and of periactive zone proteins (for example, RPM-1 (highwire)) is not dependent upon KIF1A[32–35]. Three predicted actin-binding proteins (SNN-1

(synapsin), K06A4.3 (gelsolin) and NAB-1 (neurabin-1)), an endosomal protein T18H9.7 (PACS-1a), the Y67D8C.5 HECT E3 ligase, and the SAD-1 serine/threonine protein kinase fell into this category of presynaptic proteins. F-actin forms annular structures that encircle active zones in several organisms[36,37]. These F-actin-rich regions are thought to correspond to endocytic zones, or to sites of synaptic adhesion molecules (for example, apCAM, FasII), and have been reported to regulate synaptic growth and plasticity[38,39].
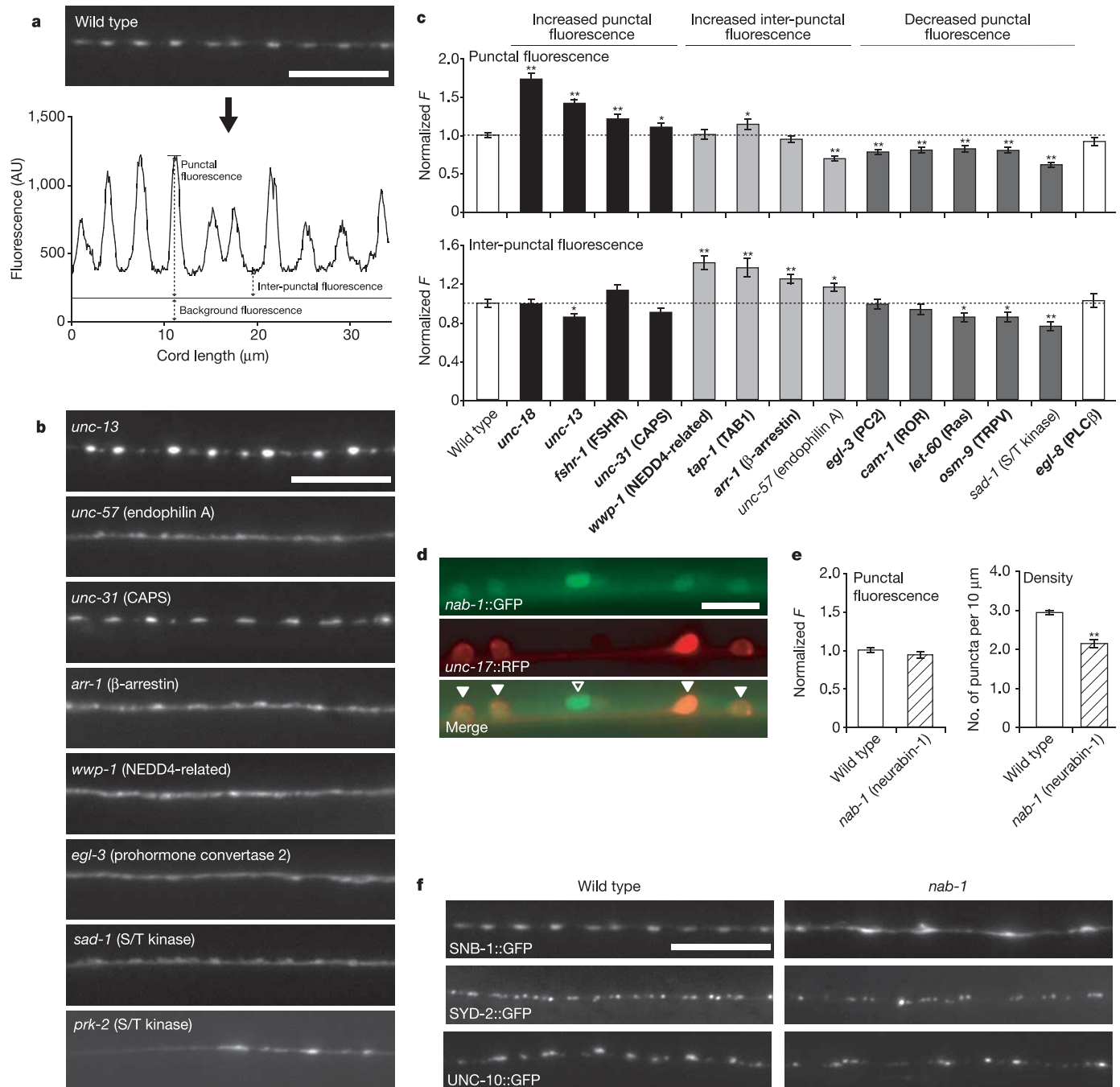
Figure 4 | Synaptobrevin distribution in mutants. a, Top: image of the dorsal cord in a wild-type animal expressing GFP::SNB-1 in DA motor neurons (*nuIs152*). Bottom: quantifying punctal and inter-punctal GFP fluorescence from line scans. AU, arbitrary units. **b,** Representative images of GFP::SNB-1 distribution in the dorsal cord of the indicated mutants. **c,** Punctal and inter-punctal fluorescence is shown for wild type and the indicated mutants. Genes not previously annotated for altered synaptobrevin distribution are indicated in bold. **d,** *nab-1* is expressed in motor neurons. Top panel: Motor neurons expressing GFP by the *nab-1* promoter (green). Middle panel: cholinergic construct (RFP expressed by the *unc-17* promoter). Bottom panel: merged image showing NAB-1 expression in cholinergic (arrowheads) and GABAergic (open arrowhead) neurons. **e,** The density of GFP::SNB-1 puncta was significantly reduced in *nab-1* mutants. **f,** Representative images of wild type (left side) and *nab-1* mutants (right side) expressing the indicated GFP fusion proteins. Double asterisks in **c, e** denote $P < 0.01$; single asterisks denote $P < 0.05$; Student's *t*-test. Scale bars, 10 μm. Error bars are standard error.

NAB-1 and SNN-1 proteins localize to regions adjacent to synaptobrevin puncta, consistent with a periactive zone localization (Fig. 3f; see also Supplementary Fig. S3). Neurabin-1 is a PDZ-containing scaffolding protein that has been implicated in regulating postsynaptic filopodial motility and spine formation in rat hippocampal neurons[40,41]. Synapsin has been proposed to tether synaptic vesicles to F-actin, constituting a reserve pool of vesicles. Rat synapsin is also localized to periactive zones[42]. The Y67D8C.5 (HECT) protein, a putative E3 ligase, was also apparently localized to periactive zones and was KIF1A-independent (Fig. 3f). A second E3 ligase, the RING-containing protein RPM-1 (highwire), was previously shown to localize to periactive zones, where it regulates the size and number of active zones present at NMJs[34,39]. It is possible that the Y67D8C.5 (HECT) protein has a similar function.

The SAD-1 protein kinase has been implicated in regulating the axo-dendritic polarity of sensory neurons in *C. elegans*[43]. SAD-1 had a very characteristic distribution where each SNB-1 punctum was often labelled by two overlapping SAD-1 puncta (Fig. 3f). This pattern is consistent with the idea that SAD-1 is localized to active zones, because the DA neurons form dyadic synapses in which a single nerve terminal is presynaptic to two different postsynaptic elements (a body muscle and a VD motor neuron)[26].

Fluorescence profiling of synaptic genes

To determine whether genes identified in our screen regulate synapse structure or development, we examined the distribution of green fluorescent protein (GFP)-tagged synaptobrevin (GFP::SNB-1) in 19 strains carrying loss-of-function mutations in genes identified in our screen. The distribution of GFP::SNB-1 in mutant strains was analysed by quantitative fluorescence microscopy (Fig. 4a and Table 1). Mutants lacking PRK-2 (Pim1) had a large gap in the dorsal cord GFP::SNB-1 fluorescence, and typically lacked the DA9 cell body, although other DA neurons were superficially normal (Fig. 4b). These results suggest that PRK-2 is required for differentiation of a subset of the DA neurons.

Changes in GFP::SNB-1 distribution can also be used to detect defects in synaptic vesicle recycling. Endocytic defects result in accumulation of synaptobrevin in the plasma membrane whereas exocytic defects cause aberrant accumulation of synaptic vesicles at nerve terminals. Corresponding changes in the distribution of GFP::SNB-1 are observed in exocytic and endocytic mutants. In an endocytic mutant (*unc-57*, which encodes endophilin A), the diffuse inter-punctal fluorescence increased by 17%, as would be predicted if the fraction of synaptobrevin in the plasma membrane was increased. In two exocytic mutants (*unc-13* and *unc-18*), the average intensity of GFP::SNB-1 puncta increased (41% and 72%, respectively), which is similar to the increase in synaptic vesicle numbers observed in these mutants by electron microscopy[44,45]. We detected significant GFP::SNB-1 changes in 11 mutants (Fig. 4b, c and Table 1).

Two mutants (*unc-31* (CAPS) and *fshr-1* (FSHR)) had increased puncta fluorescence, suggestive of exocytic defects. The CAPS protein has been proposed to be involved in secretion of DCVs[18,46]. Thus, the altered GFP::SNB-1 pattern observed in *unc-31* mutants may be an indirect consequence of neuropeptide secretion defects. Alternatively, DCVs may also contain SNB-1 and the increased punctal fluorescence in *unc-31* (CAPS) mutants might reflect DCV accumulation rather than synaptic vesicle accumulation. Interestingly, mutants lacking a predicted neuropeptide receptor (FSHR-1) and those lacking CAPS had similar GFP::SNB-1 defects; however, mutants lacking *egl-3* (PC2) had a distinct GFP::SNB-1 profile. The distinct phenotypes of the *egl-3* (PC2) and *unc-31* (CAPS) mutants might reflect the fact that these mutations affect different sets of neurotransmitters; for example, *unc-31* (CAPS) might be necessary for the release of both neuropeptides and monoamines, whereas *egl-3* (PC2) may be required for processing a subset of proneuropeptides.

Three mutants (*wwp-1* (NEDD4-related protein), *arr-1* (β-arrestin) and *tap-1* (TAB1)) had increased interpunctal fluorescence, suggestive of endocytic defects. The NEDD4 ubiquitin ligase is required for targeting many membrane proteins to lysosomes[47,48]. β-Arrestin is required for endocytosis of G-protein-coupled receptors, which in some cases is dependent upon mono-ubiquitination of β-arrestin[49]. TAB1 binds to the transforming growth factor-β activated kinase (TAK1), which in turn activates p38 MAP kinase[50]. Notably, p38 MAP kinase has recently been shown to regulate SNB-1 localization at NMJs[51]. Our results suggest that these proteins are also required for endocytosis of synaptobrevin, and perhaps for synaptic vesicle recycling.

Neurabin-1 regulates synapse density

In rodents, neurabin-1 has primarily been described in dendritic spines[41,52,53]. By contrast, our results suggest that neurabin-1 is localized to presynaptic elements at NMJs, perhaps to F-actin-rich periactive zones (Fig. 3f). Consistent with our localization data, a *nab-1* transcriptional reporter construct was expressed in ventral cord motor neurons, but not in body muscles (Fig. 4d). Overexpression of rodent neurabin-1 has been shown to cause an increase in the density of dendritic spines in cultured hippocampal neurons[40,41]. We found that mutants lacking NAB-1 had a significantly reduced density of GFP::SNB-1 puncta (Fig. 4e, f). Similar reductions in puncta densities were observed for two other active zone proteins: SYD-2 (α-liprin) and UNC-10 (RIM1) (Fig. 4f). Thus, our results suggest that NAB-1 can act both pre- and postsynaptically to regulate synapse density.

Summary

Identification of synaptic proteins has been the subject of intensive efforts over the past 20 years. These efforts have been largely restricted to biochemical strategies[54]. Here we use a functional genomic screen to identify 132 genes for which there was no previous annotated function in synaptic transmission. Thus, further screens based on gene function are likely to identify new components required for synaptic transmission, and will therefore be complementary to ongoing biochemical strategies. Although our data provide an initial glimpse of gene functions, precise conclusions about processes regulated by these genes will require detailed analysis of each gene.

METHODS

A detailed description of all Methods is provided in the Supplementary Information.

***C. elegans* strains and phenotypic analysis.** Strains were maintained at 20 °C, except where noted. Aldicarb, phorbol and levamisole responses of young adults were measured in triplicate as described[12]. Full time courses of drug responses were performed on all mutants (data not shown). For temperature-sensitive mutants, assays were done after overnight incubation at 25 °C.

RNAi feeding experiments and functional profiling. RNAi feeding screens were performed in a 24-well format as described[2]. A total of 100 embryos were placed per well. After 3 days of RNAi treatment at 20 °C, 25 young adults per well were transferred into wells containing 1 mM aldicarb, and scored for paralysis after 100 min. For each 24-well plate assayed, empty vector (negative) and *egl-3* (positive) control wells were randomly included. Two screens were performed in parallel using *eri-1; lin-15B* double mutants or *eri-1; dgk-1 lin-15B* triple mutants. Genes were considered positive based on the following criteria: significant ($P \le 0.0001$) Ric or Dgk RNAi scores, an available loss of function mutation that is aldicarb resistant, or protein localization to punctate structures in the DA motor neurons.

For functional profiling, RNAi treatments were done for 6 days (that is, two generations). Quadruplicate assays were done in seven conditions total, four genotypes and three drug treatments. Significant aldicarb resistance ($P < 0.05$) in each condition is reported. Profiling results were clustered using the hclust package in R language for statistical computing.

1. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).

2. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
3. Simmer, F. *et al.* Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* **1**, E12 (2003).
4. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
5. Simmer, F. *et al.* Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12**, 1317–1319 (2002).
6. Miller, K. G. *et al.* A genetic selection for *Caenorhabditis elegans* synaptic transmission mutants. *Proc. Natl Acad. Sci. USA* **93**, 12593–12598 (1996).
7. Nguyen, M., Alfonso, A., Johnson, C. D. & Rand, J. B. *Caenorhabditis elegans* mutants resistant to inhibitors of acetylcholinesterase. *Genetics* **140**, 527–535 (1995).
8. Nurrish, S., Segalat, L. & Kaplan, J. M. Serotonin inhibition of synaptic transmission: $G\alpha_O$ decreases the abundance of UNC-13 at release sites. *Neuron* **24**, 231–242 (1999).
9. Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
10. Boulton, S. J. *et al.* Combined functional genomic maps of the *C. elegans* DNA damage response. *Science* **295**, 127–131 (2002).
11. Robatzek, M. & Thomas, J. H. Calcium/calmodulin-dependent protein kinase II regulates *Caenorhabditis elegans* locomotion in concert with a G_o/G_q signaling network. *Genetics* **156**, 1069–1082 (2000).
12. Lackner, M. R., Nurrish, S. J. & Kaplan, J. M. Facilitation of synaptic transmission by EGL-30 Gqα and EGL-8 PLCβ: DAG binding to UNC-13 is required to stimulate acetylcholine release. *Neuron* **24**, 335–346 (1999).
13. Greener, T., Zhao, X., Nojima, H., Eisenberg, E. & Greene, L. E. Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *J. Biol. Chem.* **275**, 1365–1370 (2000).
14. Kass, J., Jacob, T. C., Kim, P. & Kaplan, J. M. The EGL-3 proprotein convertase regulates mechanosensory responses of *Caenorhabditis elegans*. *J. Neurosci.* **21**, 9265–9272 (2001).
15. Jacob, T. C. & Kaplan, J. M. The EGL-21 carboxypeptidase E facilitates acetylcholine release at *Caenorhabditis elegans* neuromuscular junctions. *J. Neurosci.* **23**, 2122–2130 (2003).
16. Jia, K., Chen, D. & Riddle, D. L. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. *Development* **131**, 3897–3906 (2004).
17. Tatar, M., Bartke, A. & Antebi, A. The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346–1351 (2003).
18. Ann, K., Kowalchyk, J. A., Loyet, K. M. & Martin, T. F. Novel Ca^{2+}-binding protein (CAPS) related to UNC-31 required for Ca^{2+}-activated exocytosis. *J. Biol. Chem.* **272**, 19637–19640 (1997).
19. Lindberg, I., Tu, B., Muller, L. & Dickerson, I. M. Cloning and functional analysis of *C. elegans* 7B2. *DNA Cell Biol.* **17**, 727–734 (1998).
20. Nathoo, A. N., Moeller, R. A., Westlund, B. A. & Hart, A. C. Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proc. Natl Acad. Sci. USA* **98**, 14000–14005 (2001).
21. Rhee, J. S. *et al.* Beta phorbol ester- and diacylglycerol-induced augmentation of transmitter release is mediated by Munc13s and not by PKCs. *Cell* **108**, 121–133 (2002).
22. Gillis, K. D., Mossner, R. & Neher, E. Protein kinase C enhances exocytosis from chromaffin cells by increasing the size of the readily releasable pool of secretory granules. *Neuron* **16**, 1209–1220 (1996).
23. Colbert, H. A., Smith, T. L. & Bargmann, C. I. OSM-9, a novel protein with structural similarity to channels, is required for olfaction, mechanosensation, and olfactory adaptation in *Caenorhabditis elegans*. *J. Neurosci.* **17**, 8259–8269 (1997).
24. Premkumar, L. S. & Ahern, G. P. Induction of vanilloid receptor channel activity by protein kinase C. *Nature* **408**, 985–990 (2000).
25. Reboul, J. *et al. C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genet.* **34**, 35–41 (2003).
26. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Tran. R. Soc. Lond. B* **314**, 1–340 (1986).
27. Paganoni, S. & Ferreira, A. Expression and subcellular localization of Ror tyrosine kinase receptors are developmentally regulated in cultured hippocampal neurons. *J. Neurosci. Res.* **73**, 429–440 (2003).
28. Hall, D. H. & Hedgecock, E. M. Kinesin-related gene *unc-104* is required for axonal transport of synaptic vesicles in *C. elegans*. *Cell* **65**, 837–847 (1991).
29. Zahn, T. R. *et al.* Dense core vesicle dynamics in *Caenorhabditis elegans* neurons and the role of kinesin UNC-104. *Traffic* **5**, 544–559 (2004).
30. Schuske, K. R. *et al.* Endophilin is required for synaptic vesicle endocytosis by localizing synaptojanin. *Neuron* **40**, 749–762 (2003).
31. Fuchs, F. & Westermann, B. Role of Unc104/KIF1-related motor proteins in mitochondrial transport in *Neurospora crassa*. *Mol. Biol. Cell* **16**, 153–161 (2004).
32. Kohn, R. *et al.* Expression of multiple UNC-13 proteins in the *C. elegans* nervous system. *Mol. Biol. Cell* **11**, 3441–3452 (2000).
33. Zhen, M. & Jin, Y. The liprin protein SYD-2 regulates the differentiation of presynaptic termini in *C. elegans*. *Nature* **401**, 371–375 (1999).
34. Zhen, M., Huang, X., Bamber, B. & Jin, Y. Regulation of presynaptic terminal organization by *C. elegans* RPM-1, a putative guanine nucleotide exchanger with a RING-H2 finger domain. *Neuron* **26**, 331–343 (2000).
35. Koushika, S. P. *et al.* A post-docking role for active zone protein Rim. *Nature Neurosci.* **4**, 997–1005 (2001).
36. Sankaranarayanan, S., Atluri, P. P. & Ryan, T. A. Actin has a molecular scaffolding, not propulsive, role in presynaptic function. *Nature Neurosci.* **6**, 127–135 (2003).
37. Dunaevsky, A. & Connor, E. A. F-actin is concentrated in nonrelease domains at frog neuromuscular junctions. *J. Neurosci.* **20**, 6007–6012 (2000).
38. Sone, M. *et al.* Synaptic development is controlled by the periactive zones of *Drosophila* synapses. *Development* **127**, 4157–4168 (2000).
39. Wan, H. I. *et al.* Highwire regulates synaptic growth in *Drosophila*. *Neuron* **26**, 313–329 (2000).
40. Oliver, C. J. *et al.* Targeting protein phosphatase 1 (PP1) to the actin cytoskeleton: the neurabin I/PP1 complex regulates cell morphology. *Mol. Cell. Biol.* **22**, 4690–4701 (2002).
41. Zito, K., Knott, G., Shepherd, G. M., Shenolikar, S. & Svoboda, K. Induction of spine growth and synapse formation by regulation of the spine actin cytoskeleton. *Neuron* **44**, 321–334 (2004).
42. Bloom, O. *et al.* Colocalization of synapsin and actin during synaptic vesicle recycling. *J. Cell Biol.* **161**, 737–747 (2003).
43. Crump, J. G., Zhen, M., Jin, Y. & Bargmann, C. I. The SAD-1 kinase regulates presynaptic vesicle clustering and axon termination. *Neuron* **29**, 115–129 (2001).
44. Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2**, 959–964 (1999).
45. Weimer, R. M. *et al.* Defects in synaptic vesicle docking in unc-18 mutants. *Nature Neurosci.* **6**, 1023–1030 (2003).
46. Renden, R. *et al. Drosophila* CAPS is an essential gene that regulates dense-core vesicle release and synaptic vesicle fusion. *Neuron* **31**, 421–437 (2001).
47. Helliwell, S. B., Losko, S. & Kaiser, C. A. Components of a ubiquitin ligase complex specify polyubiquitination and intracellular trafficking of the general amino acid permease. *J. Cell Biol.* **153**, 649–662 (2001).
48. Rotin, D., Staub, O. & Haguenauer-Tsapis, R. Ubiquitination and endocytosis of plasma membrane proteins: role of Nedd4/Rsp5p family of ubiquitin-protein ligases. *J. Membr. Biol.* **176**, 1–17 (2000).
49. Shenoy, S. K., McDonald, P. H., Kohout, T. A. & Lefkowitz, R. J. Regulation of receptor fate by ubiquitination of activated β2-adrenergic receptor and β-arrestin. *Science* **294**, 1307–1313 (2001).
50. Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
51. Nakata, K. *et al.* Regulation of a DLK-1 and p38 MAP kinase pathway by the ubiquitin ligase RPM-1 is required for presynaptic development. *Cell* **120**, 407–420 (2005).
52. Allen, P. B., Ouimet, C. C. & Greengard, P. Spinophilin, a novel protein phosphatase 1 binding protein localized to dendritic spines. *Proc. Natl Acad. Sci. USA* **94**, 9956–9961 (1997).
53. Muly, E. C., Smith, Y., Allen, P. & Greengard, P. Subcellular distribution of spinophilin immunolabeling in primate prefrontal cortex: localization to and within dendritic spines. *J. Comp. Neurol.* **469**, 185–197 (2004).
54. Grant, S. G. Systems biology in neuroscience: bridging genes to cognition. *Curr. Opin. Neurobiol.* **13**, 577–582 (2003).
55. Wang, D. *et al.* Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. *Nature* doi:10.1038/nature04010 (this issue).

ARTICLES

Genes that mediate breast cancer metastasis to lung

Andy J. Minn[1,2]*, Gaorav P. Gupta[1]*, Peter M. Siegel[1]†, Paula D. Bos[1], Weiping Shu[1], Dilip D. Giri[3]†, Agnes Viale[5], Adam B. Olshen[4], William L. Gerald[3] & Joan Massagué[1,6]

By means of *in vivo* selection, transcriptomic analysis, functional verification and clinical validation, here we identify a set of genes that marks and mediates breast cancer metastasis to the lungs. Some of these genes serve dual functions, providing growth advantages both in the primary tumour and in the lung microenvironment. Others contribute to aggressive growth selectively in the lung. Many encode extracellular proteins and are of previously unknown relevance to cancer metastasis.

Metastasis is frequently a final and fatal step in the progression of solid malignancies. Tumour cell intravasation, survival in circulation, extravasation into a distant organ, angiogenesis and uninhibited growth constitute the metastatic process[1]. The molecular requirements for some of these steps may be tissue specific. Indeed, the proclivity that tumours have for specific organs, such as breast carcinomas for bone and lung, was noted more than a century ago[2].

The identity and time of onset of the changes that endow tumour cells with these metastatic functions are largely unknown and are a subject of debate. It is believed that genomic instability generates large-scale cellular heterogeneity within tumour populations, from which rare cellular variants with augmented metastatic abilities evolve through a darwinian selection process[2,3]. Work on experimental metastasis with tumour cell lines has demonstrated that reinjection of metastatic cell populations can lead to enrichment in the metastatic phenotype[4–6]. Recently, however, the existence of genes expressed by rare cellular variants that specifically mediate metastasis has been challenged[7]. Transcriptomic profiling of primary human carcinomas has identified gene expression patterns that, when present in the bulk primary tumour population, predict a poor prognosis for patients[8–10]. The existence of such signatures has been interpreted to mean that genetic lesions acquired early in tumorigenesis are sufficient for the metastatic process, and that consequently no metastasis-specific genes exist. However, it is unclear whether these genes predicting metastatic recurrence are also functional mediators.

The lungs and bones are frequent sites of breast cancer metastasis, and metastases to these sites differ in terms of their evolution, treatment, morbidity and mortality[11]. Reasoning that each organ places different demands on circulating cancer cells for the establishment of metastases, we sought to identify genes expressed in breast cancer cells that selectively mediate lung metastasis and that are correlated with the propensity of primary human breast cancers to relapse to the lungs.

Selection of cells metastatic to the lungs
The cell line MDA-MB-231 was derived from the pleural effusion of a breast cancer patient suffering from widespread metastasis years after removal of her primary tumour[12]. Individual MDA-MB-231 cells grown and tested as single-cell-derived progenies (SCPs) have distinct metastatic abilities and tissue tropisms[13] despite having similar expression levels of genes constituting a validated Rosetta-type poor prognosis signature[9] (Supplementary Fig. S1). These different metastatic behaviours, including different tropisms to bone and lung, are associated with discrete variation in overall gene expression patterns (Supplementary Fig. S1; ref. 13). We therefore proposed that organ-specific metastasis must be determined by genes that are distinct from a Rosetta-type poor prognosis signature and are differentially expressed within the MDA-MB-231 population. Indeed, previous work has shown this to be true for most of the genes linked to the activity of bone metastatic subpopulations[4,13].

To identify genes that mediate lung metastasis we tested parental MDA-MB-231 cells and the 1834 sub-line (an *in vivo* isolate with no enhancement in bone metastatic behaviour[4]; Fig. 1a) by injection into the tail vein of immunodeficient mice (Fig. 1b). Metastatic activity was assayed by bioluminescence imaging (BLI) of luciferase-transduced cells as well as gross examination of the lungs at necropsy. The 1834 cells exhibited limited but significant lung metastatic activity compared with the parental population (Fig. 1b). When 1834-derived lung lesions were expanded in culture and reinoculated into mice, these cells (denoted LM1 subpopulations; Fig. 1a) showed increased lung metastatic activity. Another round of selection *in vivo* yielded second-generation populations (denoted LM2) that were rapidly and efficiently metastatic to the lungs (Fig. 1b). Histological analysis confirmed that LM2 lesions replaced large areas of the lung parenchyma, whereas 1834 cells exhibited intravascular growth with less extensive extravasation and parenchymal involvement (Fig. 1c). Inoculation of as few as 2×10^3 LM2 cells was sufficient for the emergence of aggressive lung metastases, whereas inoculation of 2×10^5 parental cells left only a residual, indolent population in the lungs (Fig. 1d). Furthermore, the enhancement in lung metastatic activity was tissue specific. When LM2 populations were inoculated into the left cardiac ventricle to facilitate bone metastasis, their metastatic activity was comparable to that of the parental and 1834 populations, and it was markedly inferior to that of a previously described, highly aggressive bone metastatic population (Fig. 1b).
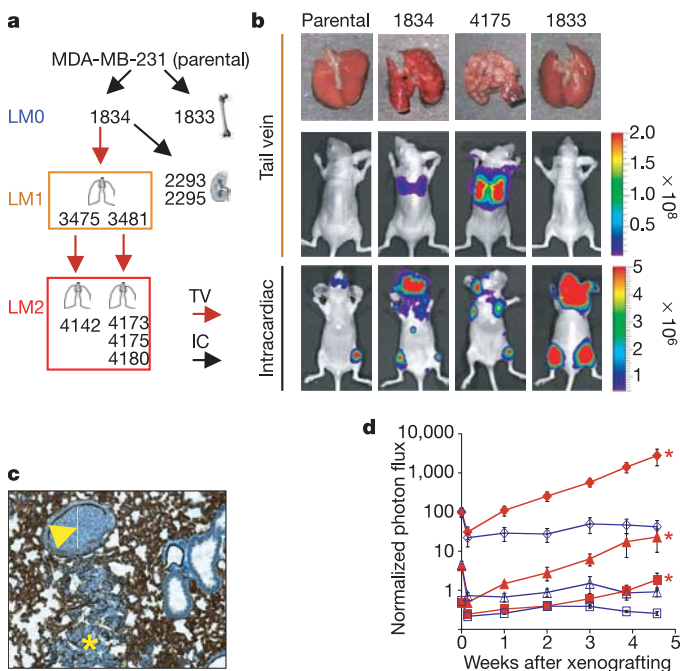
[1]Cancer Biology and Genetics Program, Departments of [2]Radiation Oncology, [3]Pathology and [4]Epidemiology and Biostatistics, [5]Genomics Core Laboratory, and [6]Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. †Present addresses: McGill University Health Centre, Montreal, Quebec, Canada H34 1A4 (P.M.S.); Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island 02912, USA (D.D.G.).
*These authors contributed equally to this work.

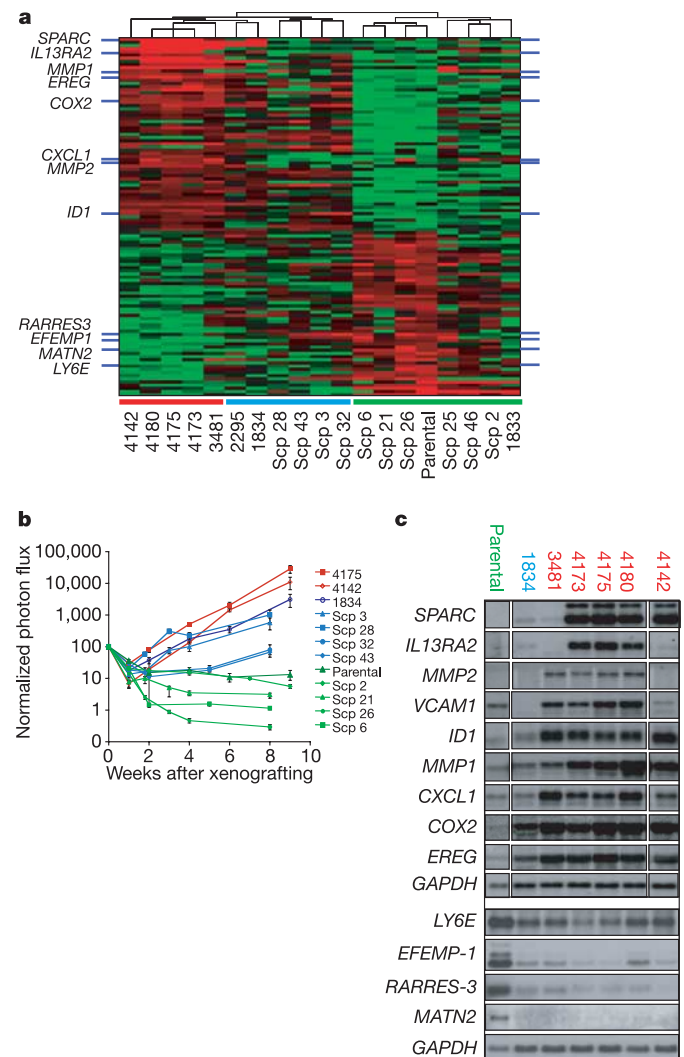## Establishing a lung metastasis signature

To identify patterns of gene expression associated with aggressive lung metastatic behaviour, we performed a transcriptomic microarray analysis of the highly and weakly lung-metastatic cell populations. The gene list obtained from a class comparison between parental and LM2 populations was filtered to exclude genes that were expressed at low levels in a majority of samples and to ensure a threefold or higher change in expression level between the two groups. A total of 95 unique genes (113 probe sets) met these criteria: 48 were overexpressed and 47 underexpressed in cell populations most metastatic to the lungs (Fig. 2a and Supplementary Table 2). This gene set was largely distinct from the bone metastasis gene-expression signature previously identified in bone metastatic isolates derived from the same parental cell line[4]. In fact, only six genes overlapped with concordant expression patterns between the two groups (Supplementary Table 3).

Hierarchical clustering with the 95-gene list confirmed a robust relationship between this gene expression signature and the lung-specific metastatic activity of cell populations selected *in vivo* (Fig. 2a). In addition, this gene expression signature segregated the SCPs (which were not used in generating the gene list) into two major groups, one transcriptomically resembling the parental cells, the other more similar to the lung-metastatic populations selected *in vivo*. This latter group of SCPs was also more metastatic to lung than the former group (Fig. 2b). However, unlike the LM2 populations, none of the lung-metastatic SCPs concordantly expressed all of the genes in the lung metastasis signature (Fig. 2a). Consistent with this was our observation that the lung metastatic activity of the LM2 populations was about one order of magnitude greater than the most aggressive SCPs (Fig. 2b). We postulated that the subset of genes from the 95-gene signature that are uniformly expressed by all lung-metastatic SCPs and populations selected *in vivo* might confer baseline lung-metastatic functions, which we define as lung meta-stagenicity. Genes expressed exclusively in the most aggressive LM2 populations may serve specialized, lung-restricted functions, which we collectively describe as lung-metastatic virulence. A final list of 54

Figure 1 | Selection of breast cancer cells metastatic to lung. a, Flow chart of the selection of organ-specific metastatic subpopulations *in vivo*, indicating the organs from which these subpopulations were isolated. Each subsequent lung-metastatic generation is designated LM0, LM1 and LM2. The LM2 cells were further analysed for metastasis by either tail-vein (TV) or intracardiac (IC) xenografting. Metastatic propensities for all cell lines used in this study are listed in Supplementary Table 1. **b,** Representative lungs harvested at necropsy and BLI of the indicated cell lines are shown after tail-vein or intracardiac injection. The colour scale depicts the photon flux (photons per second) emitted from xenografted mice **c,** A representative image of haematoxylin staining of lung cryosections from mice injected with moderately metastatic 1834 cells showing an invading lesion (asterisk) and an embolus within the vascular space (arrowhead). Vascular walls are stained with the endothelial cell marker CD31. **d,** Parental cells (red) and 4175 (LM2) cells (blue) were tested for lung metastatic activity. Numbers of cells injected were as follows: diamonds, 2×10^5; triangles, 2×10^4; squares, 2×10^3. Plots show a quantification of the luminescence signal as a function of time. Results are means ± s.e.m. for each cohort. Asterisks, $P < 0.05$ with a one-sided rank test, compared with mice injected with an equivalent number of parental cells.

Figure 2 | Gene-expression signature associated with lung metastasis. a, Comparison of gene expression profiles of LM2 populations with parental cells identifies 113 probe sets that are correlated with lung metastatic activity. This signature clusters populations selected *in vivo* and SCPs into groups that resemble the LM2 cell lines (red bar along the bottom), the parental MDA-MB-231 cell line (green bar) or an intermediate group (blue bar). **b,** LM2 populations 4175 and 4142 were assayed for lung metastatic activity as measured by BLI and were compared with parental populations and various SCPs[13]. Plots show a quantification of the luminescence signal as a function of time. Results are means ± s.e.m. for each cohort. Colour-coding is as in **a. c,** Northern blot analysis of parental, LM0, LM1 and LM2 cell lines with a set of nine lung metastasis genes selected for functional validation, as well as four genes underexpressed in the lung-metastatic populations.

candidate lung metastagenicity and virulence genes was selected for further evaluation (Supplementary Methods and Supplementary Table 4).
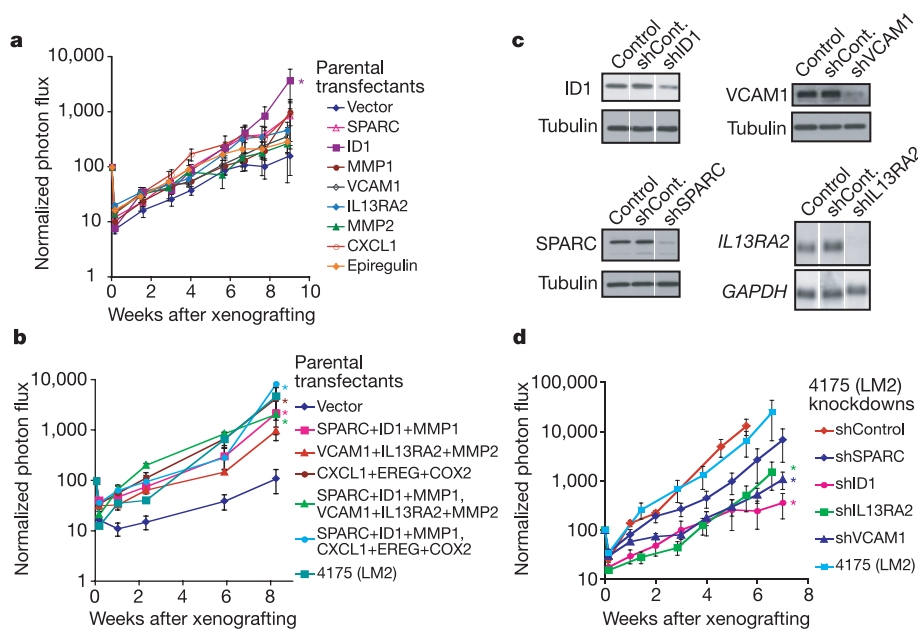
Genes that mediate lung metastasis

A subset of biologically interesting genes overexpressed in the 54 gene list was selected for functional validation. These genes include those encoding the epidermal-growth-factor family member epiregulin (*EREG*), which is a broad-specificity ligand for the HER/ErbB family of receptors[14,15], the chemokine *GRO1/CXCL1* (ref. 16), the matrix metalloproteinases *MMP1* (collagenase 1)[17] and *MMP2* (gelatinase A)[18], the cell adhesion molecule *SPARC*[19], the interleukin-13 decoy receptor *IL13Rα2* (ref. 20) and the cell adhesion receptor *VCAM1* (refs 21, 22) (Fig. 2a). These genes encode secretory or receptor proteins, indicating possible roles in the tumour cell microenvironment. In addition to these genes, we included the transcriptional inhibitor of cell differentiation and senescence *ID1* (refs 23, 24) and the prostaglandin-endoperoxide synthase *PTGS2/COX2* (ref. 25). Northern blot analysis of the various cell populations selected *in vivo* revealed expression patterns for these genes that were correlated with metastatic behaviour (Fig. 2c). *SPARC*, *IL13RA2*, *VCAM1* and *MMP2* belong to the subset of genes whose expression is generally restricted to aggressive lung-metastatic populations and are rarely expressed (less than 10% prevalence for *VCAM1* and *IL13Rα2*, and less than 2% prevalence for *SPARC* and *MMP2*) in randomly picked SCPs (data not shown). In contrast, the expression of *ID1*, *CXCL1*, *COX2*, *EREG* and *MMP1* is not restricted to aggressive lung metastasis populations but increases with lung metastatic ability. Analysis of protein expression for these genes confirmed that the differences in mRNA levels translated into significant alterations in protein levels (Supplementary Fig. S2).

To determine whether these genes have a causal function in lung metastasis, they were overexpressed by retroviral infection in the parental population either individually, in groups of three, or in groups of six (Supplementary Fig. S3). Only cells overexpressing ID1 alone were modestly more active at forming lung metastases than cells infected with vector controls (Fig. 3a). Consistent with the hypothesis that metastasis requires the concerted action of multiple effectors was our observation that combinations of these genes invariably led to more aggressive metastatic activity and that some combinations recapitulated the aggressiveness of the 4175 LM2 population (Fig. 3b). Triple combinations of lung metastasis genes in parental cells did not enhance bone metastatic activity (Supplementary Fig. S4), supporting their identity as tissue-specific mediators of metastasis. The requirement for some of these genes was tested by stably decreasing their expression in 4175 (LM2) cells with short-hairpin RNA-mediated interference (RNAi) vectors (Fig. 3c). A decrease in ID1, VCAM1 or IL13Rα2 levels decreased the lung metastatic activity of 4175 cells more than tenfold (Fig. 3d). These effects were not due to activation of the RNAi machinery, because efficient knockdown of another gene, *ROBO1*, did not inhibit lung metastasis formation (data not shown). Collectively, the results show that these nine genes are not only markers but also functional mediators of lung-specific metastasis.

Lung metastasis signature in primary tumours

A biologically meaningful and clinically relevant gene profile that mediates lung metastasis might be expressed uniquely by a subgroup of patients that suffered relapse to the lung and it should be associated with the clinical outcome. To test this, a cohort of 82 breast cancer patients treated at our institution was used in a univariate Cox proportional hazards model to relate the expression level of each lung metastasis signature gene with clinical outcome. Twelve of the 54 genes are significantly associated with lung-metastasis-free survival, including *MMP1*, *CXCL1* and *PTGS2* (Supplementary Table 5). A cross-validated multivariate analysis using a linear combination of each of the 54 genes weighted by the univariate results[26] distinguished between patients with a high risk and those with a low risk for developing lung metastasis (10-year lung-metastasis-free survival of 56% versus 89%, $P = 0.0018$; see

Figure 3 | Genes in the expression signature mediate lung metastasis. a, **b**, Retrovirus-mediated expression of selected genes from the lung metastasis signature in weakly metastatic parental MDA-MB-231 cells. Genes were tested individually (**a**) or in groups of three or six genes (**b**). **c**, Stable short hairpin (sh) RNAi constructs were introduced retrovirally into 4175 lung-metastatic cells, and their effectiveness at knocking down the expression of their intended target was validated at the protein level (ID1, VCAM1, SPARC) or at the mRNA level (*IL13RA2*). Controls were uninfected 4175 (LM2) cells, and shCont refers to 4175 cells transduced with a non-functional shRNAi. **d**, 4175 knockdown cell lines were xenografted through the tail vein to assess lung metastatic activity. One shRNA vector against ID1 that was ineffective at decreasing expression of this gene served as a negative control. Results are means ± s.e.m. for each cohort. Asterisks, $P < 0.05$ with a one-sided rank test.

Supplementary Fig. S5) but not bone metastasis (70% versus 79%, $P = 0.31$). When a similar multivariate analysis was performed by weighting each gene by a t-statistic derived from a comparison of its expression between the LM2 cell lines with that of the parental MDA-MD-231 cells, the 54 genes again distinguished patients at high risk for developing lung metastasis (62% versus 88%, $P = 0.01$; see Supplementary Fig. S5) but not bone metastasis (75% versus 79%, $P = 0.49$). These results indicate that a clinically relevant subgroup of patients might express certain combinations of lung metastasis signature genes.

To determine directly the extent to which breast cancers express the lung metastasis signature in a manner resembling the LM2 cell lines, the 54 genes were used to cluster the Memorial Sloan-Kettering Cancer Center (MSKCC) data set hierarchically. Manual inspection of branches in the dendogram revealed a group of primary tumours that concordantly expressed many elements of this signature (Fig. 4a, dashed red box). In particular, a subgroup of primary tumours expressed to various degrees most of the nine genes that were functionally validated. Many patients who developed lung metastasis were among this group. Tumours in this group predominantly expressed markers of clinically aggressive disease, including negative

oestrogen receptor (ER)/progesterone receptor status, a Rosetta-type poor-prognosis signature[8], and a basal cell subtype of breast cancer[27]. There was no association of our signature with a high expression of HER2. A molecularly similar subgroup of breast cancer was identified when the clustering analysis was repeated on a previously published Rosetta microarray data set of breast cancer patients[9] (Supplementary Fig. S6), indicating that the findings might not be unique to our cohort of patients.

Although the results of the hierarchical clustering are indicative, this approach can lead to arbitrary class assignments and is generally not ideal for class prediction[28]. We therefore took advantage of the repeated observation of our signature in two independent data sets. For training purposes the Rosetta data set was used to define a group of patients expressing the lung metastasis signature most resembling the LM2 cell lines (Supplementary Fig. S7). All 48 of the 54 lung metastasis genes that were shared between the MSKCC and Rosetta data set microarray platforms were subsequently used to generate a classifier to distinguish these tumours from the remaining tumours in the cohort (Supplementary Table 6). This classifier was then applied to the MSKCC cohort to identify tumours expressing the lung metastasis signature in a manner resembling the LM2 cell
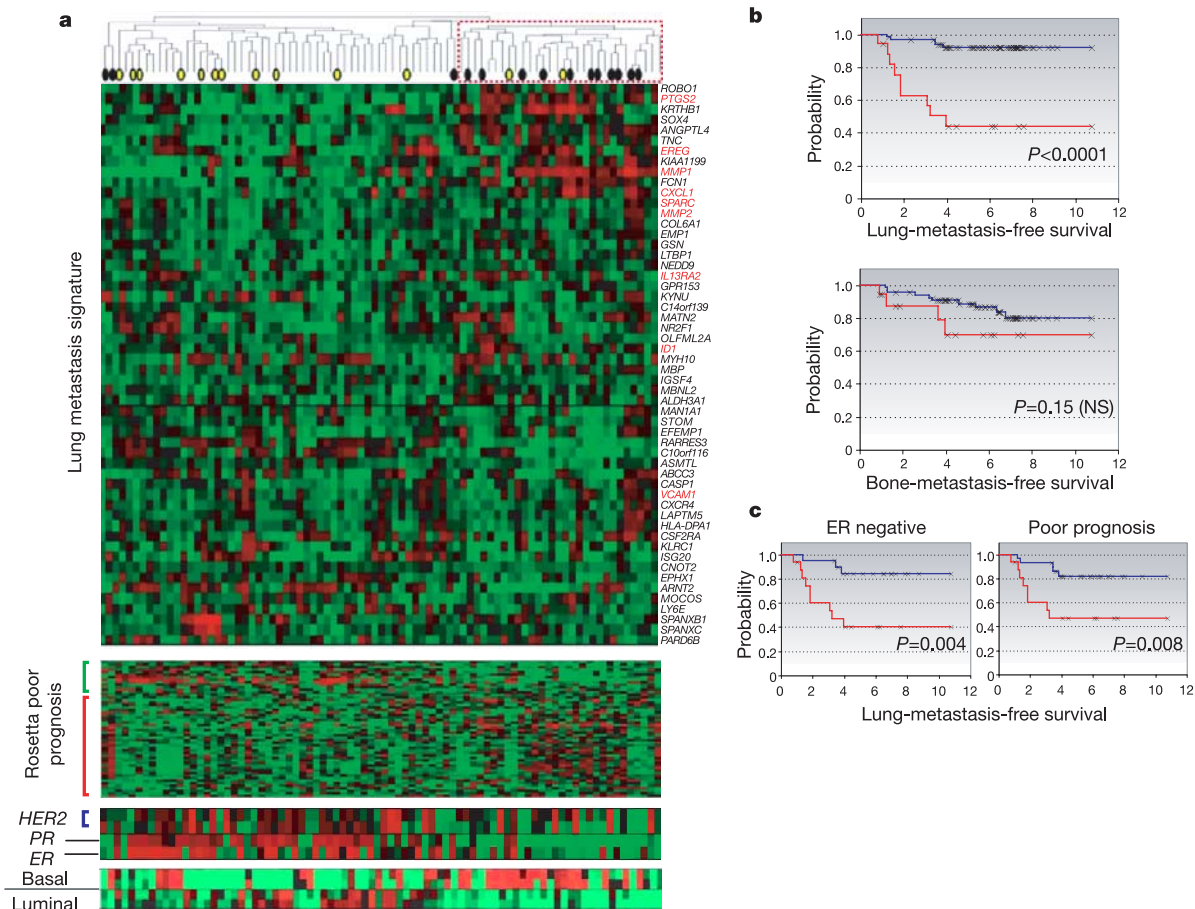
Figure 4 | Lung metastasis signature in human primary breast tumours.
a, Hierarchical clustering of primary breast carcinomas from a cohort of 82 breast cancer patients was performed with the 54 lung metastasis signature genes. A dendrogram of the tumours is shown at the top, with tumours from patients who developed lung metastasis (black circles) or metastasis at non-pulmonary sites (yellow circles) denoted. A subcluster with a reproducibility index of 0.71 (dashed red box) groups tumours that tended to express the lung metastasis signature in a manner resembling the LM2 cell lines. The genes were also clustered; gene names are on the right. Functionally validated genes are in red. The Rosetta poor-prognosis signature is displayed with the genes underexpressed (green bar) and

overexpressed (red bar) in poor-prognosis tumours indicated at the left. The expression of *HER2*, progesterone receptor (*PR*), oestrogen receptor (*ER*) and basal and luminal keratins is shown. Expression of the lung metastasis signature was confirmed in the independent Rosetta breast cancer cohort (Supplementary Fig. S6). **b,** Lung-metastasis-free survival and bone-metastasis-free survival for MSKCC patients who either expressed (red line) or did not express (blue line) the lung metastasis signature based on a classifier trained with the Rosetta cohort (Supplementary Fig. S7 and Supplementary Methods). The *P* value for each survival curve is shown. **c,** Lung-metastasis-free survival restricted to patients with ER-negative tumours or Rosetta-type poor-prognosis tumours.

lines. These patients had a markedly worse lung-metastasis-free survival ($P < 0.001$; Fig. 4b) but not bone-metastasis-free survival ($P = 0.15$; Fig. 4b). These results were independent of ER status and classification as a Rosetta-type poor-prognosis tumour (Fig. 4c). Six of the nine genes that we tested in functional validation studies (*MMP1*, *CXCL1*, *PTGS2*, *ID1*, *VCAM1* and *EREG*) were among the 18 most univariately significant ($P < 0.05$) genes that distinguished the patients used to train the classifier (Table 1 and Supplementary Fig. S7, cluster 3), and classification using only these 18 genes gave similar results (data not shown). The three remaining genes (*SPARC*, *IL13RA2* and *MMP2*) are members of the lung metastasis virulence subset and were expressed only in the most highly metastatic cell lines in our model system (Fig. 2c).

Breast tumorigenicity versus lung metastagenicity

It is unknown how and when metastasis genes are activated[29]. One explanation for the expression of a lung metastasis signature in a subgroup of primary breast cancers is that these genes may confer a growth advantage on the primary tumour while allowing growth at distant sites[7]. To test this hypothesis, MDA-MB-231 cells were injected orthotopically into the mammary fat pad of immuno-deficient mice. We found that the 1834 (LM0) and 4175 (LM2) cell populations were progressively more aggressive at growing in the mammary fat pad than in the parental cell line. This was correlated with expression of lung metastagenicity genes (Figs 2c and 5a) and was not due to a general enhancement of growth because the 4175, 1834 and parental populations had comparable abilities to meta-stasize to bone (refer to Fig. 1b). Furthermore, the 4175 and 1834 populations were also more metastatic to the lungs from the orthotopic site after primary tumour resection, recapitulating the phenotypes observed with the tail vein metastasis assay (Fig. 5b). In contrast, the virulently bone-metastatic population 1833 (ref. 4) was only marginally more aggressive in the mammary fat pad than the parental cells and did not metastasize to lung after primary tumour resection (Fig. 5a, b).

To identify which of the genes in the lung metastasis signature might be conferring growth at the primary tumour site, we quanti-fied mammary-fat-pad tumour growth of 4175 cell populations with stable knockdown of various lung metastasis genes that were pre-viously assayed for effects on metastatic behaviour (refer to Fig. 3c, d). Whereas knockdown of IL13Rα2, SPARC and VCAM1 decreased lung metastatic ability but not orthotopic tumour growth, knock-down of ID1 resulted in a statistically significant reduction in both (Figs 3d and 5c). These data indicate that some lung metastasis genes

might facilitate both breast tumorigenicity and lung metastagenicity, whereas others confer growth advantages exclusively in the lung microenvironment.

Discussion

We have identified a set of genes that mediates breast cancer metastasis to lung and is clinically correlated with the development of lung metastasis when expressed in primary breast cancers. Many of the genes in this signature have not previously been linked to metastasis. Together with bone, the lung is one of the most frequent targets of breast cancer metastasis in humans. We provide evidence that these two sites impose different requirements for the establish-ment of metastases by circulating cancer cells. In addition to providing clinical validation, potential prognostic tools and possible targets for cancer treatment, the present findings shed new light on the biology of breast cancer metastasis.

Many of the genes in the lung metastasis signature are frequently expressed in all MDA-MB-231 subpopulations that metastasize to the lungs, regardless of whether these cells were randomly picked from the parental cell line or selected *in vivo*. Most of these genes, which we denote as promoting lung metastagenicity, encode extra-cellular products including growth and survival factors (for example the HER/ErbB receptor ligand epiregulin), chemokines (CXCL1), cell adhesion receptors (for example ROBO1) and extracellular proteases (MMP1). They also include intracellular enzymes (for example COX2) and transcriptional regulators (for example ID1), as well as several other downregulated genes. Their expression pattern is tightly correlated with lung metastatic activity. When tested by overexpression in poorly metastatic cells or by RNAi-mediated knockdown in highly metastatic cells, several genes in this group function as mediators of lung metastasis but not bone metastasis. Furthermore, in the cohort of human breast cancer primary tumours examined, those expressing the lung metastasis signature had a significantly poorer lung-metastasis-free survival but not bone-metastasis-free survival. This signature therefore seems to include a set of clinically relevant genes that mediate a metastagenicity function[30,31] with selectivity for the lung.

Beside our data, other recent findings reveal the existence of metastasis gene signatures expressed by primary tumours. It is unclear at what point these metastasis gene signatures are acquired during the process of tumorigenesis because the selection pressure for this acquisition is unknown. One possibility is that elements of metastasis gene signatures might have a function in primary tumour growth. Consistent with this idea is the observation that the *in vivo*

Table 1 | Partial list of lung metastasis signature genes used to classify primary breast cancers expressing the lung metastasis signature

UG cluster	Gene symbol	Description	P
Hs.118400	FSCN1	Fascin homologue 1, actin-bundling protein (*Strongylocentrotus purpuratus*)	<0.000001
Hs.83169	MMP1	Matrix metalloproteinase 1 (interstitial collagenase)	<0.000001
Hs.9613	ANGPTL4	Angiopoietin-like 4	<0.000001
Hs.74120	C10orf116	Chromosome 10 open reading frame 116	0.000006
Hs.789	CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth-stimulating activity, alpha)	0.00002
Hs.196384	PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	0.000355
Hs.185568	KRTHB1	Keratin, hair, basic, 1	0.000444
Hs.109225	VCAM1	Vascular cell adhesion molecule 1	0.000506
Hs.17466	RARRES3	Retinoic acid receptor responder (tazarotene induced) 3	0.000627
Hs.368256	LTBP1	Latent transforming growth factor beta binding protein 1	0.001263
Hs.444471	KYNU	Kynureninase (L-kynurenine hydrolase)	0.004365
Hs.421986	CXCR4	Chemokine (C-X-C motif) receptor 4	0.005179
Hs.77667	LY6E	Lymphocyte antigen 6 complex, locus E	0.006426
Hs.410900	ID1	Inhibitor of DNA binding 1, dominant-negative helix–loop–helix protein	0.007153
Hs.255149	MAN1A1	Mannosidase, alpha, class 1A, member 1	0.010871
Hs.388589	NEDD9	Neural precursor cell expressed, developmentally downregulated 9	0.032361
Hs.115263	EREG	Epiregulin	0.03713
Hs.98998	TNC	Tenascin C (hexabrachion)	0.046859

There are 48 unique genes shared between MSKCC and Rosetta microarray platforms. Patients from the Rosetta training set were used to define a class label for patients who either expressed or did not express the lung metastasis signature. Shown is the *P* value of a *t*-test comparing the difference in gene expression between these two classes (Supplementary Fig. S7, cluster 3). Only 18 genes with *P* < 0.05 are shown.
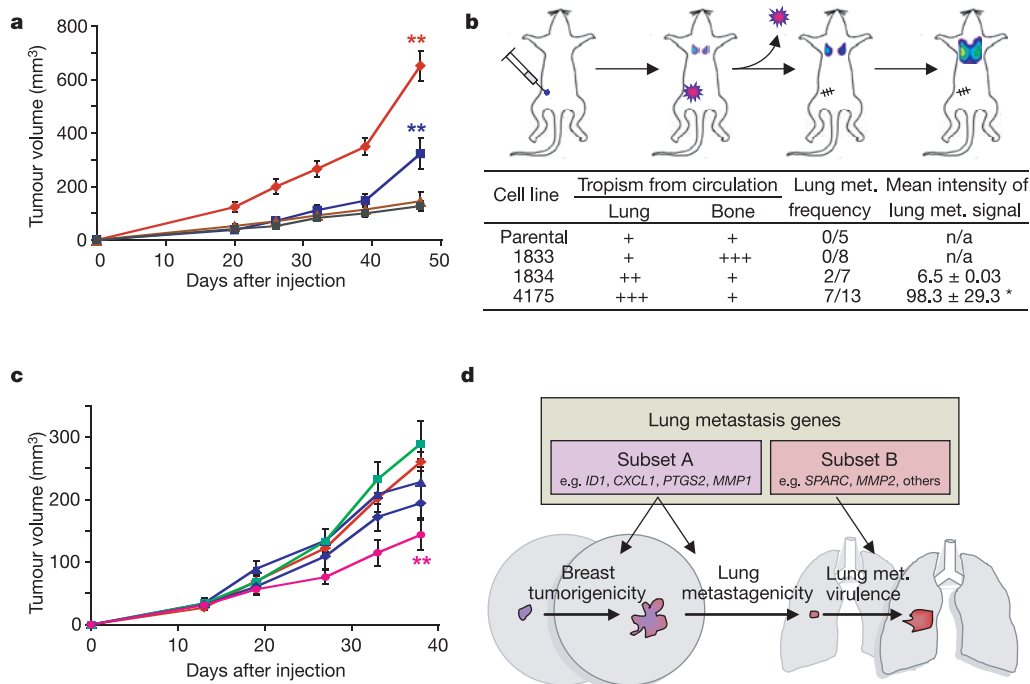
Figure 5 | Breast tumorigenicity and lung metastagenicity partially overlap. a, Representative MDA-MB-231 cell populations were injected into the mammary fat pad of immunodeficient mice and monitored for tumour growth. Red diamonds, 4175 cells ($n = 9$, where n is the number of mice in each cohort); blue squares, 1834 cells ($n = 10$); brown triangles, 1833 cells ($n = 5$); black squares, parental cells ($n = 5$). Each curve shows tumour volumes in cubic millimetres (means ± s.e.m.). **b**, As shown in the diagram, mice were inoculated with the indicated MDA-MB-231 cells into the mammary fat pad and tumours were removed after reaching a volume of 300 mm³. Lung metastasis was monitored with BLI, and normalized photon flux was measured 2 weeks after removal of the primary tumour. Asterisk, a mouse in the 4175 cohort with an unusually high normalized photon flux of 36,400 was excluded. **c**, Growth in mammary fat pad of highly lung-

metastatic 4175 (LM2) cells after stable shRNA knockdown of the following gene products: red diamonds, shControl; blue triangles, shVCAM; green squares, shIL13RA2; blue diamonds, shSPARC; pink circles, shID1. shControl refers to a cell line transduced with a short hairpin construct that did not result in effective knockdown of its target gene. Two asterisks, $P < 0.01$ by a one-sided rank test. Each curve shows tumour volumes in cubic millimetres (means ± s.e.m.). **d**, A model of two classes of genes contained within the lung metastasis signature. The first class (subset A) confers both breast tumorigenicity and basal lung metastagenicity. Examples may include *ID1*, *CXCL1*, *PTGS2* and *MMP1*. The second class (subset B) confers functions specific to the lung microenvironment, facilitating lung metastatic virulence. Examples may include *SPARC* and *MMP2*.

selected cell lines expressing the lung metastagenicity signature are more tumorigenic when implanted in the mammary glands of mice. Despite promoting growth in the mammary gland and in the lung, these genes are not general mediators of neoplastic growth. Many lung metastasis signature genes therefore seem to enhance growth both within the breast and the lung (Fig. 5d). These overlapping functions might explain how cells expressing genes involved in metastasis can be selected for in the primary tumour, providing insight into the interpretation of primary tumour microarray data.

Another subset of the lung metastasis genes is overexpressed only in rare, virulently metastatic cells selected *in vivo*. Several of these genes mediate lung metastasis in our functional assays. Many in this class encode extracellular proteins (for example *SPARC* and *MMP2*). With some exceptions (for example the receptors *IL13RA2* and *VCAM1*) this group of genes is sporadically expressed in human primary breast tumours. We propose that these genes act mainly as virulence genes[30,31] that may allow tumours to aggressively invade, colonize and grow in the lungs without markedly contributing to primary tumour growth (Fig. 5d). Thus, their expression may be rare in primary tumours but strongly selected for once such cells reach the lung. Supporting this model, a recent study analysing *MMP2* expression in matched primary breast cancers and pleural effusions found that *MMP2* levels are specifically enriched at the metastatic site[32].

Breast cancer is a heterogeneous disease with diverse metastatic behaviour. As a consequence, patients differ widely in prognosis and survival. Attempts to classify this disease molecularly have yielded several useful markers of poor prognosis. However, to our knowledge

none of these markers have yet been shown to act as functional mediators that account for the diversity of breast cancer metastases. In contrast, our lung metastasis signature seems to identify poor-prognosis patients who are at high risk of developing lung metastasis, which is consistent with the functional testing done experimentally. Further studies with additional patient cohorts, and a delineation of the role of these genes in specific steps of the metastatic process, should lead to a better understanding of the biology of metastasis and its susceptibilities to treatment.

METHODS
Cell lines. The parental MDA-MB-231 cell line was obtained from the American Type Tissue Collection. Its derivative cell lines and SCPs were described previously[4]. Cells were grown in high-glucose DMEM medium with 10% fetal bovine serum. For bioluminescent tracking, cell lines were retrovirally infected with a triple-fusion protein reporter construct encoding herpes simplex virus thymidine kinase 1, green fluorescent protein (GFP) and firefly luciferase[13,33]. GFP-positive cells were enriched by fluorescence-activated cell sorting.
Animal studies. All animal work was done in accordance with a protocol approved by the Institutional Animal Care and Use Committee. Balb/c nude mice (NCI) 4–6 weeks old were used for all xenografting studies. For lung metastasis formation, 2×10^5 viable cells were washed and harvested in PBS and subsequently injected into the lateral tail vein in a volume of 0.1 ml. Endpoint assays were conducted at 15 weeks after injection unless significant morbidity required that the mouse be euthanized earlier. For bone metastasis, 10^5 cells in PBS were injected into the left ventricle of anaesthetized mice (100 mg kg^{-1} ketamine, 10 mg kg^{-1} xylazine)[4]. Mice were imaged for luciferase activity immediately after injection to exclude any that were not successfully xenografted.

For mammary-fat-pad tumour assays, cells were harvested by trypsinization,

washed twice in PBS and counted. Cells were then resuspended (10^7 cells ml^{-1}) in a 50:50 solution of PBS and Matrigel. Mice were anaesthetized, a small incision was made to reveal the mammary gland and 10^6 cells were injected directly into the mammary fat pad. The incision was closed with wound clips and primary tumour outgrowth was monitored weekly by taking measurements of the tumour length (L) and width (W). Tumour volume was calculated as $\pi LW^2/6$. For metastasis assays, tumours were surgically resected when they reached a volume greater than $300\,\text{mm}^3$. After resection, the mice were monitored by bioluminescent imaging for the development of metastases.

Bioluminescent imaging and analysis. Mice were anaesthetized and injected retro-orbitally with 1.5 mg of D-luciferin (15 mg ml^{-1} in PBS). Imaging was completed between 2 and 5 min after injection with a Xenogen IVIS system coupled to Living Image acquisition and analysis software (Xenogen). For BLI plots, photon flux was calculated for each mouse by using a rectangular region of interest encompassing the thorax of the mouse in a prone position. This value was scaled to a comparable background value (from a luciferin-injected mouse with no tumour cells), and then normalized to the value obtained immediately after xenografting (day 0), so that all mice had an arbitrary starting BLI signal of 100.

RNA isolation, labelling and microarray hybridization. Methods for RNA extraction, labelling and hybridization for DNA microarray analysis of the cell lines have been described previously[4]. For the primary breast tumour data, tissues from primary breast cancers were obtained from therapeutic procedures performed as part of routine clinical management. Samples were snap-frozen in liquid nitrogen and stored at $-80\,°\text{C}$. Each sample was examined histologically in cryostat sections stained with hematoxylin and eosin. Regions were dissected manually from the frozen block to provide a consistent tumour cell content of greater than 70% in tissues used for analysis. All studies were conducted under protocols approved by the MSKCC Institutional Review Board. RNA was extracted from frozen tissues by homogenization in TRIzol reagent (Gibco/BRL) and evaluated for integrity. Complementary DNA was synthesized from total RNA by using a dT primer tagged with a T7 promoter. The RNA target was synthesized by transcription *in vitro* and labelled with biotinylated nucleotides (Enzo Biochem). The labelled target was assessed by hybridization to Test3 arrays (Affymetrix). All gene expression analysis was performed with an HG-U133A GeneChip (Affymetrix). Gene expression was quantified with MAS 5.0 or GCOS (Affymetrix).

Statistical analysis. The Kaplan–Meier method was used to estimate survival curves, and the log-rank test was used to test for differences between curves using WinSTAT (R. Fitch Software). The site of distant metastasis for the patients in the MSKCC data set was determined from patient records. Patients with lung metastasis developed metastasis to the lung only or to the lung within months of metastasis to other sites. A detailed description of analytical methods used in the paper is provided in Supplementary Methods.

Additional procedures. Descriptions of additional experimental procedures used are given in Supplementary Methods.

1. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2**, 563–572 (2002).
2. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).
3. Yokota, J. Tumor progression and metastasis. *Carcinogenesis* **21**, 497–503 (2000).
4. Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
5. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532–535 (2000).
6. Yang, J. et al. Twist, a master regulator of morphogenesis, plays an essential role in tumour metastasis. *Cell* **117**, 927–939 (2004).
7. Bernards, R. & Weinberg, R. A. A progression puzzle. *Nature* **418**, 823 (2002).
8. van de Vijver, M. J. et al. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
9. van 't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
10. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33**, 49–54 (2003).
11. Solomayer, E. F., Diel, I. J., Meyberg, G. C., Gollan, C. & Bastert, G. Metastatic breast cancer: clinical course, prognosis and therapy related to the first site of metastasis. *Breast Cancer Res. Treat.* **59**, 271–278 (2000).
12. Cailleau, R., Olive, M. & Cruciger, Q. V. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* **14**, 911–915 (1978).
13. Minn, A. J. et al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115**, 44–55 (2005).
14. Shelly, M. et al. Epiregulin is a potent pan-ErbB ligand that preferentially activates heterodimeric receptor complexes. *J. Biol. Chem.* **273**, 10496–10505 (1998).
15. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
16. Balkwill, F. Cancer and the chemokine network. *Nature Rev. Cancer* **4**, 540–550 (2004).
17. Egeblad, M. & Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nature Rev. Cancer* **2**, 161–174 (2002).
18. Duffy, M. J., Maguire, T. M., Hill, A., McDermott, E. & O'Higgins, N. Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. *Breast Cancer Res.* **2**, 252–257 (2000).
19. Framson, P. E. & Sage, E. H. SPARC and tumour growth: where the seed meets the soil? *J. Cell. Biochem.* **92**, 679–690 (2004).
20. Wood, N. et al. Enhanced interleukin (IL)-13 responses in mice lacking IL-13 receptor alpha 2. *J. Exp. Med.* **197**, 703–709 (2003).
21. Amatschek, S. et al. Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumour-specific genes. *Cancer Res.* **64**, 844–856 (2004).
22. O'Hanlon, D. M. et al. Soluble adhesion molecules (E-selectin, ICAM-1 and VCAM-1) in breast carcinoma. *Eur. J. Cancer* **38**, 2252–2257 (2002).
23. Desprez, P. Y., Sumida, T. & Coppe, J. P. Helix–loop–helix proteins in mammary gland development and breast cancer. *J. Mammary Gland Biol. Neoplasia* **8**, 225–239 (2003).
24. Ruzinova, M. B. & Benezra, R. Id proteins in development, cell cycle and cancer. *Trends Cell Biol.* **13**, 410–418 (2003).
25. Arun, B. & Goss, P. The role of COX-2 inhibition in breast cancer treatment and prevention. *Semin. Oncol.* **31**, 22–29 (2004).
26. Beer, D. G. et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Med.* **8**, 816–824 (2002).
27. Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
28. Simon, R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. *Br. J. Cancer* **89**, 1599–1604 (2003).
29. Hynes, R. O. Metastatic potential: generic predisposition of the primary tumour or rare, metastatic variants-or both? *Cell* **113**, 821–823 (2003).
30. Heimann, R. & Hellman, S. Clinical progression of breast cancer malignant behaviour: what to expect and when to expect it. *J. Clin. Oncol.* **18**, 591–599 (2000).
31. Schairer, C., Mink, P. J., Carroll, L. & Devesa, S. S. Probabilities of death from breast cancer and other causes among female breast cancer patients. *J. Natl. Cancer Inst.* **96**, 1311–1321 (2004).
32. Davidson, B. et al. Altered expression of metastasis-associated and regulatory molecules in effusions from breast cancer patients: a novel model for tumour progression. *Clin. Cancer Res.* **10**, 7335–7346 (2004).
33. Ponomarev, V. et al. A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* **31**, 740–751 (2004).

The 'solar model problem' solved by the abundance of neon in nearby stars

Jeremy J. Drake[1] & Paola Testa[2]

The interior structure of the Sun can be studied with great accuracy using observations of its oscillations, similar to seismology of the Earth. Precise agreement between helioseismological measurements and predictions of theoretical solar models[1] has been a triumph of modern astrophysics. A recent downward revision by 25–35 per cent of the solar abundances of light elements such as C, N, O and Ne (ref. 2) has, however, broken this accordance: models adopting the new abundances incorrectly predict the depth of the convection zone, the depth profiles of sound speed and density, and the helium abundance[1,3]. The discrepancies are far beyond the uncertainties in either the data or the model predictions[4]. Here we report neon-to-oxygen ratios measured in a sample of nearby solar-like stars, using their X-ray spectra. The abundance ratios are all very similar and substantially larger than the recently revised solar value. The neon abundance in the Sun is quite poorly determined. If the Ne/O abundance in these stars is adopted for the Sun, the models are brought back into agreement with helioseismology measurements[5,6].

The role of the Sun as a fundamental benchmark of stellar evolution theory, which itself underpins much of astrophysics, renders the 'solar model problem' described above one of some importance. The schism between helioseismology and models with a revised composition has arisen because abundant elements such as C, N, O and Ne provide major contributions to the opacity of the solar interior, which in turn influences its internal structure and the depth at which the interior becomes convective. Uncertainties in the calculated opacities themselves appear insufficient to bridge the gap[5,7], and the propriety of the recent abundance revisions has therefore been questioned.

The revised abundances are demanded by new analyses of the visible solar spectrum that take into account convection and associated velocity fields and temperature inhomogeneities through three-dimensional hydrodynamic modelling, and that relax the assumption of local thermodynamic equilibrium for computing atomic level populations[8]. The solar abundances of C, N and O can be measured accurately on the basis of their absorption lines. However, Ne lacks detectable photospheric lines in cool stars like the Sun, and the Ne abundance is therefore much less certain. The solar Ne content is assessed using observations of neon ions in nebular and hot-star spectra, and measurements of solar energetic particles[2,9]. Measurements are generally made relative to a reference element such as O; the downward revision of the solar O abundance therefore required a commensurate lowering of the Ne abundance for consistency[2]. Expressed as the ratio of the number of Ne atoms in the gas to those of O, this abundance is $A_{Ne}/A_{O} = 0.15$, which is very similar to values adopted in earlier studies[10,11]. However, it has recently been pointed out that the solar model problem could be solved were the true solar Ne abundance to be at least a factor of 2.5 times higher than recently assessed[4,5].

We are motivated by the solar model problem to investigate Ne abundances in nearby stars. Although Ne is not detected in the optical spectra of cool stars, emission lines of highly ionized Ne are prominent features of the X-ray spectra of such cool stars[12]. The Ne/O abundance ratio can be derived directly from the ratio of observed fluxes of the hydrogen-like and helium-like ions of O and Ne (see Methods for details); we adopt a slightly refined version of the method applied to the analysis of earlier solar X-ray spectra[13]. In this way, we have obtained the Ne/O abundance ratios for a sample of 21 stars lying within 100 pc of the Sun that have been observed by the Chandra X-ray Observatory using the High Energy Transmission
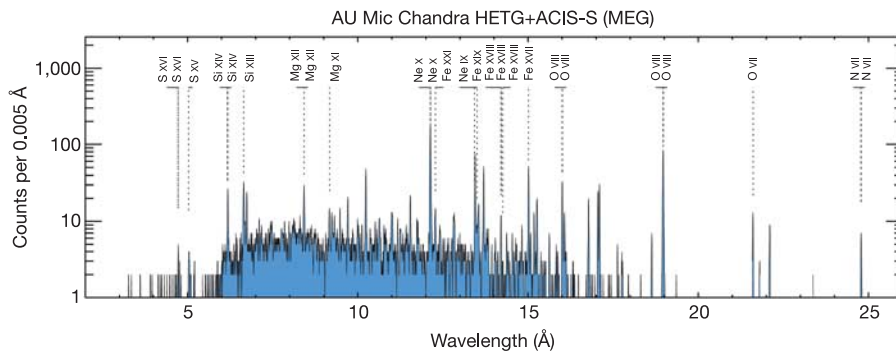
Figure 1 | A Chandra Medium Energy Grating X-ray spectrum of the M1 V star AU Mic. These data are typical of the Chandra MEG spectra upon which this study is based. Emission line features superimposed on the bremsstrahlung continuum are formed predominantly by radiative decay of transitions excited by electron impact. Prominent lines are labelled, including the lines of O VIII $2p\,^2P_{3/2,1/2} \rightarrow 1s\,^2S_{1/2}$ (18.97 Å), Ne IX $1s2p\,^1P_1 \rightarrow 1s^2\,^1S_0$ (13.45 Å) and Ne X $2p\,^2P_{3/2,1/2} \rightarrow 1s\,^2S_{1/2}$ (12.13 Å), used here to determine the Ne/O abundance ratio.

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [2]MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 70 Vassar Street, Cambridge, Massachusetts 02139, USA.
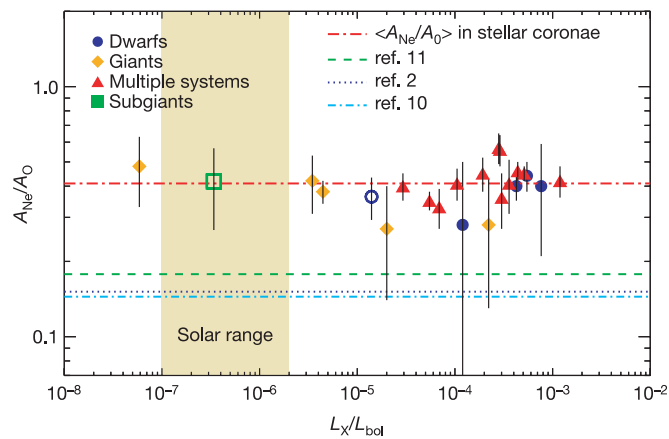
Table 1 | Star spectral line fluxes and derived Ne/O abundance ratios

Star	Type	Distance (pc)	Log (L_X/L_{bol})	$F_{Ne\ IX}$	$F_{Ne\ X}$	$F_{O\ VIII}$	A_{Ne}/A_O
					Line fluxes ($10^{-13}\,erg\,cm^{-2}\,s^{-1}$)		
AU Mic	M1V	9.9	−3.37	3.08 ± 0.20	5.60 ± 0.24	10.4 ± 0.4	0.38 ± 0.05
Prox Cen	M5Ve	1.3	−3.92	0.70 ± 0.14	1.22 ± 0.4	3.2 ± 0.3	0.28 ± 0.22
EV Lac	M1.5V	5.1	−3.11	3.09 ± 0.10	3.99 ± 1.0	9.16 ± 0.20	0.40 ± 0.19
AB Dor	K0V	15	−3.26	4.9 ± 0.4	10.7 ± 0.4	14.6 ± 0.5	0.44 ± 0.06
HR9024	G1III	135	−3.65	0.30 ± 0.11	2.09 ± 0.18	2.40 ± 0.23	0.26 ± 0.15
31 Com	G0III	94	−4.70	0.21 ± 0.05	0.84 ± 0.08	1.30 ± 0.17	0.27 ± 0.13
β Cet	K0III	29.4	−5.35	2.15 ± 0.09	5.30 ± 0.14	8.0 ± 0.3	0.37 ± 0.04
Canopus	F0II	96	−7.23	0.52 ± 0.10	0.76 ± 0.05	1.33 ± 0.25	0.48 ± 0.15
μ Vel	G5III/…	35.5	−5.46	1.40 ± 0.13	1.80 ± 0.24	4.2 ± 0.3	0.40 ± 0.11
Algol	B8V/K1IV	28	−3.5	2.9 ± 0.3	12.8 ± 0.7	13.8 ± 0.5	0.35 ± 0.09
ER Vul	G0V/G5V	50	−3.45	1.10 ± 0.11	3.0 ± 0.19	3.8 ± 0.3	0.41 ± 0.10
44 Boo	G1V/G2V	13	−4.16	3.75 ± 0.29	8.1 ± 0.4	15.2 ± 0.6	0.33 ± 0.06
TZ CrB	G0V/G0V	22	−3.36	7.63 ± 0.30	15.5 ± 0.4	22.1 ± 0.5	0.45 ± 0.04
UX Ari	G5V/K0IV	50	−3.63	3.62 ± 0.27	12.4 ± 0.3	9.7 ± 0.5	0.56 ± 0.08
ξ UMa	G0V/G0V	7.7	−4.26	4.56 ± 0.24	5.41 ± 0.22	15.8 ± 0.6	0.34 ± 0.03
II Peg	K2V/…	42	−2.92	5.26 ± 0.30	19.4 ± 0.5	19.4 ± 0.8	0.42 ± 0.06
λ And	G8III/…	26	−4.53	2.79 ± 0.20	8.82 ± 0.23	10.3 ± 0.4	0.40 ± 0.05
TY Pyx	G5IV/G5IV	56	−3.55	2.09 ± 0.22	4.78 ± 0.09	4.9 ± 0.5	0.57 ± 0.08
AR Lac	G2IV/K0IV	42	−3.72	2.7 ± 0.3	9.70 ± 0.28	9.3 ± 0.5	0.45 ± 0.07
HR1099	G5IV/K1IV	29	−3.29	8.36 ± 0.26	27.6 ± 0.3	27.9 ± 0.5	0.45 ± 0.03
IM Peg	K2III-II/…	97	−3.98	3.00 ± 0.29	15.8 ± 0.3	13.0 ± 0.7	0.41 ± 0.06

Line fluxes were measured from the Medium Energy Grating (MEG) component of Chandra HETG spectra by line profile fitting using the Package for Interactive Analysis of Line Emission (PINTofALE) software[28] (freely available from http://hea-www.harvard.edu/PINTofALE). The effective collecting area of the instrument was accounted for using standard Chandra calibration products and techniques (see http://cxc.harvard.edu/ciao/ for details). Ne/O abundance ratios were obtained assuming the O/Ne line emissivity ratio of $\int \Phi G_O dT / \int \Phi G_{Ne} dT = 1.2 \pm 0.1$, as described in the Methods. Stated flux and abundance ratio uncertainties correspond to 1σ limits.

Grating Spectrometer[14]. A representative X-ray spectrum, that of the M1 V star AU Mic, is presented in Fig. 1. The stars studied, together with observed Ne and O line fluxes and derived Ne/O abundance ratios are listed in Table 1. The Ne/O abundance ratios are illustrated as a function of the ratio of logarithmic X-ray and bolometric luminosities, L_X/L_{bol}—a commonly used index of stellar coronal activity—in Fig. 2. Because our star sample contains more objects toward higher L_X/L_{bol}, we have added Ne/O ratios for two stars of somewhat lower activity level (Procyon, an F5 subgiant, and ε Eri, a K2 dwarf) from the literature[15]. There is no trend in the Ne/O abundance ratio with L_X/L_{bol}, and the error-weighted mean ratio is $A_{Ne}/A_O = 0.41$.
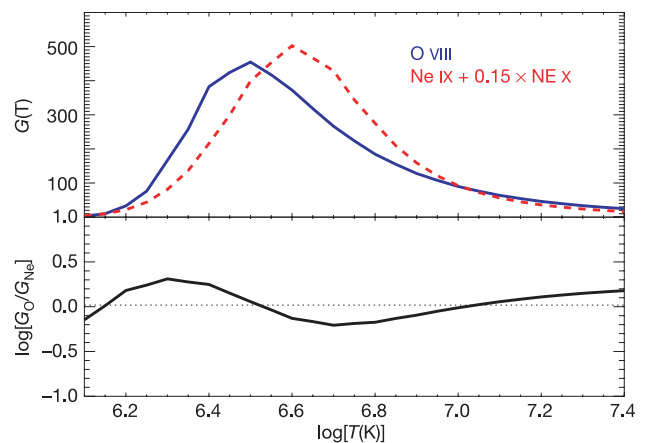
Our derived Ne/O abundance ratio is 2.7 times higher than the

currently recommended solar value[2] but is consistent with the abundance inferred from helioseismology[5,6]. Solving the solar model problem by raising the Ne abundance alone would require a minimum ratio $A_{Ne}/A_O = 0.52$, or 3.44 times larger than recommended[5]. However, raising the C, N, O and Fe abundances upward within their estimated uncertainty range of ±12% (adjusting them all together is not unreasonable because the recent downward revisions are correlated) would require an Ne abundance higher by only a factor of 2.5 (refs 5, 6)—quite within our estimated range.

Extensive review articles on the coronae of the Sun and stars document evidence that coronal chemical compositions often appear different to those thought to characterize the underlying photospheres[9,16–18]. The differences appear to relate to element first ionization potentials (FIPs): the solar corona appears enhanced in elements with low FIP (≤10 eV; such as Mg, Si Fe; refs 9, 16), whereas much

Figure 2 | Derived Ne/O abundance ratios by number, A_{Ne}/A_O, versus the coronal activity index L_X/L_{bol}. Error bars represent quadrature addition of 1σ uncertainties of line flux measurement. Also shown using open symbols are literature values[15] for the stars Procyon (F5 IV) and ε Eri (K2 V) observed using the Chandra Low Energy Transmission Grating Spectrometer (LETGS) to represent better the lower ranges of coronal activity. The error-weighted mean Ne/O abundance ratio is $A_{Ne}/A_O = 0.41$, or 2.7 times the currently assessed value[2], which is illustrated by the dark blue dotted horizontal line. The recommended value from comprehensive earlier assessments in common usage[10,11] are also illustrated.

Figure 3 | The temperature-insensitive O/Ne emissivity ratio. The upper panel illustrates the emissivities G_O, of the O VIII $2p\,^2P_{3/2,1/2} \rightarrow 1s\,^2S_{1/2}$ line, and G_{Ne} of the Ne IX $1s2p\,^1P_1 \rightarrow 1s^2\,^1S_0$ and Ne X $2p\,^2P_{3/2,1/2} \rightarrow 1s\,^2S_{1/2}$ lines combined as $G_{Ne} = G_{Ne\ IX} + 0.15G_{Ne\ X}$. The lower panel shows the logarithmic ratio G_O/G_{Ne}. Emissivities are based on electron excitation rates and ion populations[29] compiled in the CHIANTI database[30], as implemented in PINTofALE[28].

more coronally active stars appear to have depletions in low FIP elements and perhaps enrichments of high FIP elements ($\lesssim 10\,\mathrm{eV}$; such as O, Ne, Ar; refs 16–18). At first sight, this chemical fractionation might render direct interpretation of coronal abundance ratios in terms of the composition of the underlying star problematic. However, that we are seeing the same Ne/O ratio in a wide variety of stars sampling a large range of different coronal activity levels indicates that there is no significant fractionation between O and Ne in disk-integrated light from stellar coronae. We thus conclude that the results represent the true Ne/O abundance ratios of these stars. This conclusion is bolstered by findings of a constant Ne/O abundance ratio, in good agreement with our value, for a small sample of single and active binary stars observed by XMM-Newton[18,19]. In view of the consistency of the Ne/O abundance ratio in nearby stars, it seems likely that the solar ratio should be similar.

There are no recent full-disk integrated light Ne/O measurements for the Sun; existing studies are instead based on observations of particular regions and structures of the solar outer atmosphere. We provide a tabular summary and bibliography of some of the different Ne/O estimates since 1974 in the Supplementary Information. Although most of the estimates appear consistent with current assessments, the underlying observations do not sample photospheric material. The Ne/O ratio is in fact observed to differ substantially between the different observations. In particular, the highest measured Ne/O ratios based on X-ray lines are 2–3 times the accepted solar value and are compatible with the abundance ratio we find for nearby stars. These higher Ne/O ratios were obtained for hotter active regions[13,20] that are likely to dominate the solar full-disk X-ray emission. These measurements are the most directly compatible with the ones presented here based on full-disk integrated light X-ray spectra of stars.

Similarly 'high' Ne/O ratios have also been seen in γ-ray observations of flares, [3]He-rich solar energetic particle events, and in the decay phase of long duration soft X-ray events[21–23]. The γ-ray measurements probe material that is irradiated by downward-flowing accelerated particles[21,22]. The particle beams penetrate through to the chromosphere, which is probably more representative of the underlying photospheric material than coronal regions.

Recent sophisticated models of heliospheric pickup ion and anomalous cosmic ray populations from the local interstellar medium are also inconsistent with the current solar Ne/O abundance ratio, but could be reconciled by raising this to $A_{\mathrm{Ne}}/A_{\mathrm{O}} \approx 0.53$ (ref. 24), which is similar to our values for nearby stars. This Ne/O ratio is somewhat higher than the mean obtained from studies of ionized nebulae in the Milky Way and other galaxies, but falls within the scatter of results at solar metallicity[25].

The implication of our study, then, is that the higher of the observed solar Ne/O abundance ratios are the ones representative of the underlying solar composition. This scenario is in accordance with our observations of Ne/O in nearby stars and reconciles solar models with helioseismology.

METHODS

We use the resonance lines of H-like O and of H-like and He-like Ne to estimate the Ne/O abundance ratio. In hot (10^6–10^7 K) coronal plasma these lines are formed predominantly by radiative de-excitation of levels excited by collisions with thermal electrons. The flux, F_{ji}, from such a transition $j \to i$ in an ion of an element with abundance A can be written as:

$$F_{ji} = A \int_{\Delta T_{ji}} G_{ji}(T)\Phi(T)\mathrm{d}T \ \ (\text{in } \mathrm{erg\,cm^{-2}\,s^{-1}}), \quad (1)$$

where $G_{ji}(T)$ describes the line emissivity—the product of the relative population of the ion in question and the excitation rate of the transition as a function of temperature T. The kernel $\Phi(T)$—the emission measure distribution—describes the excitation power of the plasma as a function of temperature, which is proportional to the mean of the square of the electron density, N_e, and the emitting volume V, such that $\Phi(T) = N_e^2(T)\frac{\mathrm{d}V(T)}{\mathrm{d}T}$.

If we can choose O and Ne lines whose $G_{ji}(T)$ functions have very similar temperature dependence, an abundance ratio by number, $A_{\mathrm{Ne}}/A_{\mathrm{O}}$, can be derived simply from the ratio of their observed line fluxes, F_{O} and F_{Ne}, because all the temperature-dependent terms in equation (1) cancel:

$$\frac{A_{\mathrm{Ne}}}{A_{\mathrm{O}}} = \overline{\left(\frac{G_{\mathrm{O}}}{G_{\mathrm{Ne}}}\right)}\frac{F_{\mathrm{Ne}}}{F_{\mathrm{O}}} \quad (2)$$

An early study of Ne/O ratios in solar active regions[13] used the ratio of Ne IX $1s2p\,^1P_1 \to 1s^2\,^1S_0$ to O VIII $2p\,^2P_{3/2,1/2} \to 1s\,^2S_{1/2}$. This ratio does, however, have some significant residual dependence on temperature[13]. Here we remove much of this temperature dependence by addition of Ne X $2p\,^2P_{3/2,1/2} \to 1s\,^2S_{1/2}$; our combined Ne $G_{ji}(T)$ function is $G_{\mathrm{Ne}} = G_{\mathrm{Ne\ IX}} + 0.15 G_{\mathrm{Ne\ X}}$. The functions G_{O} and G_{Ne} and their ratio are illustrated as a function of temperature in Fig. 3. We have verified the small residual temperature sensitivity evident in the lower panel of Fig. 3 to be negligible for our analysis by integrating the products of G_{O} and G_{Ne} with empirically derived emission measure distributions, $\Phi(T)$, for different stars[12,26], and for functions $\Phi \propto T^a$, with $1 \le a \le 4$: the integrated emissivity ratio from these tests was $\int \Phi G_{\mathrm{O}}\,\mathrm{d}T / \int \Phi G_{\mathrm{Ne}}\,\mathrm{d}T = 1.2 \pm 0.1$. We conclude that the line ratio method is robust and the higher Ne/O abundance ratio found here will not be significantly changed through performing full emission measure distribution modelling.

We measured Ne and O line fluxes (listed in Table 1) from Chandra HETG X-ray spectra obtained directly from the Chandra public data archive (http://cda.harvard.edu). Final listed fluxes for Ne X include small reductions ($\lesssim 15\%$ for 17 out of 21 of the stars we studied, and 25–37% for the remainder) to account for a weak blend of Fe XVII at 12.12 Å. The Fe XVII 12.12 Å contribution was estimated by scaling the observed strengths of unblended Fe XVII lines at 15.26, 16.77, 17.05 and 17.09 Å (the strong resonance line at 15.01 Å was omitted to avoid potential problems with its depletion through resonance scattering) by their theoretical line strengths relative to the 12.12 Å line as predicted by the CHIANTI database. Minor blending in the wings of the Ne IX 13.447 Å line was accounted for by fitting simultaneously with the neighbouring weaker lines, comprised of a Fe XIX–FeXXI blend at 13.424 Å and Fe XIX 13.465 Å, following a detailed study of these features in the Capella binary system[27]. These blend corrections are generally very small, so the uncertainties in these procedures have negligible ($<10\%$) influence on the derived Ne/O abundance ratios.

1. Bahcall, J. N., Basu, S., Pinsonneault, M. & Serenelli, A. M. Helioseismological implications of recent solar abundance determinations. *Astrophys. J.* **618**, 1049–1056 (2005).
2. Asplund, M., Grevesse, N., Sauval, J. The solar chemical composition. Preprint at ⟨http://arXiv.org/abs/astro-ph/0410214⟩ (2004).
3. Basu, S. & Antia, H. M. Constraining solar abundances using helioseismology. *Astrophys. J. Lett.* **606**, L85–L88 (2004).
4. Bahcall, J. N., Serenelli, A. M. & Basu, S. New solar opacities, abundances, helioseismology, and neutrino fluxes. *Astrophys. J. Lett.* **621**, L85–L88 (2005).
5. Antia, H. M. & Basu, S. The discrepancy between solar abundances and helioseismology. *Astrophys. J. Lett.* **620**, L129–L132 (2005).
6. Bahcall, J. N., Basu, S., Serenelli, A. M. What is the neon abundance of the Sun? Preprint at ⟨http://arXiv.org/abs/astro-ph/0502563⟩ (2005).
7. Bahcall, J. N., Serenelli, A. M. & Pinsonneault, M. How accurately can we calculate the depth of the solar convective zone? *Astrophys. J.* **614**, 464–471 (2004).
8. Asplund, M., Nordlund, Å., Trampedach, R. & Stein, R. F. Line formation in solar granulation II. The photospheric Fe abundance. *Astron. Astrophys.* **359**, 743–754 (2000).
9. Meyer, J.-P. The baseline composition of solar energetic particles. *Astrophys. J. Suppl.* **57**, 151–171 (1985).
10. Anders, E. & Grevesse, N. Abundances of the elements—meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
11. Grevesse, N. & Sauval, A. J. Standard solar composition. *Space Sci. Rev.* **85**, 161–174 (1998).
12. Drake, J. J. *et al.* Enhanced noble gases in the coronae of active stars. *Astrophys. J. Lett.* **548**, L81–L85 (2001).
13. Acton, L. W., Catura, R. C. & Joki, E. G. Oxygen-to-neon abundance ratio in the solar corona. *Astrophys. J. Lett.* **195**, L93–L95 (1975).
14. Canizares, C. R. *et al.* High-resolution X-ray spectra of Capella: initial results from the Chandra high-energy transmission grating spectrometer. *Astrophys. J. Lett.* **539**, L41–L44 (2000).
15. Sanz-Forcada, J., Favata, F. & Micela, G. Coronal versus photospheric abundances of stars with different activity levels. *Astron. Astrophys.* **416**, 281–290 (2004).
16. Feldman, U. & Laming, J. M. Element abundances in the upper atmospheres of the Sun and stars: update of observational results. *Phys. Scr.* **61**, 222–252 (2000).
17. Drake, J. J. Chemical fractionation and abundances in coronal plasma. *Adv. Space Res.* **32**, 945–954 (2003).

18. Güdel, M. X-ray astronomy of stellar coronae. *Astron. Astrophys. Rev.* **12**, 71–237 (2004).

19. Audard, M., Güdel, M., Sres, A., Raassen, A. J. J. & Mewe, R. A study of coronal abundances in RS CVn binaries. *Astron. Astrophys.* **398**, 1137–1149 (2003).

20. Schmelz, J. T., Saba, J. L. R., Ghosh, D. & Strong, K. T. Anomalous coronal neon abundances in quiescent solar active regions. *Astrophys. J.* **473**, 519–532 (1996).

21. Murphy, R. J., Ramaty, R., Reames, D. V. & Kozlovsky, B. Solar abundances from gamma-ray spectroscopy—comparisons with energetic particle, photospheric, and coronal abundances. *Astrophys. J.* **371**, 793–803 (1991).

22. Ramaty, R., Mandzhavidze, N., Kozlovsky, B. & Murphy, R. J. Solar atmospheric abundances and energy content in flare accelerated ions from gamma-ray spectroscopy. *Astrophys. J. Lett.* **455**, L193–L196 (1995).

23. Schmelz, J. T. Elemental abundances of flaring solar plasma—Enhanced neon and sulfur. *Astrophys. J.* **408**, 373–381 (1993).

24. Frisch, P. C. & Slavin, J. D. The chemical composition and gas-to-dust mass ratio of nearby interstellar matter. *Astrophys. J.* **594**, 844–858 (2003).

25. Henry, R. B. C. & Worthey, G. The distribution of heavy elements in spiral and elliptical galaxies. *Publ. Astron. Soc. Pacif.* **111**, 919–945 (1999).

26. Garcia-Alvarez, D. *et al.* The FIP effect on late-type stellar coronae: from dwarfs to giants. *Astron. Astrophys. Soc. High Energy Astrophysics Division* **8**, 10.03 (2004).

27. Ness, J., Brickhouse, N. S., Drake, J. J. & Huenemoerder, D. P. Modeling the Ne IX triplet spectral region of Capella with the Chandra and XMM-Newton gratings. *Astrophys. J.* **598**, 1277–1289 (2003).

28. Kashyap, V. & Drake, J. J. PINTofALE: Package for the interactive analysis of line emission. *Bull. Astron. Soc. India* **28**, 475–476 (2000).

29. Mazzotta, P., Mazzitelli, G., Colafrancesco, S. & Vittorio, N. Ionization balance for optically thin plasmas: rate coefficients for all atoms and ions of the elements H to NI. *Astron. Astrophys. Suppl.* **133**, 403–409 (1998).

30. Young, P. R. *et al.* CHIANTI-An atomic database for emission lines. VI. proton rates and other improvements. *Astrophys. J. Suppl.* **144**, 135–152 (2003).

LETTERS

Crossover from 'mesoscopic' to 'universal' phase for electron transmission in quantum dots

M. Avinun-Kalish[1], M. Heiblum[1], O. Zarchin[1], D. Mahalu[1] & V. Umansky[1]

The measurement of phase in coherent electron systems—that is, 'mesoscopic' systems such as quantum dots—can yield information about fundamental transport properties that is not readily apparent from conductance measurements. Phase measurements on relatively large quantum dots[1] recently revealed that the phase evolution for electrons traversing the dots exhibits a 'universal' behaviour, independent of dot size, shape, and electron occupancy[2,3]. Specifically, for quantum dots in the Coulomb blockade regime, the transmission phase increases monotonically by π throughout each conductance peak; in the conductance valleys, the phase returns sharply to its starting value. The expected mesoscopic features in the phase evolution—related to the dot's shape, spin degeneracy or to exchange effects—have not been observed, and there is at present no satisfactory explanation for the observed universality in phase behaviour[4]. Here we report the results of phase measurements on a series of small quantum dots, having occupancies of between only 1–20 electrons, where the phase behaviour for electron transmission should in principle be easier to interpret. In contrast to the universal behaviour observed thus far only in the larger dots, we see clear mesoscopic features in the phase measurements when the dot occupancy is less than ~ 10 electrons. As the occupancy increases, the manner of phase evolution changes and universal behaviour is recovered for some 14 electrons or more. The identification of a transition from the expected mesoscopic behaviour to universal phase evolution should help to direct and constrain theoretical models for the latter.

The experimental configuration is shown in Fig. 1. The interferometer, formed in a two-dimensional electron gas (a GaAs-AlGaAs heterostructure), consists of emitter E and collector C constrictions, each formed by a quantum point contact (QPC), and a few base regions B in between. The grounded base regions (chemical potential $\mu_B = 0$) serve as draining reservoirs, ensuring that interference is only between two paths[5]. A centre island separates the incoming electrons into two paths and embedded within it is the plunger gate of the quantum dot, which controls the occupancy of the dot. The transmission phase of the quantum dot adds to the phase gain in the left arm and is then compared with the phase of the right arm. Threading a weak magnetic field through the area of the interferometer adds an Aharonov–Bohm (AB) phase to one arm, $\Delta\varphi_{AB} = 2\pi\,\Phi/\Phi_0$, where Φ is the magnetic flux enclosed by the two arms and $\Phi_0 = h/e$ is the flux quantum[6,7]. A QPC is coupled capacitively to the quantum dot, hence sensing its potential—enabling counting of the electrons in the dot[8,9].

A simplified model of the quantum dot is a resonant tunnelling device: a potential well confined between two tunnelling barriers with quasi-bound resonant states, E_n. The transmission amplitude exhibits maxima whenever the Fermi energy in the leads, E_F, coincides with E_n. Owing to the finite coupling between well and leads, the levels are broadened to Γ_n (dwell time $\sim h/\Gamma_n$), with the

transmission through the quantum dot, t_{QD}, described by the Breit–Wigner expression[10] $t_{QD} = C_n(i\Gamma/2)/(E_F - E_n + i\Gamma/2)$. Here, the C_ns are positive or negative real numbers (since at a zero magnetic field the hamiltonian is real), depending on the parity of the nth wavefunction with respect to the 'in' and 'out' openings of the quantum dot. The phase of the transmission amplitude evolves by
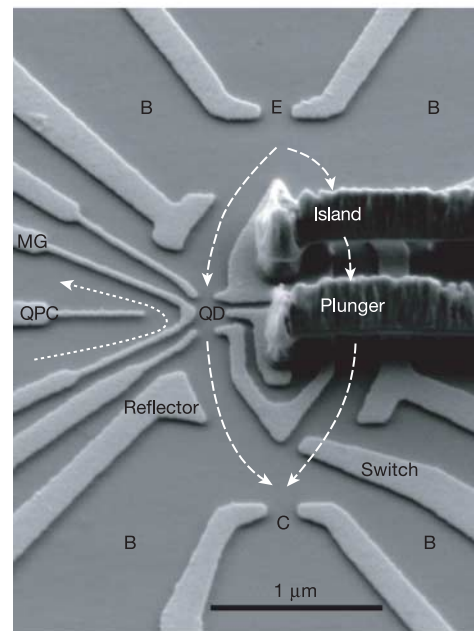
Figure 1 | SEM micrograph of the device. The two-path interferometer consists of a patterned high mobility two-dimensional electron gas formed ~ 53 nm below the surface of a GaAs-AlGaAs heterostructure, with areal carrier density $n_s = 3.3 \times 10^{11}$ cm^{-2} and mobility $\mu = 1.6 \times 10^6$ cm^2 V^{-1} s^{-1} at $T = 4.2$ K, resulting in an elastic mean free path $l = 14.5\,\mu$m. To ensure only two interfering paths from emitter E to collector C, the reflected and scattered paths are collected by the base B regions through openings formed between the gates of the reflectors and the quantum point contacts (QPCs). The 'switch' gate turns off the right path to allow tuning of the quantum dot (QD). The quantum dot is composed of two QPCs and a middle gate (MG) in between, allowing formation of 'small' and 'large' dots. The 'plunger' gate is embedded in the centre island, and is connected via a metallic air bridge to the outside to allow biasing without crossing the right (reference) arm of the interferometer. Similarly, another bridge biases the centre island. The magnetic flux is contained in the area between the two paths (shown by the dashed lines). The QPC counting detector, separated from the quantum dot by MG, has its own current path (shown by the dotted line). The transmission from E to C oscillates with the AB flux with phase dependent on the transmission phase of the quantum dot.

[1]Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

π through each resonance, while the relative phase between adjacent resonances, determined by C_n, can be 0 or π (ref. 11). However, in contrast, all previous measurements, conducted with relatively large occupancy quantum dots, consistently led only to positive C_ns (refs 2, 3, and our unpublished data).

A large number of theoretical publications have been devoted to the puzzle of phase evolution (see recent summary[4]). They may be grouped in three main classes. The first class questions whether the measured phase is the 'intrinsic transmission phase' of the quantum dot or a modified phase due to multiple paths traversing the interferometer[12,13]. The second class considers transport that is mediated by interplay of more than one quantum state. A common scenario assumes an existence of a dominant level strongly coupled to the leads, responsible for shuttling the electrons[14,15]. After occupation the electron is unloaded to a localized level, weakly coupled to the leads, allowing the dominant level to be free again to transfer another electron. Hence, the observed phase is only that of the dominant level. Based on this idea other models examined only two levels, with one of the levels dominant, adding spins, adding interactions, or assuming a finite temperature[16–18]. Interaction between two levels was invoked also in a quantum dot where the plunger gate couples

with different strengths to different energy states, leading thus to avoided level crossing and charge shuttling between levels[4]. The third class deals with specific energies where both the imaginary and the real parts of the transmission coefficient vanish[19]. These singular points, which explain the phase slips in the valleys, might result from a deviation from a strictly zero-dimensional system[4] or from the existence of Fano resonances in the dot[20]; but they can not explain the 'in phase' behaviour of all peaks. Naturally, one would expect the breakdown of every model for some tuning parameters, which we, thus far, have never observed. Still, some models may predict an 'in phase' behaviour for a very large sequence of peaks but not universal behaviour. We return to this issue later.

At low enough temperatures, both the phase coherence length and the elastic mean free path exceed the sample size. The current in a grounded collector is $I_C = (2e^2/h)T_{EC}V_{EB}$, where V_{EB} is the voltage applied between emitter and base, and T_{EC} is the transmission probability from emitter to collector[21]. The transmission T_{EC} results from a coherent sum of the two trajectory amplitudes that traverse the two arms of the interferometer, $|t_{EC}|^2 = T_{EC}$ and $t_{EC} = t_L + t_R$, where $t_{L,R}$ is the transmission amplitude associated with the left and right paths, respectively. The phase dependent part is $\tilde{T}_{EC} \propto |t_L||t_R|\cos[\Delta\varphi_{AB} + \varphi(t_R) - \varphi(t_L)]$ with φ ($t_{L,R}$) the corresponding phases, and $\varphi(t_L) = \varphi_0(t_L) + \varphi_{QD}$ the accumulated phase in the left arm. Hence, $\tilde{T}_{EC}$ oscillates as a function of magnetic flux with period Φ_0, with any change in φ_{QD} leading to a similar change in the phase of the oscillating collector current.

Measurements were done in a dilution refrigerator ($T_{lattice} \approx 20$ mK, $T_{electrons} \approx 30$ mK) with an a.c. excitation voltage $V_{EB} = 2$–$20\,\mu V$
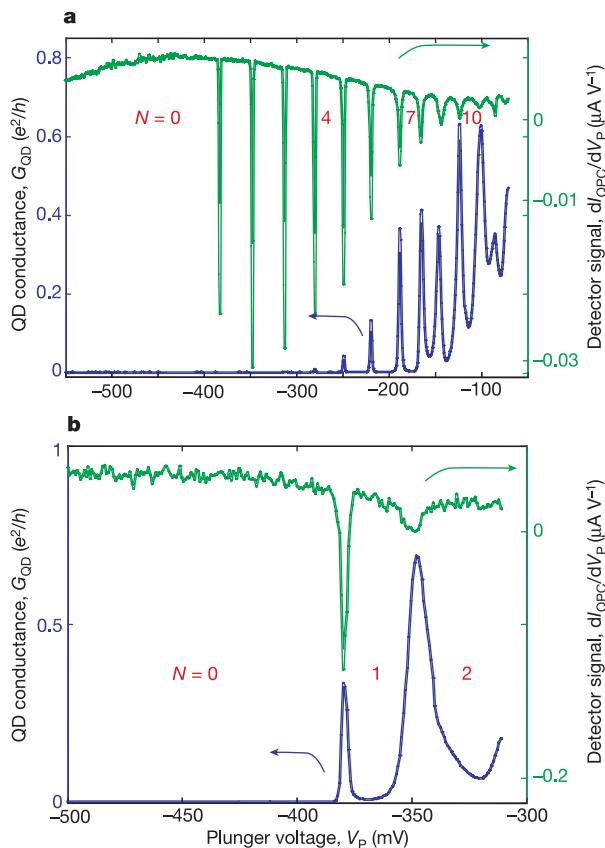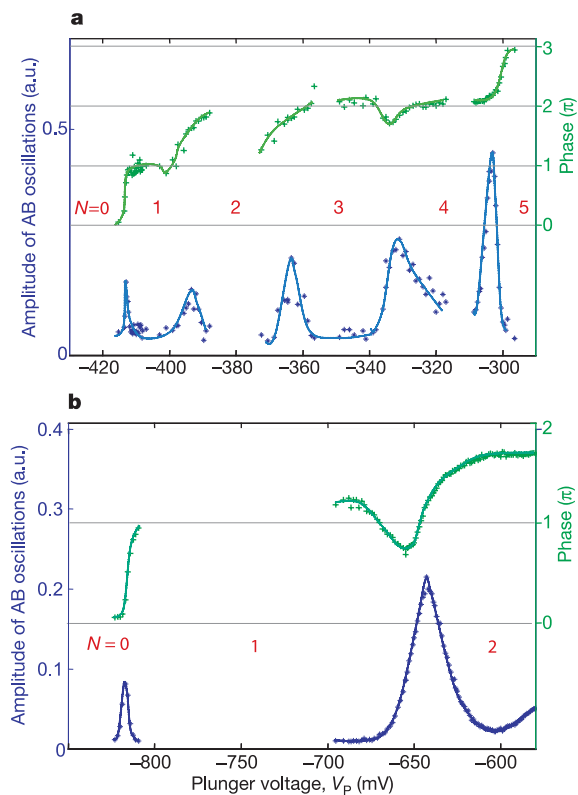

Figure 2 | Electron counting in the quantum dot by a QPC detector. In the Coulomb blockade regime we expect V_{QD} to increase as the plunger gate voltage V_P increases, eventually reaching $V_{QD} = (e/2C_{QD}) + \Delta E/e$, with C_{QD} the self capacitance of the dot and $e/2C_{QD}$ the charging energy of the dot. At this potential, an electron enters the quantum dot and screens the positive potential induced by the plunger, resulting in a potential drop. Hence, the quantum dot potential, and consequently the conductance of the QPC counting detector, exhibit a saw-tooth-like behaviour. The derivative dI_{QPC}/dV_P is a series of negative dips. **a**, Conductance peaks of the quantum dot (blue) and the corresponding detector dips (green). The detector proves the presence of electrons, even though the quantum dot conductance is too small to be measured. **b**, The quantum dot is retuned by opening the 'in' and 'out' QPCs, and changing the voltage on MG in order to allow measurable conductance of the first few electrons.
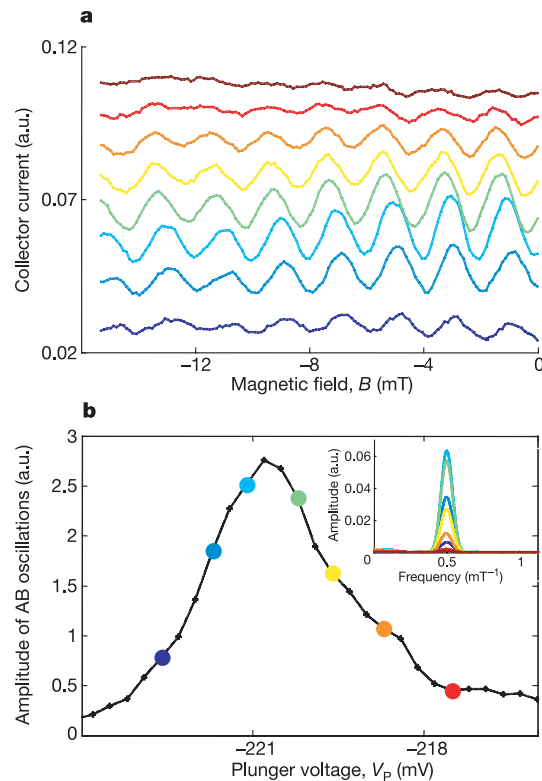
Figure 3 | Phase measurement procedure. a, The quantum dot is first tuned to conduct, and the AB oscillations (with V_P as a parameter) are monitored throughout a conductance peak at the chosen values shown in **b**. Note the change in amplitude and in phase among the different traces (that are shifted vertically for clarity). **b**, The amplitude of the oscillations plotted as a function of V_P (the coloured points correspond to the coloured traces in **a**). The amplitude was determined by a fast Fourier transform. Inset, Fourier transforms of the oscillations indicating a single AB period corresponding to an addition of a flux quantum h/e to the area enclosed by the two paths.

(frequency $\sim 23\,Hz$). The integrity of the 'two-path' interferometer was verified by observing a single period $\Delta\Phi = \Phi_0 = h/e$ in the interference signal (see below), indicating only two-path interference. Higher orders, with period Φ_0/n, were at least four orders of magnitude smaller. The quantum dot was formed by adjusting the resistance of its 'in' and 'out' QPCs to be greater than $h/2e^2$, namely, in the Coulomb blockade regime. To ensure transport mainly through one level we tuned the dot to $\Gamma < \Delta E$, with $\Delta E \approx 0.5\,meV$ and $\Gamma = 30$–$300\,\mu eV$, with the temperature being the smallest energy, $k_B T < 3\,\mu eV$. In the 'constant interaction' model, the complex interaction among the electrons is represented by a capacitor C_{QD}, leading to a charging energy $U = e/2C_{QD} \approx 1$–$3\,meV$. Varying the plunger gate voltage V_P changes the potential landscape in the quantum dot and consequently the occupation, N. Classically, for a certain V_P degeneracy takes place, $E(N) = E(N+1)$; allowing the number of electrons to fluctuate between N and $N+1$ with no energy cost, allowing current flow.

The electron counting detector is a separately biased QPC, capacitively coupled to the quantum dot[8,9]. The induced potential in the QPC is $V_{QPC} = V_{QD}\,C_{QPC-QD}/C_{QPC}$, where C_{QPC} is the self capacitance of the QPC and C_{QPC-QD} is the mutual capacitance between quantum dot and QPC detector. Charging the quantum dot affects V_{QPC} and the conductance of the QPC, which in turn is analysed by a small current (20–80 nA). The potential energy of the quantum dot rises linearly with plunger voltage, reaching eventually $\Delta E + e^2/2C_{QD}$, followed by a sharp drop when an electron enters the dot. Indeed, the conductance of the QPC detector exhibited repetitive saw-tooth-like oscillations as a function of V_P, with one period for every additional electron entering the dot. We measured the derivative dI_{QPC}/dV_P (via applying a small a.c. voltage to the plunger gate), resulting in easily identifiable dips in the derivative (see Fig. 2). Note that determining the occupancy of the quantum dot via a separate detector is necessary since for sufficiently negative V_P the quantum dot inadvertently pinches off (via the mutual coupling among the different gates). Consequently, the dot's conductance peaks weaken and are impossible to resolve. As seen in Fig. 2a, the conductance dips of the detector persisted down to $V_P \approx -400\,mV$, much beyond $V_P \approx -300\,mV$ where the conductance peaks of the quantum dot were too small to resolve. We carefully retuned the quantum dot without changing its occupancy, in order to maximize the conductance peaks, allowing reliable phase measurements (for example, Fig. 2b).

Varying the magnetic field in a range 0–30 mT, after the quantum dot was tuned to conduct, resulted in relatively high visibility AB oscillations with a single period $\Delta B = \Phi_0/area \approx 2\,mT$ (see Fig. 3; in a different structure $\Delta B \approx 3.5\,mT$). The coherent part of the transmission (Fig. 3b) and the transmission phase (Figs 4–6) were both determined from data such as shown in Fig. 3a by performing a fast complex Fourier transform of the AB oscillations as a function of V_P.

We studied two different configurations of interferometers and quantum dots, as well as thermally recycled structures (which behave as different devices after thermal recycling). A single tuning allowed the addition of only 2–3 electrons without changing drastically the coupling of the dot to the leads or the symmetry of the interferometer. Hence, the quantum dot and the interferometer were
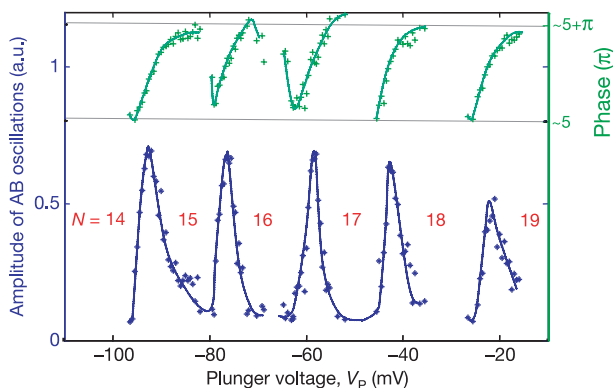
Figure 4 | **Phase evolution and coherent conductance for the first few electrons in the quantum dot.** The dotted lines are a guide to the eye. **a**, Typical behaviour of the 'amplitude of AB oscillations' and the phase for the first five electrons ($N = 1\ldots5$). Whereas the first two electrons enter different orbital states, the following two share the same orbital parity with the second electron. **b**, Detailed behaviour of the phase for the first two electrons in a different device. Note the two-orbital singlet, which is robust for all tuning parameters. Before the entry of the second electron, the phase always exhibits a drop—as large as $\pi/2$—independent of how much the quantum dot is being pinched. It may be related to an onset of Kondo correlation.

Figure 5 | **Two examples of phase evolution for $N = 5$–8 with different tuning parameters of the quantum dot and the interferometer.** This is to demonstrate that the features are mesoscopic, that is, sensitive to the dot's configuration. Whereas the fifth and sixth electrons remain in similar parity states for the two tunings, the seventh and eighth electrons are in different parity orbital states in **a** and in similar parity orbital states in **b**.

retuned after every few added electrons, keeping the occupation checked with the QPC detector, in order to optimize the visibility and CB conditions. The measured phase in different occupation regimes was then patched together in order to obtain a continuous phase evolution over a wide range of electron occupation. We present in Figs 4–6 examples of phase and amplitude of the coherent part of the transmission coefficient for an increasing electron number in the quantum dot. We did not subtract any extraneous phase that may be induced in the arms of the interferometer by the varied plunger voltage since this phase is difficult to determine accurately. However, we estimated it to be a weak function of V_P and quite small for the addition of 2–3 electrons at a time.

The phase evolution across the first two conductance peaks and valleys (the first two electrons entering the dot) already exhibited a marked deviation from universal behaviour. As demonstrated in the results of two different dot and interferometer designs, the phase climbed by π for each of the first two added electrons (Fig. 4). Moreover, this dependence was robust and independent of a dot's tunings. Evidently, the different phase of the first two conductance peaks suggests that the first two electrons occupy two opposite parity orbital states, and not one state as assumed thus far[22–24]. Since the ground state of a two-electron system must be a singlet[25], two opposite spins occupy the lowest two states. This is not surprising, since our dots are likely to have a very shallow potential well, and hence a relatively small single particle level spacing, favouring an occupation of two levels in order to minimize the Coulomb repulsion[26]. We will not speculate here on the reproducible dip in the phase before the entering of the second electron, which is as large as $\pi/2$ (Fig. 4b). Adding the third and fourth electrons (Fig. 4a), the phase evolves from π to 2π—a similar range to the second electron. Consequently, these data suggest that the second to the fourth electrons all occupy similar parity orbital levels. Similarly, the fifth electron evolves from $\sim 2\pi$ to $\sim 3\pi$, which is indistinguishable from a phase evolving from 0 to π of the first electron.

Figure 5a and b shows the sensitivity of the phase to the tuning parameters. While the phase in the fifth, sixth and seventh conductance peaks is independent of tuning parameters, the phase of the eighth electron depends on the tuning parameters. For somewhat different parameters of the quantum dot, the eighth electron has a different parity from that of the seventh electron. The above described examples (Figs 4 and 5) clearly demonstrate a reasonable phase behaviour, sensitive to details of the potential, by showing spin degeneracy, exchange interaction, or Cooper channel interaction[26].

This is distinctly different from the repetitive, universal-like, phase rigidity observed in many electron dots[2,3].

With the addition of more electrons the phase evolves through a transition region, which resembles universal behaviour (not shown). The phase climbs throughout each conductance peak and drops in the valleys, riding though on a rising background phase. However, for dot occupation of some 14 electrons and higher (Fig. 6), the phase 'locks' into the universal behaviour—being then insensitive to the dot's tuning. In other words, the intricacies of shape-dependent level occupation and parity effects disappear altogether.

Since our aim was to explore the validity of the bizarre phase evolution over a large range of parameters, we may now ask: what have we learnt from this new set of measurements? First, that there are two distinct regimes of phase evolution. (1) For occupation 1 to about 10, the phase evolution is highly sensitive to the dot's configuration and occupation. (2) For occupation higher than about 14, phase evolution is universal-like, with phase ranging only between 0 and π, and is independent of the dot's configuration and occupation. Whereas regime (1) can be explained by current understanding of quantum dots, regime (2) presents difficulties, especially since screening is expected to be more effective at higher occupations (possibly leading to single-particle-like behaviour). However, the absence of phase universality in regime (1) is gratifying for two main reasons: first, it confirms that our measured phase is indeed the intrinsic phase of the quantum dot. Otherwise, we would not observe such distinct differences between dilute and highly populated dots. Second, we can now comment on some aspects of the existing theories. An outstanding feature of larger quantum dots is the smaller level spacing, which might lead to levels overlapping. This will favour models that invoke transmission mediated through more than one quantum state.

An illuminating example can be that of a dominant orbital state strongly coupled to the leads. Such a level can be simply a solution of Schrödinger's equation[15], or alternatively, the higher of any two sequential levels. The higher level is broader, since it experiences a lower potential barrier to the leads (Y. Oreg, personal communication). Consequently, a broader level is likely to be occupied, and hence responsible for electron shuttling and the similar phase of many peaks. Still, a theory has to be developed that explains the robustness of the effect for an extremely large number of conductance peaks.

1. Van Houten, H., Beenakker, C. W. J. & Staring, A. A. W. in *Single Charge Tunnelling - Coulomb Blockade Phenomena in Nanostructures* (eds Grabert, H. & Devoret, M. H.) 167–216 (Plenum, New York, 1992).
2. Schuster, R. *et al.* Phase measurement in a quantum dot via a double slit interference experiment. *Nature* **385**, 417–420 (1997).
3. Ji, Y., Heiblum, M., Sprinzak, D., Mahalu, D. & Shtrikman, H. Phase evolution in a Kondo-correlated system. *Science* **290**, 779–783 (2000).
4. Hackenbroich, G. Phase coherent transmission through interacting mesoscopic systems. *Phys. Rep.* **343**, 463–538 (2001).
5. Yacoby, A., Schuster, R. & Heiblum, M. Phase rigidity and *h/2e* oscillations in a single-ring Aharonov-Bohm experiment. *Phys. Rev. B* **53**, 9583–9586 (1996).
6. Aronov, A. G. & Sharvin, Yu. V. Magnetic flux effects in disordered conductors. *Rev. Mod. Phys.* **59**, 755–779 (1987).
7. Aharonov, Y. & Bohm, D. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491 (1959).
8. Field, M. *et al.* Measurements of Coulomb blockade with a noninvasive voltage probe. *Phys. Rev. Lett.* **70**, 1311–1314 (1993).
9. Sprinzak, D., Ji, Y., Heiblum, M., Mahalu, D. & Shtrikman, H. Charge distribution in a Kondo correlated quantum dot. *Phys. Rev. Lett.* **88**, 176805 (2002).
10. Breit, G. & Wigner, E. Capture of slow neutrons. *Phys. Rev.* **49**, 519–531 (1936).
11. Hackenbroich, G. & Weidenmüller, H. A. Transmission through a quantum dot in an Aharonov-Bohm ring. *Phys. Rev. Lett.* **76**, 110–113 (1996).
12. Weidenmüller, H. A. Transmission phase of an isolated CB resonance. *Phys. Rev. B* **65**, 245322 (2002).
13. Aharony, A., Entin-Wohlman, O., Halperin, B. I. & Imry, Y. Phase measurement in the mesoscopic AB interferometer. *Phys. Rev. B* **66**, 115311 (2002).
14. Levy Yeyati, A. & Büttiker, M. Scattering phases in quantum dots: An analysis based on lattice models. *Phys. Rev. B* **62**, 7307–7315 (2000).

Figure 6 | The quantum dot undergoes universal phase evolution after fourteen electrons have entered. This behaviour is independent of tuning and is ubiquitous to all measured many-electron dots. Note that the absolute value of phase (with respect to the phase of the first electron) is difficult to determine accurately owing to an accumulated phase in the leads or a distortion of the interferometer, so only its approximate value is noted: however, the phase span is always π.

15. Silvestrov, P. G. & Imry, Y. Towards an explanation of the mesoscopic double-slit experiment: A new model for charging of a quantum dot. *Phys. Rev. Lett.* **85,** 2565–2568 (2000).

16. Silvestrov, P. G. & Imry, Y. Spin effects and transport in QD with overlapping resonances. *Phys. Rev. B* **65,** 035309 (2002).

17. Silva, A., Oreg, Y. & Gefen, Y. The signs of QD lead matrix elements: The effect on transport vs. spectral properties. *Phys. Rev. B* **66,** 195316 (2002).

18. Oreg, Y. & Gefen, Y. Electron scattering through a quantum dot: A phase lapse mechanism. *Phys. Rev. B* **55,** 13726–13729 (1997).

19. Lee, H. W. Generic transmission zeros and in-phase resonances in time-reversal symmetric single channel transport. *Phys. Rev. Lett.* **82,** 2358–2361 (1999).

20. Entin-Wohlman, O., Aharony, A., Imry, Y. & Levinson, Y. The Fano effect in AB interferometers. *J. Low Temp. Phys.* **126,** 1251–1273 (2002).

21. Büttiker, M. 4-Terminal phase-coherent conductance. *Phys. Rev. Lett.* **57,** 1761–1764 (1986).

22. Kouwenhoven, L. P., Oosterkamp, T. H., Tarucha, S., Austing, D. G. & Honda, H. Coulomb oscillations in few-electron quantum dot. *Physica B* **249–251,** 191–196 (1998).

23. Zumbuhl, D. M., Marcus, C. M., Hanson, M. P. & Gossard, A. C. Voltage-tunable singlet-triplet transition in lateral quantum dots. *Phys. Rev. B* **66,** 035320 (2002).

24. Kyriakidis, J., Pioro-Ladriere, M., Ciorga, M., Sachrajda, A. S. & Hawrylak, P. Cotunneling spectroscopy in few electron QDs. *Phys. Rev. Lett.* **93,** 256801 (2004).

25. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* Ch. 32 (Holt, Rinehart and Winston, Orlando, 1976).

26. Oreg, Y., Brouwer, P. W., Waintal, X. & Halperin, B. I. in *Nano-Physics and Bio-Electronics* (eds Chakraborty, T., Peeters, F. & Sivan, U.) Ch. 5 (Elsevier, Amsterdam, 2002).

LETTERS

Unconventional critical behaviour in a quasi-two-dimensional organic conductor

F. Kagawa[1], K. Miyagawa[1,2] & K. Kanoda[1,2]

Changing the interactions between particles in an ensemble—by varying the temperature or pressure, for example—can lead to phase transitions whose critical behaviour depends on the collective nature of the many-body system. Despite the diversity of ingredients, which include atoms, molecules, electrons and their spins, the collective behaviour can be grouped into several families (called 'universality classes') represented by canonical spin models[1]. One kind of transition, the Mott transition[2], occurs when the repulsive Coulomb interaction between electrons is increased, causing wave-like electrons to behave as particles. In two dimensions, the attractive behaviour responsible for the superconductivity in high-transition temperature copper oxide[3,4] and organic[5–7] compounds appears near the Mott transition, but the universality class to which two-dimensional, repulsive electronic systems belongs remains unknown. Here we present an observation of the critical phenomena at the pressure-induced Mott transition in a quasi-two-dimensional organic conductor using conductance measurements as a probe. We find that the Mott transition in two dimensions is not consistent with known universality classes, as the observed collective behaviour has previously not been seen. This peculiarity must be involved in any emergent behaviour near the Mott transition in two dimensions.

Using the pressure-sweep technique to vary the effective electronic interactions in $(V_{1-x}Cr_x)_2O_3$, Limelette et al. successfully revealed the critical phenomena of the Mott transition in three dimensions (3D; ref. 8). The critical exponents are consistent with those found using mean-field theory. For the study of the Mott transition in two dimensions (2D), organic conductors, which are low-dimensional and highly compressible, are suitable. The organic family with a half-filled band[9], κ-(BEDT-TTF)$_2$X, are good candidates, where BEDT-TTF is bis(ethylenedithio)tetrathiafulvalene and X stands for various kinds of anions. The κ-(BEDT-TTF)$_2$X family has a quasi-2D layered structure (Fig. 1a) and shows similar behaviour to high-transition temperature (high-T_c) copper oxides[10]. In the conceptual phase diagram of the family[11] (Fig. 1b), unconventional superconductivity appears in the marginal metallic phase near the antiferromagnetic insulator, like high-T_c copper oxides[3,4], although the Mott transition is induced by pressure in the organics but by carrier doping in the copper oxides. Furthermore, the exotic pseudogap behaviour is emergent near the Mott transition in both the organics[12] and copper oxides[3,4]. These similarities suggest that they share common physics—the Mott transition in 2D correlated electrons.

The Mott insulator κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl (denoted by
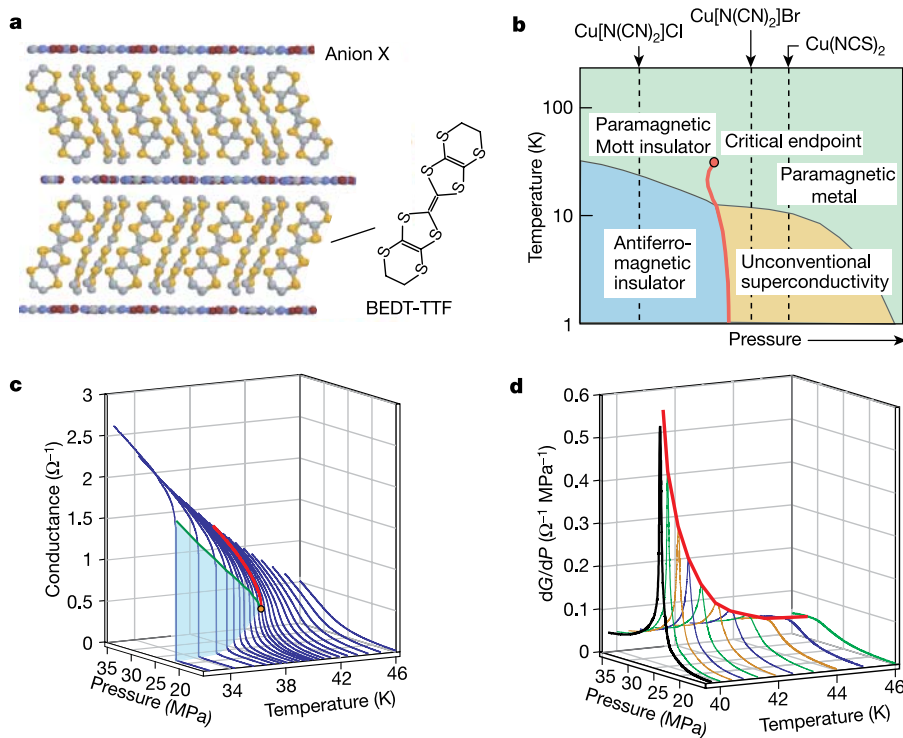
Figure 1 | Crystal structure and conductance profile around the critical endpoint of a Mott transition. a, Crystal structure of κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl. 2D conducting layers of BEDT-TTF dimers are separated by insulating anion layers, like high-T_c copper oxides. The dimer has one hole, and thus the band is half-filled. **b**, Generic pressure–temperature phase diagram of κ-(BEDT-TTF)$_2$X, of which three members (X = Cu[N(CN)$_2$]Cl, Cu[N(CN)$_2$]Br and Cu(NCS)$_2$) are indicated along the pressure axis. The red line represents the first-order transition, which terminates at a finite-temperature critical endpoint (filled circle). **c**, Pressure dependence of conductance $G_T(P)$ around the critical endpoint (filled circle). The shaded area indicates the conductance jump. The red and green curves represent the critical behaviour at $T = T_c \approx 39.7$ K and $T < T_c$, which give the critical exponents δ and β, respectively (see Fig. 2b and c). The hysteresis of the conductance jump (for example, ~0.2 MPa at ~32 K) is not appreciable at this scale. **d**, Pressure derivative of conductance $G'_T(P) \equiv \partial G_T(P)/\partial P$ as a function of pressure at temperatures above $T_c \approx 39.7$ K. The black curve shows the data at $T = T_c$. The red curve represents the critical divergence of the pressure derivative, which gives the critical exponent γ (see Fig. 2d).

[1]Department of Applied Physics, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. [2]CREST, Japan Science and Technology Corporation, Kawaguchi 332-0012, Japan.

κ-Cl hereafter) undergoes an insulator to metal transition when pressure is applied and thus allows us to approach the critical endpoint of the first-order Mott transition[13–16] (Fig. 1b). This feature provides the foundation of our research because the critical phenomena, which are direct clues to identifying the universality class, are generally found around the endpoint, in analogy to the liquid–gas transition. In the Mott transition, the conductance in the metallic region was found to describe the critical phenomena very well by Limelette et al.[8]. To characterize the Mott critical phenomena of κ-Cl in detail, we measured the in-plane conductance G around the endpoint with the standard d.c. four-probe method under isothermal pressure sweep ($\sim$3 MPa h^{-1} in ascending and descending processes) using helium gas as a pressure medium. Throughout the experiment, the conductance was independent of the applied current within two orders of magnitude. The experiments were performed on three single crystals of κ-Cl, which were grown by the conventional electrochemical oxidation of BEDT-TTF[17].

Figure 1c shows the conductance profile as a function of temperature T and pressure P. Below the critical temperature T_c, the conductance jump (shaded area) is seen at well-defined pressures, where the first-order metal–insulator transition occurs. Above T_c, however, the pressure dependence of conductance $G_T(P)$ at a given temperature is continuous, corresponding to the metal–insulator crossover. The characteristic 'crossover pressure' $P_{cross}(T)$ is defined as that giving the peak in the pressure derivative of conductance, $G'_T(P) \equiv \partial G_T(P)/\partial P$ (Fig. 1d). The first-order transition line and the crossover line are represented in the pressure–temperature phase diagram (Fig. 2a).

Around the critical endpoint, we found three regions of critical behaviour in the conductance, which we call regions (1)–(3). Two of

them are visible in Fig. 1c: in region (1) at $T = T_c$, the $G_{T=Tc}(P)$ curve is continuous but has a vertical gradient at the critical pressure P_c (see the red curve in Fig. 1c); and in region (2) below T_c, the conductance jump continuously vanishes as $T \rightarrow T_c$ from below (see the green curve in Fig. 1c). Region (3) of critical behaviour is found in Fig. 1d, where the peak value in the $G'_T(P)$ curve grows continuously and diverges as $T \rightarrow T_c$ from above (see the red curve in Fig. 1d). The fixed parameters of the critical endpoint are $T_c \approx 39.7 \pm 0.1$ K and $P_c \approx 24.8 \pm 0.1$ MPa, which are sample-independent.

Generally, the critical phenomena show power-law behaviour and are characterized quantitatively by the exponents of the power laws (called 'critical exponents'). As phase transitions belonging to the same universality class are characterized by the same set of critical exponents, we can identify the universality class from the values of the exponents. Below we present the results of the κ-Cl crystal investigated in the greatest detail. As seen in Fig. 2b–d (see also Methods), the conductance behaviour of regions (1)–(3) follows power laws around the critical endpoint. From the slopes in the logarithmic plots, we determined that the critical exponents (δ, β, γ) are nearly equal to (2, 1, 1). In the data analysis, we avoided the region very close to the endpoint, $|T - T_c|/T_c <\sim 2\times 10^{-2}$ and $P - P_c < \sim 0.2$ MPa, because the analysis in this region is easily affected by ambiguities in the determination of T_c and $G_c \equiv G_{T=Tc}(P_c)$, the pressure accuracy (± 0.05 MPa) and the sample quality. The reproducibility of the power-law behaviour was ensured for two other crystals ($\sim$1.9 $< \delta < \sim$2, $\sim$0.9 $< \beta < \sim$1 and $\sim$0.9 $< \gamma < \sim$1) over the wide pressure–temperature range except in the close vicinity of the endpoint, although the prefactor of each power law (normalized by the value of G_c) showed some sample dependence ($\pm 30\%$).

It is surprising that the observed exponents, $(\delta, \beta, \gamma) \approx (2, 1, 1)$, are not only previously unknown but also far from any existing values (Fig. 3). The widely known universality classes have in common $\delta \geq 3$ and $\beta \leq 0.5$ (note that $(\delta, \beta) = (3, 0.5)$ are the mean-field values). This is a natural consequence of collective (non-local) fluctuation effects, which generally cause more rapid ordering than the mean-field scheme dealing with the single-site (local) problem. In the present case, however, the deviation of δ and β from the mean-field values is in the opposite direction (that is $\delta < 3$ and $\beta > 0.5$) to the conventional cases, that is, the ordering is more moderate than the mean-field scheme. This feature is obviously anomalous, suggesting the discovery of novel critical behaviour out of the conventional framework.
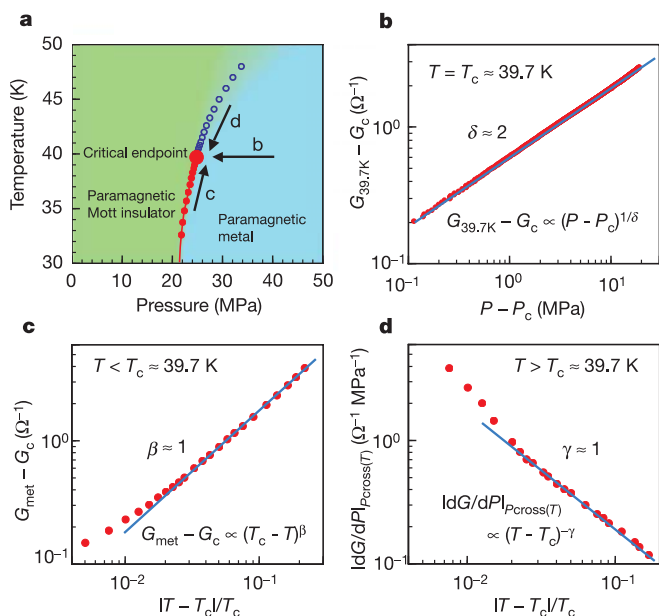
Figure 2 | Critical exponents of the Mott transition in κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl. a, Pressure–temperature phase diagram of κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl around the critical endpoint. Filled and open circles represent the first-order Mott transition points where the conductance jumps and the crossover points ($P_{cross}(T),T$) defined in the text, respectively. Three arrows stand for the direction along which we evaluated the critical exponents δ, β and γ, as shown in b–d, respectively. b–d, Power-law fittings (blue lines) of the critical phenomena of conductance (see Fig. 1c and d) on logarithmic scales, where the critical values (G_c, P_c, T_c) $\approx$ (0.77 Ω^{-1}, 24.8 MPa, 39.7 K) are used for the analysis (Supplementary Fig. 1). The fittings give the critical exponents, $(\delta, \beta, \gamma) \approx (2, 1, 1)$.
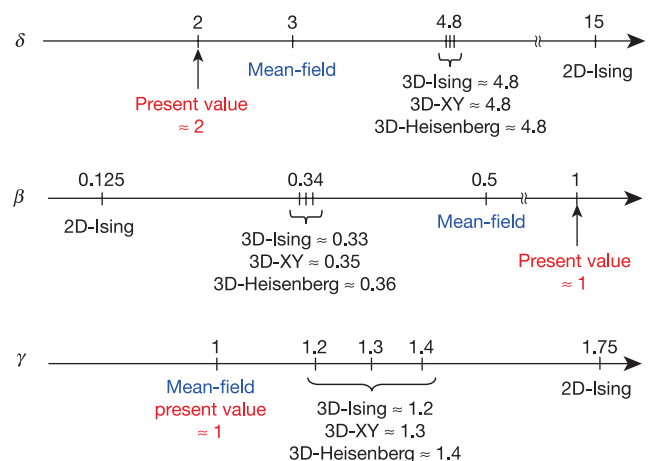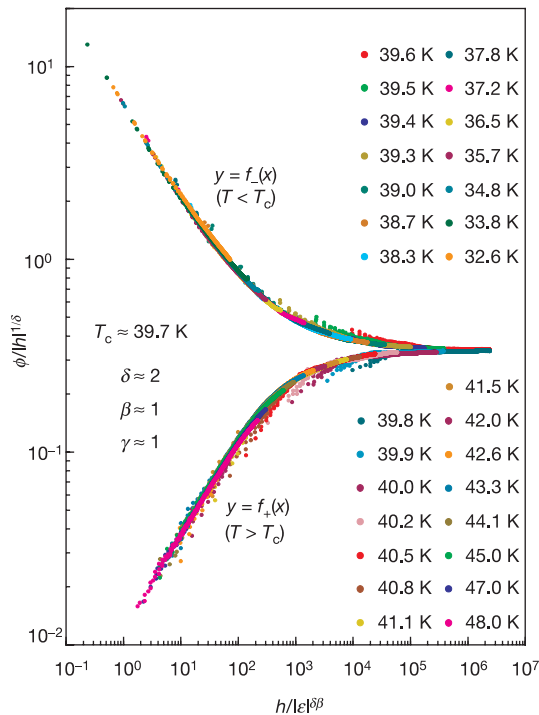

Figure 3 | Comparison of the critical exponents (δ, β, γ) of the present case with those of the known universality classes (mean-field, Ising model, XY model and Heisenberg model). The examples of phase transitions belonging to each universality class are tabulated in Supplementary Table 1.

Below we examine whether the present values correctly correspond to critical behaviour, using two methods based on a scaling hypothesis[1]. One is the scaling relation $\delta = 1 + (\gamma/\beta)$, which should be fulfilled by any critical phenomenon. It is evident that the present values, $(\delta, \beta, \gamma) \approx (2, 1, 1)$, satisfy the relation. The other prerequisite is the existence of a universal form of the equation of state[1,18], which in the present case is expected to follow (see Methods):

$$\{G_T(P) - G_T(P_{\mathrm{cross}}(T))\}/\{P - P_{\mathrm{cross}}(T)\}^{1/\delta}$$

$$= f_+\left(\frac{P - P_{\mathrm{cross}}(T)}{|T - T_c|^{\delta\beta}}\right) \text{ for } T > T_c, \quad (1)$$

$$\{G_T(P) - G_c\}/\{P - P_I(T)\}^{1/\delta} = f_-\left(\frac{P - P_I(T)}{|T - T_c|^{\delta\beta}}\right) \text{ for } T < T_c, \quad (2)$$

where $P_I(T)$ represents the pressure of the conductance jump in the ascending-pressure process. Using the present value, $(\delta, \beta) \approx (2, 1)$, and the whole data set of $G_T(P)$ in the metallic region $(P > P_I(T)$ for $T < T_c$ and $P > P_{\mathrm{cross}}(T)$ for $T > T_c$), we plotted $\{G_T(P) - G_T(P_{\mathrm{cross}}(T))\}/\{P - P_{\mathrm{cross}}(T)\}^{1/\delta}$ versus $\{P - P_{\mathrm{cross}}(T)\}/|T - T_c|^{\delta\beta}$ for $T > T_c$ and $\{G_T(P) - G_c\}/\{P - P_I(T)\}^{1/\delta}$ versus $\{P - P_I(T)\}/|T - T_c|^{\delta\beta}$ for $T < T_c$. The resulting plot reasonably leads to the two scaling curves over a wide range (Fig. 4). The consistency between Figs 2 and 4 justifies the present exponents and the above analysis of the data (see also Supplementary Discussion). It is also confirmed that the scaling plot with $(\delta, \beta) \approx (2, 1)$ is well reproduced for two other crystals. Therefore we conclude that the observed exponents, $(\delta, \beta, \gamma) \approx (2, 1, 1)$, characterize correctly the critical behaviour of the (pressure-induced) Mott transition in a quasi-2D system; they are inconsistent with known universality classes and suggest novel critical behaviour. As seen above, the conductance describes the Mott criticality in κ-Cl well, as in $(V_{1-x}Cr_x)_2O_3$ (ref. 8), in that it shows the critical behaviour and

satisfies the scaling hypothesis. This suggests that conductance is an essential quantity characterizing the Mott criticality.

The present results mean that 2D correlated electrons form a many-body system with anomalous collective behaviour, which cannot be understood from known spin models. Generally, the universality class depends on the system dimensionality, order-parameter dimensionality, and interaction range. We note that the Mott transition in the 3D system $(V_{1-x}Cr_x)_2O_3$ was reported[8] to be equivalent to the predicted conventional Ising transition in 3D[19,20]. Thus the two-dimensionality of κ-Cl apparently seems to be responsible for the unconventionality ($\delta < 3$ and $\beta > 0.5$). However, at least within the Ising scheme, two-dimensionality would give merely the conventional 2D-Ising (or mean-field) values. Therefore some new framework beyond the conventional ones should be invoked to explain the unconventional exponents, and it will be a challenging future problem. We note that recent theoretical work highlighting the quantum nature of electrons predicted Mott criticality with unconventional exponents in 2D (refs 21, 22).

Our conductance results suggest that novel criticality—perhaps belonging to a new universality class—is inherent in quasi-2D, giving rise to exotic properties such as unconventional superconductivity and pseudogap behaviour. This raises a fundamental problem of how this peculiarity of the system background is involved in the emergence of the macroscopic condensate near the Mott transition.

METHODS

Order parameter and scaling variables of the Mott transition. Following the scaling hypothesis[1], we introduce three variables for the analysis of the critical phenomena: order parameter, ϕ, and two scaling variables, h and ε, where h is the conjugate field of ϕ and ε represents the distance to the critical endpoint. In the scaling framework, the critical exponents (δ, β, γ) around the endpoint, $(h, \varepsilon) = (0, 0)$ in the h–ε plane, are defined as $\phi(h, 0) \propto h^{1/\delta}$, $\phi(0, \varepsilon) \propto |\varepsilon|^\beta$ and $\partial\phi(h, \varepsilon)/\partial h|_{h=0} \propto |\varepsilon|^{-\gamma}$. Note that these critical exponents are not independent of each other, but the scaling relation $\delta = 1 + (\gamma/\beta)$ should be fulfilled among them. To analyse the Mott critical phenomena, one should relate (ϕ, h, ε) of the Mott transition to the observable quantities. According to the theoretical framework of strongly correlated electrons, dynamical mean-field theory (DMFT)[18,23,24], the order parameter of the Mott transition has a scalar nature[18,20,24] and can be measured by the conductance in the metallic region, $G–G_c$ (refs 18, 24, 25). This has been demonstrated by an experiment on a 3D oxide[8]. In the (pressure-induced) Mott transition, the experimental variables are pressure and temperature. However, the scaling valuables h and ε do not generally have direct correspondence to pressure and temperature, respectively, but are approximated well by a linear combination of $(T - T_c)$ and $(P - P_c)$ near the endpoint (linear-mixing approximation, see Supplementary Discussion): $h \approx (P - P_c)$ with some admixture of $(T - T_c)$ and $\varepsilon \approx (T - T_c)$ with some admixture of $(P - P_c)$. Following the approach adopted in the case of the liquid–gas transitions[1] or the Mott transition in a 3D system[8], we assume tentatively that the admixtures are negligibly small (non-mixing approximation). Thus we evaluated the critical exponents (δ, β, γ) by taking $(G–G_c, P - P_c, T - T_c)$ for (ϕ, h, ε), that is, the exponents are determined from the power-law behaviour of conductance (1)–(3) (see main text) as follows: (1) $G_{T=Tc}(P) - G_c \propto (P - P_c)^{1/\delta}$, (2) $G_{\mathrm{met}}(T) - G_c \propto (T_c - T)^\beta$ (along the first-order transition line), and (3) $\partial G_T(P)/\partial P|_{P=P_{\mathrm{cross}}(T)} \propto (T - T_c)^{-\gamma}$ (along the crossover line), where $G_{\mathrm{met}}(T)$ is the conductance value of the metallic regime just after the conductance jump in the ascending-pressure process. As seen in the text, the Mott criticality is well described by using this non-mixing approximation. In the Supplementary Discussion, we prove that the non-mixing approximation does not influence the present evaluation of the exponents.

Universal form of equation of state. Around the critical endpoint, the order parameter ϕ should follow a universal form of the equation of state including the critical exponents: $\phi(h, \varepsilon)/h^{1/\delta} = f_\pm(h/|\varepsilon|^{\delta\beta})$, where $f_+(x)$ for $\varepsilon > 0$ and $f_-(x)$ for $\varepsilon < 0$ are called 'scaling functions' depending on the details of the system. Within the non-mixing approximation, $\varepsilon \approx T - T_c$. Since $h = 0$ on the crossover line and the first-order transition line, h for $T \neq T_c$ is given as follows: $h \approx (P - P_{\mathrm{cross}}(T))$ for $T > T_c$ and $h \approx (P - P_I(T))$ for $T < T_c$, where $P_I(T)$ represents the pressure of the conductance jump in the ascending-pressure process. Above T_c, ϕ should be zero at $h = 0$. Thus ϕ for $T > T_c$ is given as $\phi \approx (G_T(P) - G_T(P_{\mathrm{cross}}(T)))$. As a result, the universal form of equation of state is approximated as equations (1) and (2) (see main text). In fact, the

Figure 4 | Scaling plot of the whole data set of $G_T(P)$ in the metallic region. The plot of $\phi(h, \varepsilon)/h^{1/\delta}$ versus $h/|\varepsilon|^{\delta\beta}$ with the present values, $(\delta, \beta) \approx (2, 1)$, falls into two scaling curves corresponding to the scaling function $f_\pm(x)$ over a wide range. $[\phi, h, \varepsilon]$ in the labels of the axes corresponds to $[G_T(P) - G_T(P_{\mathrm{cross}}(T)), P - P_{\mathrm{cross}}(T), |T - T_c|]$ for $T > T_c$ and $[G_T(P) - G_c, P - P_I(T), |T - T_c|]$ for $T < T_c$ (see Methods).

536

successful scaling behaviour shown in Fig. 4 assures us that the non-mixing approximation is valid in the present case.

1. Kadanoff, L. P. et al. Static phenomena near critical points: Theory and experiment. Rev. Mod. Phys. 39, 395–431 (1967).
2. Mott, N. F. Metal-Insulator Transitions (Taylor & Francis, London, 1990).
3. Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. Rev. Mod. Phys. 70, 1039–1263 (1998).
4. Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. Rev. Mod. Phys. 75, 473–541 (2003).
5. Mayaffre, H., Wzietek, P., Jérome, D., Lenoir, C. & Batail, P. Superconducting state of κ-(ET)$_2$CU[N(CN)$_2$]Br studied by ^{13}C NMR: Evidence for vortex-core-induced nuclear relaxation and unconventional pairing. Phys. Rev. Lett. 75, 4122–4125 (1995).
6. De Soto, S. M. et al. ^{13}C NMR studies of the normal and superconducting states of the organic superconductor κ-(ET)$_2$Cu[N(CN)$_2$]Br. Phys. Rev. B 52, 10364–10368 (1995).
7. Kanoda, K., Miyagawa, K., Kawamoto, A. & Nakazawa, Y. NMR relaxation rate in the superconducting state of the organic conductor κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Br. Phys. Rev. B 54, 76–79 (1996).
8. Limelette, P. et al. Universality and critical behavior at the Mott transition. Science 302, 89–92 (2003).
9. Kino, H. & Fukuyama, H. Phase diagram of two-dimensional organic conductors: (BEDT-TTF) $_2$X. J. Phys. Soc. Jpn 65, 2158–2169 (1996).
10. Mckenzie, R. H. Similarities between organic and cuprate superconductors. Science 278, 820–821 (1997).
11. Kanoda, K. Electron correlation, metal-insulator transition and superconductivity in quasi-2D organic systems, (ET)$_2$X. Physica C 282, 299–302 (1997).
12. Miyagawa, K., Kawamoto, A. & Kanoda, K. Proximity of pseudogapped superconductor and commensurate antiferromagnet in a quasi-two-dimensional organic system. Phys. Rev. Lett. 89, 017003 (2002).
13. Lefebvre, S. et al. Mott transition, antiferromagnetism, and unconventional superconductivity in layered organic superconductors. Phys. Rev. Lett. 85, 5420–5423 (2000).
14. Fournier, D., Poirier, M., Castonguay, M. & Truong, K. Mott transition, compressibility divergence, and the P-T phase diagram of layered organic superconductors: An ultrasonic investigation. Phys. Rev. Lett. 90, 127002 (2003).
15. Limelette, P. et al. Mott transition and transport crossovers in the organic compound κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl. Phys. Rev. Lett. 91, 016401 (2003).
16. Kagawa, F., Itou, T., Miyagawa, K. & Kanoda, K. Transport criticality of the first-order Mott transition in the quasi-two-dimensional organic conductor κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl. Phys. Rev. B 69, 064511 (2004).
17. Williams, J. M. et al. From semiconductor-semiconductor transition (42 K) to the highest-T_C organic superconductor, κ-(ET)$_2$Cu[N(CN)$_2$]Cl (T_C = 12.5 K). Inorg. Chem. 29, 3272–3274 (1990).
18. Georges, A. Strongly correlated electron materials: Dynamical mean-field theory and electronic structure. AIP Conf. Proc. 715, 3–74 (2004).
19. Castellani, C., DiCastro, C., Feinberg, D. & Ranninger, J. New model Hamiltonian for the metal-insulator transition. Phys. Rev. Lett. 43, 1957–1960 (1979).
20. Kotliar, G., Lange, E. & Rozenberg, M. J. Landau theory of the finite temperature Mott transition. Phys. Rev. Lett. 84, 5180–5183 (2000).
21. Imada, M. Quantum Mott transition and multi-furcating criticality. J. Phys. Soc. Jpn 73, 1851–1863 (2004).
22. Imada, M. Quantum Mott transition and superconductivity. J. Phys. Soc. Jpn 74, 859–862 (2005).
23. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev. Mod. Phys. 68, 13–125 (1996).
24. Kotliar, G. & Vollhardt, D. Strongly correlated materials: Insights from dynamical mean-field theory. Phys. Today 57(3), 53–59 (2004).
25. Onoda, S. Ginzburg-Landau theory and classical critical phenomena of Mott transition. Preprint at ⟨http://xxx.lanl.gov/abs/cond-mat/0408207⟩ 2004.

LETTERS

# Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers

Emilio Mayorga[1]*, Anthony K. Aufdenkampe[2]*, Caroline A. Masiello[3], Alex V. Krusche[4], John I. Hedges[1]‡,
Paul D. Quay[1], Jeffrey E. Richey[1] & Thomas A. Brown[5]

Rivers are generally supersaturated with respect to carbon dioxide, resulting in large gas evasion fluxes that can be a significant component of regional net carbon budgets[1,2]. Amazonian rivers were recently shown to outgas more than ten times the amount of carbon exported to the ocean in the form of total organic carbon or dissolved inorganic carbon[1]. High carbon dioxide concentrations in rivers originate largely from *in situ* respiration of organic carbon[1–3], but little agreement exists about the sources or turnover times of this carbon[2,4,5]. Here we present results of an extensive survey of the carbon isotope composition (^{13}C and ^{14}C) of dissolved inorganic carbon and three size-fractions of organic carbon across the Amazonian river system. We find that respiration of contemporary organic matter (less than five years old) originating on land and near rivers is the dominant source of excess carbon dioxide that drives outgassing in medium to large rivers, although we find that bulk organic carbon fractions transported by these rivers range from tens to thousands of years in age. We therefore suggest that a small, rapidly cycling pool of organic carbon is responsible for the large carbon fluxes from land to water to atmosphere in the humid tropics.

Riverine CO_2 concentrations in Amazonian lowlands are 5–30 times supersaturated with respect to atmospheric equilibrium[1]; such conditions may be prevalent throughout the humid tropics. *In situ* respiration is the primary source of CO_2 sustaining supersaturation in rivers, although inputs from groundwater supersaturated by soil respiration can be important in small systems and from submerged root respiration in floodplain-influenced systems[1–3,6–8]. Although air–water gas exchange is a bi-directional process, atmospheric CO_2 invasion has a negligible role compared to the large CO_2 evasion fluxes, except at low supersaturation[2,3,6,7]. ^{13}C and ^{14}C isotopes can provide constraints on sources and turnover times of organic carbon fuelling river respiration, but no previous tropical study has used a dual-isotope approach to address these questions. Studies in temperate eastern USA provide contrasting findings. In the Hudson River, up to 70% of the centuries-old terrestrial organic carbon entering the river is respired in transit, and the average age of riverine organic carbon decreases downstream[2]. However, the youngest components of dissolved organic carbon (DOC) are preferentially respired in the York River[5], and modern dissolved inorganic carbon (DIC) in the Parker River may be explained by respiration of young DOC produced within the estuary[4]. Documenting key patterns and controls on CO_2 sources in diverse ecosystems is critical to advance our understanding of CO_2 outgassing from rivers and its contribution to regional net carbon budgets.

To identify dominant sources and turnover times of riverine carbon throughout the Amazon basin, we analysed ^{14}C and ^{13}C of
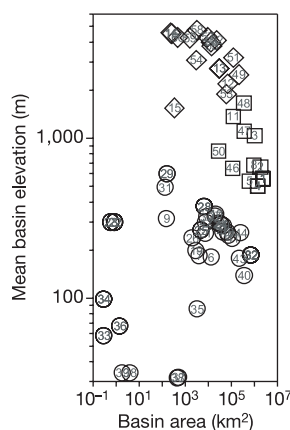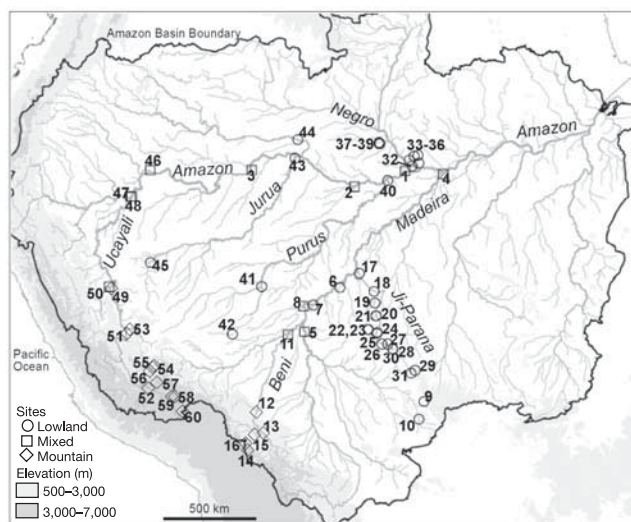

Figure 1 | Amazon basin and river sites sampled for carbon isotopes. We used GTOPO30 elevation[28] and a regional river network data set[29] to categorize each site by topography according to the percentage of the drainage area above 1,000 m elevation: mountain (diamond), ≥50% (16 sites); mixed (square), ≥10% (11 sites); and lowland (circle), <10% (33 sites). Mountain sites are found only in the Andean Cordillera, while mixed sites are large rivers draining both mountain and lowland areas. Site numbers are displayed. Distribution of sites by drainage area (river size) and mean basin elevation is shown. Additional information is in Supplementary Table S1.

[1]School of Oceanography, University of Washington, Seattle, Washington 98195, USA. [2]Stroud Water Research Center, Avondale, Pennsylvania 19311, USA. [3]Department of Earth Science, Rice University, Houston, Texas 77005, USA. [4]Laboratório de Ecologia Isotópica, CENA-USP, 13400-970 Piracicaba SP, Brazil. [5]Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, California 94551, USA.
*These authors contributed equally to this work.
‡Deceased.
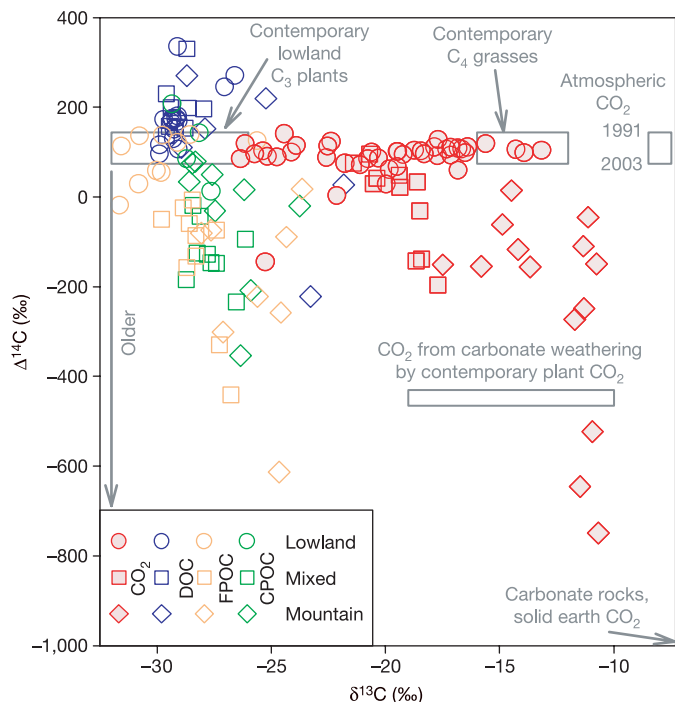
Figure 2 | Distribution of ^{14}C and ^{13}C isotopes. Lower $\Delta^{14}C$ values indicate older carbon. Symbol shapes are as in Fig. 1; Red (shaded), blue, orange, and green symbols represent CO_2, DOC, FPOC and CPOC respectively. Isotopic ranges of carbon sources are shown by grey boxes. Plant $\delta^{13}C$ ranges from −32‰ to −26‰ for lowland plant material fixed via the C_3 photosynthetic pathway dominant among plants, to −16‰ to −12‰ for tropical C_4 grasses[24,30]. Phytoplankton take up H_2CO_3 and impose ^{13}C fractionations of ∼−20‰, leading to biomass $\delta^{13}C$ values of −32 to −45‰ in mixed and lowland rivers[8,10,30], beyond the range found in our lowland organic carbon and CO_2 observations. Carbonate rocks and CO_2 from solid earth degassing are ^{14}C-free ($\Delta^{14}C = -1,000$‰; ref. 10). $\delta^{13}C$ of carbonate rocks is ∼0‰; degassed CO_2 ranges from −6‰ for volcanic to 10‰ for metamorphic sources[10]. The carbonate weathering region represents CO_2 in equilibrium with HCO_3^- resulting from weathering by CO_2 derived from respired, contemporary C_3 or C_4 plant material[10].

DIC, DOC, and suspended fine and coarse particulate organic carbon fractions (FPOC and CPOC), grouping sites topographically (Fig. 1). This survey represents the most extensive dual carbon isotope inventory so far in a large, diverse basin, and the first ^{14}C analysis of DIC in Amazonian rivers. It complements but greatly exceeds previous carbon isotope surveys[5,7,9], enabling an integrated assessment of carbon cycling.

DIC is composed of dissolved CO_2 and bicarbonate and carbonate ions in pH-dependent chemical and isotopic equilibrium[10]. In studies of marine and homogeneous river systems, where pH is nearly uniform, it has been the convention to report the isotopic composition of total DIC, which is directly measured. However, when the turnover of DIC by CO_2 fluxes is as rapid as in many of these tropical rivers, a quasi-steady-state condition constrains the isotopic composition of outputs by CO_2 outgassing to equal that of inputs by CO_2 production from respiration[7,8]. In response to this pre-eminence of CO_2 sources and the large range of observed pH (3.8 to 8.7), we base our analysis on the calculated isotopic values of CO_2 gas in equilibrium with DIC, avoiding the need to discuss the isotopic variation in ^{13}C of DIC caused by changes in pH alone (see Methods and Supplementary Discussion). We report measured $\delta^{13}C$-DIC in Supplementary Table S3. For ^{14}C we use $\Delta^{14}C$ notation, which includes a correction for ^{13}C such that $\Delta^{14}C$ of CO_2 and DIC are equal at any pH (ref. 11).

CO_2 in nearly all lowland rivers had a $\Delta^{14}C$ signature bound within the range of local atmospheric $\Delta^{14}C$-CO_2 estimated for the lowland
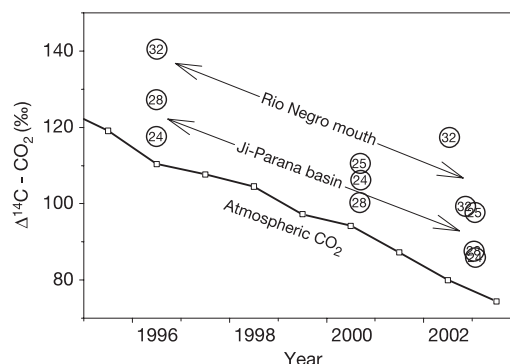
Figure 3 | Temporal evolution of ^{14}C-CO_2 at four lowland sites from medium to large rivers in the Ji-Parana basin and Rio Negro. Legend is as in Fig. 1. These sites drain continental shields and were analysed for ^{14}C-DIC 2–3 times between 1996 and 2003.

DIC sampling period (1996–2003, Fig. 2; Methods). Testing of thermonuclear bombs resulted in a large increase in atmospheric $^{14}CO_2$, reaching a peak in the late 1960s and steadily decreasing afterwards[12,13]. Continual changes in atmospheric radiocarbon content means that we need to assess the mean age for modern carbon samples by the offset between riverine $\Delta^{14}C$ and the sampling year's annual mean atmospheric $\Delta^{14}C$-CO_2 (ref. 14). From 1996 to 2003, $\Delta^{14}C$-CO_2 at four supersaturated sites in medium to large lowland rivers decreased by 32–42‰, equivalent to the annually averaged atmospheric $\Delta^{14}C$-CO_2 decrease of 36‰ (-5.6 ± 2.3‰ yr^{-1}) (Fig. 3). Thus, atmospheric offsets remained roughly constant, suggesting constant respiratory organic carbon turnover times for each basin. The range of these offsets suggests rapid turnover of photosynthetically sequestered atmospheric CO_2; outgassed CO_2 is derived from atmospheric CO_2 sequestered within <4 yr in the medium-sized Ji-Parana basin ($\Delta^{14}C$ offsets: 14 ± 6‰, $n = 8$) and 4–7 yr previously at the Rio Negro mouth ($\Delta^{14}C$ offsets: 29 ± 9‰, $n = 3$).

Rapid carbon turnover is probably widespread across Amazonian rivers. However, DIC from carbonate mineral weathering, with its typically enriched $\delta^{13}C$ and highly depleted $\Delta^{14}C$ (Fig. 2), commonly obscures the influence of respiration. To focus on sites where DIC originates largely as respired CO_2, we identified sites draining substantial carbonate lithologies through their inorganic solute composition[15] (Methods and Supplementary Discussion). Only lowland sites had little potential for direct, substantial carbonate contributions to DIC (38 samples in 25 lowland sites, Supplementary Fig. S1). Observations in carbonate-free lowlands contained supersaturated, predominantly contemporary CO_2 (Table 1, Fig. 2) and could be divided into two groups based on atmospheric $\Delta^{14}C$-CO_2 offsets. The largest group (32 samples from 21 sites) encompassed first-order streams and large rivers carrying contemporary CO_2 with atmospheric $\Delta^{14}C$-CO_2 offsets ranging from −3 to 38‰ (14 ± 11‰), indicating a mean CO_2 age of approximately 2 yr (and <5 yr in 87% of the observations); enriched $\delta^{13}C$-CO_2 in the Ji-Parana region (-17.5 ± 2.2‰, $n = 21$) suggest a C_4 plant influence. The second group of carbonate-free observations consisted of six samples in three small and two medium-sized rivers in the Ji-Parana region having considerable negative $\Delta^{14}C$ offsets (-32 ± 21‰), indicating a mean source age of several decades. Relatively depleted $\delta^{13}C$-CO_2 values (-20.0 ± 1.7‰) compared to Ji-Parana rivers from the first group suggest an influence from groundwater influx of aged soil CO_2 with a significant terrestrial C_3 plant source relative to other rivers in that highly deforested region. In both groups, respiration of submerged tree and grass roots can be excluded as important CO_2 sources because our data set was predominantly collected during low water.

Isotopic signatures in carbonate-free lowland rivers demonstrate

Table 1 | ^{14}C and ^{13}C isotopic composition for each carbon fraction

Site	CO₂		DIC	DOC		FPOC		CPOC	
Category	$\Delta^{14}C$	$\delta^{13}C$	$\delta^{13}C$	$\Delta^{14}C$	$\delta^{13}C$	$\Delta^{14}C$	$\delta^{13}C$	$\Delta^{14}C$	$\delta^{13}C$
Mountain	−240 ± 233 (14)	−12.9 ± 2.2	−4.9 ± 2.7	94 ± 176 (6)	−26.0 ± 3.0	−202 ± 198 (8)	−25.7 ± 1.7	−39 ± 146 (9)	−27.0 ± 1.6
Mixed	−14 ± 99 (11)	−19.4 ± 1.2	−14.2 ± 2.9	196 ± 59 (9)	−29.0 ± 0.6	−135 ± 141 (10)	−28.2 ± 0.9	−124 ± 66 (9)	−27.7 ± 0.9
Lowland	89 ± 44 (43)	−20.1 ± 3.6	−17.0 ± 5.9	177 ± 64 (15)	−29.0 ± 0.9	90 ± 55 (10)	−29.8 ± 1.8	112 ± 83 (4)	−28.5 ± 0.7
Carbonate-free	98 ± 20 (38)	−19.8 ± 3.7	−17.1 ± 6.2	175 ± 67 (11)	−29.1 ± 0.7	129 ± 10 (6)	−29.2 ± 2.1	121 ± 100 (3)	−28.4 ± 0.9

All carbonate-free sites are in the lowlands. Data are reported as mean ± s.d. (number of samples), in per mil (‰). The number of samples for ^{13}C is the same as that shown for ^{14}C. The number of samples for DIC is the same as for CO₂.

that CO_2 originated from heterotrophic respiration of contemporary C_3 and C_4 organic carbon (Fig. 2), yet CO_2 appears isotopically distinct from the associated bulk organic carbon load. All organic carbon fractions were considerably depleted in ^{13}C relative to CO_2 (Table 1, Fig. 2 and Supplementary Discussion). Furthermore, DOC was generally older than CO_2, FPOC showed a bimodal distribution with ages similar to CO_2 in non-turbid rivers and older carbon in high-sediment rivers, and CPOC exhibited a wide range of ages. We conclude that *in situ* respiration is fuelled largely by an unmeasured organic subfraction that cycles over a period of less than about five years and typically makes up a small component of the riverine organic carbon load. ^{13}C enrichment of CO_2 relative to organic carbon indicates that this missing organic carbon source fuelling river respiration is disproportionately composed of river-corridor and floodplain C_4 grasses, which may be intrinsically more biodegradable[7,16].

Mountain and mixed rivers contain older dissolved CO_2 ($\Delta^{14}C = −749$ to 96‰) with clear carbonate mineral dissolution signatures. However, observed CO_2 supersaturation in these rivers must be generated by CO_2 sources other than carbonates (Supplementary Discussion). These CO_2 fluxes gradually flush out geologically derived DIC, replacing its isotopic signature. Indeed, a CO_2 trend of increasing $\Delta^{14}C$ and decreasing $\delta^{13}C$ is observed from the Peruvian Andes to the Amazon mainstem, with $\Delta^{14}C$-CO_2 reaching 30–76‰ in the central mainstem, still below atmospheric levels (Fig. 2). Along the Ucayali and western-central mainstem, inputs of probably young CO_2 from *in situ* respiration and lowland tributaries drive large evasion fluxes of ^{14}C-depleted CO_2 derived from carbonate dissolution. In the more arid Ucayali mountain headwaters, highly depleted $\Delta^{14}C$-CO_2 ($<−500‰$), enriched $\delta^{13}C$-CO_2 ($\sim −11‰$), and considerable CO_2 supersaturation point to solid-earth degassing as a large dry-season CO_2 source, as documented in other tectonically active mountain ranges[17].

Control of respiration by a small fraction of organic carbon does not imply that bulk organic carbon is unreactive. On the contrary, measured organic carbon fractions appear to be mineralized throughout the river system. The strongest evidence is that FPOC generally becomes younger and more depleted in ^{13}C downstream from mountain sites (Fig. 2, Table 1), where all organic carbon fractions mirror the high-altitude ^{13}C enrichment in plants of ~1‰ per 1,000 m elevation[18,19]. Within 1,000 km from mountain head-waters, $\delta^{13}C$-FPOC becomes nearly indistinguishable from lowland carbon. With FPOC tightly associated with mineral surfaces[19,20], no downstream changes in the ratio of FPOC to fine suspended sediment concentration (0.8–1.5%, Supplementary Table S2), and $>85\%$ of the Amazon mainstem mineral load ultimately originating in the Andes[21], this observation implies nearly complete mineralization of old Andean FPOC and replacement with new lowland organic carbon during transit within the river or during long-term floodplain storage[7]. Even if all FPOC leaving the Andes were mineralized within the river channel, the resulting CO_2 flux would be 1/40 of total CO_2 evasion fluxes[1], hardly affecting the isotopic signature of CO_2. CPOC and DOC descending from the Andes follow identical trends of gradual ^{13}C depletion to lowland values (Fig. 2), although without a conservative mineral carrier as for FPOC, simple dilution by lowland

organic carbon cannot be ruled out. DOC is generally modern (<50 yr old) everywhere, demonstrating that old DOC does not escape from the basin.

Isotopic evidence for dominance of respiration fluxes by a rapidly cycling, typically small fraction of total organic carbon confirms the hypothesis previously posited for the Amazon mainstem from respiration[8,16], ^{13}C (ref. 7), and mass balance studies[1,7]. It implies that gradual consumption or replacement of old fractions in bulk organic carbon can occur in parallel with high rates of respiration of a highly labile organic carbon subfraction. This paradigm has been advanced for tropical and temperate soils[22], providing strong conceptual linkage between aquatic and terrestrial carbon dynamics[23]. It suggests that deforestation in the Amazon leads to immediate changes to the organic sources of riverine heterotrophic energy and argues that such impact is not inconsistent with apparent lag times observed in bulk organic carbon composition[24]. Although the mechanism proposed here may be widespread across the humid tropics[1] and appears consistent with radiocarbon observations from temperate rivers[2,4,5], it is probable that certain river systems, such as those draining eroding peats, are fuelled by old organic carbon. Explicit accounting of isotopic signatures of CO_2 outgassing from different river types may be required to accurately interpret isotope-based regional tropospheric CO_2 inversions.

METHODS

Sample collection and analysis. Samples analysed for ^{14}C-DIC were collected between 1991 and 2003, whereas organic-carbon ^{14}C samples are from 1995–1996. All samples were preserved with mercuric chloride immediately after collection at mid-depth from the deepest section of the channel. DIC samples were prepared as described in ref. 7 and stored in tightly capped glass bottles for up to 24 months; in the lab, the top half of the bottle was drawn into a vacuum line (eliminating particles) and stripped of CO_2 after acidification[7]. CPOC (63–2,000 µm) was isolated either by sieving or with a plankton net, FPOC (0.1–63 µm) by tangential flow microfiltration, and DOC (1,000 atomic mass units to 0.1 µm) by tangential flow ultrafiltration[19]. Final concentration and drying was achieved by centrifugal evaporation or freeze drying[19], and the dried powder stored in the dark at ambient temperature for up to 6 yr. Ultrafiltration yields ranged from 40% in the Andes to 80% in the lowlands[19,25]. Organic samples were combusted as in ref. 7.

Cryogenically purified CO_2 from organic carbon and DIC was analysed for stable isotope and radiocarbon by dual-inlet Isotope Ratio Mass Spectrometry and Accelerator Mass Spectrometry (AMS)[26], respectively; $>90\%$ of ^{14}C analyses were carried out at the Lawrence Livermore National Laboratory's Center for Accelerator Mass Spectrometry, and the rest at the University of Arizona AMS Laboratory. CO_2 extracted from DIC was stored in sealed glass ampoules for up to 8 yr. ^{13}C is reported in $\delta^{13}C$ notation versus the PDB standard[10]. Radiocarbon values are reported as age-corrected $\Delta^{14}C$ adjusted for sample $\delta^{13}C$ (ref. 11); carbon is defined as modern when it originates after 1890 (ref. 11). Absolute $\Delta^{14}C$ and $\delta^{13}C$ analysis errors (1σ) are typically $<6‰$ and $<0.2‰$, respectively. Isotopes for all carbon fractions were not always analysed at each site. Additional analyses include pH, major ions, alkalinity and total carbon fraction concentrations (Supplementary Table S2). Major ions were quantified by ion chromatography. Alkalinity was measured by Gran titration, or estimated from temperature, pH and DIC when not measured. CO_2 concentrations were either measured directly by headspace equilibration (A.K.A., manuscript in preparation) or estimated from temperature, pressure, pH, DIC, and alkalinity. Sampling sites are grouped by topographic characteristics (Fig. 1). In the discussion, grouped observations are reported as mean ± standard deviation

(number of samples) and compared only qualitatively owing to low number of samples per group.

Isotopic calculation of CO_2 gas in equilibrium with DIC. DIC is composed of dissolved carbonate species ($H_2CO_3(aq)$, $HCO_3^-(aq)$, and $CO_3^{2-}(aq)$) in temperature- and pH-dependent equilibrium with one another. Isotopic fractionation occurs during conversion from one species to another and dissolution of CO_2 gas[10]; CO_2 gas is hereafter referred to as simply CO_2. $\delta^{13}C$ of CO_2 gas in equilibrium with DIC is calculated from measured $\delta^{13}C$-DIC and pH, and from temperature-dependent isotopic equilibrium fractionations between CO_2 and DIC species[10,27]. pH can vary dramatically in a large basin and is largely a function of weathering lithologies (Supplementary Table S2). $\Delta^{14}C$ is defined to be insensitive to mass-dependent fractionation[11]; as a result $\Delta^{14}C$-CO_2 = $\Delta^{14}C$-DIC. Although a focus on isotopes of CO_2 instead of DIC is unconventional, it yields more straightforward assessments of the impact of respiration and air–water gas exchange on DIC across geochemically diverse rivers.

Radiocarbon trends in atmospheric CO_2. Measurements from Schauinsland Station, Germany, were used to characterize atmospheric $\Delta^{14}C$-CO_2 from 1991 to 2003 (refs 12, 14). The uncharacterized effect of seasonal and short-term atmospheric variability is minimized by comparing river ^{14}C only against time-weighted annual means[14]. A constant +8‰ offset was added to Schauinsland annual means to account for a 5‰ depletion from regional fossil-fuel emissions at Schauinsland relative to the well-mixed, mid-latitude European troposphere (Jungfraujoch site[14]), and approximately 3‰ further depletion at the mid-latitude troposphere relative to tropical South America[12,13]. Atmospheric $\Delta^{14}C$-CO_2 composition within the Amazon basin is unknown, but seasonal and regional variability may be as large as 10‰ (ref. 13). Riverine $\Delta^{14}C$ values within 5‰ of our estimated atmospheric annual average for the sampling year probably represent carbon turnover times of one year or less. Mid-1990s tropospheric CO_2 can be characterized by a partial pressure (p_{CO_2}) of 370 p.p.m. and $\delta^{13}C$ composition of −8‰ (refs 10, 12, 13).

Carbonate mineral dissolution and DIC. Carbonate weathering is a source of high DIC concentrations enriched in $\delta^{13}C$ and highly depleted in $\Delta^{14}C$ when, as is most common, it involves the dissolution of ^{14}C-dead carbonate minerals ($CaCO_3(s)$) by carbonic acid ($H_2CO_3(aq)$) from respired modern plant matter in soils (Fig. 2 and Supplementary Discussion). We used alkalinity and dissolved inorganic cation composition to identify sites draining substantial carbonate lithologies[15] (Supplementary Fig. S1). Although exchange with the atmosphere or input of respired CO_2 may erase the isotopic signature of carbonate dissolution, we identified and excluded carbonate-influenced sites in order to focus unambiguously on the isotopic relationship between organic carbon and DIC.

1. Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, M. V. & Hess, L. L. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO_2. *Nature* **416**, 617–620 (2002).
2. Cole, J. J. & Caraco, N. F. Carbon in catchments: Connecting terrestrial carbon losses with aquatic metabolism. *Mar. Freshwat. Res.* **52**, 101–110 (2001).
3. Mulholland, P. J. *et al.* Inter-biome comparison of factors controlling stream metabolism. *Freshwat. Biol.* **46**, 1503–1517 (2001).
4. Raymond, P. A. & Hopkinson, C. S. Ecosystem modulation of dissolved carbon age in a temperate marsh-dominated estuary. *Ecosystems* **6**, 694–705 (2003).
5. Raymond, P. A. & Bauer, J. E. Riverine export of aged terrestrial organic matter to the North Atlantic Ocean. *Nature* **409**, 497–500 (2001).
6. Jones, J. B. & Mulholland, P. J. Carbon dioxide variation in a hardwood forest stream: An integrative measure of whole catchment soil respiration. *Ecosystems* **1**, 183–196 (1998).
7. Quay, P. D. *et al.* Carbon cycling in the Amazon River: Implications from the ^{13}C compositions of particles and solutes. *Limnol. Oceanogr.* **37**, 857–871 (1992).
8. Devol, A. H. & Hedges, J. I. in *The Biogeochemistry of the Amazon Basin* (eds McClain, M. E., Victoria, R. L. & Richey, J. E.) 275–306 (Oxford Univ. Press, New York, 2001).
9. Hedges, J. I. *et al.* Organic carbon-14 in the Amazon River system. *Science* **231**, 1129–1131 (1986).
10. Clark, I. & Fritz, P. *Environmental Isotopes in Hydrogeology* (Lewis Publishers, Boca Raton, 1997).
11. Stuiver, M. & Polach, H. A. Discussion: reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
12. Levin, I. & Hesshaimer, V. Radiocarbon—A unique tracer of global carbon cycle dynamics. *Radiocarbon* **42**, 69–80 (2000).
13. Randerson, J. T., Enting, I. G., Schuur, E. A. G., Caldeira, K. & Fung, I. Y. Seasonal and latitudinal variability of troposphere $\Delta^{14}CO_2$: Post bomb contributions from fossil fuels, oceans, the stratosphere, and the terrestrial biosphere. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2002GB001876 (2002).
14. Levin, I. & Kromer, B. The tropospheric $^{14}CO_2$ level in mid-latitudes of the Northern Hemisphere (1959–2003). *Radiocarbon* **46**, 1261–1272 (2004).
15. Stallard, R. F. & Edmond, J. M. Geochemistry of the Amazon. 2. The influence of geology and weathering environment on the dissolved load. *J. Geophys. Res.* **88**, 9671–9688 (1983).
16. Melack, J. M. & Forsberg, B. R. in *The Biogeochemistry of the Amazon Basin* (eds McClain, M. E., Victoria, R. L. & Richey, J. E.) 235–274 (Oxford Univ. Press, New York, 2001).
17. Kerrick, D. M. Present and past nonanthropogenic CO_2 degassing from the solid earth. *Rev. Geophysics* **39**, 565–585 (2001).
18. Körner, C., Farquhar, G. D. & Wong, S. C. Carbon isotope discrimination by plants follows latitudinal and altitudinal trends. *Oecologia* **88**, 30–40 (1991).
19. Hedges, J. I. *et al.* Organic matter in Bolivian tributaries of the Amazon River: A comparison to the lower mainstem. *Limnol. Oceanogr.* **45**, 1449–1466 (2000).
20. Aufdenkampe, A. K., Hedges, J. I., Richey, J. E., Krusche, A. V. & Llerena, C. A. Sorptive fractionation of dissolved organic nitrogen and amino acids onto fine sediments within the Amazon Basin. *Limnol. Oceanogr.* **46**, 1921–1935 (2001).
21. Gibbs, R. J. The geochemistry of the Amazon River system: Part I. The factors that control the salinity and the composition and concentration of the suspended solids. *Geol. Soc. Am. Bull.* **78**, 1203–1232 (1967).
22. Trumbore, S. E. Age of soil organic matter and soil respiration: Radiocarbon constraints on belowground C dynamics. *Ecol. Appl.* **10**, 399–411 (2000).
23. Grimm, N. B. *et al.* Merging aquatic and terrestrial perspectives of nutrient biogeochemistry. *Oecologia* **442**, 485–501 (2003).
24. Bernardes, M. C. *et al.* Riverine organic matter composition as a function of land use changes, Southwest Amazon. *Ecol. Appl.* **14**, S263–S279 (2004).
25. Aufdenkampe, A. K. *et al.* Organic matter in the Peruvian headwaters of the Amazon: A comparison to Bolivian tributaries and the lowland Amazon mainstem. *Org. Geochem.* (in the press).
26. Vogel, J. S., Nelson, D. E. & Southon, J. R. ^{14}C background levels in an accelerator mass spectrometry system. *Radiocarbon* **29**, 323–333 (1987).
27. Zhang, J., Quay, P. D. & Wilbur, D. O. Carbon isotope fractionation during gas-water exchange and dissolution of CO_2. *Geochim. Cosmochim. Acta* **59**, 107–114 (1995).
28. Gesch, D. B., Verdin, K. L. & Greenlee, S. K. New land surface digital elevation model covers the Earth. *Eos* **80**, 70–71 (1999).
29. Mayorga, E., Logsdon, M. G., Ballester, M. V. R. & Richey, J. E. Estimating cell-to-cell land surface flow paths from digital channel networks, with an application to the Amazon basin. *J. Hydrol.* (in the press).
30. Hedges, J. I. *et al.* Compositions and fluxes of particulate organic material in the Amazon River. *Limnol. Oceanogr.* **31**, 717–738 (1986).

Author Contributions E.M. and A.K.A. contributed equally to all stages of this work. C.A.M. and T.A.B. supported the ^{14}C analyses and contributed to discussions and manuscript preparation. A.V.K. collected and pre-processed many of the samples and contributed to discussions and manuscript preparation. J.I.H., P.D.Q. and J.E.R. contributed to project design and execution, including manuscript preparation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.M. (emiliomayorga@alum.mit.edu).

LETTERS

A sharp lithosphere–asthenosphere boundary imaged beneath eastern North America

Catherine A. Rychert[1], Karen M. Fischer[1] & Stéphane Rondenay[2]

Plate tectonic theory hinges on the concept of a relatively rigid lithosphere moving over a weaker asthenosphere, yet the nature of the lithosphere–asthenosphere boundary remains poorly understood. The gradient in seismic velocity that occurs at this boundary is central to constraining the physical and chemical properties that create differences in mechanical strength between the two layers. For example, if the lithosphere is simply a thermal boundary layer that is more rigid owing to colder temperatures, mantle flow models[1,2] indicate that the velocity gradient at its base would occur over tens of kilometres. In contrast, if the asthenosphere is weak owing to volatile enrichment[3–6] or the presence of partial melt[7], the lithosphere–asthenosphere boundary could occur over a much smaller depth range. Here we use converted seismic phases in eastern North America to image a very sharp seismic velocity gradient at the base of the lithosphere—a 3–11 per cent drop in shear-wave velocity over a depth range of 11 km or less at 90–110 km depth. Such a strong, sharp boundary cannot be reconciled with a purely thermal gradient, but could be explained by an asthenosphere that contains a few per cent partial melt[7] or that is enriched in volatiles relative to the lithosphere[3–6].

Precise determination of the thickness of the continental lithosphere and the magnitude and depth range of the velocity gradient at its base has proved challenging with existing seismic techniques. Discontinuities in the 80–210 km depth range have been observed on a regional scale in continental settings by reflection and refraction experiments and by teleseismic body-wave phases[4,8–16], and in some cases these features have been interpreted as the lithosphere–asthenosphere boundary[11–15]. Yet, perhaps because of significant variations in the depth of the discontinuity, an omnipresent discontinuity has yet to be illuminated in global or continental stacks of seismic phases[17]. More significantly, even where the base of the continental lithosphere has been identified[11–15], constraints on the velocity gradient at this boundary have been too loose to permit a clear assessment of its physical and chemical properties.

In this study, we imaged the base of the lithosphere in eastern North America using P-to-S (Ps) conversions produced by scattering of teleseismic P waves at this discontinuity. Ps arrival times may be used to determine discontinuity depths. For example, Ps and S-to-P (Sp) conversions recorded on and near the Hawaiian Islands have successfully identified the base of the lithosphere at depths of 50 km to 110 km (refs 18–20). Furthermore, Ps amplitudes as a function of wave period directly reflect the gradient of velocity with depth, allowing us to probe the mechanisms responsible for the lithosphere–asthenosphere boundary.

Our study region spans the Palaeozoic Appalachian orogen and the eastern coastal margin of the United States (Fig. 1). Following the collision of the proto-African and proto-North American plates, the last major tectonic events to affect the lithosphere in this area were Triassic and Jurassic rifting in the east related to the opening of the Atlantic Ocean[21], and passage of the lithosphere over a plume ~ 100–120 Myr ago[22]. Surface waves image a seismically fast lithospheric lid in eastern North America that extends to roughly 200 km depth beneath the craton just to the west of our study region, and thins to 80–90 km depth at the continental margin in the east[23,24].

We analysed SV components of Ps conversions recorded by six permanent broadband stations that have been operating for over 5 years (Fig. 1). For each station, the recorded waveforms were decomposed into P and S components using a free-surface transformation matrix and then simultaneously deconvolved in the
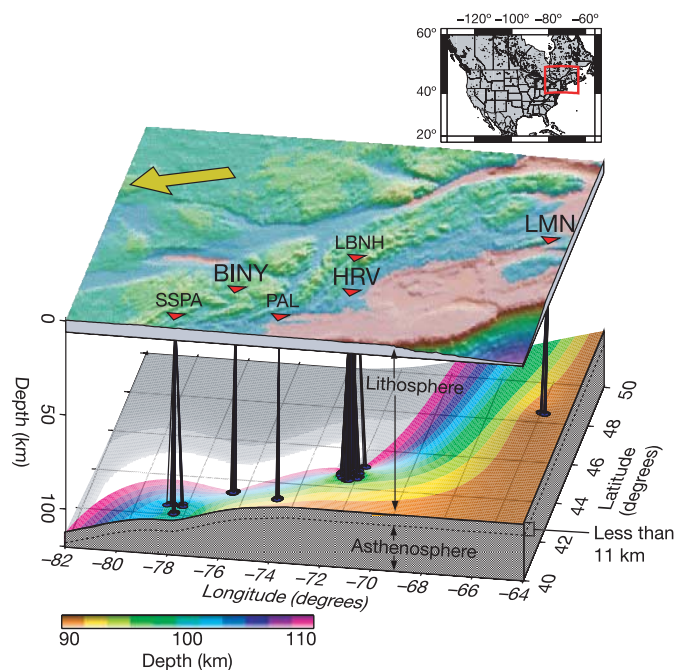

Figure 1 | Three-dimensional view of the lithosphere–asthenosphere boundary and surface topography. Red box in the inset map highlights the location of the study region within North America. Shading on the top surface indicates topography. Yellow arrow points in the direction of absolute plate motion; plate velocity is 2.5 cm yr^{-1}. Red triangles denote station locations. The lower surface represents the location of the base of the lithosphere interpolated from migrated Ps waveform images at the six labelled stations. The larger text corresponds to the stations (HRV, LMN, BINY) where this phase is most clearly observed. This surface ranges from 90 km (orange) to 110 km (pink) depth. Each colour band covers 2 km in depth. Blue circles on the discontinuity surface indicate the conversion points of the Ps phases. Black lines connect piercing points to the station at which the conversion is observed.

[1]Department of Geological Sciences, Brown University, Box 1846, Providence, Rhode Island 02912, USA. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

frequency domain and migrated to depth, thus illuminating the shear component of energy scattered from velocity discontinuities[8].

A negative phase from depths comparable to the base of the lithosphere is clearly observed at stations HRV, LMN and BINY, and is also apparent at stations PAL, SSPA and LBNH (Fig. 2; Supplementary Information, section 1.1). Synthetic seismograms calculated for the crustal models that best fit the Moho and the first Moho reverberations confirm that crustal phases and their interference cannot mimic the apparent arrival from the base of the lithosphere (Supplementary Information, section 1.1). In addition, the arrival time of the phase at different stations does not consistently correlate with variations in crustal thickness, as would be the case if this phase were related to Moho reverberations. For
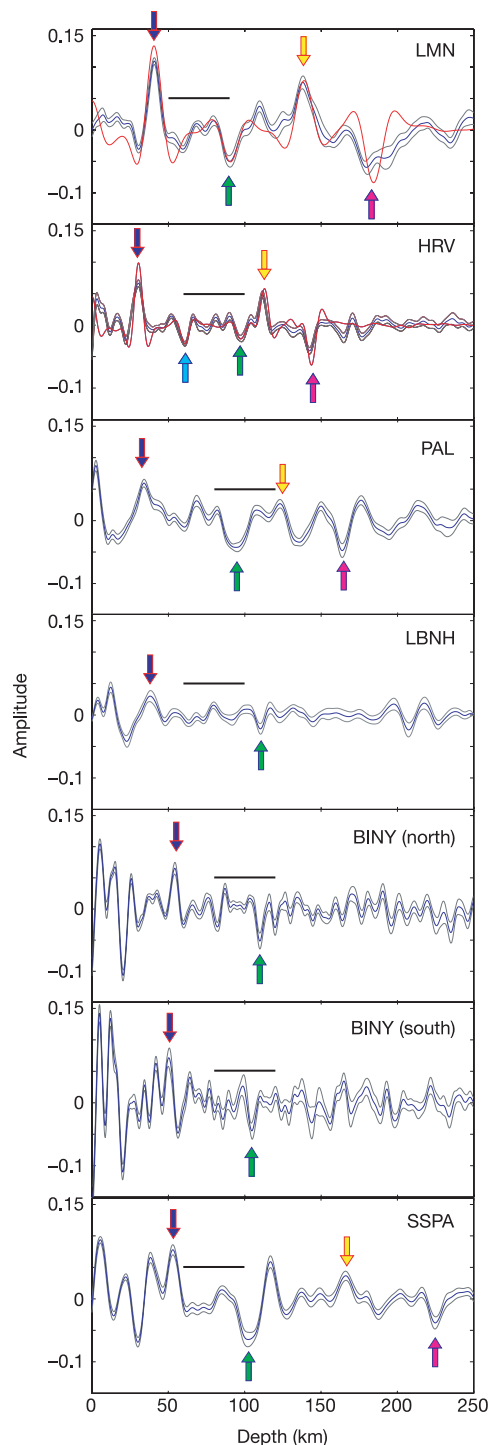
example, the Moho at station LMN is ~10 km deeper than the Moho at HRV, yet the phase from the base of the lithosphere at LMN migrates to a depth ~7 km shallower than the phase at HRV (Fig. 2). Although variations in the character of the phase occur between stations, they correspond in general to the character of the entire impulse response, and much of this variation is unrelated to the velocity gradient at the discontinuity (Supplementary Information, section 1.1).

The depth of the observed discontinuity ranges from 90 km (LMN) to 110 km (LBNH), increasing to the northwest beneath the Appalachian orogen (Fig. 1) with a landward rather than a seaward[25] dip. Uncertainties on the absolute depth of the discontinuity at each station are less than 1.6 km (Supplementary Information, section 1.1). We associate this phase with the base of the lithosphere, because its depth lies within the transition from fast seismic lid to slow low-velocity zone seen in surface wave tomography[23,24] (Fig. 2).

The Ps phase scattered from the base of the lithosphere requires a very sharp transition between the lithosphere and the asthenosphere. The velocity gradient at the lithosphere–asthenosphere boundary was determined by inverting the migrated waveforms at stations HRV and LMN using a damped least-squares method, assuming a linear gradient (Supplementary Information, section 2). Shallower discontinuities were also modelled to avoid biasing estimates of the lithosphere–asthenosphere gradient, and trade-offs between the absolute depth, depth extent, and magnitude of the lithosphere–asthenosphere gradient were carefully assessed (Supplementary Information, section 2 and Supplementary Table 1). The data at HRV resolve a 3.1–5.7% drop in shear-wave velocity over 5 km or less centred at ~97.0 km depth. The best-fitting models for LMN indicate that the gradient occurs over 5 km or less and the velocity contrast ranges from 6.8% to 7.4% centred at ~90.7 km (Supplementary Table 2). However, thicknesses of almost 11 km cannot be ruled out by 95% confidence limits at LMN, and a thickness of 10 km implies a velocity contrast of 10.7%. Although the depth ranges over which these strong velocity contrasts occur are surprisingly sharp (<11 km), they are well-resolved owing to the relatively short-period energy that dominates the scattered waves (Supplementary Information, section 2).

At station HRV, in addition to the phase from the base of the lithosphere, we observe a phase that is consistent with a negative velocity contrast of 5.4 ± 0.6% at 60.9 ± 0.4 km depth (Supplementary Information, section 2), whereas no discontinuity internal to the mantle lithosphere is observed at LMN. If the observed phase does in fact originate from a discontinuity at 61 km depth below HRV, it may be explained by greater hydration or chemical depletion in the lower lithosphere at HRV than at LMN. However, because the existence of a

Figure 2 | Imaging discontinuities with waveforms from individual stations. Blue lines show SV waveforms deconvolved and migrated in single bins for stations LMN, HRV, PAL, LBNH, BINY in two back-azimuth bins, and SSPA. A positive phase corresponds to a velocity increase with depth, while a negative phase indicates a velocity decrease with depth. Coloured arrows indicate phases from the base of the lithosphere (green), the Moho (dark blue), crustal reverberations (yellow and magenta), and, for station HRV, a discontinuity at 61 km depth (light blue). Error bars corresponding to two standard deviations (grey lines) were calculated with bootstrap tests in which a random 20% of the events in the bin were randomly replaced by another 20%, and the deconvolved, migrated waveforms were recalculated 100 times. Horizontal black lines correspond to the lithosphere–asthenosphere boundary as determined by surface wave models, defined as the greatest negative velocity gradient (LMN[23], all others[24]). Synthetic waveforms corresponding to the models obtained by inverting the data are shown for HRV and LMN (red lines). Crustal phase amplitudes were not included in these inversions and are therefore the one aspect of the synthetic waveforms that do not match the data (Supplementary Information, section 2).

61 km discontinuity is ambiguous (Supplementary Information, section 1.2), we modelled the velocity drop at the lithosphere–asthenosphere boundary beneath HRV with and without a 61 km discontinuity, and found that a similar velocity contrast is required (3.3–5.7% and 3.1–4.9%, respectively; Supplementary Table 2).

At the lithosphere–asthenosphere boundary, the shear-wave velocity contrasts that best fit the Ps scattering (3.1–7.4) are comparable to surface wave studies in which the total drop in velocity from the lithospheric lid to the slower asthenosphere varies from 3% to 11% across the region[23,24]. However, these same surface wave studies cannot clearly resolve the difference between sharp velocity gradients and those that occur over ~40–50 km. Our results require that the velocity contrast occurs over less than 11 km. In addition, the shear-wave velocity drop at the base of the lithosphere in this continental orogen/margin environment is similar to that found in old Pacific oceanic lithosphere using joint surface-wave and body-wave inversions (6.4%); yet, again the Ps scattering requires a sharper gradient (<11 km versus <30 km; ref. 4). Reflection studies in localized zones at the base of the continental or passive margin lithosphere have invoked gradients over comparable depth ranges[15], but they did not model the magnitude of the velocity drop. Another Ps study[26] that included station HRV did not notice a phase from the base of the lithosphere because the elimination of higher frequencies in the study caused crustal reverberations to destructively interfere with the phase from the base of the lithosphere (Supplementary Information, section 1.1).

What mechanisms could be responsible for the strong, sharp velocity gradient associated with the base of the lithosphere? Experimental studies[27] suggest that if the velocity gradient is caused solely by temperature, temperature increases of at least 220 °C and 120 °C are required to explain the velocity contrasts at LMN and HRV, respectively (Supplementary Information, section 3). However, numerical models of mantle flow in which viscosity depends only on temperature and pressure indicate that the thermal gradient at the base of the lithosphere is typically less than 5 °C km^{-1}, and definitely less than 10 °C km^{-1} (refs 1, 2). These models are not able to match the greater than 20 °C km^{-1} temperature gradient that would be required at the base of the lithosphere at LMN and HRV, and another mechanism besides temperature is required to explain the observed velocity contrast.

A vertical variation in mantle chemical composition—for example, a lithosphere dehydrated and depleted in Fe, Ca and Al owing to past melt extraction versus an undepleted, volatile-rich asthenosphere[3–6]—could produce a sharp enough discontinuity. Here we assume that mantle temperatures are sub-solidus. The maximum possible shear-wave velocity variation due to depletion is ~1.5% (ref. 28). Because hydration reduces velocity primarily by increasing attenuation[5], its effect is limited by reasonable attenuation contrasts between the lithosphere and asthenosphere and is no more than roughly 4.3% (Supplementary Information, section 3). Depletion and hydration are therefore individually too small to explain the 6.8–7.4% velocity drop inferred from the best-fitting models at LMN. However, a velocity contrast caused by the combination of dehydration and depletion encompasses the HRV values (3.3–5.7%) and could reach the LMN range with a modest contribution from temperature. A sharp decrease in grain size with depth could also reduce velocity[27], but because grain size, like hydration, affects velocity via attenuation, the total contribution from hydration and grain size still cannot exceed ~4.3%.

Another possibility is that the strong, sharp lithosphere–asthenosphere boundary is produced by the presence of small amounts of partial melt in the asthenosphere, which could easily produce the roughly 7% velocity drop inferred for the best-fitting LMN models and even the 11% velocity drop associated with the largest gradient thickness permitted by the LMN data[29] (Supplementary Information, section 3). In this scenario, the lithosphere–asthenosphere boundary is defined by the solidus, preventing melt from rising into

the lithosphere. One mechanism for generating melt beneath eastern North America is decompression of mildly hydrated[3] (an atomic ratio of 800–1,000 H per 10^6 Si) asthenospheric material as it flows upward along the contours of the more rigid, shallowing lithosphere in response to the west-southwest absolute motion of the North American plate[30] (Fig. 1). Melt would be continually regenerated by rising hydrated asthenosphere, replacing any melt carried out of the region by solid mantle flow or lost via melt migration and cooling. Alternatively, temperatures may exceed the damp-solidus within the asthenosphere without the need for decompression. In these scenarios involving asthenospheric melting, or those in the preceding paragraph at sub-solidus conditions, temperature will also contribute to the observed velocity gradient, and because melt and/or compositional boundaries would produce rapid decreases in viscosity with depth[3], temperature gradients could be sharper than in models where viscosity depends solely on temperature and pressure[1,2].

In the future, comparing well-resolved velocity gradients at the base of the lithosphere between a variety of tectonic environments should help to isolate the relative roles of temperature, hydration, depletion and melting. On the basis of the present study, however, we are able to conclude that the strong, sharp lithosphere–asthenosphere boundary beneath this continental region cannot be defined solely by temperature, but is consistent with a rapid rheological transition produced by melt or volatiles in the asthenosphere.

1. King, S. D. & Ritsema, J. African hot spot volcanism: small scale convection in the upper mantle beneath cratons. *Science* 290, 1137–1140 (2000).
2. Zaranek, S. E., Parmentier, E. M. & Fischer, K. M. Effects of basal drag and ablation on the evolution of cratonic lithosphere. *Eos* 85(17), T41C–03 (2004).
3. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle; implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* 144, 93–108 (1996).
4. Gaherty, J. B., Kato, M. & Jordan, T. H. Seismological structure of the upper mantle; a regional comparison of seismic layering. *Phys. Earth Planet. Inter.* 110, 21–41 (1999).
5. Karato, S.-I. & Jung, H. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth Planet. Sci. Lett.* 157, 193–207 (1998).
6. Hirth, G., Evans, R. L. & Chave, A. D. Comparison of continental and oceanic mantle electrical conductivity: Is the Archean lithosphere dry? *Geochem. Geophys. Geosyst.* 1, 2000GC00048 (2000).
7. Anderson, D. L. *Theory of the Earth*, (Blackwell Scientific, Boston, Massachusetts, 1989).
8. Bostock, M. G. Mantle stratigraphy and evolution of the Slave province. *J. Geophys. Res.* 103, 21183–21200 (1998).
9. Dueker, K., Yuan, H. & Zurek, B. Thick-structured Proterozoic lithosphere of the Rocky Mountain region. *GSA Today* 11, 4–9 (2001).
10. Revenaugh, J. & Jordan, T. H. Mantle layering from ScS reverberations; 3, The upper mantle. *J. Geophys. Res.* 96, 19781–19810 (1991).
11. MONA LISA Working Group, MONA LISA—Deep seismic investigations of the lithosphere in the southeastern North Sea. *Tectonophysics* 269, 1–19 (1997).
12. Morozova, E. A., Morozov, I. B., Smithson, S. B. & Solodilov, L. N. Heterogeneity of the uppermost mantle beneath Russian Eurasia from the ultra-long-range profile QUARTZ. *J. Geophys. Res.* 104, 20329–20348 (1999).
13. Ryberg, T. *et al.* Two-dimensional velocity structure beneath northern Eurasia derived from the super long-range seismic profile Quartz. *Bull. Seismol. Soc. Am.* 86, 857–867 (1996).
14. Steer, D. N., Knapp, J. H. & Brown, D. L. Super-deep reflection profiling: exploring the continental mantle lid. *Tectonophysics* 286, 111–121 (1998).
15. Steer, D. N. *et al.* Deep structure of the continental lithosphere in an unextended orogen; an explosive-source seismic reflection profile in the Urals (Urals Seismic Experiment and Integrated Studies (URSEIS 1995)). *Tectonics* 17, 143–157 (1998).
16. Thybo, H. & Perchuc, E. The seismic 8° discontinuity and partial melting in continental mantle. *Science* 275, 1626–1629 (1997).
17. Shearer, P. M. Constraints on upper mantle discontinuities from observations of long-period reflected and converted phases. *J. Geophys. Res.* 96, 18147–18182 (1991).
18. Li, X. *et al.* Mapping the Hawaiian plume conduit with converted seismic waves. *Nature* 405, 938–941 (2000).
19. Collins, J. A. *et al.* Broadband seismology in the oceans; lessons from the ocean seismic network pilot experiment. *Geophys. Res. Lett.* 28, 49–52 (2001).
20. Li, X., Kind, R., Yuan, X., Wolbern, I. & Hanka, W. Rejuvenation of the lithosphere by the Hawaiian plume. *Nature* 427, 827–829 (2004).

21. Hatcher, R. D. Jr in *The Appalachian-Ouachita Orogen in the United States* (eds Hatcher, R. D. Jr, Thomas, W. A. & Viele, G. W.) 511–519 (Geological Society of America, Boulder, Colorado, 1989).

22. Heaman, L. M. & Kjarsgaard, B. A. Timing of eastern North American kimberlite magmatism; continental extension of the Great Meteor Hotspot track? *Earth Planet. Sci. Lett.* **178**, 253–268 (2000).

23. Van der Lee, S. High-resolution estimates of lithospheric thickness from Missouri to Massachusetts, USA. *Earth Planet. Sci. Lett.* **203**, 15–23 (2002).

24. Li, A., Forsyth, D. W. & Fischer, K. M. Shear velocity structure and azimuthal anisotropy beneath eastern North America from Rayleigh wave inversion. *J. Geophys. Res.* **108**, doi:10.1029/2002JB02259 (2003).

25. Menke, W. & Levin, V. Anomalous seaward dip of the lithosphere-asthenosphere boundary beneath northeastern USA detected using differential-array measurements of Rayleigh waves. *Geophys. J. Int.* **149**, 413–421 (2002).

26. Li, A., Fischer, K. M., van der Lee, S. & Wysession, M. E. Crust and upper mantle discontinuity structure in eastern North America. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000190 (2002).

27. Jackson, I., Fitz Gerald, J. D., Faul, U. H. & Tan, B. H. Grain size sensitive sesmic wave attenuation in polycrystalline olivine. *J. Geophys. Res.* **107**, doi:10.1029/2001JB001225 (2002).

28. Lee, C.-T. A. Compositional variation of density and seismic velocities in natural peridotites at STP conditions: Implications for seismic imaging of compositional heterogeneities in the upper mantle. *J. Geophys. Res.* **108**, doi:10.1029/2003JB002413 (2003).

29. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity; effects of realistic partial melt geometries. *J. Geophys. Res.* **105**, 10975–10986 (2000).

30. Fouch, M. J., Fischer, K. M., Parmentier, E. M., Wysession, M. E. & Clarke, T. J. Shear wave splitting, continental keels, and patterns of mantle flow. *J. Geophys. Res.* **105**, 6255–6275 (2000).

LETTERS

Host shift to an invasive plant triggers rapid animal hybrid speciation

Dietmar Schwarz[1], Benjamin M. Matta[1]†, Nicole L. Shakir-Botteri[1] & Bruce A. McPheron[1]

Speciation in animals is almost always envisioned as the split of an existing lineage into an ancestral and a derived species. An alternative speciation route is homoploid hybrid speciation[1] in which two ancestral taxa give rise to a third, derived, species by hybridization without a change in chromosome number. Although theoretically possible it has been regarded as rare[1] and hence of little importance in animals. On the basis of molecular and chromosomal evidence, hybridization is the best explanation for the origin of a handful of extant diploid bisexual animal taxa[2–6]. Here we report the first case in which hybridization between two host-specific animals (tephritid fruitflies) is clearly associated with the shift to a new resource. Such a hybrid host shift presents an ecologically robust scenario for animal hybrid speciation because it offers a potential mechanism for reproductive isolation through differential adaptation to a new ecological niche[7]. The necessary conditions for this mechanism of speciation[7] are common in parasitic animals, which represent much of animal diversity[8]. The frequency of homoploid hybrid speciation in animals may therefore be higher than previously assumed.

Although uncommon, homoploid hybrid speciation has been well described in plants[9]. In animals, potential examples of homoploid hybrid speciation are even rarer, far less conclusive and the evolutionary mechanism remains elusive[1]. In general, homoploid hybrid speciation faces two obstacles. Hybrids have to be both fit and reproductively isolated in order to form an independent population[10]. Fitness not only includes the viability and fertility of hybrid offspring but also its ability to avoid competition and backcrossing with the parental taxa. By definition, homoploid hybrid speciation has to occur in sympatry or parapatry. This requires that hybrids need to occupy a hybrid-specific ecological niche[10], otherwise hybrid genotypes will either be outcompeted in parental habitats or lead to

introgression between the parental species without the formation of an independent hybrid-origin lineage. The other major challenge is the evolution of reproductive isolation in the face of gene flow from the parental taxa in sympatry or parapatry. The host-specific life history of many parasitic taxa offers a robust solution to the problem of ecological and reproductive isolation, particularly if there is habitat choice[7]. A shift to a new host could provide a hybrid population with a separate resource that is free from competition from the parental taxa. At the same time, a host shift offers a potential mechanism of reproductive isolation, especially if mating occurs on the host. Such an epistatic combination of mate and host choice facilitates host-driven speciation in theoretical models[7].

In 1997 we discovered the infestation of non-native, brushy honeysuckle forms, *Lonicera* spp., by tephritid fruitflies within the *Rhagoletis pomonella* species complex in the northeastern United States (see Supplementary Information). All taxa within the *R. pomonella* species complex are specialized fruit parasites—each occupies only a very limited range of host plants[11]. The *Lonicera* plants that were infested represent a mixture of parentals (*L. morrowii*), described hybrids (*L. × bella* and *L. × amoena*) and introgressed forms that originated from Asian introductions to North America during the past 250 yr (ref. 12). These honeysuckle forms are widely distributed and abundant invasive weeds throughout the northeastern United States[13]. Although the introduced honeysuckle taxa serve as hosts for *Rhagoletis* in Asia and Europe[14], no infestation of introduced or native *Lonicera* by *Rhagoletis* has been described in North America[15]. The new insect colonists, however, belong to a monophyletic group of *Rhagoletis* that consists entirely of native North American taxa, most of which overlap in distribution with our newly discovered infestation of *Lonicera*[11]. This suggests that the infestation of invasive honeysuckle forms is the result of a recent

Table 1 | Pooled allele frequencies of diagnostic alleles for the described species within the *R. pomonella* species complex and the *Lonicera* fly

|	R. pomonella (all pop.)	R. pomonella (central PA)	R. mendax (all pop.)	R. mendax (central PA)	Lonicera fly (central PA)	R. zephyria (all pop.)	R. zephyria (central PA)
N_{pop} (N_{ind})	87 (7,581)	2 (105)	26 (992)	1 (36)	3 (243)	11 (578)	2 (143)
Had^{100}	0.759	0.772	0.094	0.125	0.157	0.000	0.000
Had^{111}	0.002*‡	0.000	0.000	0.000	0.673	0.992	0.986
N_{pop} (N_{ind})	26 (1,982)	1 (36)	29 (1,089)	1 (32)	2 (101)	3 (207)	2 (82)
Fum^{158}	0.001	0.000	0.849	0.923	0.178	0.013†	0.000
N_{pop} (N_{ind})	77 (6,357)	2 (89)	26 (959)	1 (34)	3 (146)	8 (503)	2 (123)
$Dia\text{-}2^{100}$	0.721	0.826	0.001	0.000	0.000	0.024*	0.000
N_{pop} (N_{ind})	79 (6,516)	2 (98)	26 (855)	1 (36)	3 (240)	8 (470)	2 (134)
$Aat\text{-}2^{100}$	0.353	0.470	0.011	0.000	0.000	0.000	0.000

R. cornivora, a member of the *R. pomonella* species complex, is excluded from the data. Central PA indicates populations surveyed in this study; all pop. indicates previously reported data from populations throughout the ranges of the described taxa (for literature sources see Supplementary Table 1) except *R. pomonella*, for which only the northern populations are represented. N_{pop} indicates the number of surveyed populations; N_{ind} indicates the number of surveyed individuals.
* Allele observed only in zones of described hybridization between *R. pomonella* and *R. zephyria* in the Pacific northwest and Minnesota.
† Allele observed only in the Pacific northwest.
‡ Had^{111} has been observed at a frequency of 0.007 in one population ($n = 76$) of the flowering dogwood fly, an undescribed species in the *R. pomonella* complex.

[1]Department of Entomology, The Pennsylvania State University, 501 ASI Building, University Park, Pennsylvania 16803, USA. †Present address: Department of Surgery, University of Pittsburgh, 200 Lothrop St, Pittsburgh, Pennsylvania 15213, USA.
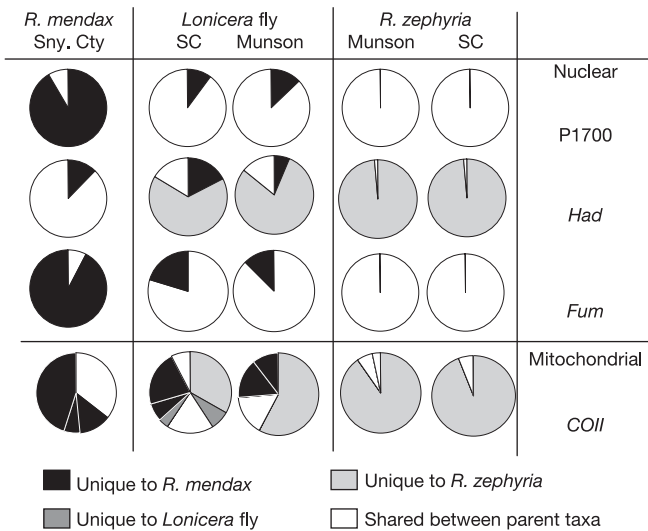
Figure 1 | **Allele frequencies at hybrid diagnostic nuclear and mitochondrial loci in *Lonicera* fly and parental populations from central Pennsylvania.** For sample sizes, see Supplementary Table 2 (for mitochondrial *COII* (bottom row) $n = 31, 27, 19, 31, 17$ left to right). There is a posterior probability > 0.05 that either one of the two *Lonicera* fly-specific haplotypes was missed in at least one of the two parental species (see Supplementary Information). Sny. Cty, Snyder County; SC, State College.

host shift, which we tested by analysing them for private alleles known to characterize other *Rhagoletis* species.

Previous studies that sampled hundreds (sometimes thousands) of individuals of the *Rhagoletis* species from multiple locations within their range showed that these taxa are characterized by unique high-frequency (private) alleles[11]. Given this previous information, we expected to assign unequivocally this as-yet-undescribed population on *Lonicera* to a known *Rhagoletis* taxon. Instead we found that the flies infesting *Lonicera* (subsequently referred to as the '*Lonicera* fly') showed a unique mixture of species-specific allozyme alleles that indicate that the *Lonicera* fly formed through the hybridization of the blueberry maggot (*R. mendax)* and the snowberry maggot (*R. zephyria*). The *Lonicera* fly samples exhibit mixtures of private alleles for both *R. zephyria* (*Had[111]*) and *R. mendax* (*Fum[158]*), and lack two common *R. pomonella* alleles (*Dia-2[100]* and *Aat-2[100]*; Table 1). The hybrid origin of the *Lonicera* fly is further supported by additional allozyme and sequence-based markers from five of the six nuclear linkage groups in the *R. pomonella* species complex[16] (see Supplementary Information) and the mitochondrial genome (Fig. 1). In addition to *Had* and *Fum* in linkage group III, the *Lonicera* fly possesses alleles that are private in the parental species at both P1700 (ref. 16) (linkage group V) and the mitochondrial cytochrome

oxidase II (*COII*) gene. At *Aat-2*, *Idh*, *Pgm* (linkage group I), *Mpi* (linkage group II) and P2963 (linkage group IV) the *Lonicera* fly shows intermediate allele frequencies (Supplementary Fig. 2 and Table 2). Our analysis of multilocus genotypes using a model-based clustering approach[17] shows that all *Lonicera* fly individuals are consistently classified as recombinants of the two extremes of the ancestry gradient between *R. mendax* and *R. zephyria* (Fig. 2). The *Lonicera* fly individuals show a high variance in their ancestry coefficient and span almost the entire spectrum between these putative ancestors, with more individuals being similar to *R. zephyria* than to *R. mendax*. Consistent intermediate allele frequencies across independent linkage groups offer strong evidence for the hybrid origin of the *Lonicera* fly against the competing hypothesis of incomplete lineage sorting (see Supplementary Information). The probability is very small ($P < 0.0005$) that such a pattern would have arisen by drift or selection in an old, undescribed non-hybrid species that shares a common ancestor with *R. mendax* and *R. zephyria*. Incomplete lineage sorting is also contrary with the known history of non-native *Lonicera* introduction to North America, which suggests a recent *Lonicera* fly origin after a host shift. We considered the alternative hypothesis of host shift followed by non-hybrid speciation by either one of the two putative parental species. Given the sampling of *Rhagoletis* (Table 1), an *R. mendax*/*R. zephyria* hybrid origin is a much more parsimonious explanation than the drift- or selection-induced increase of extremely rare or previously undetected alleles on honeysuckle (see Supplementary Information). The 95% credible intervals for the frequency of an unobserved *R. zephyria*-specific *Had* allele in *R. mendax* and vice versa are 0–0.0029 and 0–0.0041, respectively. Additional evidence is provided by the observed consistent intermediate pattern over all examined nuclear linkage groups and the mitochondrial genome. It is unlikely that drift or selection in a single parental origin population would have prompted the consistent shift in the frequency of unlinked loci towards the frequency of a second species that did not contribute to the host shift ($P < 0.008$ considering only nuclear loci; see Supplementary Information).

Do the flies on *Lonicera* comprise a hybrid zone[18] that is maintained by continued, substantial immigration from both parental species? If so, we would expect a high incidence of F_1 hybrids (and backcrosses) and strong deviations from Hardy–Weinberg and linkage equilibrium[18]. We assigned individual multilocus genotypes to six different classes of offspring that would result from two generations of hybridization (pure parentals, F_1 and F_2 hybrids, and first generation backcrosses). From this analysis we found no evidence that any of 50 sampled *Lonicera* fly individuals represents an F_1 genotype. In contrast, we clearly identified simulated F_1 hybrid genotypes in both a hybrid swarm with continued immigration from both parental species and a population consisting solely of F_1 hybrid genotypes (Fig. 3; see also Supplementary Fig. 1a–e for the assignment to other hybrid classes). In addition, no significant deviations
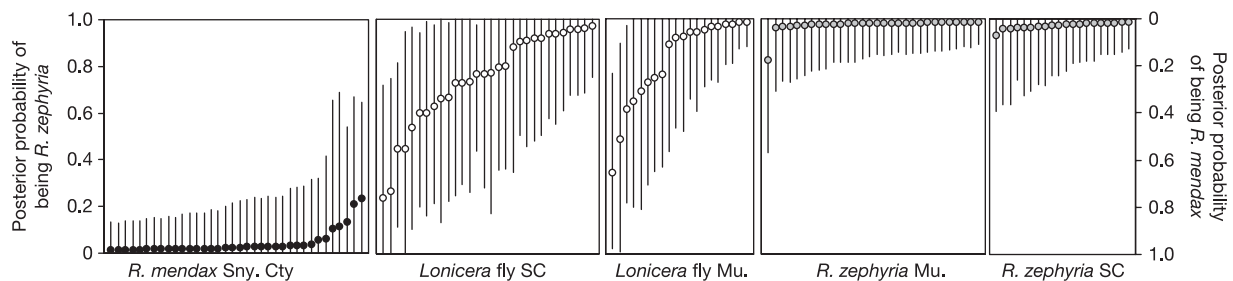
Figure 2 | **Assignment of ancestry from two parental populations to individuals of the *Lonicera* fly and its parents.** Assignment was without any prior information of population membership (as implemented in STRUCTURE[17]). Data for *R. mendax* (black circles), *Lonicera* fly (white circles) and *R. zephyria* (grey circles) are shown. Collection sites: Snyder County (Sny. Cty), State College (SC) and Munson (Mu.). Error bars represent 95% credible intervals.

from Hardy–Weinberg and linkage equilibrium in the observed *Lonicera* fly populations were observed (Supplementary Tables 3 and 4). We only observed significant linkage between *Had* and *Fum*, two genes that are known to be in the same linkage group[19] and might be part of a chromosomal inversion[20]. Given the postulated recent origin of the *Lonicera* fly by hybridization and host shift, the overall absence of detectable admixture linkage disequilibrium might be surprising. However, even admixture linkage disequilibrium between *Aat* and *Idh* (the two loci in our study separated by the shortest map distance) is expected to decay below detectable levels in <100 yr under complete isolation (see Supplementary Information). Given our results, it is unlikely that the *Lonicera* fly currently experiences substantial immigration from both parental taxa. We cannot, however, exclude asymmetric immigration from *R. zephyria*. With the current available genetic resolution it is not possible to distinguish between *R. zephyria* immigrants and *Lonicera* flies with *R. zephyria*-like genotypes formed by random segregation in an isolated population (Fig. 2).

The shift to honeysuckle probably freed the hybrid origin *Lonicera* fly from parental species competition, and provided a potential mechanism for reproductive isolation, as hypothesized above (our ability to produce apparently fit *R. mendax*/*R. zephyria* F$_1$ hybrids in the laboratory demonstrates the potential for natural hybridization in principle). *Rhagoletis* mates on the host, leading to an epistatic interaction between mate and host choice[7]. In *R. pomonella* a recent shift from native hawthorn to introduced apple has led to the evolution of partially reproductively isolated host races[7]. Studies on host choice, mate choice and host-specific fitness trade-offs of the *Lonicera* fly and its parental species are needed to validate our ecological assumptions, but the population genetic evidence presented above is consistent with the expected outcome of our proposed model. Given that hybridization seems to be frequent in animals[21] and that host-specific lifestyles could represent as much as 50% of animal diversity[8], the acquisition of hybrid-specific niches through host shift may be a phenomenon in need of deeper consideration. Anthropogenic changes offer new opportunities for hybridization, because previously geographically separated organisms come into contact due to human-mediated introductions[22]. At the same time, these community alterations provide access to potential new hosts and thereby the opportunity for habitat shifts. In this context it is important to note that the formation of the *Lonicera* fly is a novel example of how invasive weeds can influence the evolution of native fauna. Hybridization between parasites could widen the spectrum of potential new hosts by generating new phenotypes[9,22,23], a mechanism akin to the idea that plant hybrids

serve as a bridge for the acquisition of new hosts by herbivores[24]. Hybridization can be difficult to detect, especially in morphologically cryptic species, such as *Rhagoletis*, that may account for much of insect diversity. The availability of extensive genetic data for the *R. pomonella* species complex (including reports of local introgression between *R. pomonella* and *R. zephyria*[25]) is unusual compared to many host-specific organisms. It is likely that specialized parasite hybrid origin populations, such as the *Lonicera* fly, have gone undetected because of methodological difficulties and the traditional bias against hybridization as an evolutionary force in zoology[2]. Hybridization should be considered as a viable hypothesis for the origin of other host-specific animals, and we predict that future studies will discover more populations with a *Lonicera* fly-like evolutionary history.

METHODS

Sample collection. All samples were collected in central Pennsylvania, USA, between 2000 and 2002. The *Lonicera* fly was collected at one location in State College, Centre County (allozyme data from a second State College *Lonicera* fly sample are included in Table 1) and one location in Munson, Clearfield County. *R. zephyria* samples were taken from one location in State College and one location in Munson. One *R. mendax* sample was collected near Middleburg, Snyder County. Two local reference populations for *R. pomonella* were collected from two different hawthorn trees in State College. The five sampling sites in State College were separated by a maximum distance of 4 km from each other. Munson is located 30 km northwest of State College and the *Lonicera* fly and *R. zephyria* samples were separated by 0.7 km. The *R. mendax* sample location is 60 km east of State College. All samples consisted of larvae extracted from infested fruit, or pupae or adults reared from fruit. Larvae, pupae and adult flies were stored at $-80\,^{\circ}C$ until further analysis.

Population genetic data collection. Total genomic DNA was extracted from the head or one-third of each individual, whereas the remaining parts were used for allozyme analysis. Seven allozyme loci were examined by standard starch gel techniques[11,26]: β-hydroxyacid dehydrogenase (*Had*; E.C. 1.1.1.30), isocitrate dehydrogenase (*Idh*; E.C. 1.1.1.42), NADH-dependent diaphorase-2 (*Dia-2*; E.C. 1.6.2.2), aspartate aminotransferase-2 (*Aat-2*; E.C. 2.6.1.1), phosphoglucomutase (*Pgm*; E.C. 5.4.2.2), mannose-6-phosphate isomerase (*Mpi*; E.C. 5.3.1.8) and fumarate hydratase (*Fum*; E.C. 4.2.1.2). Two additional nuclear loci, developed from a complementary DNA library[16], were amplified by polymerase chain reaction (PCR) and scored for restriction length polymorphism. To confirm heterozygote genotypes we cloned a limited number of heterozygote individuals during the development of our protocol for the restriction digest. We further used restriction enzymes in excess and added previously scored individuals as positive controls to each new batch of digestions. P1700 (T3, 5′-ACATACATTCTGCATCTTGCGAAAG-3′; T7, 5′-TTAAGCCGACTTCTTC TTGAAACC-3′) was polymorphic at one restriction site for *Rsa*1. P2963 (T3, 5′-AGTCAACGACCTGCTTATTT-3′; T7, 5′-TGCACCTTAATTCACGAAAA TC-3′) was cut with *Alu*1, *Ase*1 and *Tsp*5091 at four restriction sites, and the haplotype inference software HAPLOTYPER was used to determine genotypes[27]. A 636-base pair (bp) piece of the mitochondrial *COII* coding region (C2-J-3136, 5′-CAAAATAGTGCCTCTCCC-3′; TK-N-3772, 5′-GAGACCATTACTTGCTTT CAGTCA-3′)[28] was amplified and sequenced. The variation within a 423-bp subsection of this sequence was used to characterize individual haplotypes.

Data analysis. Allele and haplotype frequencies were calculated using Arlequin ver. 2000 (ref. 29). To compare the allelic composition of the *Lonicera* fly and its parental species in central Pennsylvania, we collected all available published allozyme data for the parental species of the *Lonicera* fly as well as for *R. pomonella* (see references in Fig. 1). No population data were available for the two nuclear sequence-based markers and the mitochondrial DNA. Published allele frequencies at each locus were converted into allele counts, and a single allele frequency was recalculated for the North American superpopulation of each taxon. All samples gathered as part of this study were combined in a similar fashion. Multilocus genotypes representing *R. mendax* (Snyder County, $n = 36$), *Lonicera* fly (State College, $n = 30$; Munson, $n = 20$) and *R. zephyria* (State College, $n = 30$; Munson, $n = 20$) were analysed using STRUCTURE version 2 (ref. 17). All nuclear markers and mitochondrial DNA haplotypes were included except for the monomorphic *Dia-2* and *Fum*. *Fum* was excluded because it is in strong linkage disequilibrium with *Had* (Supplementary Table 2), and no map distance between these two markers is reported. The known map distances between *Aat-2*, *Pgm* and *Idh* on linkage group I were incorporated into the analysis by using the linkage model in STRUCTURE[17] (1,000,000 Markov chain Monte Carlo (MCMC) replicates and independent allele frequency model).
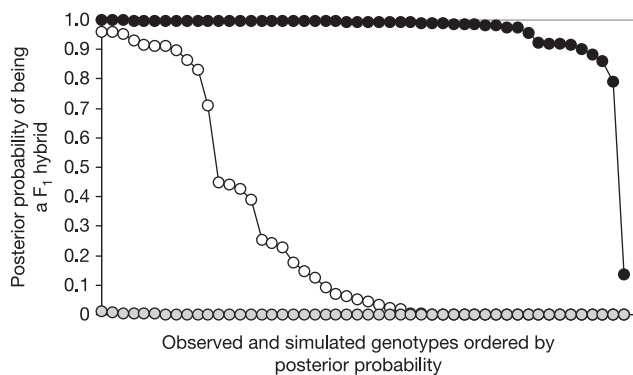
Figure 3 | Posterior probability for the assignment of observed *Lonicera* fly individuals as F$_1$ hybrids between *R. mendax* and *R. zephyria*. Data for observed *Lonicera* fly individuals (grey circles), simulated hybrid swarm (white circles) and simulated population of F$_1$ crosses between *R. mendax* and *R. zephyria* (black circles; see Supplementary Information for details on simulation) are shown.

Ancestry coefficients for membership in two populations were estimated without any prior knowledge of host-plant origin (that is, population membership). The same data set, with the exception of the mtDNA haplotypes, was used in the analysis with NewHybrids version 1.1 beta[30]. In the assignment of genotypes to hybrid classes that result from two generations of hybridization, observed *R. mendax* and *R. zephyria* genotypes were treated as being of known parental origin (100,000 MCMC replicates). A simulated population of 50 F_1 genotypes and 50 genotypes from a simulated hybrid zone were generated from the observed parental genotypes from *R. mendax* (Snyder County) and *R. zephyria* (State College and Munson; see Supplementary Information). Hardy–Weinberg and linkage equilibrium was tested using 100,000 permutations in Arlequin[29]. We had sample sizes large enough to allow for a meaningful test of linkage equilibrium between allozyme loci only in the *Lonicera* fly State College sample.

1. Coyne, J. A. & Orr, A. H. *Speciation* 337–342 (Sinauer, Sunderland, 2004).
2. Dowling, T. E. & Secor, C. L. The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* **28**, 593–619 (1997).
3. Salzburger, W., Baric, S. & Sturmbauer, C. Speciation via introgressive hybridization in East African cichlids? *Mol. Ecol.* **11**, 619–625 (2002).
4. Taylor, D. J., Hebert, P. D. & Colbourne, J. K. Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Mol. Phylogenet. Evol.* **5**, 495–510 (1996).
5. Hirai, H. *et al.* Chromosomal differentiation of the *Schistosoma japonicum* complex. *Int. J. Parasitol.* **30**, 441–452 (2000).
6. Pialek, J., Hauffe, H. C., Rodruigez-Clark, K. M. & Searle, J. B. Raciation and speciation in house mice from the Alps: the role of chromosomes. *Mol. Ecol.* **10**, 613–625 (2001).
7. Berlocher, S. H. & Feder, J. L. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu. Rev. Entomol.* **47**, 773–815 (2002).
8. Price, P. W. *Evolutionary Biology of Parasites* (Princeton Univ. Press, Princeton, 1980).
9. Rieseberg, L. H. *et al.* Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216 (2003).
10. Buerkle, A. C., Morris, R. J., Asmussen, M. A. & Rieseberg, L. H. The likelihood of homoploid hybrid speciation. *Heredity* **84**, 441–451 (2000).
11. Berlocher, S. H. Radiation and divergence in the *Rhagoletis pomonella* species group: Inferences from allozymes. *Evolution* **54**, 543–557 (2000).
12. Rehder, A. *Manual of Cultivated Trees and Shrubs Hardy in North America* 852–876 (Macmillan, New York, 1947).
13. United States Department of Agriculture. PLANTS Database ⟨http://plants.usda.gov/⟩ (2005).
14. Smith, J. J. & Bush, G. L. in *Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior* (eds Aluja, M. & Norrbom, A. L.) 187–217 (CRC Press, Boca Raton, 2000).
15. Foote, R. H., Blanc, F. L. & Norrbom, A. L. *Handbook of the Fruit Flies (Diptera: Tephritidae) of America North of Mexico* 322–364 (Comstock, Ithaca, 1993).
16. Roethele, J. B., Romero-Severson, J. & Feder, J. L. Evidence for broad-scale conservation of linkage map relationships between *Rhagoletis pomonella* (Diptera: Tephritidae) and *Drosophila melanogaster* (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am.* **94**, 936–947 (2001).
17. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
18. Barton, N. H. & Hewitt, G. M. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**, 113–148 (1985).
19. McPheron, B. A. & Berlocher, S. H. Segregation and linkage of allozymes of *Rhagoletis tabellaria. J. Hered.* **76**, 218–219 (1985).
20. Feder, J. L., Roethele, J. B., Filchak, K. E., Niedbalski, J. & Romero-Severson, J. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella. Genetics* **163**, 939–953 (2003).
21. Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005).
22. Fonseca, D. M. *et al.* Emerging vectors in the *Culex pipiens* complex. *Science* **303**, 1535–1538 (2004).
23. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
24. Floate, K. D. & Whitham, T. G. The "hybrid bridge" hypothesis: host shifting via plant hybrid swarms. *Am. Nat.* **141**, 651–662 (1993).
25. Feder, J. L., Williams, S. M., Berlocher, S. H., McPheron, B. A. & Bush, G. L. The population genetics of the apple maggot fly, *Rhagoletis pomonella* and the snowberry maggot, *R. zephyria*: Implications for models of sympatric speciation. *Entomol. Exp. Appl.* **90**, 9–24 (1999).
26. Shaw, C. R. & Prasad, R. Starch gel electrophoresis of enzymes—a compilation of recipes. *Biochem. Genet.* **4**, 297–320 (1970).
27. Niu, T., Qin, Z. S., Xu, X. & Liu, J. S. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **70**, 157–169 (2002).
28. Simon, C., Bekenbach, A., Crespi, B., Liu, H. & Flook, P. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
29. Schneider, S., Roessli, D. & Excoffier, L. *Arlequin ver. 2000 A Software for Genetic Data Analysis* (Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000).
30. Anderson, E. C. & Thompson, E. A. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229 (2002).

LETTERS

Robust habit learning in the absence of awareness and independent of the medial temporal lobe

Peter J. Bayley[1], Jennifer C. Frascino[1] & Larry R. Squire[1,2,3,4]

Habit memory is thought to involve slowly acquired associations between stimuli and responses and to depend on the basal ganglia[1]. Habit memory has been well studied in experimental animals but is poorly understood in humans because of their strong tendency to acquire information as conscious (declarative) knowledge. Here we show that humans have a robust capacity for gradual trial-and-error learning that operates outside awareness for what is learned and independently of the medial temporal lobe. We tested two patients with large medial temporal lobe lesions and no capacity for declarative memory. Both patients gradually acquired a standard eight-pair object discrimination task over many weeks but at the start of each session could not describe the task, the instructions or the objects. The acquired knowledge was rigidly organized, and performance collapsed when the task format was altered.

Declarative memory affords the capacity for conscious recollection of facts and events and depends on the integrity of the memory system based in the medial temporal lobe[2,3]. Declarative memory can be contrasted with a collection of phylogenetically early, nondeclarative (and unconscious) memory abilities, including skills and habits, simple forms of conditioning, and other forms of experience-dependent behaviour that are expressed through performance rather than conscious recollection. A large body of literature involving humans, monkeys and rodents can be understood by recognizing that tasks requiring declarative memory depend on the integrity of the hippocampus and related structures, whereas other tasks can be learned nondeclaratively and are supported by other brain systems[4–6].

A continuing uncertainty about this taxonomy, and about the organization of mammalian memory systems, follows from the observation that some memory tasks, which humans acquire as declarative knowledge through memorization, can be learned nondeclaratively by experimental animals[7]. In this instance, the same task that is performed poorly by amnesic patients with hippocampal lesions is acquired at a normal rate by monkeys with hippocampal lesions. These findings raise questions about the extent to which one memory system can substitute for another, and about whether humans have the same capacity for nondeclarative memory that nonhuman animals have. In humans, perhaps some forms of nondeclarative memory are not well developed or are so overridden by the tendency to engage conscious memory strategies that nondeclarative memory is ineffective.

The problem is well illustrated by concurrent discrimination learning, a standard task that has been used to study mammalian memory for more than 50 years[8,9]. In a common version of the task, eight pairs of objects are presented five times each day, one pair at a time in a mixed order, for a total of 40 trials. One object in each pair is always correct, and a choice of the correct object yields a reward. Humans readily learn this task, performing at about 90% correct after only one or two days of training. That the task ordinarily

depends on declarative memory, and on memorizing which object is correct in each pair, is demonstrated by the fact that task performance was correlated highly with the ability to describe the objects and by the fact that amnesic patients exhibited little learning during the period that their controls mastered the task[10,11].

In contrast to the findings from humans, monkeys learned the same concurrent discrimination task gradually in several hundred trials, and after hippocampal lesions they learned this task (or a related version) at a normal rate[7,12,13]. A standard interpretation of the monkey data is that monkeys acquire the concurrent discrimination task by trial-and-error (habit) learning with the support of the basal ganglia[1,13,14]. That is, they acquire a disposition to respond appropriately to each object pair. Habit memory is proposed to involve slowly acquired associations between stimuli and responses that develop outside awareness and are rigidly organized, with the result that what is learned is not readily expressed except when the task is presented just as it was during training.

One possibility is that habit learning is only weakly developed in humans, and some amount of declarative memory must be available to guide the learning. Indeed, successful learning of habit-like tasks has been reported only for moderately impaired amnesic patients who retain a considerable capacity for declarative memory[15,16]. Moreover, some amnesic patients do not learn such tasks[17]. A related view is that the concept of an independent habit system is unnecessary, because habit learning and other kinds of learning depend on similar mechanisms[18]. Alternatively, if the capacity for habit learning is as well developed in humans as it appears to be, for example, in the monkey, then patients with large medial temporal lobe lesions and no capacity for declarative memory should be able to acquire the concurrent learning task gradually and to a high level of proficiency in the same manner that monkeys learn the task.

We tested two males (patients EP and GP) who have large medial temporal lobe lesions[19] (see also Supplementary Information) and profound memory impairment as a result of herpes simplex encephalitis (Fig. 1). Four controls were also tested. Both patients fail altogether at formal tests of new learning ability, for example recall or recognition of word lists, stories and diagrams[20,21]. Further, they have acquired little, if any, new declarative knowledge about the world since the onset of their amnesia, and neither of them can draw an accurate floor plan of his current residence[22]. Their capacity for habit learning has not been explored.

Figure 2 shows that the controls learned the task easily during three days of testing. EP and GP learned gradually during 36 and 28 sessions, respectively. EP improved at a nearly linear rate from 45.0% correct to 85.0% correct (linear trend: $F_{1,35} = 54.7$, $P < 0.001$). GP improved from 55.0% to 92.5% correct (linear trend: $F_{1,27} = 13.6$, $P = 0.001$). During training, participants gave verbal labels to some of the objects.

The successful learning exhibited by the patients across weeks was
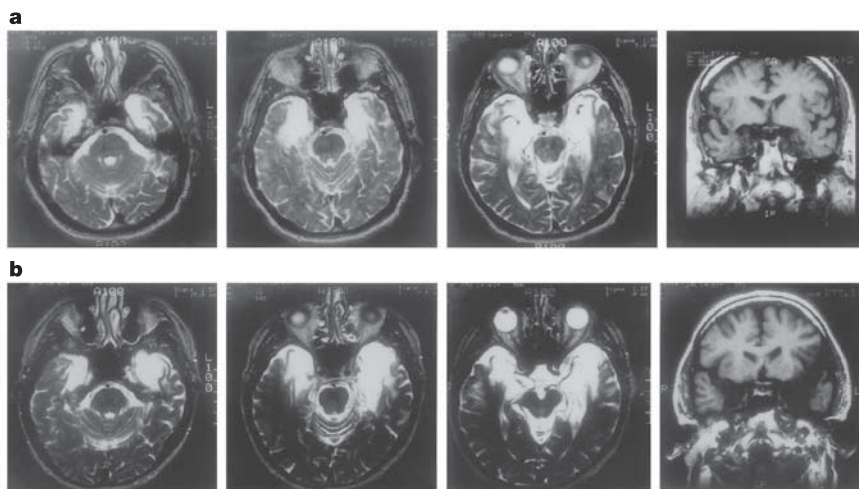
Figure 1 | **Magnetic resonance images showing the extent of damage to the temporal lobe in amnesic patients.** Images are shown for patient EP (**a**) and patient GP (**b**). The first three images in each row are T$_2$-weighted axial images through the temporal lobe. The images are continuous 5-mm sections (with 2.5-mm gaps) and are arranged from ventral to dorsal (left to right). Damaged tissue is indicated by a bright signal. The last image in each row is a coronal T$_1$-weighted image at the level of the amygdala. Damaged tissue is indicated by a dark signal. See ref. 19 and Supplementary Information for a detailed description of the lesions.

not accompanied by reportable knowledge about the nature of the task. Even during testing, they never recognized that they had encountered the same task in previous sessions. By the fourth session both patients, without specific instruction, began to pick up the objects and turn them over (to look for the word "correct"), but they never acquired declarative memory of what they did each day. Before each session, both patients were told that they had been participating in testing on a regular basis, and they were asked to describe the testing. EP consistently replied that he would be asked questions and have discussions, but he never mentioned objects or described a learning task. When he was then told that the testing involved objects, EP regularly stated that he was to describe the objects and tell what they were used for. GP was able to report a little more about the testing. Before every session except one, he stated that he would look at photographs and answer questions such as "What's missing?" or "What doesn't belong?" On 11 occasions he said that he was to "pick the correct one". On 10 occasions he also mentioned objects, but usually misdescribed what he was to do (for example, "choose the one that doesn't belong"; "point out what's missing"). On four different occasions he correctly stated that he was to pick the correct object, but he was unable to describe any of them.

Comments offered by the patients during and at the end of each test day described an automatic kind of responding. For example, EP was regularly asked at the conclusion of testing whether he had a strategy for choosing the correct object. After session 34, he said "No. It's just up here [pointing to head] from the memory. It seems like it's up there, and it comes down and out." [Examiner: "Do you say to yourself, 'I remember seeing that one?'"]. "No. It just seems that's the one. It's here (pointing to head) somehow or another and the hand goes for it." After session 32, when asked if he had a strategy, EP replied, "No. It seems [pause] like from memory, or I'm just recalling this is the one and I picked it up. I can't say memory. I just feel this is the one." During session 23, GP stated, "I'm missing the same one." [Examiner: "Why are you selecting that one?"] "It's just jumping out at me. 'I'm the one. I'm the one.' I keep wanting to pick it."

To test whether the knowledge that had been acquired was rigidly organized, as is thought to occur for habit learning[1,23], or whether it could be used flexibly, we administered a sorting task 3–6 days after the conclusion of formal training. Controls succeeded with no difficulty, scoring 95.3% correct (Fig. 2a). In contrast, EP and GP failed the sorting task altogether (56.3% and 50.0% correct,

respectively; chance performance = 50%; Fig. 2b, c). EP placed nine objects in the group meant to contain correct objects and seven objects in the other group. GP placed eight objects in each group. Both patients expressed uncertainty and made exclamations when they first saw the array of 16 objects. EP said, "Gosh sakes. How to remember this?" When he moved to look underneath one of the objects, and was stopped by the examiner, he said "That's just a habit, I think. It's just, what's underneath?"

Immediately after the sorting task, the familiar task format was reinstated (pairs of objects one at a time for 40 trials) but now participants were instructed to verbalize their choices ("left" or "right") rather than pick up the objects. All participants scored 90% correct or better (Fig. 2). Immediately thereafter, each correct object was paired with an incorrect object, different from the one it had been paired with during training, and 40 additional test trials were given (re-pairing). The performance of EP and GP declined only a little, from 90.0% to 85.0% correct and from 92.5% to 75.0% correct, respectively (controls = 99.4% correct). These findings agree with the results of a recent study[24] in which rats with hippocampal lesions, trained on three visual discrimination problems, exhibited perfect transfer when the stimuli were recombined into novel pairs. It seems that re-pairing of stimuli in such tasks does not effectively challenge the ability to express knowledge flexibly, and simple associative processes can guide choice behaviour (for example, choose the object in each pair with the greatest positive value or least negative value).

Three days later, both patients were tested again with the standard 40-trial procedure but now no instructions were given. The first pair of objects was presented, and the examiner simply said that participants should do what comes naturally. EP picked up an object, turned it over to reveal the word "correct", and said, "that's the natural thing to me." Similarly, GP said, "the correct one", then picked up the correct object and turned it over. EP and GP went on to obtain scores of 95% and 100% correct, respectively. After the test was concluded, EP was asked how he knew what to do. "It seems to be automatic [pointing to head and then to table]. My mind just seemed to tell me, 'just pick it up, it's the right one.'" Finally 17 days after its first administration, each patient was again given the sorting task, followed immediately by 40 trials of the standard task (Fig. 2). The results were the same as before. EP and GP failed the sorting task (56.3% and 50.0% correct, respectively), but succeeded at the
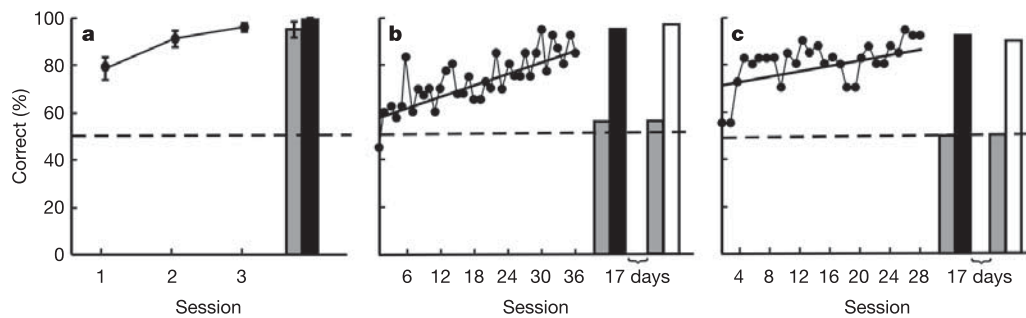
Figure 2 | Performance on the concurrent discrimination task. a, Controls ($n = 4$) learned the task easily within three sessions and performed well on the sorting task 3–6 days later (grey bar). The black bar shows performance immediately afterwards when participants were asked to verbalize their choices rather than reach for objects. Results are means ± s.e.m. **b,** EP gradually learned the object pairs across 18 weeks. Five days later, he failed the sorting task (grey bar) but then, immediately afterwards, performed well in the standard task format while verbalizing his responses (black bar). Seventeen days later, EP again failed the sorting task (grey bar) but performed perfectly when 40 trials were given exactly as in original training (white bar). **c,** GP learned the object pairs gradually during 14 weeks. Like EP, he failed the sorting task on two different occasions, 5 days after training and again 17 days later. In both instances he performed well immediately afterwards when the original task format was reinstated (black bar, verbalizing; white bar, standard task). The dashed line indicates the results expected by chance.

standard task (97.5% and 90.0% correct, respectively). EP placed 11 objects in the group meant to contain correct objects and 5 objects in the other group. GP placed 8 objects in each group. Whereas the two patients performed well on the standard task and at chance on the sorting task, the four controls performed nearly perfectly on both tasks. Control scores were also similar on the two tasks in a second study when performance was reduced by testing 4–7 days after only 1 day of training (40 trials) on a new set of eight object pairs (standard task 87.5% correct; sorting task 79.7% correct; $t_3 = 1.7$, $P = 0.19$).

These findings show directly the existence of a robust capacity for habit learning that can operate outside awareness and independently of declarative memory. The information acquired by both patients was rigidly organized and most accessible when the task was structured just as it was during training. Neither the instructions nor the response requirements (verbalizing versus reaching for objects) were critical, but performance did depend on the presentation of pairs of objects and on the requirement to choose one object in each pair. Thus, performance collapsed entirely in the sorting task when all the objects were encountered at once.

If the patients did learn the concurrent discrimination task as a habit in the manner that monkeys learn the task (that is, very gradually through reinforcement and with the support of the basal ganglia rather than the medial temporal lobe), then one might expect that they should have learned in about the same number of trials as monkeys with large medial temporal lobe lesions. The data support this expectation. Four monkeys with large medial temporal lobe lesions learned this same concurrent discrimination task (five sessions/week rather than two) in a mean of 1,100 trials[7]. EP and GP first reached a score of 95% correct after 1,200 and 1,040 trials, respectively. This comparison is not straightforward because normal monkeys acquired the task in about 500 trials, and monkeys that required more trials, including the four animals just mentioned, had damage that extended laterally to include some lateral temporal cortex[7]. Still, both patients also have some lateral temporal damage (in the fusiform and insular cortices) that could have slowed their learning. Accordingly, it seems reasonable to suppose that the rate of learning shown by the patients approximated the learning rate of monkeys with similar lesions.

The present results provide particularly useful evidence for the distinction between conscious (declarative) and unconscious (non-declarative) learning systems. The biological reality of this distinction has sometimes been challenged[25,26]. For example, amnesic patients might have conscious knowledge about a successfully performed task but forget it by the time they are queried. Further, task information that is acquired successfully might be easier to acquire or be presented more often than task information that is not acquired. The present findings sweep aside these objections, because towards the end of training both patients regularly performed at high levels immediately after failing to describe the task, the instructions and the objects. The findings therefore affirm the validity of unconscious learning and show that humans possess a robust capacity for gradual, trial-and-error (habit) learning that can operate outside awareness for what is learned and independently of the medial temporal lobe. It seems likely that many tasks, including concurrent discrimination, which humans ordinarily acquire rapidly as declarative knowledge (through memorization) can also be acquired more slowly as habit memory. However, in such cases, what is acquired is rigidly organized and altogether different from declarative memory.

METHODS

Subjects. EP was born in 1922, had 12 years of education, and developed amnesia in 1992. GP was born in 1946, had 16 years of education, and developed amnesia in 1987. Estimates of brain damage were based on a quantitative analysis of magnetic resonance images[19] (also see Supplementary Information), following published procedures for segmenting the medial temporal lobe[27–29]. Volume estimates were also calculated for the lateral temporal, frontal, parietal and occipital lobes. Four controls averaged 70.2 years of age and had 13.2 years of education.

The concurrent discrimination task. Eight pairs of junk objects (miscellaneous pieces of plastic or metal that were not readily nameable, each mounted on 8.9 cm × 12.7 cm cardboard) were first presented one at a time. Participants were asked to choose one of the objects in each pair, and the object not chosen was then designated the 'correct' object for the duration of testing. Formal testing began 6 or 7 days later with the presentation of all eight pairs of objects five times each (session 1). Participants were told that the same object in each pair would always be correct and that they should try to learn the correct objects by trial and error. Specifically, they were told to make a choice between each pair of objects and then pick up the object, turn it over, and discover whether the word "correct" appeared under the cardboard base. Testing continued in this manner twice each week on nonconsecutive days (40 trials per session). The order in which the object pairs were presented varied each day with the constraint that all eight pairs were presented within each block of eight trials. The left–right position of the correct object also varied pseudorandomly.

The sorting task. All 16 objects were placed together on a table, and participants were told that some of the objects had been consistently designated as correct. They were then asked to sort the objects, placing the correct objects to one side of the table and the other objects to the other side.

1. Mishkin, M. & Petri, H. L. in *Neuropsychology of Memory* (eds Squire, L. R. & Butters, N.) 287–296 (Guilford, New York, 1984).
2. Squire, L. R. & Zola-Morgan, S. The medial temporal lobe memory system. *Science* **253**, 1380–1386 (1991).

3. Eichenbaum, H. & Cohen, N. J. *From Conditioning to Conscious Recollection: Memory Systems of the Brain* (Oxford Univ. Press, New York, 2001).
4. Squire, L. R. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
5. Squire, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* **82**, 171–177 (2004).
6. Schacter, D. L. & Tulving, E. *Memory Systems 1994* (MIT Press, Cambridge, Massachusetts, 1994).
7. Buffalo, E. A., Stefanacci, L., Squire, L. R. & Zola, S. M. A reexamination of the concurrent discrimination learning task: the importance of anterior inferotemporal cortex, area TE. *Behav. Neurosci.* **112**, 3–14 (1998).
8. Hayes, K. J., Thompson, R. & Hayes, C. Concurrent discrimination learning in chimpanzees. *J. Comp. Physiol. Psychol.* **46**, 105–107 (1953).
9. Correll, R. E. & Scoville, W. B. Effects of medial temporal lesions on visual discrimination performance. *J. Comp. Physiol. Psychol.* **60**, 175–181 (1965).
10. Squire, L. R., Zola-Morgan, S. & Chen, K. Human amnesia and animal models of amnesia: performance of amnesic patients on tests designed for the monkey. *Behav. Neurosci.* **11**, 210–221 (1988).
11. Hood, K. L., Postle, B. R. & Corkin, S. An evaluation of the concurrent discrimination task as a measure of habit learning: performance of amnesic subjects. *Neuropsychologia* **37**, 1375–1386 (1999).
12. Malamut, B. L., Saunders, R. C. & Mishkin, M. Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behav. Neurosci.* **98**, 759–769 (1984).
13. Teng, E., Stefanacci, L., Squire, L. R. & Zola, S. M. Contrasting effects on discrimination learning following hippocampal lesions or conjoint hippocampal-caudate lesions in monkeys. *J. Neurosci.* **20**, 3853–3863 (2000).
14. Fernandez-Ruiz, J., Wang, J., Aigner, T. G. & Mishkin, M. Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. *Proc. Natl Acad. Sci. USA* **98**, 4196–4201 (2001).
15. Knowlton, B. J., Squire, L. R. & Gluck, M. Probabilistic classification learning in amnesia. *Learn. Mem.* **1**, 106–120 (1994).
16. Myers, C. E. *et al.* Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J. Cogn. Neurosci.* **15**, 185–193 (2003).
17. Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S. & Gluck, M. A. Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia* **42**, 524–535 (2004).
18. Gaffan, D. Memory, action and the corpus striatum: current developments in the memory-habit distinction. *Semin. Neurosci.* **8**, 33–38 (1996).
19. Bayley, P. J., Gold, J. J., Hopkins, R. O. & Squire, L. R. The neuroanatomy of remote memory. *Neuron* **46**, 799–810 (2005).
20. Stefanacci, L., Buffalo, E. A., Schmolck, H. & Squire, L. R. Profound amnesia following damage to the medial temporal lobe: a neuroanatomical and neuropsychological profile of patient E.P. *J. Neurosci.* **20**, 7024–7036 (2000).
21. Levy, D. A., Bayley, P. J. & Squire, L. R. The anatomy of semantic knowledge: medial vs. lateral temporal lobe. *Proc. Natl Acad. Sci. USA* **101**, 6710–6715 (2004).
22. Bayley, P. J. & Squire, L. R. Failure to acquire new semantic knowledge in patients with large medial temporal lobe lesions. *Hippocampus* **15**, 273–280 (2005).
23. Wise, S. P. The role of the basal ganglia in procedural memory. *Sem Neurosci.* **8**, 39–46 (1996).
24. Driscoll, I., Sutherland, R. J., Prusky, G. T. & Rudy, J. W. Damage to the hippocampal formation does not disrupt representational flexibility as measured by a novelty transfer test. *Behav. Neurosci.* **118**, 1427–1432 (2004).
25. Shanks, D. R. & St.John, M. F. Characteristics of dissociable human learning systems. *Behav. Brain Sci.* **17**, 367–447 (1994).
26. Nosofsky, R. M. & Zaki, S. R. Math modelling, neuropsychology, and category learning: Response to B. Knowlton (1999). *Trends Cogn. Sci.* **3**, 125–126 (1999).
27. Insausti, R. *et al.* MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR Am. J. Neuroradiol.* **19**, 659–671 (1998).
28. Amaral, D. G. & Insausti, R. in *The Human Nervous System* (ed. Paxinos, G.) 711–755 (Academic, San Diego, 1990).
29. Gold, J. J. & Squire, L. R. Quantifying medial temporal lobe damage in memory-impaired patients. *Hippocampus* **15**, 79–85 (2005).

LETTERS

Molecular recycling within amyloid fibrils

Natàlia Carulla[1]†, Gemma L. Caddy[1], Damien R. Hall[1], Jesús Zurdo[1], Margarida Gairí[2], Miguel Feliz[2], Ernest Giralt[3,4], Carol V. Robinson[1] & Christopher M. Dobson[1]

Amyloid fibrils are thread-like protein aggregates with a core region formed from repetitive arrays of β-sheets oriented parallel to the fibril axis[1,2]. Such structures were first recognized in clinical disorders[1,3], but more recently have also been linked to a variety of non-pathogenic phenomena ranging from the transfer of genetic information to synaptic changes associated with memory[4–7]. The observation that many proteins can convert into similar structures *in vitro* has suggested that this ability is a generic feature of polypeptide chains[1,8]. Here we have probed the nature of the amyloid structure by monitoring hydrogen/deuterium exchange in fibrils formed from an SH3 domain[9–12] using a combination of nuclear magnetic resonance spectroscopy and electrospray ionization mass spectrometry. The results reveal that under the conditions used in this study, exchange is dominated by a mechanism of dissociation and re-association that results in the recycling of molecules within the fibril population. This insight into the dynamic nature of amyloid fibrils, and the ability to determine the parameters that define this behaviour, have important implications for the design of therapeutic strategies directed against amyloid disease.

Hydrogen/deuterium (H/D) exchange, which measures the degree of protection of labile hydrogens against exchange with solvent deuterons as a function of time, is a powerful method for studying the stability, dynamics and folding of globular proteins[13]. Recent applications of this approach to amyloid fibrils have shown that valuable structural information can be obtained about regions of the polypeptide chains that are involved in persistent hydrogen-bonded structure or are excluded from solvent, providing a way to map the core structure of fibrillar assemblies[14–16].

We used this method with amyloid fibrils formed from the SH3 domain of the α-subunit of bovine phosphatidylinositol-3′-kinase (PI(3)K-SH3)[9–12]. Homogeneous populations of PI(3)K-SH3 amyloid fibrils were prepared using a protocol based on seeded growth (Methods; see also Supplementary Methods and Supplementary Fig. S1)[17]. The fibril preparations were exposed to deuterated buffer for varying periods of time and then solubilized into monomers by transfer to a dimethylsulphoxide (DMSO) solution. The site-specific hydrogen exchange behaviour of PI(3)K-SH3 within amyloid fibrils was analysed by nuclear magnetic resonance spectroscopy (NMR). After a 4-day incubation in D_2O at pH* 1.6 (the pH meter electrode reading in the deuterated solvent without correction for the isotope effect), the average extent of exchange of the amide hydrogens is $30 \pm 15\%$; this value rises only to $38 \pm 16\%$ after 15 days (Fig. 1). This result shows that PI(3)K-SH3 molecules are highly resistant to exchange over time when incorporated into fibrils; by contrast, monomeric PI(3)K-SH3 molecules are unfolded under these conditions and exchange fully with the solvent within minutes. The amide hydrogens from the majority of residues within the fibrils

exchange to a similar degree (ranging from 25% to 45% after 15 days), but two residues at the amino terminus (Met 3 and Ser 4) and six at the carboxy terminus (all but Lys 82 of the residues from Ile 79 to Ser 85) show a significantly greater extent of exchange (Fig. 1). In addition, several other amide hydrogens distributed throughout the sequence appear to exchange at faster than the average rate.

This pattern of exchange suggests that the most stable region of the amyloid fibrils encompasses much of the central region of the protein sequence (Ala 5 to Tyr 78). However, Fourier transform infrared (FTIR) studies on PI(3)K-SH3 amyloid fibrils[11] and an outline structure put forward on the basis of cryo-electron microscopy and X-ray fibre diffraction data[9] indicate that only about 40% of the protein sequence is contained within the fibril β-sheet core; the remainder forms long connecting loops and turns. Because most of the amide hydrogens are protected from solvent exchange, the present data suggest that even such regions of structure are tightly packed within the fibrillar structure, a conclusion consistent with
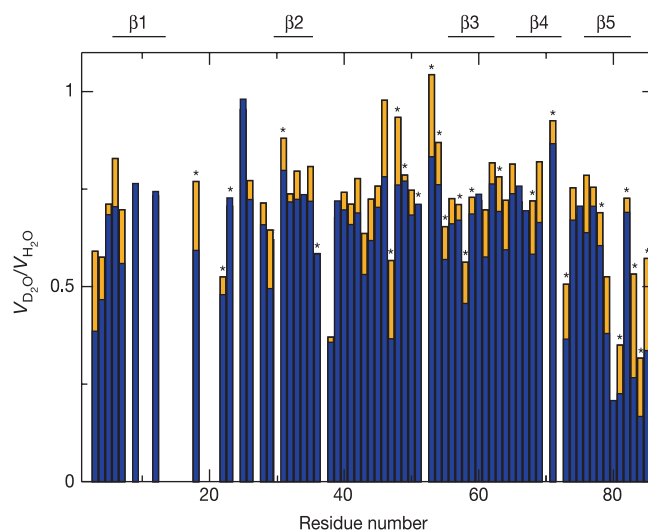
Figure 1 | H/D exchange of PI(3)K-SH3 amyloid fibrils monitored by NMR spectroscopy. The peak volumes in ^1H-^{15}N heteronuclear single quantum correlation (HSQC) spectra recorded after H/D exchange (V_{D_2O}) relative to those before exchange (V_{H_2O}) after dissolution of the fibrils (see main text) are plotted against residue number. The bars correspond to the relative peak volumes in the spectra of fibrils exchanged for 4 (orange) and 15 (blue) days. An asterisk above a bar indicates a residue whose resonance is not fully resolved; the absence of a bar indicates that the resonance of the residue is not detectable (Supplementary Fig. 2S). The locations of the β-sheet strands in the native state of PI(3)K-SH3 are indicated.

[1]University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK. [2]Serveis Cientificotècnics, Parc Científic de Barcelona, Baldiri Reixac 10-12, Barcelona 08028, Spain. [3]Institut de Recerca Biomèdica de Barcelona, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain. [4]Universitat de Barcelona, Departament de Química Orgànica, Martí i Franqués 1, Barcelona 08028, Spain. †Present address: Institut de Recerca Biomèdica de Barcelona, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain.

the finding that PI(3)K-SH3 molecules within amyloid fibrils are essentially completely resistant to cleavage by proteolytic enzymes[11,18]. We observed no clear correlation between the location of β-sheet structure in the native state and the degree of protection of


specific residues of the protein within the amyloid fibrils. In fact, amide hydrogens in the C-terminal region that exchange most readily in the fibrils are located in one of the five β-strands (β5) of the native structure (Fig. 1).

To analyse the H/D exchange behaviour in the fibrils further, we used electrospray ionization mass spectrometry (ESI-MS), a complementary method to NMR because it can detect and characterize populations of molecules with different degrees of exchange[19]. The mass spectra show two extremely well-resolved peaks after solubilization of each sample of amyloid fibrils that was exposed to exchange (Fig. 2a). This result indicates that two distinct, isotopically labelled populations of PI(3)K-SH3 are present within the amyloid fibrils. The lower mass peak (P_{pd}, for partially deuterated) has an average mass that is 74.1 ± 3.3 Da greater than that of the protein in the absence of deuterium, representing a population of molecules in which about one-half the total number of hydrogens at labile sites has exchanged for deuterons. The higher mass peak (P_{fd}, for fully deuterated) has an average mass 133.7 ± 3.7 Da greater than that of the fully protonated protein; this species has therefore exchanged most of its labile hydrogens. Although small variations in the masses of the two peaks occur over the time course of the experiment, the average mass difference (59.6 ± 2.2 Da) is constant.

When the same experiment was carried out with non-exchanged PI(3)K-SH3 amyloid fibrils, a single peak was observed that had an average mass that was increased by 65.0 ± 3.4 Da as a result of the rapid exchange of very labile side-chain hydrogens of the protein molecules with the deuterons of the solubilizing solvent, 95% DMSO-d_6/5% D_2O. The average mass of the P_{pd} species discussed above is therefore approximately 9 Da more than the average mass obtained in this experiment, indicating that about nine of the amide hydrogens of the protein molecule within the amyloid structure are directly accessible to solvent. These data are highly consistent with the NMR results that showed that a small number of amides located at the N and C termini of the protein in particular are able to exchange faster than the rest.

When the fibrils were left to exchange for longer periods of time before mass spectrometric analysis, the intensity of the P_{fd} species increases relative to that of the P_{pd} species (Fig. 2). After 19 days, $32.9 \pm 0.5\%$ of the molecules within the fibrils are fully deuterated, a result broadly consistent with the NMR data (Fig. 1); after 49 days, about 40% of the molecules appear as P_{fd} species. The observation of two highly resolved peaks in the mass spectrum of each sample (Fig. 2) provides direct evidence that the fibril population is remarkably homogeneous; if there had been significantly different types of aggregates within a sample, the differences in their exchange behaviour[20] would have caused broadening of the peaks representing P_{pd} and P_{fd} and/or the appearance of peaks corresponding to masses intermediate between P_{pd} and P_{fd}[21]. This conclusion is supported by electron micrographs that show no evidence of species other than the well-defined fibrils of the type studied by cryo-electron microscopy and X-ray fibre diffraction[9] either before (Fig. 3a) or after (Fig. 3b) the exchange reaction was performed.

To account for the presence of distinct isotopic populations in the
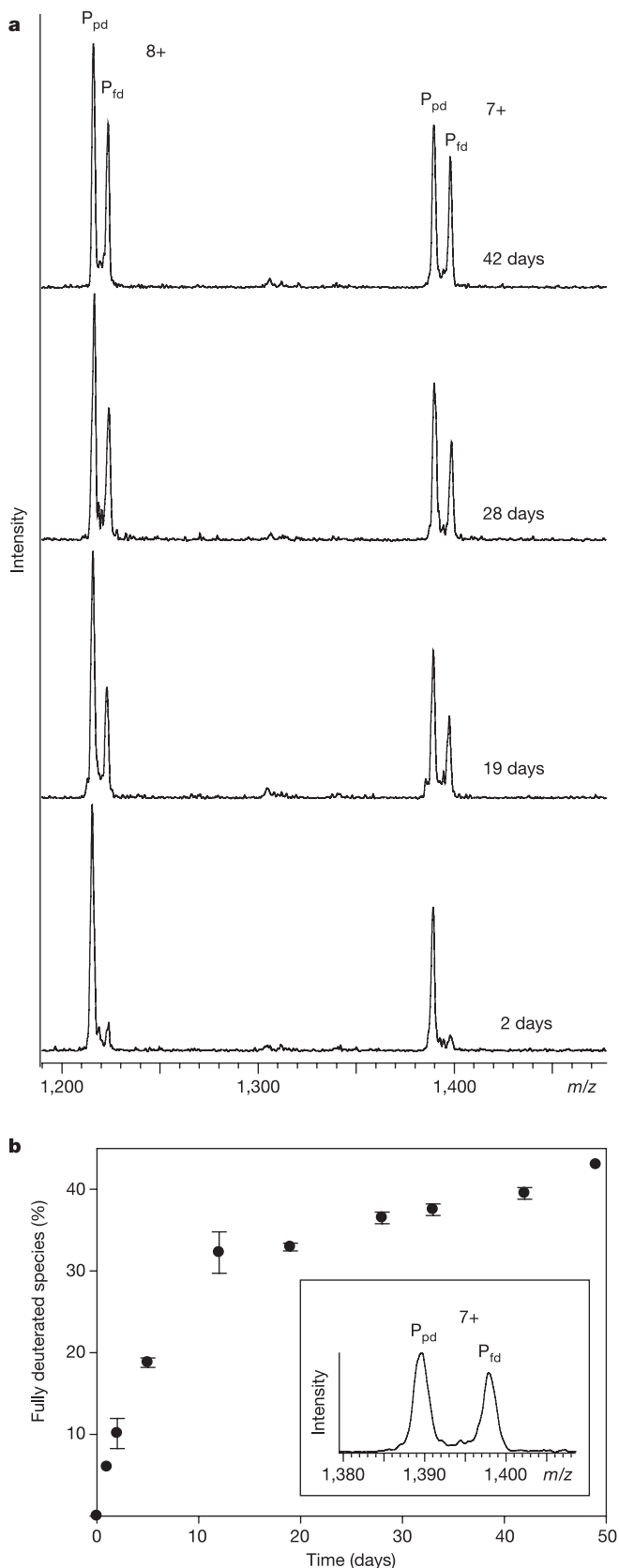
Figure 2 | **H/D exchange of PI(3)K-SH3 amyloid fibrils analysed by ESI-MS. a**, Mass spectra (8+ and 7+ charge states) showing the relative populations of P_{pd} and P_{fd} species after exchange for different times. The spectral intensities are all plotted relative to the same intensity of the P_{pd} 8+ charge state. **b**, Plot of the relative fraction of P_{fd} molecules in the sample as a function of time of exchange. Error bars indicate standard deviations resulting from duplicate samples. The inset shows the expansion of the spectrum corresponding to the 7+ charge states of the P_{pd} and P_{fd} species after 42 days. The peak widths of the P_{pd} and P_{fd} species (3.0 ± 0.14 and 2.9 ± 0.14 Da, respectively) are comparable to that observed in spectra of solutions in which the soluble monomeric PI(3)K-SH3 protein was allowed to exchange in 95% DMSO-d_6/5% D_2O solvent (2.9 ± 0.07 Da).

mass spectra, we considered a variety of models (for discussion see Supplementary Information). The only mechanism that seems to be consistent with all the data is one in which the ensemble of molecules within the fibrils is in dynamic equilibrium with a pool of soluble protein molecules. In the simplest form of this recycling model, protein molecules dissociate from the fibrils with a rate constant k_{on} and once in solution hydrogen exchange occurs rapidly with a rate constant k_{ex}. The protein molecules are then reincorporated into the fibrils with a rate constant k_{on} (Fig. 4a). In a mass spectrometric analysis of the solubilized fibrils, two peaks would then be observed: P_{pd}, which represents the population of molecules that has not yet dissociated from the fibrils, and P_{fd}, which represents the population that has dissociated from the fibrils, exchanged and then been reincorporated. In support of this mechanism we have shown, using ESI-MS, that there is always a significant quantity of soluble protein in the fibril-containing solution (see Supplementary Methods and Supplementary Fig. S3).

In this model, the initial slope of a plot of isotope exchange against time (Fig. 2b) represents k_{off}, the average dissociation rate constant of molecules from the fibrils, because in the initial stages of the reaction all molecules that dissociate from the fibrils will undergo isotope exchange. From the data shown in Fig. 2b, this value is estimated to be $1.4 \times 10^{-4} s^{-1}$, assuming that the average fibril length is 1 μm, a value consistent with the electron microscopic data (Fig. 3b), and, for simplicity, that exchange occurs only from one end of each fibril. As the extent of exchange increases, however, an increasing fraction of molecules within any given amyloid fibril will already have undergone exchange and reincorporation in a deuterated form. Dissociation of these molecules does not increase the overall fraction of P_{fd} species and is manifest as a marked slowing of the net rate of isotope exchange (Fig. 4b).

A distinguishing feature of the recycling model is that the kinetic parameters defining the rates of association and dissociation are directly related to the distribution of fibril lengths. As the latter can be estimated from electron microscopy images, this relationship enables the essential features of the recycling mechanism to be tested. To do so, a model of the exchange process was constructed and a series of simulations was carried out (see Supplementary Discussion and Supplementary Table S1). The simulations reproduce the form of the experimental time course of the exchange reaction well, although they underestimate the extent of deuterium incorporation (Fig. 4d). Likely reasons for the low estimate of the degree of exchange on this highly simplified model include: a lack of precision in estimating the average fibril length; the assumption that exchange occurs at just one end of the fibril[22] when it may occur at both[23]; the possibility that fibrils may break to create new fibril ends[24]; and the possibility that the fibril distribution may change during the course of the H/D exchange reaction. All of these assumptions are discussed in the Supplementary Information.

As an additional test of the validity of our conclusions, fibril preparations were subjected to procedures designed to generate a more uniform distribution of fibril lengths. A single batch of solution containing amyloid fibrils was divided into three portions, and each one was exposed to H/D exchange conditions. One portion (A) was left undisturbed during exchange except when samples were removed for analysis at varying time points. A second portion (B) was subjected to mild shear forces through repetitive pipetting at these time points (data shown in Fig. 2); the third portion (C) was sonicated at the various time points. Comparison of electron micrographs of each of the fibril suspensions before (Fig. 3a) and after the incubation process in which H/D exchange was allowed to take place, reveals that little fibril breakage results from mechanical manipulation (B) (Fig. 3b) but that sonication (C) results in very extensive fibril breakage, which generates a population of fibrils with very uniform dimensions (Fig. 3c). Moreover, the H/D exchange results show that at all time points the extent of deuteration is significantly greater in the fibril suspensions where

breakage was achieved by sonication, particularly at longer times (Fig. 3g).

The average length of the fibrils after sonication is approximately 100 nm (Fig. 3c), a value corresponding to about 40 molecules per filament (Fig. 4c). In this case, the experimental exchange rates agree well with the predictions of the model using the range of k_{on} and k_{off} values that best describe the PI(3)K-SH3 system (for example, after 30 days, the predicted degree of exchange with $k_{on} = 1.4 M^{-1} s^{-1}$ and $k_{off} = 7.1 \times 10^{-5} s^{-1}$ is 55%, compared with the experimental value of $58 \pm 2.1\%$; Fig. 4d). We attribute this good agreement to the fact that sonication produces a uniform distribution of fibril lengths. Additional support in favour of the recycling model comes from the observation that the fibrils in the sonicated samples increase
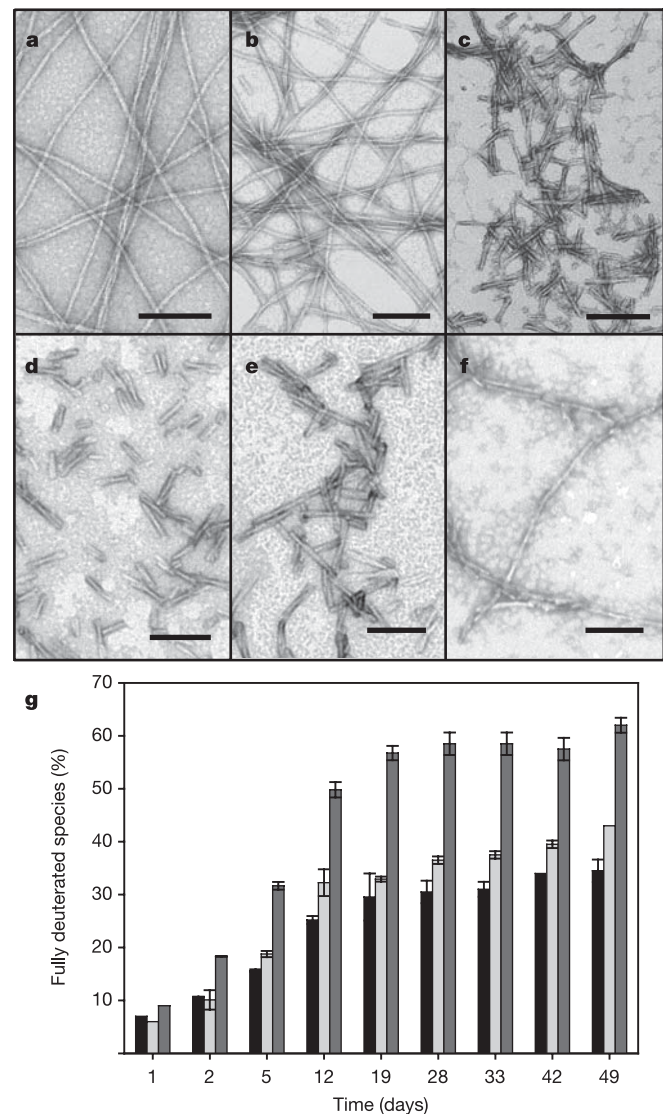
Figure 3 | Electron micrographs and H/D exchange data for fibril suspensions subjected to different processing strategies. a–c, Images of fibrils collected before exchange (**a**) and after 49 days of exchange during which the samples were subjected to repetitive pipetting (**b**) or sonication (**c**). **d–f,** Images of fibrils collected after 7 days of repeated sonication (**d**), and 7 days (**e**) and 48 days (**f**) after the end of the sonication procedure. Scale bars, 100 nm. **g,** Histogram of the percentage of fully deuterated species, P_{fd}, as a function of time of exchange for fibril suspensions that were untreated (black bars), subjected to repetitive pipetting (light grey bars) or sonication (dark grey bars). Note that the x axis represents the times at which different samples were analysed and therefore is nonlinear. Error bars indicate standard deviations resulting from duplicate samples.

markedly in length when left undisturbed for prolonged periods of time (Figs 3d–f) even though no additional monomeric protein has been added to the fibril suspension. We therefore conclude that hydrogen exchange under the conditions used here occurs predominantly when the protein molecules are at least transiently dissociated from the fibrils.

The rate constants for molecular association and dissociation determined in the present study indicate that the time for recycling half of the molecules within an ensemble of fibrils with an average length of 100 nm is between 2 and 20 days. An important aspect of this conclusion is that these dynamic processes are taking place on timescales that are relevant to biology. If the rates of molecular association and dissociation had been much faster, then thermodynamic considerations alone would be completely adequate for understanding the nature of fibril deposition under physiological conditions. If the rates had been very much slower, the exchange process would not have been relevant to the behaviour of living systems. The present findings suggest that it will be important to explore both the kinetics and thermodynamics of fibril formation

and dissolution to understand their biological properties[1,25,26]. Such considerations are of particular importance in the development of therapeutic strategies for prevention or treatment of protein aggregation diseases, as most of these strategies are ultimately dependent on the relative magnitudes of the rates at which the aggregates form or are degraded[27].

The ability to observe the dynamical behaviour of amyloid fibrils by combining NMR and ESI-MS not only offers new insight into the properties of these intriguing molecular assemblies but provides an opportunity to obtain quantitative estimates of the various rate constants that determine fibril growth and dissolution, and hence to investigate the factors that determine such properties. Measurements of the type described here have the potential to permit detailed investigation of the specific factors that determine the nature of fibril dynamics under particular solution conditions. As well as their intrinsic interest, these factors represent fundamental parameters that underlie a wide range of phenomena including the production and behaviour of novel biomaterials[28,29], the exploitation of biomolecular assemblies in living systems[6,7], the onset and development of
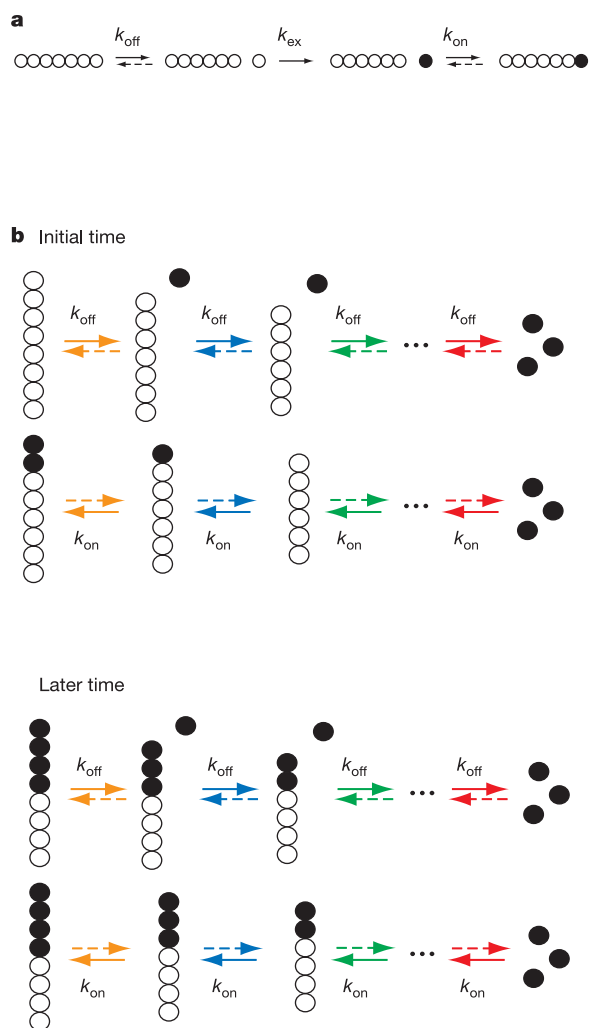
Figure 4 | The molecular recycling model. a, Schematic representation of the recycling process involving a single amyloid fibril. A protonated protein molecule (open circle) dissociates from the fibril. Once in solution, hydrogen exchange takes place rapidly and the molecule is subsequently reincorporated in a fully deuterated state (filled circle). **b**, Schematic representation of the recycling mechanism for a distribution of amyloid fibrils at different times of exchange. At a given time point the dissociation of a molecule from a fibril (above) is counteracted by the re-association of that molecule in another fibril (below). **c**, Simulated equilibrium distributions
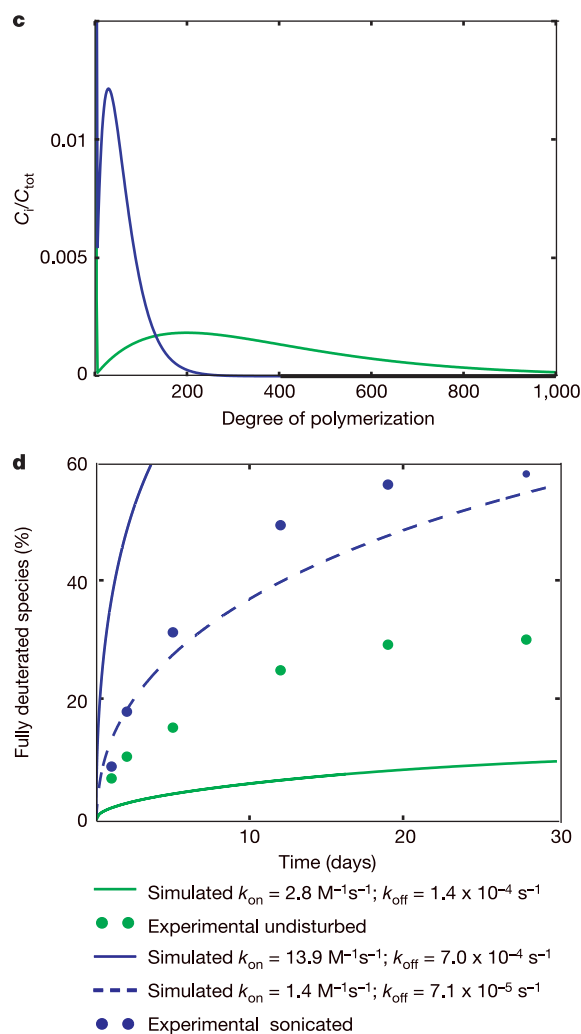
corresponding to average degrees of polymerization (that is, average number of molecules in the fibrils) of 40 (blue) and 400 (green), approximately the numbers of molecules that are expected in a single filament of length 100 nm and 1 μm, respectively (see Supplementary Discussion and Supplementary Table S1). C_i/C_{tot} represents the relative number of molecules with a given degree of polymerization. **d**, Simulations of the time dependence of the incorporation of deuterated molecules into fibrils using the molecular recycling model as a function of the degree of polymerization shown in **c**; the simulated behaviour is compared to the experimental data shown in Fig. 3g.

amyloid disorders[1,3], and the efficacy of therapeutic strategies designed to combat the increasingly widespread family of debilitating human diseases[27,30].

METHODS

Amyloid fibrils with a high degree of morphological homogeneity were prepared by incubating 0.5 mM of freshly prepared PI(3)K-SH3 in H_2O at pH 2.0 after seeding by addition of 2% (v/v) of a solution of pre-formed fibrils for 7 days at 35 °C. Immediately after addition of the pre-formed fibrils, the solution was sonicated for 2 min to fragment the added fibrils and create additional polymerization surfaces, and the homogeneity of the samples was assessed by electron microscopy (Fig. 3) and FTIR spectroscopy (Fig. S1). The fibrils prepared in this manner were allowed to exchange in a deuterated buffer at pH* 1.6 and 28 °C. The nature and degree of exchange was monitored using solution NMR spectroscopy and ESI-MS after dissolution of the fibril samples in 95% DMSO-d_6/5% D_2O mixtures[15]. Simulations of the H/D exchange data were carried out using a series of models based on different possible mechanisms of the exchange process. Full details of these methods and the various simulations are given in the Supplementary Information.

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
2. Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Advan. Protein Chem.* **50**, 123–159 (1997).
3. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).
4. Wickner, R. B. *et al.* Prions: proteins as genes and infectious entities. *Genes Dev.* **18**, 470–485 (2004).
5. Chapman, M. R. *et al.* Role of *Escherichia coli curli* operons in directing amyloid fiber formation. *Science* **295**, 851–855 (2002).
6. Kelly, J. W. & Balch, W. E. Amyloid as a natural product. *J. Cell Biol.* **161**, 461–462 (2003).
7. Si, K., Lindquist, S. & Kandel, E. R. A neuronal isoform of the *Aplysia* CPEB has prion-like properties. *Cell* **115**, 879–891 (2003).
8. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678–699 (2003).
9. Jiménez, J. L. *et al.* Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **18**, 815–821 (1999).
10. Zurdo, J., Guijarro, J. I., Jiménez, J. L., Saibil, H. R. & Dobson, C. M. Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J. Mol. Biol.* **311**, 325–340 (2001).
11. Zurdo, J., Guijarro, J. I. & Dobson, C. M. Preparation and characterization of purified amyloid fibrils. *J. Am. Chem. Soc.* **123**, 8141–8142 (2001).
12. Ventura, S. *et al.* Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl Acad. Sci. USA* **101**, 7258–7263 (2004).
13. Englander, S. W. & Krishna, M. M. G. Hydrogen exchange. *Nature Struct. Biol.* **8**, 741–742 (2001).
14. Kheterpal, I., Zhou, S., Cook, K. D. & Wetzel, R. Aβ amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proc. Natl Acad. Sci. USA* **97**, 13597–13601 (2000).
15. Hoshino, M. *et al.* Mapping the core of the β2-microglobulin amyloid fibril by H/D exchange. *Nature Struct. Biol.* **9**, 332–336 (2002).
16. Olofsson, A., Ippel, J. H., Wijmenga, S. S., Lundgren, E. & Oehman, A. Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J. Biol. Chem.* **279**, 5699–5707 (2004).
17. Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307**, 262–265 (2005).
18. Polverino de Laureto, P. *et al.* Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J. Mol. Biol.* **334**, 129–141 (2003).
19. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. Detection of transient protein folding populations by mass spectrometry. *Science* **262**, 896–900 (1993).
20. Yamaguchi, K. *et al.* Core and heterogeneity of β2-microglobulin amyloid fibrils as revealed by H/D exchange. *J. Mol. Biol.* **338**, 559–571 (2004).
21. Kheterpal, I. *et al.* Aβ protofibrils possess a stable core structure resistant to hydrogen exchange. *Biochemistry* **42**, 14092–14098 (2003).
22. Ban, T., Hamada, D., Hasegawa, K., Naiki, H. & Goto, Y. Direct observation of amyloid fibril growth monitored by thioflavin T fluorescence. *J. Biol. Chem.* **278**, 16462–16465 (2003).
23. Goldsbury, C., Kistler, J., Aebi, U., Arvinte, T. & Cooper, G. J. S. Watching amyloid fibrils grow by time-lapse atomic force microscopy. *J. Mol. Biol.* **285**, 33–39 (1999).
24. Hall, D. & Edskes, H. Silent prions lying in wait: a two-hit model of prion/amyloid formation and infection. *J. Mol. Biol.* **336**, 775–786 (2004).
25. Hammarstroem, P., Wiseman, R. L., Powers, E. T. & Kelly, J. W. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**, 713–716 (2003).
26. Sekijima, Y. *et al.* The biological and chemical basis for tissue-selective amyloid disease. *Cell* **121**, 73–85 (2005).
27. Cohen, F. E. & Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905–909 (2003).
28. Hamada, D., Yanagihara, I. & Tsumoto, K. Engineering amyloidogenicity towards the development of nanofibrillar materials. *Trends Biotechnol.* **22**, 93–97 (2004).
29. MacPhee, C. E. & Woolfson, D. N. Engineered and designed peptide-based fibrous biomaterials. *Curr. Opin. Solid State Mater. Sci.* **8**, 141–149 (2004).
30. Dobson, C. M. In the footsteps of alchemists. *Science* **304**, 1261–1262 (2004).

LETTERS

Action potential refractory period in ureter smooth muscle is set by Ca sparks and BK channels

T. Burdyga[1] & Susan Wray[1]

In excitable tissues the refractory period is a critical control mechanism preventing hyperactivity and undesirable tetani, by preventing subsequent stimuli eliciting action potentials and Ca^{2+} entry. In ureteric smooth muscle, peristaltic waves that occur as invading pacemaker potentials produce long-lasting action potentials (300–800 ms) and extraordinarily long (more than 10 s) refractory periods[1–6], which prevent urine reflux and kidney damage[2]. For smooth muscles neither the mechanisms underlying the refractory period nor the link between excitability and refractoriness are properly understood. Here we show that a negative feedback process, which depends on Ca^{2+} loading the sarcoplasmic reticulum (SR) during the action potential and on the subsequent activation of local releases of Ca^{2+} from the SR (sparks[7]), stimulating plasmalemmal Ca^{2+}-sensitive K^{+} (BK) channels, determines the refractory period of the action potential. As sparks gradually reduce the Ca^{2+} load in the SR, electrical inhibition is released, the refractory period is terminated and peristaltic contractions occur again. The refractory period can be manipulated, for example from 10 s to 100 s, by altering the Ca^{2+} content of the SR or release mechanism or by inhibiting BK channels. This insight into the control of excitability and hence function provides a focus for therapies directed at pathologies of smooth muscle.

Suprathreshold stimulation of ureteric muscle bundles every 10 s elicited a propagating Ca^{2+} transient associated with the action potential in response to the first pulse (Fig. 1a–c), but, despite the repolarization of the action potential (Fig. 1d), subsequent stimuli evoked only small depolarizations, a non-propagating Ca^{2+} transient and no action potentials. Not until the ninth stimulus was a full-size increase in propagating Ca^{2+} and action potential triggered in the example shown (see also Supplementary movie 1), and the mean refractory period was 75.7 ± 6.8 s ($n = 9$).

The mechanism underlying this period of inexcitability is unclear. Although an after-hyperpolarization follows the action potential[3] it is transient, and membrane potential is restored before termination of the refractory period (Fig. 1d); decreased Ca^{2+} channel availability has therefore been suggested[6]. We therefore investigated whether the abolition of the action potential is because there is no Ca^{2+} current as such or whether the membrane is not being depolarized and hence there is no Ca^{2+} current. We recorded Ca^{2+} currents and transients in voltage-clamped single cells at inter-stimulus intervals between 2 and 20 s. With inter-pulse intervals of 10 s or more there was no diminution of Ca^{2+} current or transient (Fig. 2). Even with intervals of 5 s the Ca^{2+} current was 70% of the amplitude of the control ($n = 5$). Inactivation/desensitization of Ca^{2+} channels cannot therefore account for the decreased excitability and refractory period. Figure 2 also makes the point that in voltage-clamped preparations the refractory period is much shorter, namely 10 s, indicating that *in vivo*, when the membrane is not clamped, changes in the membrane potential and hence action potential might produce the

refractory period. The subsequent experiments were designed to test this.

Ca^{2+} sparks—small, transient, local increases in intracellular Ca^{2+} concentration (Fig. 3a)—occur as the SR releases Ca^{2+} through Ca^{2+} (ryanodine)-gated channels[7,8] (guinea-pig ureter has no $Ins(1,4,5)P_3$ channels[9]). In smooth muscles[10] including guinea-pig ureter[11,12], Ca^{2+} sparks activate plasmalemmal BK channels, generating spontaneous transient outward currents (STOCs)[10]. It has previously been
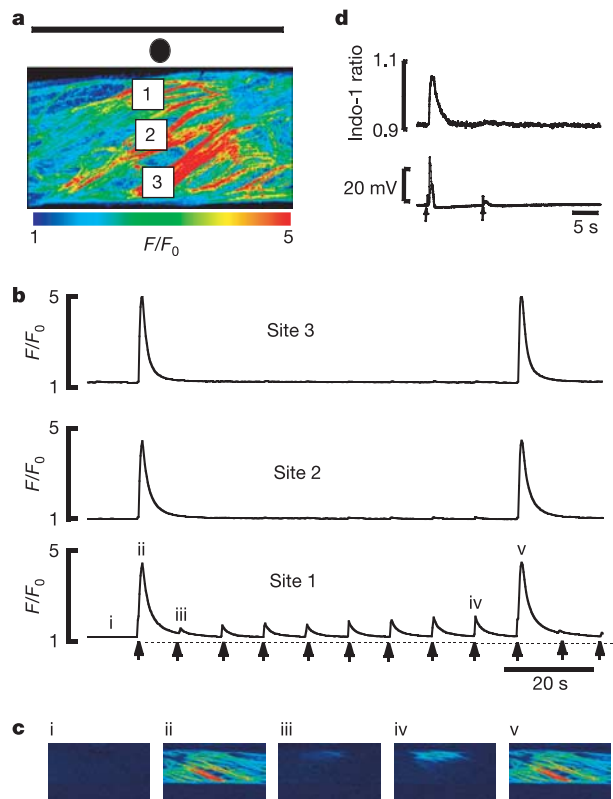
Figure 1 | Refractory period in intact guinea-pig ureteric smooth muscle. **a**, Pseudo-colour confocal image of Fluo-4-loaded ureteric muscle bundles during a global Ca^{2+} signal. Three sites are identified and the black circle indicates the stimulation electrode. Scale bar at top, 2 mm. **b**, Global and regional Ca^{2+} transients evoked by suprathreshold stimulation, every 10 s (also see Supplementary movie 1), at the sites identified in **a**; arrows indicate stimuli. **c**, X–Y Ca^{2+} images (original magnification ×4) of bundles, using sequential subtraction to highlight differences, at rest (i), during a global Ca^{2+} transient (ii and v) and regional Ca^{2+} transients, recorded at the beginning (iii) and end (iv) of the refractory period. **d**, Simultaneous Ca^{2+} (Indo-1, top) and electrical (sucrose gap, bottom) responses to two stimuli, 10 s apart, in guinea-pig ureter.

[1]Department of Physiology, The University of Liverpool, Liverpool L69 3BX, UK.

shown that local Ca^{2+} signalling to BK channels regulates blood pressure[13]; the SR and its Ca^{2+} load therefore have a key function in smooth muscle activity[14]. We therefore tested the hypothesis that Ca^{2+} release from the SR, by activating BK channels, controls the refractory period.

We first investigated whether Ca^{2+} spark and BK activity were influenced by the occurrence of action potentials. As shown in Fig. 3b, initially few sparks were present, but after depolarization and a global increase in Ca^{2+} there was vigorous Ca^{2+} spark activity, especially from frequently discharging sites[15] (see Supplementary movie 2). The frequency of Ca^{2+} sparks was highest after the action potential and lowest at the end of the refractory period (Fig. 3d). STOC activity was also significantly increased after the action potential in voltage-clamped ureteric myocytes (Fig. 3b, bottom). Sparks ($n = 19$) and STOCs ($n = 14$) were potentiated by 1 mM caffeine (a releaser of Ca^{2+} from the SR; Fig. 3b, c) and were abolished by pretreatment with ryanodine ($n = 6$; Fig. 3d, top) or cyclopiazonic acid (CPA; $n = 10$; Fig. 3d, bottom), both of which deplete Ca^{2+} in the SR. Furthermore this spark and STOC activity can directly decrease membrane excitability (Fig. 3d). Simultaneous measurement of current and potential show that after an action potential small hyperpolarizations occur as a consequence of STOCs ($n = 5$). Thus, immediately after the action potential, when the tissue is refractory, there is a transient but pronounced activation of Ca^{2+} sparks and BK channels, which hyperpolarize the membrane. As

shown in Fig. 3e with a double-pulse depolarization, Ca^{2+} sparks are greatly increased in number and amplitude on the second stimulation, indicating that loading of the SR has occurred at a time when the global Ca^{2+} transient and action potential fail, as shown in Fig. 1.

If the refractory period is due to increased Ca^{2+} spark activity, generating STOCS and decreasing excitability, then abolishing sparks or blocking BK channels should increase excitability and shorten the refractory period; this did indeed occur. Tetraethylammonium (TEA) and iberiotoxin, blockers of BK channels, produced profound changes in electrical activity. Both inhibitors abolished outward current, including STOCs ($n = 7$ and $n = 5$, respectively; Fig. 4a), and excitability was restored ($n = 11$; Fig. 4b). Emptying the SR with CPA also restored excitability ($n = 13$; Fig. 4b). There was also inhibition of the after-hyperpolarization and a small (3–6 mV) depolarization of the membrane potential in both cases. All these changes in electrical activity culminated in cells being able to produce a full-sized action potential when stimulated with a pulse at 10-s intervals. Iberiotoxin ($n = 5$; Fig. 4c) and TEA (8 of 8; Fig. 4d) reduced the refractory period such that every stimulation produced a full-sized, propagating Ca^{2+} signal.

Manipulation of the content of Ca^{2+} in the SR or of its release from the SR had profound effects on the refractory period; this is consistent with a previous report that blocking the SR Ca^{2+} pump (SERCA) shortened the refractory period[16]. Disabling the SR with
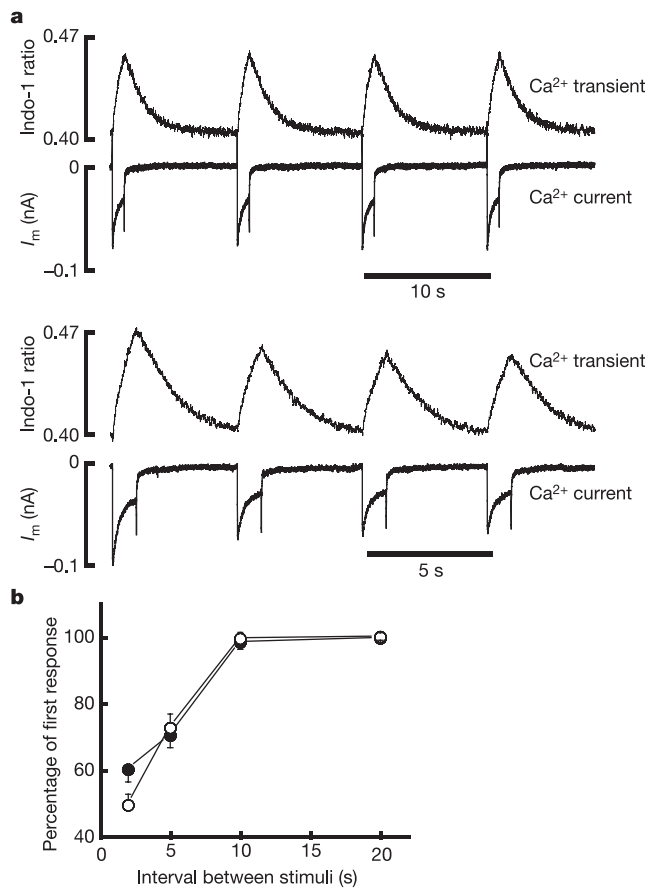
Figure 2 | Effects of stimulus interval on Ca^{2+} current. a, Ca^{2+} current (bottom traces) and Ca^{2+} transient (top traces) evoked by 1 s depolarizing voltage steps from −50 to 0 mV applied at 10-s (top panel) or 5-s (bottom panel) intervals. **b,** The effect of inter-stimulus interval on the amplitude of Ca^{2+} transient (open symbols) and Ca^{2+} current (filled symbols), compared with first response (100%). Results are shown as means and s.e.m. All data were obtained in voltage-clamped guinea-pig ureteric myocytes.
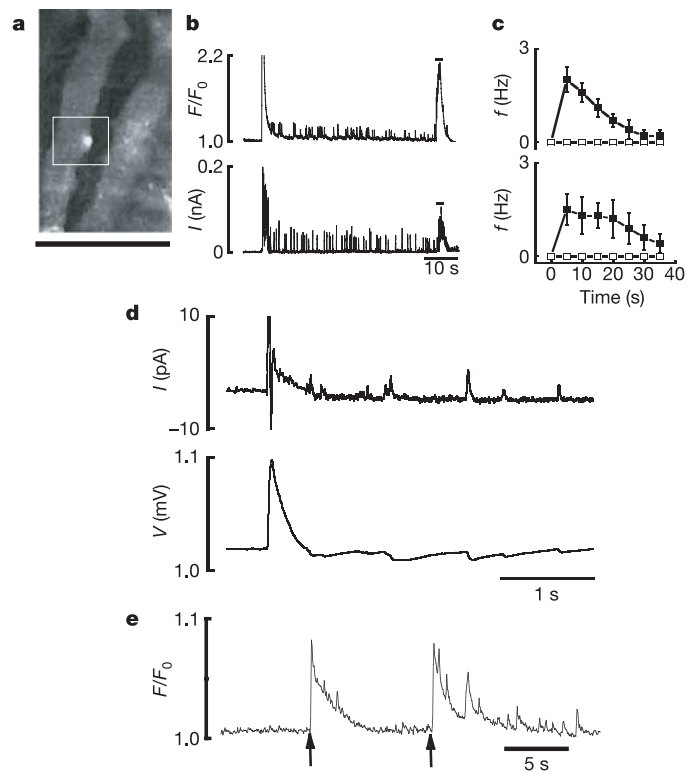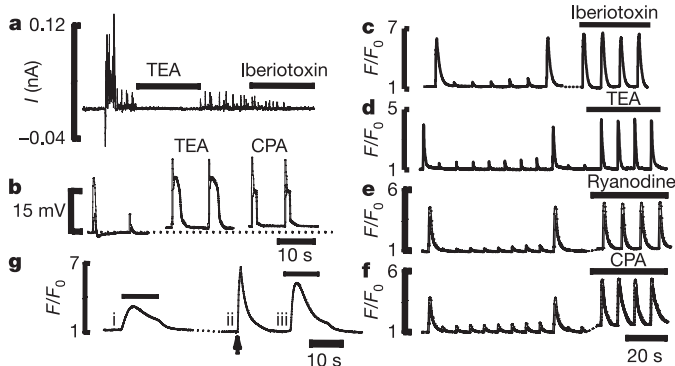

Figure 3 | Ca^{2+} sparks, STOCs and the refractory period. a, $X–Y$ image of Ca^{2+} sparks (box), in myocyte *in situ*. Scale bar at bottom, 35 μm. **b,** Top: global Ca^{2+} transient with depolarization (off-scale peak at left) followed by evoked sparks during the refractory period (see Supplementary movie 2). Caffeine (1 mM; application indicated by the horizontal bars) potentiates Ca^{2+} sparks. Bottom: STOCs after depolarization and their potentiation by caffeine, in voltage-clamped myocytes. **c,** Frequency of sparks (filled symbols, top) and STOCs (filled symbols, bottom) after a Ca^{2+} transient and a depolarizing step, respectively. Open symbols, effects of ryanodine (50 μM) (top) and cyclopiazonic acid (CPA; 20 μM) (bottom) respectively. Error bars indicate s.e.m. **d,** Simultaneous current (top) and membrane potential (bottom) recordings after action potential. **e,** Sparks recorded locally in response to a double-pulse depolarization.

ryanodine abolished sparks and the refractory period (6 of 6; Fig. 4e). CPA reduced the refractory period to less than 10 s in 9 of 10 preparations (Fig. 4f). Ca^{2+} sparks produce hyperpolarization by activating BK channels, and the refractory period is associated with a transient after-hyperpolarization[3]. We therefore increased the external K^+ concentration, $[K^+]_o$, from 5.9 to 15 mM, to change the membrane potential by about 8 mV. This small elevation overcame the refractory period, tested at 10 s, in 4 of 7 preparations and decreased it to 20 s in 3 of 7 preparations (not shown), which is consistent with previous findings[6].

Figures 3 and 4 indicate that the luminal Ca^{2+} content of the SR might be important in modulating Ca^{2+} spark activity[17] and refractoriness. If the Ca^{2+} content of the SR is increased when there is a global increase in intracellular Ca^{2+}, for example that accompanying action-potential activity (Figs 1–4), this could provide a mechanistic link between excitation and subsequent inhibition. The Ca^{2+} content of the SR was therefore measured with caffeine (10 mM) to release Ca^{2+} from the SR. Figure 4g shows examples ($n = 9$) of the Ca^{2+} releases from the SR before and after the action-potential-initiated global Ca^{2+} transient (position ii; 100%). The amplitudes of increases in Ca^{2+} evoked by caffeine were 38.1 ± 3.9% before the termination of the refractory period (Fig. 4g, position i) and 65.1 ± 6.9% at the beginning of the refractory period (position ii; data are statistically significantly different). Therefore action-potential-stimulated Ca^{2+} influx increases the Ca^{2+} content of the SR and thereby transiently increases Ca^{2+} spark activity, providing a direct link between excitation and the SR. In turn, the Ca^{2+} sparks stimulate BK channels, produce transient after-hyperpolarization and cause inexcitability of the membrane; that is, refractoriness. Thus, after depolarization the SR is full and there is a burst of evoked spark activity; the resultant large outward BK current causes inexcitability, as shown in Fig. 3. Any agents that modulate SR content, BK channels, STOCs or the coupling between them[18,19] will influence the refractory period, as shown here.

The refractory period in ureteric smooth muscle, a key determinant of tissue excitability, is dependent on small Ca^{2+} signals, Ca^{2+} sparks, that originate from an intracellular Ca^{2+} store. A crucial coupling mechanism between Ca^{2+} entry, produced by the action potential, and the subsequent SR filling and Ca^{2+} release events ensures that periods of excitability are followed by refractoriness. The molecular targets on the cell membrane are the BK channels;

if they are blocked, or if Ca^{2+} sparks are abolished, then the refractory period almost disappears, and Ca^{2+} channel inactivation will determine refractoriness. The refractory period in the ureters is vital to peristalsis and the movement of urine from kidneys to bladder. This fundamental breakthrough in our understanding of the molecular basis by which refractoriness is controlled can be exploited for subsequent therapeutic advantage in cases of dysfunction that ultimately threaten renal function. The above model is the first to show how excitation loads the SR and causes refractoriness. This role for the SR adds to its increasing function in excitation–contraction coupling in muscle. Thus, through BK channels it influences tonic activity in blood vessels[14] and limits contractility in uterus[20] and bladder[21], and a decrease in excitability such as that proposed here might be the underlying mechanism. As these tissues also have $Ins(1,4,5)P_3$ receptors it also indicates that the effects on excitability might not be limited to ryanodine-gated mechanisms of release from the SR, and there will be a dynamic interaction between Ca^{2+} load in the SR, membrane currents and excitability. These data contrast with the role of the SR in cardiac muscle, in which ryanodine-gated Ca^{2+} release from the SR triggered by Ca^{2+} influx is crucial to contraction[8]; recent work has shown a role for the SR in increasing excitability through Na/Ca exchange[22] and a role for $Ins(1,4,5)P_3$ and ryanodine-gated release of Ca in arrhythmogenesis[23,24].

METHODS

Force, Ca^{2+} and electrical measurements in intact tissues. Guinea-pigs were humanely killed; ureters were dissected into strips 3–4 mm long and loaded with Ca-sensitive Indo-1 (15 μM) or Fluo-4 (15 μM, for confocal studies) for 2 h at 22 °C, as detailed elsewhere[25]. Rinsed tissues were kept in physiological saline (22 °C) for at least 30 min, then placed in a 200-μl bath on an inverted microscope, with both ends fixed by aluminium clips. Tissues were electrically field-stimulated with rectangular pulses (100 ms duration; 5 V) through two silver electrodes. The electrical activity in intact tissue was recorded by a modified double sucrose gap[26]. In some experiments wortmannin (5 μM), a blocker of myosin light-chain kinase[26], was used to eliminate movement artefacts, but there was no difference in any data obtained with or without wortmannin. All experiments were performed at 22 °C.

Single isolated cells and membrane current measurements. Isolated ureteric myocytes were produced by enzymatic digestion[25], then filtered, centrifuged (2,000 r.p.m., 2 min) and placed for 30 min in KB medium (see below) containing 5 μM Fluo-4 acetoxymethyl ester (AM) or Indo-1 AM with PluronicF-127 at 22 °C. Cells were then washed, centrifuged and placed in KB medium for at least 30 min. In voltage clamp the membrane potential was held at −50 or −40 mV and depolarized to 0 mV. For current clamp and simultaneous measurement of membrane currents and potentials, an EPC9 amplifier (Heka) and software were used, allowing current-to-voltage conversion during current recording. The patch pipettes had an input resistance of 3–6 MΩ, and capacitive currents were cancelled electronically (Axopatch 200B or EPC9 amplifier); currents were low-pass filtered at 2 kHz (Clampex 8; Axon Instruments or Heka) and analysed with Origin 6 (Microcal).

Confocal imaging. Images were taken on a Perkin Elmer, Nipkow disc-based UltraView confocal scanner linked to a fast digital camera, computer workstation and Olympus inverted microscope. This system, combined with low-power (4× objective, numerical aperture (NA) 0.13) and high-power (60× water-immersion objective, NA 1.2) lenses, allowed the acquisition of images from a large area at 20–50 frames s^{-1} (full frame $2160 \times 2100 \,\mu m^2$ and $150 \times 110 \,\mu m^2$ with the 4× and 60× objectives, respectively). Background fluorescence was removed for measurements of Ca^{2+}; for illustration of dynamic changes in the fluorescence signal (Fig. 1c), the signal at the maximum rate of change during the rising phase is shown. The preparations were superfused with physiological saline at 22 °C.

Solutions. Tissues and cells were superfused with oxygenated buffered physiological solution (pH 7.4) at 7–8 ml min^{-1} with the following composition (in mM): NaCl 120.4, KCl 5.9, $MgSO_4$ 1.2, $CaCl_2$ 2, glucose 11.5, HEPES 11. The KB medium composition (pH 7.4) was (in mM): KCl 40, K_2HPO_4 10, taurine 10, TES ((N-Tris(hydroxymethyl)methyl-2-aminoethanesulphonic acid) 10, glucose 11, pyruvate 5, creatine 5, EGTA 0.04, potassium glutamate 100, adjusted with KOH. The patch pipette solution was (in mM): CsCl or KCl 130, HEPES 10, $MgCl_2$ 5, glucose 10, ATP 5, EGTA 0.1; pH was adjusted to 7.2 with NaOH. All agents were dissolved in aqueous solvents apart from CPA, nifedipine and

Figure 4 | Modulation of the refractory period. a, Inhibitory effect of TEA (2 mM) and iberiotoxin (200 nM) on STOCs, evoked by stepping from −40 to 0 mV (voltage-clamped myocytes). **b,** Action potentials evoked by suprathreshold depolarizing pulses every 10 s in the absence and the presence of TEA and CPA, both of which also produced a small depolarization and increased the duration of the action potential. **c–f,** The effects of iberiotoxin (**c**), TEA (**d**), 50 μM ryanodine (**e**) and 20 μM CPA (**f**) on the refractory period. Bars indicate the duration of application. **g,** Ca^{2+} transients recorded in response to 10 mM caffeine (application indicated by the horizontal bars) before (i) and after (iii) a global Ca^{2+} transient (ii) produced by depolarization; dotted line, break in recording.

wortmannin, which were dissolved in dimethylsulphoxide. Local application of test solutions to cells or intact preparations was by means of a pressure-regulated perfusion system (Digitimer) with a delivery tip 100 or 250 μm in diameter, attached to a three-dimensional manipulator.

Statistics. Values are means ± s.e.m. and n is the number of animals for intact tissues and cells. Differences were taken as significant at $P < 0.05$.

1. Wemyss-Holden, G. D., Rose, M. R., Payne, S. R. & Testa, H. J. Non-invasive investigation of normal individual ureteric activity in man. *Br. J. Urol.* **71**, 156–160 (1993).
2. Weiss, R. M. *Campbell's Urology* (ed. Walsh, P. C.) 94–128, (W.B. Saunders Company, Philadelphia, 1986).
3. Exintaris, B. & Lang, R. J. K. (+) channel blocker modulation of the refractory period in spontaneously active guinea-pig ureters. *Urol. Res.* **27**, 319–327 (1999).
4. Kuriyama, H., Osa, T. & Toida, N. Membrane properties of the smooth muscle of guinea-pig ureter. *J. Physiol. (Lond.)* **191**, 225–238 (1967).
5. Cuthbert, A. W. The relation between response and the interval between stimuli of the isolated guinea-pig ureter. *J. Physiol. (Lond.)* **180**, 225–238 (1965).
6. Maggi, C. A., Giuliani, S. & Santicioli, P. Effect of Bay K 8644 and ryanodine on the refractory period, action potential and mechanical response of the guinea-pig ureter to electrical stimulation. *Naunyn-Schmiedebergs Arch. Physiol.* **349**, 510–522 (1994).
7. Nelson, M. T. *et al.* Relaxation of arterial smooth muscle by calcium sparks. *Science* **270**, 633–637 (1995).
8. Cheng, H., Lederer, W. J. & Cannell, M. B. Calcium sparks: elementary events underlying excitation–contraction coupling in heart muscle. *Science* **262**, 740–744 (1993).
9. Burdyga, T. V., Taggart, M. J., Crichton, C., Smith, G. L. & Wray, S. The mechanism of Ca^{2+} release from the SR of permeabilised guinea-pig and rat ureteric smooth muscle. *Biochim. Biophys. Acta* **1402**, 109–114 (1998).
10. Benham, C. D. & Bolton, T. B. Spontaneous transient outward currents in single visceral and vascualr smooth muscle cells of the rabbit. *J. Physiol. (Lond.)* **381**, 385–406 (1986).
11. Imaizumi, Y., Muraki, K. & Watanabe, M. Ionic currents in single smooth muscle cells from the ureter of the guinea-pig. *J. Physiol. (Lond.)* **411**, 131–159 (1989).
12. Lang, R. J. Identification of the major membrane currents in freshly dispersed single smooth muscle cells of guinea-pig ureter. *J. Physiol. (Lond.)* **412**, 375–395 (1989).
13. Brenner, R. *et al.* Vasoregulation by the B1 subunit of the calcium-activated potassium channel. *Nature* **407**, 870–876 (2000).
14. Wellman, G. C., Santana, L. F., Bonev, A. D. & Nelson, M. T. Role of the phospholamban in the modulation of arterial Ca^{2+} sparks and Ca^{2+}-activated K$^+$ channels by cAMP. *Am. J. Physiol. Cell Physiol.* **281**, C1029–C1037 (2001).
15. Gordienko, D. V., Greenwood, I. A. & Bolton, T. B. Direct visualization of sarcoplasmic reticulum regions discharging Ca^{2+} sparks in vascular myocytes. *Cell Calcium* **29**, 13–28 (2001).
16. Maggi, C. A., Giuliani, S. & Santicioli, P. Effect of the Ca^{2+}-ATPase inhibitor, cyclopiazonic acid, on electromechanical coupling in the guinea-pig ureter. *Br. J. Pharmacol.* **114**, 127–137 (1995).
17. Cheranov, S. Y. & Jaggar, J. H. Sarocplasmic reticulum calcium load regulates rat arterial smooth muscle calcium sparks and transient K$_{Ca}$ currents. *J. Physiol. (Lond.)* **544**, 71–84 (2002).
18. ZhuGe, R. *et al.* Ca^{2+} spark sites in smooth muscle cells are numerous and differ in number of ryanodine receptors, large-conductance K$^+$ channels, and coupling ratio between them. *Am. J. Physiol. Cell Physiol.* **287**, C1577–C1588 (2004).
19. Kotlikoff, M. I. Calcium-induced calcium release in smooth muscle: the case for loose coupling. *Prog. Biophys. Mol. Biol.* **83**, 171–191 (2003).
20. Matthew, A. J. G., Kupittayanant, S., Burdyga, T. V. & Wray, S. Characterization of contractile activity and intracellular Ca^{2+} signalling in mouse myometrium. *J. Soc. Gynecol. Investig.* **11**, 207–212 (2004).
21. Heppner, T. J., Bonev, A. D. & Nelson, M. T. Ca^{2+}-activated K$^+$ channels regulate action potential repolarization in urinary bladder smooth muscle. *Am. J. Physiol.* **273**, C110–C117 (1997).
22. Bogdanov, K. Y., Vinogradova, T. M. & Lakatta, E. G. Sinoatrial nodal cell ryanodine receptor and Na$^+$-Ca^{2+} exchanger: molecular partners in pacemaker regulation. *Circ. Res.* **88**, 1254–1258 (2001).
23. Zima, A. V. & Blatter, L. A. Inositol-1,4,5-trisphosphate-dependent Ca^{2+} signalling in cat atrial excitation–contraction coupling and arrhythmias. *J. Physiol. (Lond.)* **555**, 607–615 (2004).
24. Hove-Madsen, L. *et al.* Atrial fibrillation is associated with increased spontaneous calcium release from the sarcoplasmic reticulum in human atrial myocytes. *Circulation* **110**, 1358–1363 (2004).
25. Shabir, S., Borisova, L., Wray, S. & Burdyga, T. Rho-kinase inhibition and electromechanical coupling in phasic smooth muscle; Ca^{2+}-dependent and independent mechanisms. *J. Physiol. (Lond.)* **560**, 839–855 (2004).
26. Burdyga, T. V. & Wray, S. Simultaneous measurements of electrical activity, intracellular [Ca^{2+}] and force in intact smooth muscle. *Pflügers Arch.* **435**, 182–184 (1997).

Somatic control of germline sexual development is mediated by the JAK/STAT pathway

Matthew Wawersik[1], Allison Milutinovich[1], Abbie L. Casper[1], Erika Matunis[2], Brian Williams[1] & Mark Van Doren[1]

Germ cells must develop along distinct male or female paths to produce the sperm or eggs required for sexual reproduction. In both mouse and *Drosophila*, the sexual identity of germ cells is influenced by the sex of the surrounding somatic tissue (for example, refs 1, 2, reviewed in refs 3, 4); however, little is known about how the soma controls germline sex determination. Here we show that the janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway provides a sex-specific signal from the soma to the germ line in *Drosophila* embryonic gonads. The somatic gonad expresses a JAK/STAT ligand, *unpaired* (*upd*), in a male-specific manner, and activates the JAK/STAT pathway in male germ cells at the time of gonad formation. Furthermore, the JAK/STAT pathway is necessary for male-specific germ cell behaviour during early gonad development, and is sufficient to activate aspects of male germ cell behaviour in female germ cells. Our findings provide direct evidence that the JAK/STAT pathway mediates a key signal from the somatic gonad that regulates male germline sexual development.

While investigating communication between the somatic gonad and germ line, we found that the JAK/STAT pathway is specifically activated in male, but not female, germ cells. In *Drosophila*, JAK/STAT signalling (reviewed in ref. 5) is initiated when a UPD or UPD-like ligand binds a transmembrane receptor (Domeless), activating the janus kinase Hopscotch (HOP), which phosphorylates the STAT92E transcription factor. STAT activation has been shown to regulate STAT gene expression[6] and can induce upregulation of the STAT92E protein, which can be used as an assay for JAK/STAT pathway activation. We found that STAT92E is upregulated specifically in male, but not female, germ cells at the time of gonad formation (Fig. 1a, b). This reflects male-specific activation of the JAK/STAT pathway because the activated form of STAT92E (phosphorylated STAT92E) was also only detected in male germ cells (Fig. 1c, d), and JAK activity is necessary and sufficient for STAT92E expression. Expression of a JAK inhibitor, *Socs36E* (ref. 7), resulted in loss of STAT92E expression in male germ cells (Fig. 1e), and expression of constitutively active JAK[8] (*hop*^{TumL}) induced STAT92E production in female germ cells (Fig. 1f). The male-specific activation of STAT92E at this time is distinct from STAT92E activation in germ cells in the early embryo, which is not sex-specific and is regulated by the mitogen-activated protein (MAP) kinase pathway[9].

We also found that STAT92E expression in male germ cells is dependent on their association with the somatic gonad. STAT92E was not detected in germ cells that were migrating to the gonad (Fig. 2a), but was detected in male germ cells after they had made contact with the somatic gonad (Fig. 2b). STAT92E expression was greatly reduced or absent in *eyes absent* (*eya*) mutants (data not shown), where somatic gonad identity is initiated but not maintained[10]. Furthermore, STAT92E was not detected in germ cells found outside the

gonad in wild-type embryos (Fig. 2b, arrows) nor in mis-localized germ cells in *wunen* (Fig. 2c) and *HMG coenzyme A reductase* (data not shown) mutants that lack guidance cues that target germ cells to the somatic gonad[11,12]. However, in these same mutants, STAT92E was detected in the few germ cells that contact the somatic gonad in male embryos (Fig. 2c, inset).

STAT92E expression in the germ line is dependent on the sex of the surrounding soma. When XX (normally female) germ cells were present in a soma that was masculinized by expression of the male form of the somatic sex-determination gene *doublesex* (*dsx*)[13], germ cells expressed STAT92E (Fig. 2d). *dsx* does not have an autonomous role in germ cells themselves[14], indicating that STAT92E induction in these embryos is caused by masculinization of the soma. Conversely, when the somatic gonad of an XY (normally male) embryo was feminized by expression of the sex determination gene *transformer*
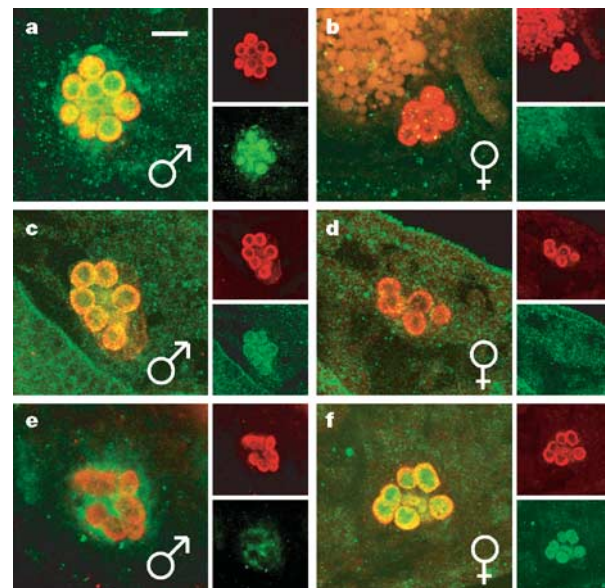
Figure 1 | The JAK/STAT pathway is activated in male germ cells.
a–f, Embryonic gonads (stage 15) co-immunolabelled with anti-Vasa (red) and either anti-STAT92E (**a**, **b**, **e**, **f**; green) or anti-phospho-STAT92E (**c**, **d**; green). Smaller panels to the right show the individual channels. Sex was determined in this and subsequent figures as described in the Methods. **a–d**, Wild-type male (**a**, **c**) and female (**b**, **d**) gonads. **e**, Male embryo with germ cells expressing *Socs36E* (UAS-*Socs36E*; *nanos*-Gal4). **f**, Female embryos with germ cells expressing *hop*^{TumL} (UAS-*hop*^{TumL}; *nanos*-Gal4). The scale bar (12.5 μm) in **a** applies to all panels.

[1]Biology Department, The Johns Hopkins University, Baltimore, Maryland 21218, USA. [2]Department of Cell Biology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

(*tra*) in the mesoderm, but not germ cells, STAT92E expression was no longer observed in XY germ cells (data not shown). Taken together, these data indicate that the male somatic gonad is necessary and sufficient to activate the JAK/STAT pathway in either XX or XY germ cells.

Consistent with this, we found that the JAK/STAT ligand *upd* is expressed specifically in the male, but not female, somatic gonad (Fig. 2e, f). Expression of STAT92E in male germ cells was no longer detected in embryos in which *upd* and two homologues, *upd2* and *upd3* (ref. 15), were deleted (*Df*(1)*os*[1a]; Fig. 2g). Because male germ cells from embryos mutant for *upd* alone still expressed STAT92E (data not shown), JAK/STAT activation in the germ line may be regulated redundantly by *upd* and one or more of its homologues. In addition, expression of *upd* in either the mesoderm (data not shown) or germ cells (Fig. 2h) was sufficient to induce STAT92E expression in

XX germ cells. Expression of *upd2* or *upd3* was also capable of inducing STAT92E expression in germ cells (data not shown).

upd is also important for embryonic patterning[16] and somatic sex determination[17,18]. Notably, *upd* promotes female identity in the soma[17,18], but promotes male development in the germ line (see below). To verify that the effects of *upd* on the germ line are not indirectly caused by other effects of *upd*, we examined indicators of embryonic segmentation (Engrailed), somatic sex determination (Sex lethal), somatic gonad identity (Eyes absent) and somatic gonad sexual identity (Sox100B) (data not shown). *Df*(1)*os*[1a] hemizygous male embryos exhibited segmentation defects as expected[16], but formed gonads that expressed normal somatic and sex-specific markers. Embryos ectopically expressing *upd* were normal in all respects examined.

We next addressed whether activation of the JAK/STAT pathway by
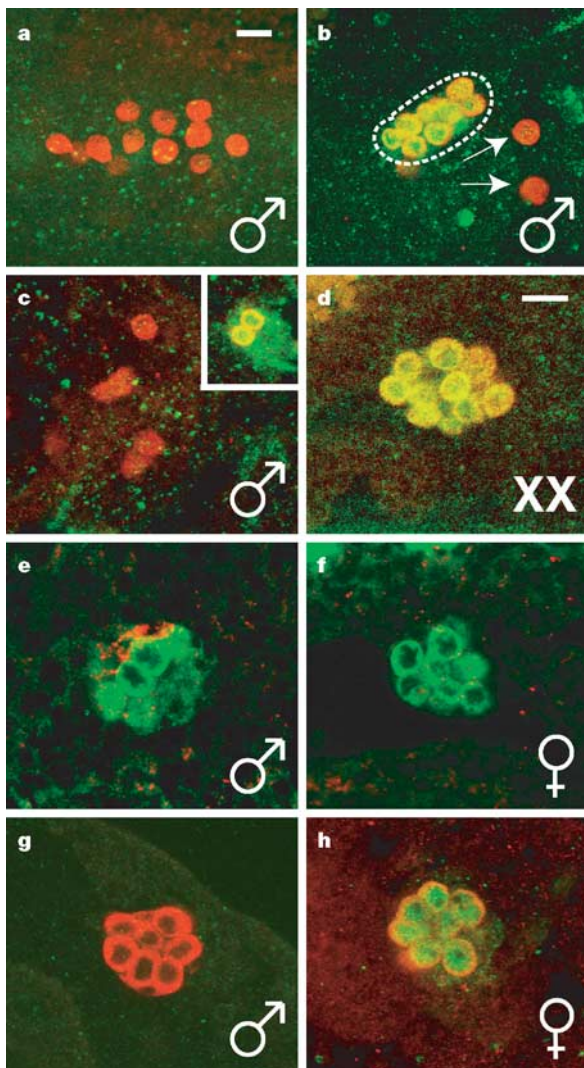
Figure 2 | The JAK/STAT pathway is activated by the male somatic gonad. Embryos are at stage 15 and wild type unless noted. **a–d**, Embryos labelled with anti-Vasa (red) and anti-STAT92E (green). **a, b**, Male embryos at stage 11 (**a**) and stage 13 (**b**). Gonad (outline) and mis-localized germ cells (arrows) are indicated. **c**, Mis-localized germ cells in male *wunen*[CE] mutant. The inset shows germ cells in the gonad of the same embryo. **d**, XX germ cells in a male soma (*dsx*[D]/*dsx*[1]). **e, f**, *upd* riboprobe (red) and anti-Vasa (green). *upd* is expressed in male (**e**) and not female (**f**) gonads. **g, h**, Embryos labelled with anti-Vasa (red) and anti-STAT92E (green). Gonads in an XY *Df*(1)*os*[1a] embryo (**g**), and a female embryo (**h**) with germ cells expressing *upd* (UAS-*upd*; *nanos*-Gal4) are shown. The scale bar (12.5 μm) in **a** and **d** applies to panels **a–c** and **d–h**, respectively.
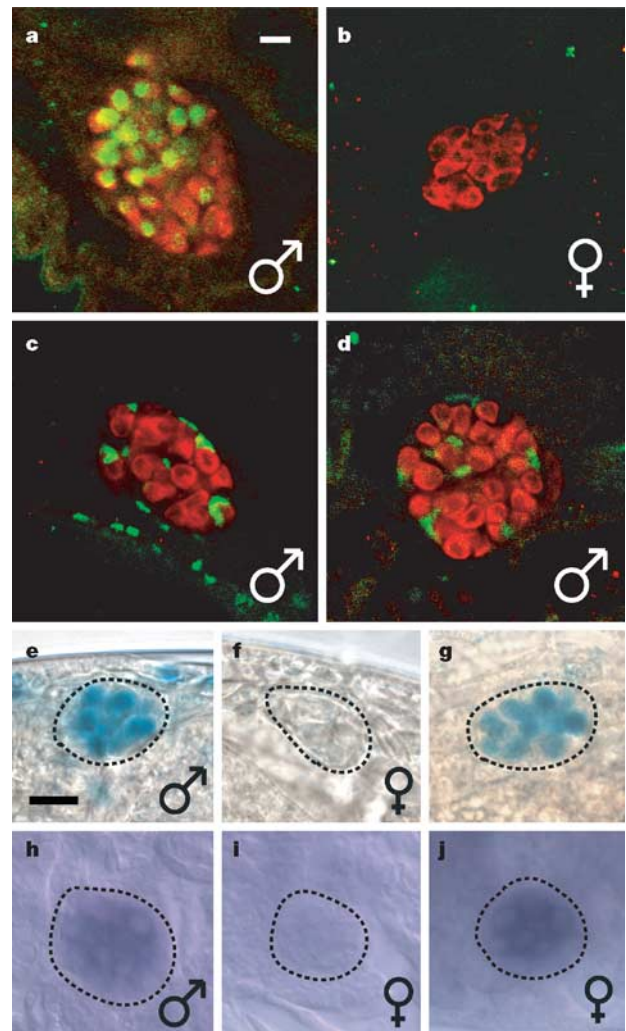
Figure 3 | JAK/STAT activation regulates male-specific germ cell behaviour. a–d, Larval (L1) gonads. Anti-Vasa (red) and anti-β-Gal (green) labelling is shown. **a, b**, *mgm-1* expression in gonads from male (**a**) and female (**b**) offspring from *mgm-1*/+ × *Kr-GFP*/Y. **c**, Male *Stat92E*[06346] mutant from mating number 1 (see Methods) with no *mgm-1* expression in the germ cells. **d**, Male gonad of a *Stat92E*[06346]/+ larva showing somatic *Stat92E*[06346] enhancer trap expression. **e–g**, *mgm-1* expression (X-gal staining) in stage 16 gonads (circled) in wild-type male (**e**) and female (**f**) embryos, and an embryo expressing ectopic *upd* (**g**) (mating number 2; see Methods). **h–j**, *dpa* expression (*in situ* hybridization) in stage 17 gonads (circled) in wild-type male (**h**) and female (**i**) embryos, and a female embryo expressing ectopic *upd* (**j**) (mating number 3; see Methods). The scale bar (12.5 μm) in **a** and **e** applies to panels **a–d** and **e–j**, respectively.

the male somatic gonad regulates male-specific development of germ cells. In adult testes, the JAK/STAT pathway is required for maintenance of germline stem cells[19,20], making it difficult to assess the role of this pathway on male germ cell identity at this stage. Instead, we examined germ cells during embryogenesis and early larval stages, when germ cell development first becomes sexually dimorphic. In the mouse, the earliest manifestation of sex determination in the germ line is differential regulation of the germline cell cycle by the soma[4]. In *Drosophila*, germ cells undergo one to two divisions after their formation, but are arrested in the cell cycle during germ cell migration and only resume division shortly after the gonad has formed[21,22]. Because larval testes contain more germ cells than larval ovaries[23], we investigated whether proliferation is regulated differently in male and female germ cells. Indeed, sex-specific analysis of a mitotic marker (phosphohistone H3) in the germ line indicated that germ cell proliferation is entirely male-specific during early stages of gonad development (Table 1). Furthermore, male-specific germ cell division is dependent on the male somatic gonad. Male germ cells did not proliferate in *eya* mutants that lack the somatic gonad, or in lost germ cells within *wunen* mutant embryos. XX germ cells in a masculinized soma (dsx^D/dsx^1) proliferated, whereas XY germ cells in a feminized soma (UAS-tra^F; *twi*-Gal4) did not. Thus, the pattern of germ cell proliferation correlates exactly with the activity of the JAK/STAT pathway in germ cells.

To assess whether JAK/STAT signalling regulates male-specific germ cell division, we examined embryos lacking zygotic *Stat92E* activity and observed a marked decrease in male germ cell proliferation (Table 1). Similar reductions in germ cell proliferation were observed in the *upd/upd-like* mutant ($Df(1)os^{1a}$) and in embryos where the JAK inhibitor *Socs36E* was expressed in germ cells. Thus, JAK/STAT activity is required within germ cells for proper male-specific germ cell division in the gonad. We also found that expression of *upd* in the germ line was sufficient to induce proliferation in female germ cells. Thus, the JAK/STAT pathway can induce XX germ cells to exhibit this male-specific germ cell behaviour.

We next investigated whether the JAK/STAT pathway regulates other aspects of male germ cell development. *male germline marker-1 (mgm-1)* is a *lacZ* enhancer trap line that is expressed in male germ cells, but not female germ cells, and therefore is a marker for male germ cell identity[24] (see Fig. 3a, b, e, f). Inhibiting the JAK/STAT pathway by removing zygotic *Stat92E* activity did not affect *mgm-1*

expression in the embryo (data not shown), which is as expected because initial *mgm-1* expression is dependent on germ cell autonomous cues[25]. However, removal of zygotic *Stat92E* activity completely abolished *mgm-1* expression in first instar larvae. In wild-type first instar male larvae (Fig. 3a), *mgm-1* expression was observed in most germ cells, which are likely to be developing male germline stem cells and spermatogonia. No *mgm-1* expression was observed in *Stat92E* mutant larvae (Fig. 3c), and β-galactosidase expression was only observed in the soma, not the germ line, in the pattern expected from the *Stat92E* P element allele (Fig. 3d). In an experiment where 25% of larvae were expected to be both male and contain the *mgm-1* enhancer trap (see Methods), 23.2% ($n = 262$) of wild-type larvae exhibited *mgm-1* expression in the germ cells, whereas no *Stat92E* mutant larvae ($n = 55$) exhibited germ cell *mgm-1* expression (Fig. 3c; this is significantly different from wild-type siblings, $P = 2.9 \times 10^{-5}$). Thus, *Stat92E* mutants exhibit a strong effect on male germline development, and some male germline cell types are either missing or have an altered identity.

Finally, we assessed the extent to which activation of the JAK/STAT pathway can masculinize female germ cells. Female germ cells expressing *upd* are not expected to be fully masculinized because, although we are activating a male-specific signal, these germ cells are otherwise still in a female somatic environment and retain female germ cell autonomous cues. Indeed, such embryos gave rise to fertile adult females, indicating that at least some germ cells retain, or revert back to, a female identity (data not shown). This may be due, in part, to the failure of the *upd* construct to be expressed in the adult female germ line[26]. However, *upd* is sufficient to induce male-specific gene expression in embryonic XX germ cells. Although *mgm-1* is normally expressed only in germ cells in males (Fig. 3e, f), *mgm-1* was expressed in all embryos (100%, $n = 71$, Fig. 3g) when *upd* was ectopically expressed. In addition, we have identified two new male germline markers—*disc proliferation abnormal (dpa)* and *Minichromosome maintenance 5 (Mcm5)* (Fig. 3h, i and data not shown)—that can also be induced by *upd*. Whereas these genes were normally expressed in germ cells only in males (*dpa*: 96% of males, $n = 56$, 0% females, $n = 63$; *Mcm5*: 89.7% males, $n = 68$, 0% females, $n = 70$), female embryos exhibited germ cell expression of these genes when *upd* was ectopically expressed (Fig. 3j and data not shown). In an experiment where only 50% of embryos are expected to express ectopic *upd* in the germ line (see Methods), 32.5% of female embryos

Table 1 | Germ cell division in embryonic gonads is male-specific and regulated by the JAK/STAT pathway

Genotype	Percentage of XY embryos with one or more dividing germ cells (n)	Percentage of XX embryos with one or more dividing germ cells (n)
Wild type*	28.0 (147)	0 (170)
eyes absent	0 (69)	ND
wunen (mis-localized germ cells)†	0 (42)	ND
wunen (gonadal germ cells)†	33.0 (42)	ND
dsx^D/dsx^1 (masculinized soma)	39.0 (41)	24 (33)
twi-Gal4/UAS-tra^F (feminized soma)	0 (42)	0 (40)
Stat92E[06346]	4.0 (101)¶	0 (90)
Control‡	28.0 (232)¶	0 (214)
UAS-*Socs36E*; *nanos*-Gal4	6.3 (111)‖	0 (103)
Control§	38.8 (85)‖	0 (88)
$Df(1)os^{1a}$	7.5 (94)#	NA**
Control‡	35.1 (94)#	ND
UAS-*upd*; *nanos*-Gal4	55.0 (40)	19.0 (58)☆
Control‡	21.7 (46)	0 (46)☆

*w^{1118} ($n = 108$) and *faf-lacZ* ($n = 209$).
†Mis-localized germ cells and germ cells in gonads examined separately.
‡Wild-type siblings.
§UAS-mCD::GFP driven by the *nanos*-Gal4 driver.
¶$P = 2.1 \times 10^{-7}$; Student's *t*-test (see Methods).
‖$P = 3.2 \times 10^{-9}$; Student's *t*-test (see Methods).
#$P = 1.1 \times 10^{-6}$; Student's *t*-test (see Methods).
☆$P = 7.8 \times 10^{-4}$; Student's *t*-test (see Methods).
**Not applicable; deficiency on X chromosome.

expressed *dpa* ($n = 135$) and 21.3% expressed *Mcm5* ($n = 47$). Therefore, *upd* expression is sufficient to activate male-specific gene expression in female germ cells.

Our data indicate that the JAK/STAT pathway mediates a critical signal from the male somatic gonad that is required for male germ cell development. This signal probably acts together with male germ cell autonomous cues to promote male germline identity and spermatogenesis. This signal is also sufficient to activate the male pattern of proliferation and gene expression in female germ cells, even when these germ cells retain female germ cell autonomous cues and are present in an otherwise female soma. It will be interesting in the future to identify additional (for example, female) somatic signals, along with germ cell autonomous cues, and to assess the relative contribution of these factors to proper germline sexual development. Because one of the earliest aspects of sex-specific germ cell behaviour in both *Drosophila* and mouse is the regulation of the germline cell cycle by the somatic gonad, it will be of further interest to determine whether the somatic signals operating in *Drosophila* have a similar role in germline sex determination in mammals.

METHODS

Fly stocks. w^{1118}, Canton-S and *ru st faf-lacZ e ca* flies were used as wild type. *twi*-Gal4, *24B*-Gal4 (mesodermal drivers) and *nanos*-Gal4::VP16 (germ cell driver) were used to express the following UAS constructs as indicated in the text: UAS-*upd* (M. Zeidler), UAS-*CG5988* (*upd2*), UAS-*CG5963* (*upd3*) and UAS-*hop*^{TumL} (D. Harrison), UAS-*Socs36E* (B. Mathey-Prevot), UAS-*tra*^{F} and UAS-*mCD8::GFP*. Other stocks include *mgm-1* (M. Steinmann-Zwicky), *Stat92E*^{06346}, *Df(1)os*^{1a} (D. Harrison), *upd*^{YM55}, *wunen*^{CE} (K. Howard), *hmgcr*^{clb1}, *eya*$^{cli-IID}$, *dsx*^{D} and *dsx*^{1}. Any unspecified stocks are from the Van Doren Laboratory or the Bloomington Stock Center (http://flybase.bio.indiana.edu/).

Immunocytochemistry, *in situ* hybridization and X-gal staining. Immunocytochemistry was performed as described for embryos[27] and larvae[28] except that 0.1% Tween-20 was substituted for Triton X-100 during sonication and immunostaining of larvae. For quantification of germ cell division, embryos were obtained for fixation from collections 6 h after laying at 25 °C then aged 21 h at 18 °C. Larval analyses were conducted using 30–48-h-old larvae aged at 25 °C. RNA *in situ* hybridization was performed as described[27] with hybridization carried out for 20–48 h and revealed using either colorimetric (NBT/BCIP) or fluorescent (HNPP, Roche) alkaline phosphatase substrates. *upd* antisense riboprobe was prepared by digesting pBS-KS$^+$-*upd* with *Hind*III, and transcribing with T7 polymerase (Promega) using digoxigenin-labelled UTP (Boehringer-Mannheim). *dpa* and *Mcm5* anti-sense riboprobes were generated by digesting pFLC-1-*dpa* or pFLC-1-*Mcm5* (clones RE04051 and RE67590 from BACPAC Resources at CHORI) with *Sac*I or *Sma*I respectively, then transcribing with T3 polymerase (Promega). X-gal staining of embryos was performed as described[29] with reactions proceeding for up to 48 h.

Antibodies. The following primary antibodies were used: chick anti-Vasa (K. Howard) at 1:10,000; rabbit anti-STAT92E (S. Hou) at 1:1,000; rabbit anti-phospho-STAT92E (Cell Signaling Technologies) at 1:50; rabbit anti-phospho-histone H3 (Upstate Biotechnology) at 1:1,000; rabbit anti-Sox100B (S. Russell) at 1:1,000; rabbit anti-GFP (Torrey Pines) at 1:3,000; rabbit anti-β-gal (Cappel) at 1:10,000; mouse anti-β-gal (Promega) at 1:1,000; mouse anti-SXL-M18 (Developmental Studies Hybridoma Bank (DSHB; P. Schedl) at 1:300; mouse anti-EN-4D9 (DSHB; C. Goodman) at 1:2; mouse anti-EYA-10H6 (DSHB; N. Bonini) at 1:25; AP-conjugated sheep anti-digoxigenin (Roche) at 1:2,000. Secondary antibodies used at 1:500 include: Alexa488- and Alexa546-goat anti-rabbit, Alexa488- and Alexa568-goat anti-mouse, Alexa546-goat anti-chick (all from Molecular Probes), Cy5-goat anti-rabbit (Amersham-Pharmacia), and Cy5-goat anti-chick (Rockland).

Analysis of whole-mount embryos and larvae. Embryos and larvae were mounted in 70% glycerol in PBS with 0.1% Triton X-100 (Sigma) or 2.5% DABCO (Sigma) in 70% glycerol. Samples were viewed with a Zeiss Axioplan light microscope or a Zeiss 510 Meta confocal microscope. Staging of embryos was conducted according to ref. 30. Genotype of embryos was determined using GFP-expressing balancer chromosomes as described[27]. Statistical analyses were conducted using a Student's *t*-test assuming equal variance to compare test versus control samples.

The sex of embryos and larvae was determined either by immunostaining with female-specific anti-SXL antibodies (Figs 1e, f and 2a–d, h) or male-specific anti-Sox100B antibodies (Fig. 3a–d), or by using an X chromosome carrying either P{Dfd-lacZ-HZ2.7} (W. McGinnis; Figs 1a–d and 3h–j) or FM7c, P{GAL4-Kr.C}, P{UAS-GFP.S65T} (Figs 2e, f and 3e–g) (channels not shown in figures). Labelled X chromosomes were only present in male parents, and therefore only inherited by female embryos.

Embryos from mating number 1 (see below) were used to examine *mgm-1* expression in *Stat92E* mutant larvae. *Stat92E* mutants were identified by lack of GFP expression in the somatic gonad. These larvae were also labelled with anti-Sox100B to make sure that a normal fraction of male larvae were obtained. Embryos from mating number 2 were used to determine the effects of *upd* expression on *mgm-1* expression in germ cells. *mgm-1*/UAS-*upd*; *nanos*-Gal4/+ embryos were identified by *mgm-1* expression in the anterior soma, combined with lack of *ftz-lacZ* expression. Embryos from mating number 3 were used to examine the effects of *upd* expression on *dpa* and *Mcm5* expression. Embryonic sex was determined by the presence or absence of *Dfd-lacZ* expression. For the mating scheme below, *Kr-GFP* indicates P{Gal4-Kr.C}, P{UAS-GFP.S65T}.

Mating 1: Male $\dfrac{Mgm\text{-}1,\ CyO}{+}$; $\dfrac{Stat92E^{06346}}{TM3\ Sb,\ Kr\text{-}GFP}$ X Female $\dfrac{Stat92E^{06346}}{Tm3\ Sb,\ Kr\text{-}GFP}$

Mating 2: Male $\dfrac{Mgm\text{-}1,\ CyO}{Sp}$; $\dfrac{nanos\text{-}Gal4}{nanos\text{-}Gal4}$ X Female $\dfrac{UAS\text{-}upd}{CyO,\ ftz\text{-}lacZ}$

Mating 3: Male $\dfrac{Dfd\text{-}lacZ}{}$; $\dfrac{UAS\text{-}upd}{+}$ X Female $\dfrac{nanos\text{-}Gal4}{nanos\text{-}Gal4}$

1. Steinmann-Zwicky, M., Schmid, H. & Nothiger, R. Cell-autonomous and inductive signals can determine the sex of the germ line of *Drosophila* by regulating the gene *Sxl*. *Cell* **57**, 157–166 (1989).
2. McLaren, A. The fate of germ cells in the testis of fetal Sex-reversed mice. *J. Reprod. Fertil.* **61**, 461–467 (1981).
3. Oliver, B. Genetic control of germline sexual dimorphism in *Drosophila*. *Int. Rev. Cytol.* **219**, 1–60 (2002).
4. McLaren, A. Primordial germ cells in the mouse. *Dev. Biol.* **262**, 1–15 (2003).
5. Hou, S. X., Zheng, Z., Chen, X. & Perrimon, N. The Jak/STAT pathway in model organisms: emerging roles in cell movement. *Dev. Cell* **3**, 765–778 (2002).
6. Ichiba, M., Nakajima, K., Yamanaka, Y., Kiuchi, N. & Hirano, T. Autoregulation of the *Stat3* gene through cooperation with a cAMP-responsive element-binding protein. *J. Biol. Chem.* **273**, 6132–6138 (1998).
7. Callus, B. A. & Mathey-Prevot, B. SOCS36E, a novel *Drosophila* SOCS protein, suppresses JAK/STAT and EGF-R signalling in the imaginal wing disc. *Oncogene* **21**, 4812–4821 (2002).
8. Harrison, D. A., Binari, R., Nahreini, T. S., Gilman, M. & Perrimon, N. Activation of a *Drosophila* Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. *EMBO J.* **14**, 2857–2865 (1995).
9. Li, J., Xia, F. & Li, W. X. Coactivation of STAT and Ras is required for germ cell proliferation and invasive migration in *Drosophila*. *Dev. Cell* **5**, 787–798 (2003).
10. Boyle, M., Bonini, N. & DiNardo, S. Expression and function of clift in the development of somatic gonadal precursors within the *Drosophila* mesoderm. *Development* **124**, 971–982 (1997).
11. Zhang, N., Zhang, J., Purcell, K. J., Cheng, Y. & Howard, K. The *Drosophila* protein Wunen repels migrating germ cells. *Nature* **385**, 64–67 (1997).
12. Van Doren, M., Broihier, H. T., Moore, L. A. & Lehmann, R. HMG-CoA reductase guides migrating primordial germ cells. *Nature* **396**, 466–469 (1998).
13. Nagoshi, R. N. & Baker, B. S. Regulation of sex-specific RNA splicing at the *Drosophila* doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes Dev.* **4**, 89–97 (1990).
14. Schupbach, T. Autosomal mutations that interfere with sex determination in somatic cells of *Drosophila* have no direct effect on the germline. *Dev. Biol.* **89**, 117–127 (1982).
15. Agaisse, H., Petersen, U. M., Boutros, M., Mathey-Prevot, B. & Perrimon, N. Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury. *Dev. Cell* **5**, 441–450 (2003).
16. Harrison, D. A., McCoon, P. E., Binari, R., Gilman, M. & Perrimon, N. *Drosophila* unpaired encodes a secreted protein that activates the JAK signalling pathway. *Genes Dev.* **12**, 3252–3263 (1998).
17. Jinks, T. M., Polydorides, A. D., Calhoun, G. & Schedl, P. The JAK/STAT signalling pathway is required for the initial choice of sexual identity in *Drosophila melanogaster*. *Mol. Cell* **5**, 581–587 (2000).
18. Sefton, L., Timmer, J. R., Zhang, Y., Beranger, F. & Cline, T. W. An extracellular activator of the *Drosophila* JAK/STAT pathway is a sex-determination signal element. *Nature* **405**, 970–973 (2000).
19. Tulina, N. & Matunis, E. Control of stem cell self-renewal in *Drosophila* spermatogenesis by JAK-STAT signalling. *Science* **294**, 2546–2549 (2001).
20. Kiger, A. A., Jones, D. L., Schultz, C., Rogers, M. B. & Fuller, M. T. Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**, 2542–2545 (2001).

21. Sonnenblick, B. P. Germ cell movements and sex differentiation of the gonads in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **26**, 373–381 (1941).

22. Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K. & Kobayashi, S. Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. *Nature Cell Biol.* **1**, 431–437 (1999).

23. Kerkis, J. The growth of the gonads in *Drosophila melanogaster. Genetics* **16**, 212–244 (1931).

24. Staab, S., Heller, A. & Steinmann-Zwicky, M. Somatic sex-determining signals act on XX germ cells in *Drosophila* embryos. *Development* **122**, 4065–4071 (1996).

25. Janzer, B. & Steinmann-Zwicky, M. Cell-autonomous and somatic signals control sex-specific gene expression in XY germ cells of *Drosophila. Mech. Dev.* **100**, 3–13 (2001).

26. Rorth, P. Gal4 in the *Drosophila* female germline. *Mech. Dev.* **78**, 113–118 (1998).

27. DeFalco, T. J. *et al.* Sex-specific apoptosis regulates sexual dimorphism in the *Drosophila* embryonic gonad. *Dev. Cell* **5**, 205–216 (2003).

28. Patel, N. in *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology* (eds Goldstein, L. S. B. & Fyrberg, E. A.) 445–487 (Academic Press, San Diego, 1994).

29. Moore, L. A., Broihier, H. T., Van Doren, M., Lunsford, L. B. & Lehmann, R. Identification of genes controlling germ cell migration and embryonic gonad formation in *Drosophila. Development* **125**, 667–678 (1998).

30. Campos-Ortega, J. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, New York, 1985).

LETTERS

The Lyme disease agent exploits a tick protein to infect the mammalian host

Nandhini Ramamoorthi[1], Sukanya Narasimhan[1], Utpal Pal[1], Fukai Bao[1], Xiaofeng F. Yang[4], Durland Fish[3], Juan Anguita[5], Michael V. Norgard[4], Fred S. Kantor[2], John F. Anderson[6], Raymond A. Koski[7] & Erol Fikrig[1,3]

The Lyme disease agent, *Borrelia burgdorferi*, is maintained in a tick–mouse cycle[1,2]. Here we show that *B. burgdorferi* usurps a tick salivary protein, Salp15 (ref. 3), to facilitate the infection of mice. The level of *salp15* expression was selectively enhanced by the presence of *B. burgdorferi* in *Ixodes scapularis*, first indicating that spirochaetes might use Salp15 during transmission. Salp15 was then shown to adhere to the spirochaete, both *in vitro* and *in vivo*, and specifically interacted with *B. burgdorferi* outer surface protein C. The binding of Salp15 protected *B. burgdorferi* from antibody-mediated killing *in vitro* and provided spirochaetes with a marked advantage when they were inoculated into naive mice or animals previously infected with *B. burgdorferi*. Moreover, RNA interference-mediated repression of *salp15* in *I. scapularis* drastically reduced the capacity of tick-borne spirochaetes to infect mice. These results show the capacity of a pathogen to use a secreted arthropod protein to help it colonize the mammalian host.

Lyme borreliosis serves as a model to examine how microbe–vector interactions influence pathogen transmission to the mammalian host. *B. burgdorferi*, the spirochetal agent of Lyme disease, is primarily maintained in the USA in *I. scapularis* ticks and *Peromyscus leucopus* mice. Spirochaetes preferentially express specific genes to survive in a complex enzootic cycle[4]. For example, *B. burgdorferi* cells entering *I. scapularis* express outer surface protein (Osp)A to colonize the vector[5]. When the arthropod engorges on a host, the spirochaetes then downregulate OspA and upregulate OspC, a lipoprotein that facilitates the migration of *B. burgdorferi* from the

I. scapularis gut to the tick salivary glands and can independently participate in the establishment of vertebrate infection[6,7]. While *B. burgdorferi* are being transmitted during tick feeding, the arthropod is also secreting saliva to aid in engorgement[8]. *I. scapularis* saliva possesses antigens with immunosuppressive, anticomplement and antihaemostatic activity, among other functions, which enable the vector to take an effective blood meal[9,10]. We now explore the hypothesis that *B. burgdorferi* in transit through the tick might use components of *I. scapularis* saliva to enhance spirochaete transmission to, and survival within, the vertebrate host.

To determine first whether *B. burgdorferi* influenced the expression of tick genes, we examined the profile of the genes encoding 14 antigenic *I. scapularis* salivary proteins that elicit strong humoral responses in the host upon tick feeding[11], in uninfected and *B. burgdorferi*-infected ticks. We found that the expression of one gene, *salp15*, which encodes an *I. scapularis* protein known to inhibit T-cell activation[3], was selectively increased in *B. burgdorferi*-infected tick salivary glands during engorgement (Fig. 1a). As an example of one control, the expression of the gene encoding another salivary antigen, *salp25D*, remained the same regardless of infection. Quantitative polymerase chain reaction (PCR) further showed that *salp15* mRNA levels were 13-fold higher ($P < 0.001$) in *B. burgdorferi*-infected engorged ticks than in engorged *I. scapularis* without *B. burgdorferi* (Fig. 1b). In contrast, the amount of *salp25D* mRNA was similar in both groups of ticks (Fig. 1b). The enhancement in *salp15* expression was specific to *B. burgdorferi*, because the levels of *salp15* mRNA in the salivary glands of engorged ticks infected with
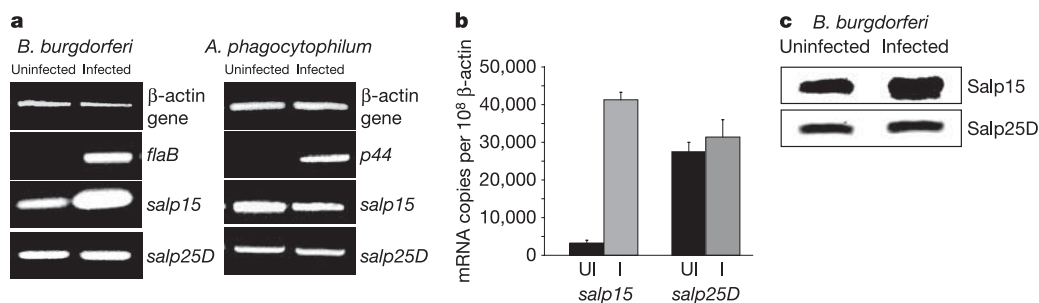
Figure 1 | **Salp15 levels are specifically enhanced in *Borrelia burgdorferi*-infected tick salivary glands. a**, RT–PCR profile of fed *Ixodes scapularis* salivary glands that were uninfected or infected with either *B. burgdorferi* or *Anaplasma phagocytophilum*. Expression of the β-actin gene from *I. scapularis* was used as a control. *flaB* and *p44* expression are indicative of *B. burgdorferi* and *A. phagocytophilum* infection, respectively. **b**, *salp15* was upregulated in *B. burgdorferi*-infected *I. scapularis* salivary glands. The difference between *salp15* mRNA levels in infected and uninfected nymphs was significant, in contrast to *salp25D* expression (Student's *t*-test). Results are means + s.e.m. from three quantitative PCR experiments. **c**, Salp15 protein levels were 1.6-fold higher in *B. burgdorferi*-infected salivary glands, as quantified by ImageJ (NIH). Salp25D served as a control.

[1]Sections of Rheumatology and [2]Allergy and Immunology, Department of Internal Medicine, and [3]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [4]Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [5]Department of Biology, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA. [6]Department of Entomology, Connecticut Agricultural Experiment Station, New Haven, Connecticut 06504, USA. [7]L2 Diagnostics, New Haven, Connecticut 06511, USA.

Anaplasma phagocytophilum, another pathogen borne by *I. scapularis*, remained unchanged (Fig. 1a). Salp15 protein levels were 1.6-fold higher in *B. burgdorferi*-infected salivary glands than in uninfected glands, as demonstrated by immunoblotting (Fig. 1c). The selective upregulation of a tick salivary antigen in the presence of *B. burgdorferi* raises the possibility that Salp15 might be used by the pathogen, either during its interim stay in the arthropod salivary gland or during its transit into the mammalian host.

To assess whether Salp15 interacts with a *B. burgdorferi* antigen, a gel overlay assay, using recombinant Salp15, was performed. Salp15 bound a 22-kDa *B. burgdorferi* antigen that was identified as OspC (Fig. 2a) when subjected to matrix-assisted laser desorption ionization mass spectrometry peptide analysis (Supplementary Table S1). Consistent with this was our observation that Salp15 did not bind to lysates of OspC-deficient *B. burgdorferi* but did adhere to OspC-deficient *B. burgdorferi* that were genetically complemented to produce OspC (Fig. 2b). In addition to binding the spirochaete lysates, we also observed that Salp15 interacted with intact wild-type *B. burgdorferi*, but not with OspC-deficient *B. burgdorferi*, *in vitro* (Fig. 2c). These observations further confirmed that the Salp15–OspC interaction was specific. Moreover, *B. burgdorferi* in infected tick salivary glands were copiously covered with native Salp15, as detected with an antibody directed against recombinant Salp15 (Fig. 2d). As a control, an antibody against another tick salivary protein, Salp25D, failed to bind to *B. burgdorferi* in the salivary gland. Uninfected salivary glands stained diffusely for both proteins (data not shown). Salp15 therefore directly associates with *B. burgdorferi* within the vector.

The enhanced expression of Salp15 in the presence of *B. burgdorferi* within ticks, and the specific adherence of Salp15 to OspC on the surface of *B. burgdorferi*, suggest a critical role for Salp15 in establishing spirochaete infection. To determine whether Salp15 influenced the ability of *B. burgdorferi* to colonize the mammalian host, spirochaetes were preincubated with Salp15 and injected into naive C3H mice. At 25 days the spirochaete load in animals that received *B. burgdorferi* and Salp15 was markedly elevated in the joints (ninefold; $P < 0.001$), skin (fivefold; $P < 0.001$) and bladder (25-fold; $P < 0.001$) as measured by quantitative PCR (Fig. 3a). The levels were higher than in mice that received *B. burgdorferi* alone (Fig. 3a) or in mice in which Salp15 was injected at a distal site from the *B. burgdorferi* inoculum (Supplementary Fig. S1). The antibody titres to *B. burgdorferi* was similar in all groups of mice. Spirochaete levels were also increased to a similar degree at earlier time points (8 and 15 days) in mice that received *B. burgdorferi* preincubated with Salp15 (data not shown). These results are consistent with previous reports that inoculation of *B. burgdorferi* together with tick salivary gland lysates enhanced the spirochaete load in mice[12].

Innate and adaptive immune responses contribute to controlling the levels of *B. burgdorferi* during infection[13,14]. In particular, humoral immunity has repeatedly been shown to destroy spirochaetes, both *in vitro* and *in vivo*[15–17]. Antibody-mediated killing assays were therefore performed to determine whether Salp15 could protect spirochaetes from the borreliacidal effects of *B. burgdorferi* antisera. Cultured wild-type *B. burgdorferi* were killed by a monoclonal antibody against OspA, sera from *B. burgdorferi*-infected mice (data not shown) or rabbit *B. burgdorferi* antisera within 18 h, whereas spirochaetes preincubated with Salp15 were significantly protected (Fig. 3b, c). This effect was not noted with OspC-deficient spirochaetes but was evident with OspC-deficient *B. burgdorferi* that were genetically complemented to produce OspC (data not shown). Microscopic observation revealed an 8.5-fold higher level of viable spirochaetes. The *B. burgdorferi* remained viable until 24 h after exposure to the antisera (Fig. 3b, c). Beyond this time, the Salp15-treated spirochaetes began dying, perhaps because dividing spirochaetes lacking the protective interaction with Salp15 might have been targeted by the borreliacidal antibodies.

In areas where Lyme disease is endemic, mice are repeatedly
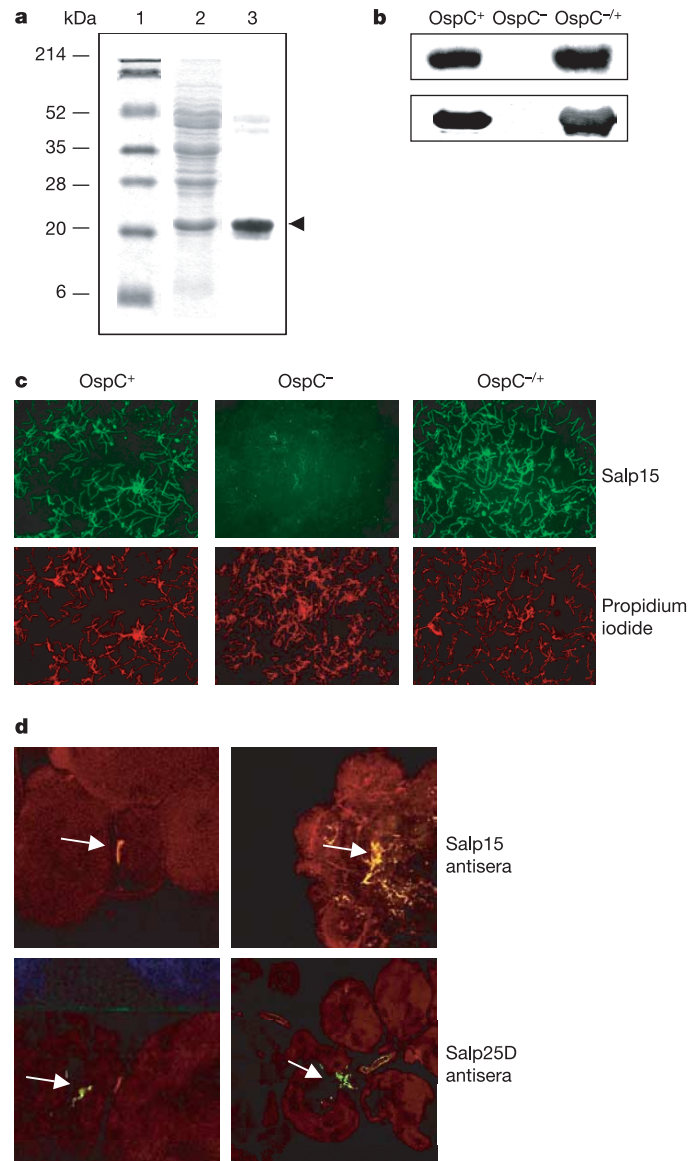
Figure 2 | Salp15 interacts with outer surface protein (Osp)C of *Borrelia burgdorferi*. **a**, Salp15 bound specifically to OspC. Lanes 1 and 2, Ponceau S stain of molecular mass markers (lane 1) and *B. burgdorferi* lysate (lane 2); lane 3, Salp15 overlay. The protein band bound by Salp15 (marked by an arrowhead) was identified as OspC. **b**, Salp15 binding to OspC was confirmed by using wild-type ($OspC^+$), OspC-deficient ($OspC^-$) and OspC-complemented ($OspC^{-/+}$) *B. burgdorferi*. Bacterial lysates of the isolates were probed with anti-Salp15 (top) and anti-OspC (bottom) antibody. **c**, Salp15 binds to the surface of intact *B. burgdorferi*. Unfixed wild-type ($OspC^+$), OspC-deficient ($OspC^-$) and OspC-complemented ($OspC^{-/+}$) *B. burgdorferi* were probed with FITC-conjugated Salp15 (green) and propidium iodide (red). Original magnification × 40. **d**, Salp15 binds *B. burgdorferi* within the tick salivary gland. Top: salivary glands from *B. burgdorferi*-infected nymphs were probed with Salp15 antibody (red). The spirochaetes (as indicated by arrows) were stained with an FITC-labelled anti-*B. burgdorferi* antibody (green). Bottom: anti-Salp25D served as a control. Co-localization (yellow) was observed with Salp15 antisera, in contrast to Salp25D. Left, salivary gland with a single spirochaete; right, a cluster of *B. burgdorferi* is harboured. Images are representative of ten independent experiments. Original magnification × 40.

exposed to ticks with *B. burgdorferi*, and often have evidence of a humoral response to the spirochaete[18]. Therefore, in nature, *B. burgdorferi* transmitted by tick bites are frequently exposed to *B. burgdorferi*-specific antibodies[19]. This is the environment—during the natural infection of mice—in which the influence of Salp15 is
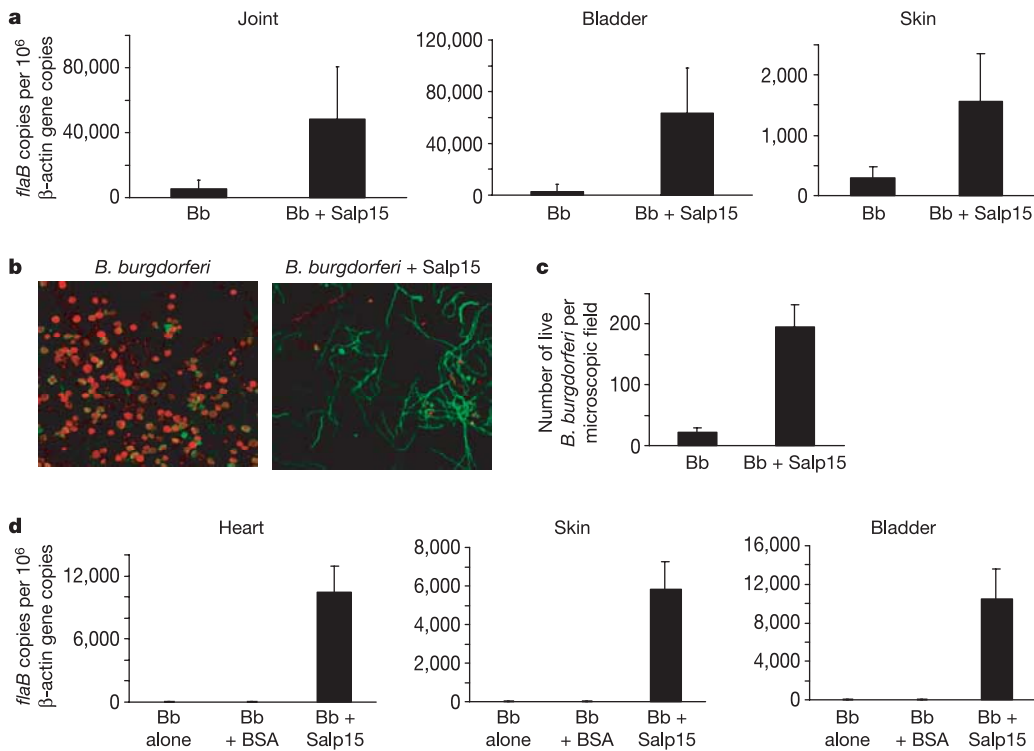
Figure 3 | Salp15 markedly enhances the *Borrelia burgdorferi* load in the murine host. a, Naive mice were inoculated with *B. burgdorferi* in the presence (Bb + Salp15) and absence (Bb) of recombinant Salp15. *B. burgdorferi flaB* was measured in the different tissues. Results are means + s.e.m. from three independent experiments. **b**, Salp15 protects *B. burgdorferi* from antibody-mediated destruction *in vitro*. *B. burgdorferi* antiserum was incubated with wild-type spirochaetes for 18 h. Live bacteria stained green (Syto 9 stain); dead bacteria stained red (propidium iodide). Left, untreated *B. burgdorferi*; right, *B. burgdorferi* preincubated with Salp15. Original magnification × 63. **c**, Quantitative assessment of the data shown in **b**. After exposure to antibody, *B. burgdorferi* preincubated with Salp15 were 8.5-fold more viable than the untreated *B. burgdorferi*. Numbers are averages of 20 random microscopic fields. Results are means + s.e.m. from one representative experiment. **d**, Protection from *in vivo* killing of *B. burgdorferi* by Salp15 in immune mice. Salp15 markedly enhanced the *B. burgdorferi* load in a previously infected murine host. Mice were treated as described in Methods. The different groups of mice received phosphate-buffered saline (Bb), bovine serum albumin (Bb + BSA) as control, or Salp15 (Bb + Salp15). *B. burgdorferi flaB* levels were quantified from the different tissues. Results are means + s.e.m. from three quantitative PCR experiments.

likely to be most important. In laboratory experiments, tick-borne spirochaetes are more resistant than cultured spirochaetes to being killed *in vivo* by immune sera[20]. It was therefore important to examine whether the interaction between Salp15 and *B. burgdorferi* could enable spirochaetes to colonize mice that had previously developed an immune response to *B. burgdorferi*. To mimic this situation, C3H mice were infected with *B. burgdorferi* for 14 days, a period during which a protective anti-*B. burgdorferi* humoral response to infection has been noted[20], and the mice were then treated with ceftriaxone, an antibiotic that eradicates spirochaete infection[21]. As expected, these animals were resistant to infection with *in vitro* cultured *B. burgdorferi* administered alone or in the presence of bovine serum albumin (Fig. 3d). In contrast, these mice were fully susceptible to infection with spirochaetes that had been preincubated with Salp15 (Fig. 3d). Quantitative PCR performed 25 days after infection detected *B. burgdorferi* only in mice that received spirochaetes with Salp15, and not in animals that received *B. burgdorferi* alone. Moreover, the spirochaete burdens were comparable to those in the non-immune mice infected with *B. burgdorferi* in the presence of Salp15. Taken together, these data indicate that Salp15 allows *B. burgdorferi* to successfully colonize mice that had been previously exposed to spirochaetes and also indicate that Salp15 might be important in spirochaete colonization and dissemination.

To further examine the role of Salp15 in *B. burgdorferi* survival *in vivo*, *salp15*-deficient *I. scapularis* nymphs were generated by RNA interference (RNAi). The successful downregulation of *salp15* expression in tick salivary glands was determined by PCR with reverse transcription (RT–PCR) (Fig. 4a). The unchanged levels of *salp25D* and tick β-actin gene confirmed the specificity of RNAi. Buffer-injected (mock) and *salp15* double-stranded RNA (dsRNA)-injected *B. burgdorferi*-infected nymphs were then fed on naive laboratory mice and on mice that had previously developed an immune response to *B. burgdorferi*. Quantitative PCR revealed a significant decrease in spirochaete levels in the skin of mice that were fed on by the *salp15*-deficient ticks in comparison with mock-injected ticks, both in naive mice (Fig. 4b; $P < 0.01$) and in mice that had an immune response (Fig. 4c; $P < 0.001$). In contrast, spirochaete levels in the salivary glands of engorged *salp15*-deficient and control tick salivary glands $(2.9 \pm 1.3) \times 10^2$ and $(3.1 \pm 2.1) \times 10^2$ *flaB* copies per 10^6 β-actin gene copies, respectively, $(P > 0.5)$ were similar. Spirochaete numbers in the host were unaffected for *salp25D*-deficient nymphs (data not shown). Studies were also performed with *P. leucopus*, a natural reservoir of *B. burgdorferi*, and similar results were obtained when *salp15* dsRNA injected *B. burgdorferi*-infected nymphs were fed on naive *P. leucopus* mice and on mice that had an immune response to *B. burgdorferi* (Supplementary Fig. S2a, b). The *salp15*-deficient ticks showed no alterations in engorgement, ruling out the possibility of reduced transmission as a result of decreased feeding. The lack of complete abrogation of pathogen transmission in *salp15*-deficient *I. scapularis* might be due to the inability of RNAi to suppress gene expression totally or due to the spirochaetes' ability to compensate for the absence of *salp15*. These *in vivo* data show conclusively that Salp15 facilitates tick-borne *B. burgdorferi* infection in the host.
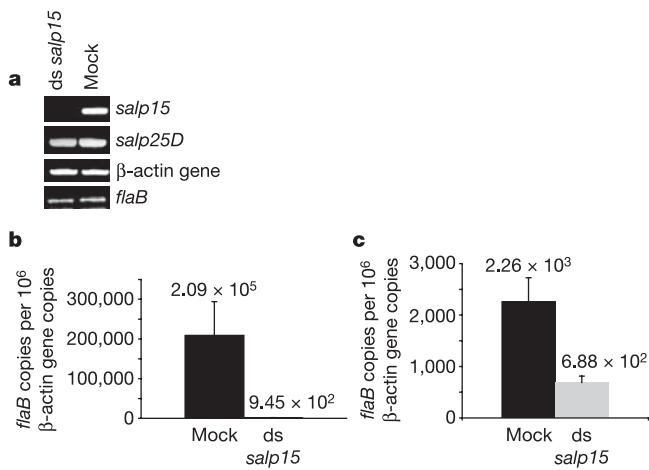
Figure 4 | Gene silencing of *salp15* expression by RNA interference.
a, *Borrelia burgdorferi*-infected nymphal ticks were microinjected with *salp15* dsRNA (ds *salp15*) or buffer alone (mock), and fed on naive mice. Levels of *salp15*, *salp25D* and β-actin gene were assessed in the salivary gland by RT–PCR. Levels of *flaB* were also measured. Data are representative of three independent experiments. **b**, **c**, *salp15* dsRNA reduced the transmission of *B. burgdorferi* to the host. Nymphal ticks were injected with *salp15* dsRNA (ds *salp15*) or buffer alone (mock), fed on naive mice or on mice previously infected with *B. burgdorferi* and then treated as described in Methods. Quantification of *flaB* revealed lower levels of spirochaete in the skin of mice that were fed upon by *salp15*-deficient ticks in both the naive mice (**b**) and the preimmune mice (**c**) in comparison with mock-injected ticks (Student's *t*-test). Results are means + s.e.m. and data are representative of three independent experiments.

Transmission of an arthropod-borne pathogen occurs at the complex interface of microbe, vector and vertebrate host. Although host–pathogen, vector–pathogen and vector–host interactions have been delineated[22,23], we now describe a triangular relationship in which an infectious agent exploits an arthropod protein to facilitate infection of the mammalian host. The use of a specific *I. scapularis* salivary protein by *B. burgdorferi* to enhance infection in mice serves as a model for the other arthropod-borne infections. In particular, the presence of *B. burgdorferi* within the vector induces the expression of Salp15 in ticks, and this pathogen–vector interaction is needed for the establishment of *B. burgdorferi* infection in mice that have previously been exposed to the spirochaetes, the most common reservoir host environment that this microbe encounters naturally. Although these present data show that Salp15 is used by *B. burgdorferi* to augment infection, the increased Salp15 levels in ticks during spirochaete infection could also be advantageous to *I. scapularis*. Perhaps *B. burgdorferi*-induced enhancement of Salp15, with its immunosuppressive properties[3], might enable ticks to more effectively engorge and/or avoid rejection by the host[24], providing a preferential survival advantage for both *I. scapularis* and *B. burgdorferi*. As microbes and vectors have evolved together over millions of years, the most successful relationship between them will be mutual, rather than parasitic or commensal, as the interaction between *B. burgdorferi* and *I. scapularis* may well be. The vector and pathogen factors that influence successful microbial infection of the mammalian host might also serve as targets for vaccines and therapeutics to combat arthropod-borne diseases.

METHODS

Spirochaetes. A low-passage clonal isolate of *Borrelia burgdorferi* strain N40 that is infectious in mice[25] was used throughout the study. Clonal isolates of OspC-deficient and OspC-complemented and wild-type *B. burgdorferi* 297 (ref. 6) were used for immunofluorescence and *in vitro* binding assays.
Mice. C3H/HeJ (C3H) mice were purchased from the Jackson Laboratory. *Peromyscus leucopus* were obtained from a colony at Yale University, maintained by D. Fish.

Ticks. *Ixodes scapularis* nymphs were obtained from The Connecticut Agricultural Experiment Station. Laboratory-infected *I. scapularis* nymphs were obtained as described previously[26]. The uninfected and infected nymphs were fed on pathogen-free C3H mice and were collected after 66 h of feeding for the dissection of salivary glands.
RNA extractions from tick salivary glands. RNA was extracted from the following tick pools: uninfected and infected nymphal tick salivary glands, in accordance with the kit manufacturer's directions (AquaPure RNA isolation kit; Bio-Rad). The RNA samples purified from these groups were used in the qualitative RT–PCR and quantitative PCR assays.
RT–PCR and quantitative PCR. The primer pairs for each assayed gene are listed in Supplementary Table S2. Tick salivary gland complementary DNA was made with the iScript cDNA synthesis kit (Bio-Rad) in accordance with the manufacturer's protocol. Quantitative PCR was performed in accordance with the manufacturer's protocol with a Bio-Rad i-Cycler. The probes used contained a 5′ reporter, 6-carboxyfluorescein (FAM), and a 3′ quencher, 6-carboxy-*N*,*N*,*N*′,*N*′-tetramathylrhodamine (TAMRA; Applied Biosystems). The tick salivary-gland cDNA levels were normalized to *I. scapularis* β-actin gene, and *salp15*, *salp25D* and *B. burgdorferi* N40 flagellin (*flaB*) were then quantified. In the murine infection studies, the mouse β-actin gene was used to normalize the amount of DNA in these samples and *flaB* was used to quantify the levels of spirochaete in the murine tissue samples.
Western blotting. Western blotting was performed as described previously[11]. The blots were probed with guinea pig Salp15 or Salp25D antisera.
Solid-phase overlay assay. *B. burgdorferi* lysates were separated by conventional SDS–PAGE and transferred by western blotting. The blots were then overlaid with purified recombinant Salp15 (ref. 3). The membranes were probed with horseradish peroxidase-conjugated monoclonal anti-V5 antibody (Invitrogen). Mouse anti-OspC monoclonal antibody was also used to probe the blots.
Confocal microscopy. Salivary glands from nymphal ticks were prepared for confocal microscopy as described previously[27]. In brief, acetone-fixed glands were incubated with the guinea pig antisera raised against Salp15 and Salp25D, followed by rhodamine-conjugated anti-guinea-pig IgG (Molecular Probes). The glands were counterstained with fluorescein isothiocyanate (FITC)-conjugated goat anti-*B. burgdorferi* antibody (Kirkegaard and Perry Laboratories), and viewed with a Zeiss LSM 510 scanning laser confocal microscope.
Immunofluorescence and *in vitro* binding assay. Wild-type *B. burgdorferi* 297, OspC-deficient and OspC-complemented isolates (10^7 spirochaetes ml^{-1}) were placed on sialyated slides (PGC Scientific). The spirochaetes were then incubated with Salp15-FITC or BSA-FITC[3], counterstained with propidium iodide and examined with a Zeiss Axioscope fluorescence microscope.
***In vitro* protection assay.** *B. burgdorferi* N40, OspC-deficient and OspC-complemented isolates (10^8 spirochaetes ml^{-1}) were used. Salp15 was incubated with the spirochaete for 1 h at 25 °C. The spirochaetes were then incubated with OspA monoclonal antibody or polyclonal mouse or rabbit *B. burgdorferi* antisera for 18, 24 or 48 h at 33 °C. The percentage of viable spirochaetes was quantified with the Live–Dead Bacterial viability kit (Molecular Probes).
***In vivo* infection of non-immune mice.** Pathogen-free C3H mice were infected with *B. burgdorferi* (10^2 spirochaetes per mouse) with and without recombinant Salp15 or BSA ($30 \mu g$ ml^{-1}) intradermally. At 25 days after infection, tissue samples were collected for DNA isolation and quantitative PCR.
***In vivo* infection of immune mice.** Pathogen-free C3H mice (five mice per group) were infected with *B. burgdorferi* (10^2 spirochaetes per mouse). At 14 days after infection, the mice were tested for antibody raised against *B. burgdorferi* by enzyme-linked immunosorbent assay. They were then treated with ceftriaxone (16 mg per kg body weight). Ear punches were taken to amplify *flaB* DNA to confirm the absence of spirochaete. The mice were then inoculated with *B. burgdorferi* (10^4 spirochaetes per mouse) with recombinant Salp15 ($30 \mu g$ per mouse) intradermally. Control mice received *B. burgdorferi* with or without BSA. Tissue samples were collected, and spirochaete levels were quantified 25 days later.
RNA interference. cDNA from nymphs was prepared as described previously and *salp15* was amplified with gene-specific primers 5′-GAGCTCGCATCAACCGCTGACAAA-3′ and 5′-GGTACCCTAACATCCGGGAATGTG-3′ containing *Sac*I and *Kpn*I restriction sites. The *salp15* fragment was cloned into the L440 double T7 Script vector, and the dsRNA was synthesized and purified with the Megascript RNAi kit in accordance with the manufacturer's protocol (Ambion). *salp15* dsRNA was injected as described[28] into nymphal *I. scapularis* infected with *B. burgdorferi*. The ticks were then placed on naive or previously infected C3H mice or *P. leucopus* (as described above) and allowed to feed until 72 h. Ticks were collected, cDNA was made from salivary glands and the expression of *salp15* was assessed. Seven days after tick feeding, skin samples from the site of a tick bite were excised and levels of *flaB* were measured with quantitative PCR.

1. Burgdorfer, W. *et al.* Lyme disease—a tick-borne spirochetosis? *Science* **216**, 1317–1319 (1982).
2. Ribeiro, J. M., Mather, T. N., Piesman, J. & Spielman, A. Dissemination and salivary delivery of Lyme disease spirochetes in vector ticks (Acari: Ixodidae). *J. Med. Entomol.* **24**, 201–205 (1987).
3. Anguita, J. *et al.* Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4$^+$ T cell activation. *Immunity* **16**, 849–859 (2002).
4. Steere, A. C., Coburn, J. & Glickstein, L. The emergence of Lyme disease. *J. Clin. Invest.* **113**, 1093–1101 (2004).
5. Schwan, T. G. & Piesman, J. Temporal changes in outer surface proteins A and C of the Lyme disease-associated spirochete, *Borrelia burgdorferi*, during the chain of infection in ticks and mice. *J. Clin. Microbiol.* **38**, 382–388 (2000).
6. Pal, U. *et al.* OspC facilitates *Borrelia burgdorferi* invasion of *Ixodes scapularis* salivary glands. *J. Clin. Invest.* **113**, 220–230 (2004).
7. Grimm, D. *et al.* Outer-surface protein C of the Lyme disease spirochete: a protein induced in ticks for infection of mammals. *Proc. Natl Acad. Sci. USA* **101**, 3142–3147 (2004).
8. Wikel, S. K. Host immunity to ticks. *Annu. Rev. Entomol.* **41**, 1–22 (1996).
9. Sauer, J. R., McSwain, J. L., Bowman, A. S. & Essenberg, R. C. Tick salivary gland physiology. *Annu. Rev. Entomol.* **40**, 245–267 (1995).
10. Ribeiro, J. M. & Francischetti, I. M. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol.* **48**, 73–88 (2003).
11. Das, S. *et al.* Salp25D, an *Ixodes scapularis* antioxidant, is 1 of 14 immunodominant antigens in engorged tick salivary glands. *J. Infect. Dis.* **184**, 1056–1064 (2001).
12. Zeidner, N. S., Schneider, B. S., Nuncio, M. S., Gern, L. & Piesman, J. Coinoculation of *Borrelia* spp. with tick salivary gland lysate enhances spirochete load in mice and is tick species-specific. *J. Parasitol.* **88**, 1276–1278 (2002).
13. Steere, A. C. & Glickstein, L. Elucidation of Lyme arthritis. *Nature Rev. Immunol.* **4**, 143–152 (2004).
14. Wooten, R. M. *et al.* Toll-like receptor 2 is required for innate, but not acquired, host defense to *Borrelia burgdorferi*. *J. Immunol.* **168**, 348–355 (2002).
15. McKisic, M. D. & Barthold, S. W. T-cell-independent responses to *Borrelia burgdorferi* are critical for protective immunity and resolution of Lyme disease. *Infect. Immun.* **68**, 5190–5197 (2000).
16. Fikrig, E., Barthold, S. W., Chen, M., Chang, C. H. & Flavell, R. A. Protective antibodies develop, and murine Lyme arthritis regresses, in the absence of MHC class II and CD4+ T cells. *J. Immunol.* **159**, 5682–5686 (1997).
17. Sadziene, A., Thompson, P. A. & Barbour, A. G. *In vitro* inhibition of *Borrelia burgdorferi* growth by antibodies. *J. Infect. Dis.* **167**, 165–172 (1993).
18. Bunikis, J. *et al. Borrelia burgdorferi* infection in a natural population of *Peromyscus leucopus* mice: a longitudinal study in an area where Lyme borreliosis is highly endemic. *J. Infect. Dis.* **189**, 1515–1523 (2004).
19. Brunet, L. R., Sellitto, C., Spielman, A. & Telford, S. R. Antibody response of the mouse reservoir of *Borrelia burgdorferi* in nature. *Infect. Immun.* **63**, 3030–3036 (1995).
20. de Silva, A. M. *et al.* Immune evasion by tickborne and host-adapted *Borrelia burgdorferi*. *J. Infect. Dis.* **177**, 395–400 (1998).
21. Malawista, S. E., Barthold, S. W. & Persing, D. H. Fate of *Borrelia burgdorferi* DNA in tissues of infected mice after antibiotic treatment. *J. Infect. Dis.* **170**, 1312–1316 (1994).
22. Coleman, J. L. *et al.* Plasminogen is required for efficient dissemination of *Borrelia burgdorferi* in ticks and for enhancement of spirochetemia in mice. *Cell* **89**, 1111–1119 (1997).
23. Nagamune, K. *et al.* Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. *J. Exp. Med.* **199**, 1445–1450 (2004).
24. Nazario, S. *et al.* Prevention of *Borrelia burgdorferi* transmission in guinea pigs by tick immunity. *Am. J. Trop. Med. Hyg.* **58**, 780–785 (1998).
25. Barthold, S. W., Beck, D. S., Hansen, G. M., Terwilliger, G. A. & Moody, K. D. Lyme borreliosis in selected strains and ages of laboratory mice. *J. Infect. Dis.* **162**, 133–138 (1990).
26. Piesman, J. Standard system for infecting ticks (Acari: Ixodidae) with the Lyme disease spirochete, *Borrelia burgdorferi*. *J. Med. Entomol.* **30**, 199–203 (1993).
27. Pal, U. *et al.* Inhibition of *Borrelia burgdorferi*-tick interactions *in vivo* by outer surface protein A antibody. *J. Immunol.* **166**, 7398–7403 (2001).
28. Narasimhan, S. *et al.* Disruption of *Ixodes scapularis* anticoagulation by using RNA interference. *Proc. Natl Acad. Sci. USA* **101**, 1141–1146 (2004).

LETTERS

T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand

Kaushik Choudhuri[1], David Wiseman[1], Marion H. Brown[1], Keith Gould[2]* & P. Anton van der Merwe[1]*

The binding of a T-cell antigen receptor (TCR) to peptide antigen presented by major histocompatibility antigens (pMHC) on antigen-presenting cells (APCs) is a central event in adaptive immune responses[1,2]. The mechanism by which TCR–pMHC ligation initiates signalling, a process termed TCR triggering, remains controversial[3–5]. It has been proposed[6–8] that TCR triggering is promoted by segregation at the T cell–APC interface of cell-surface molecules with small ectodomains (such as TCR–pMHC and accessory receptors) from molecules with large ectodomains (such as the receptor protein tyrosine phosphatases CD45 and CD148). Here we show that increasing the dimensions of the TCR–pMHC interaction by elongating the pMHC ectodomain greatly reduces TCR triggering without affecting TCR–pMHC ligation. A similar dependence on receptor–ligand complex dimensions was observed with artificial TCR–ligand systems that span the same dimensions as the TCR–pMHC complex. Interfaces between T cells and APCs expressing elongated pMHC showed an increased intermembrane separation distance and less depletion of CD45. These results show the importance of the small size of the TCR–pMHC complex and support a role for size-based segregation of cell-surface molecules in TCR triggering.

Antigen recognition by T cells involves simultaneous interactions between many T-cell surface receptors and their ligands on APCs or target cells[9]. TCR–pMHC and various accessory receptor–ligand interactions that contribute to TCR triggering span a relatively short intermembrane distance ($\sim$14 nm), whereas receptor protein tyrosine phosphatases known to inhibit TCR triggering, including CD45 and CD148, have much larger ectodomains[1,10,11]. This led to the hypothesis that size-dependent segregation of the TCR and its accessory molecules from inhibitory molecules with large ectodomains is required for TCR triggering[6–8]. A key prediction of this hypothesis is that TCR triggering is dependent on the dimensions of the TCR–pMHC interaction.

In order to test this prediction, we exploited a recently described method[12] to produce a murine pMHC class I complex as a single-chain trimer (SCT) fusion protein (Fig. 1a and Supplementary Methods). The SCT consists of the ovalbumin-derived peptide SIINFEKL (pOVA), β_2-microglobulin and the H-2K^b MHC class I heavy chain, joined by two flexible glycine/serine linker sequences (Fig. 1a). The insertion of spacers containing two, three or four immunoglobulin superfamily (IgSF) domains into the membrane-proximal stalk region of the SCT allowed us to produce elongated forms of H-2K^b molecules (SCT-CD2, SCT-CD22 and SCT-CD4; Fig. 1a). A key advantage of using SCTs is that intracellular assembly is less dependent on the *cis*-interactions in the endoplasmic reticulum (for example, with tapasin and TAP) required for assembly of native peptide–MHC I complexes (ref. 13). Consequently, assembly and peptide loading are unlikely to be affected by domain insertions.

A further advantage of this system is the availability of a monoclonal antibody (25-D1.16) specific for the pOVA/H-2K^b complex[14], which allows the detection and quantification of cell-surface pOVA/H-2K^b. As a control, we expressed a single-chain dimer (SCD) fusion protein, which contained just the β_2-microglobulin and the H-2K^b heavy chain joined by a glycine/serine linker (Fig. 1a).

The SCT construct is stably expressed on CHO cells, as determined by staining with the 25-D1.16 antibody (Supplementary Fig. 1). SCT-transfected CHO cells were used to stimulate the pOVA/H-2K^b-specific murine T-cell hybridoma B3Z, and activation was monitored by measurement of interleukin-2 (IL-2) secretion. CHO cells expressing SCTs were at least as efficient at stimulating B3Z cells as CHO cells expressing equivalent levels of SCD loaded with exogenous pOVA (pOVA/SCD), indicating that covalent linkage of pOVA to H-2K^b in the SCT does not abrogate TCR recognition (Supplementary Fig. 1). When constructs encoding the elongated SCTs were transfected into CHO cells, they produced proteins of the expected size (Supplementary Fig. 2a), all of which were recognized at the cell surface by the pOVA/H-2K^b-specific monoclonal antibody (Supplementary Fig. 2b).

To directly compare stimulation efficiencies of the SCTs, CHO cells expressing each SCT were sorted for similar surface ligand densities by flow cytometry (Supplementary Fig. 2b), and varying numbers of these CHO APCs were used to stimulate B3Z cells. CHO APCs expressing elongated SCTs were much less effective at activating B3Z cells, as measured by IL-2 secretion (Fig. 1b). Moreover, their ability to activate B3Z cells decreased with increasing ectodomain size, with SCT-CD4 being the least effective (Fig. 1b). When APCs expressing $\sim$10-fold lower surface density of SCTs or elongated SCTs were compared, the differences were even more pronounced, with only SCT-expressing CHO APCs stimulating detectable IL-2 secretion (Supplementary Fig. 3). A similar dependence on the size of pMHC was observed when the SCT-transfected APCs were used to stimulate CD8$^+$ T cells from mice transgenic for the OT1 TCR, which is identical to the B3Z TCR[15] and therefore also recognizes pOVA/H-2K^b. Only SCT- and SCT-CD2-expressing APCs stimulated IL-2 production (Fig. 1c) and cell proliferation (Supplementary Fig. 2c) in a secondary stimulation; SCT-CD22 and SCT-CD4 had no effect.

As IL-2 secretion and proliferation are very late consequences of TCR triggering, we next investigated whether proximal TCR signals were also dependent on pMHC dimensions. One of the earliest events in TCR signal transduction is increased phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAMs) of the TCR-associated homodimeric ζ-chains (TCR-ζ) (ref. 9). When B3Z T cells were incubated with APCs expressing SCT, there was a robust increase in TCR-ζ phosphorylation, which peaked within 5 min (Fig. 1d), consistent with previous studies of the kinetics
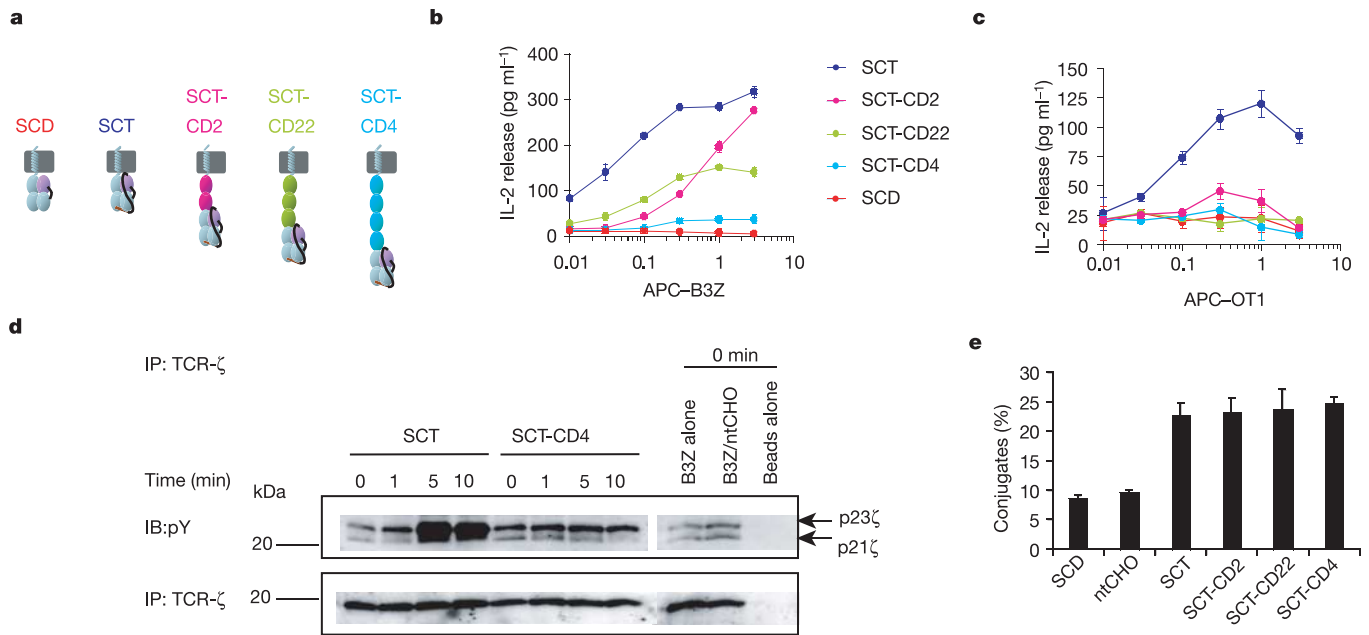
Figure 1 | Effect of elongating the SCT ectodomain on T-cell triggering.
a, Schematic representation of SCD, SCT and elongated SCTs. SCT was elongated with the two IgSF domains of human CD2 (SCT-CD2), the three amino-terminal IgSF domains of murine CD22 (SCT-CD22) or the four IgSF domains of human CD4 (SCT-CD4). Glycine/serine linkers are depicted as black lines. **b, c,** IL-2 production by B3Z T-cell hybridomas (**b**) and primed CD8$^+$ T cells from OT1 transgenic mice (**c**) following incubation with APCs expressing the indicated constructs. **d,** Analysis of TCR-ζ phosphorylation in B3Z cells following incubation with APCs expressing SCT or SCT-CD4. As controls, TCR-ζ phosphorylation was also analysed in B3Z cells incubated alone or with non-transfected (nt) CHO cells. IP, immunoprecipitate; IB, immunoblot. **e,** Conjugate formation between B3Z T cells and APCs expressing the indicated SCT constructs, expressed as a percentage of live-gated events. For all graphs, error bars represent s.d.

of TCR-ζ phosphorylation (for example, ref. 16). In contrast, incubation with APCs expressing SCT-CD4 resulted in hardly detectable increases in TCR-ζ phosphorylation (Fig. 1d). This demonstration that elongation of pMHC abrogates a very early event following TCR engagement indicates that it inhibits TCR triggering.

One possible explanation for the inability of elongated forms of pMHC to induce TCR triggering is that they bind less well to the TCR. To exclude this possibility, we assessed conjugate formation between B3Z cells and SCT-transfected APCs. We reasoned that, because of the relatively high densities of SCT on our APCs, the SCT–TCR interaction would mediate conjugate formation. We assessed conjugate formation by two-colour flow cytometry, for which B3Z cells and CHO APCs were labelled with red and green fluorescent vital dyes, respectively, and conjugates were identified as double-positive events (Supplementary Fig. 4a). Although conjugate formation between B3Z cells and SCD-expressing APCs was similar to background levels, there was an ~3-fold increase in conjugate formation with SCT-expressing APCs (Fig. 1e and Supplementary Fig. 4a). The same level of conjugate formation was observed between B3Z cells and APCs expressing elongated SCTs, indicating that TCRs bind elongated SCTs at least as effectively as SCT (Fig. 1e and Supplementary Fig. 4a). As a further assessment of TCR binding to SCT, we used immunofluorescence microscopy to measure the degree of SCT accumulation at the conjugate interface, as this would be driven by and thus reflect TCR engagement. Very few conjugates were observed between T cells and SCD-expressing APCs, and only 2% of these showed greater than twofold enrichment at the interface. In contrast, many more conjugates were observed with APCs expressing SCT and SCT-CD4, and ~30% of these showed greater than twofold enrichment at the interface (Table 1 and Supplementary Fig. 4b). Notably, there was no difference in the interface enrichment observed with SCT- or SCT-CD4-expressing APCs. Taken together, these data indicate that the abrogation of TCR triggering seen with elongated forms of pMHC is not the result of decreased TCR engagement.

To investigate the mechanism by which elongation of SCT abrogates TCR triggering, we compared the intermembrane distance in B3Z–APC conjugates using transmission electron microscopy. Intermembrane distances were measured in regions where both membranes exhibited a trilaminar appearance, as this indicated that the section was normal to the plane of both membranes (Fig. 2a, b). Although there was significant variation in the distances measured, the intermembrane distance was increased at interfaces involving APCs expressing SCT-CD4 (Fig. 2c). The mean intermembrane distance at the interface between B3Z cells and SCT-expressing APCs was 13.1 ± 3.1 nm (mean ± s.d., $n = 46$), which agrees well with the dimensions of pMHC–TCR complexes determined by structural studies[17,18]. In contrast, the intermembrane distance between B3Z cells and SCT-CD4-expressing APCs ($n = 57$) was significantly increased ($P < 0.0001$), at 17.8 ± 4.6 nm. The intermembrane distances showed greater variation with SCT-CD4 APCs, ranging from 10 to 28 nm (Fig. 2c). Thus, elongation of SCT with the 10–12-nm long[1] CD4 ectodomain increases the intermembrane distance by a variable amount, with the maximum increase consistent with the insert length. Submaximal increases in intermembrane distance are consistent with bending of the SCT chimaera and/or

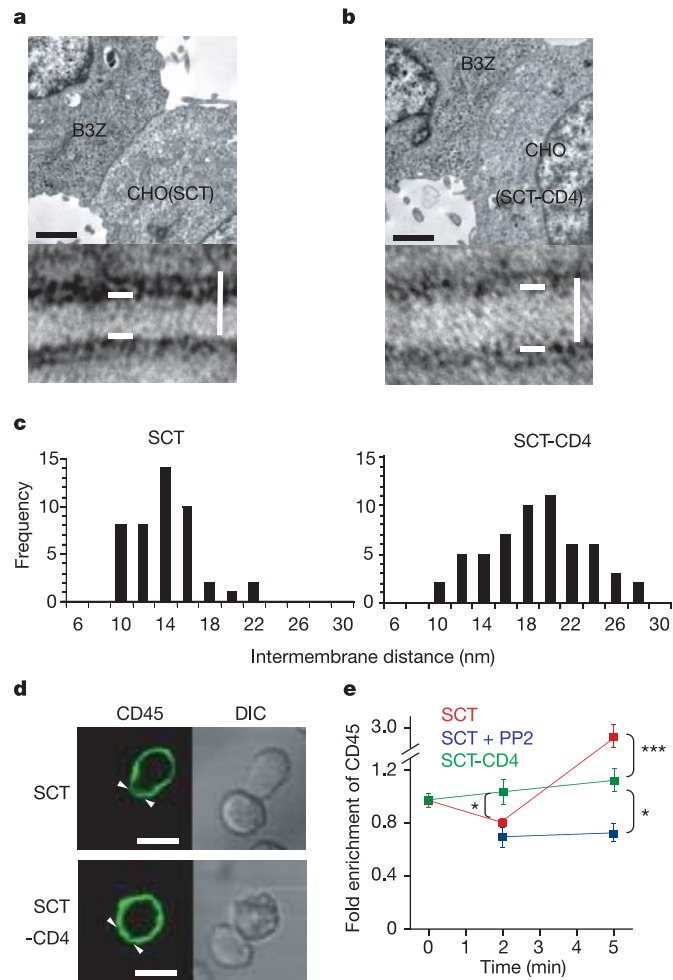**Table 1 | Enrichment of pMHC at the T cell–APC conjugate interface**

CHO APC	Fold enrichment (mean ± s.e.m.)	Conjugates with fold enrichment >2.0 (%)
SCT ($n = 96$)	1.80 ± 0.14	33*
SCT-CD4 ($n = 59$)	1.93 ± 0.13	29*
SCD ($n = 42$)	1.21 ± 0.07	2

For each conjugate (for example, see Supplementary Fig. 4b), a region of the APC containing the interface (with B3Z cell) was selected (Region 1, R1), and the remaining APC cell perimeter was defined as a non-interface region (R2). Fold enrichment was calculated using the formula (R1 MFI − background MFI)/(R2 MFI − background MFI), where MFI is the mean fluorescence intensity and the background is measured from a cell-free area of the same image.
* $P < 0.01$ compared with SCD (analysis of variance (ANOVA), with correction for multiple comparisons).

orientation of TCR–SCT complexes at angles not perpendicular to the membrane surfaces.
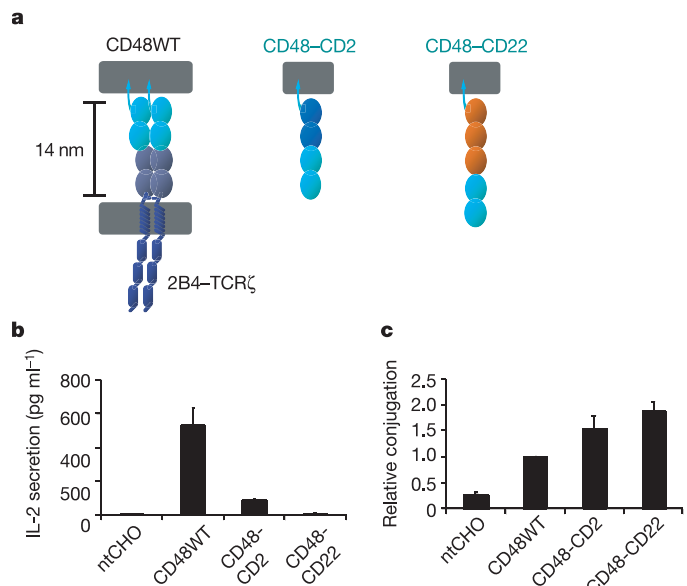
One possible mechanism by which elongation of SCT (and the resulting increase in intermembrane separation distance) could abrogate TCR triggering is by enhancing access of cell-surface molecules with large ectodomains (such as CD45) to the TCR–CD3 complex. We examined whether elongation of SCT affected the

distribution of CD45 at the T-cell–APC interface. When T cells were incubated with APCs expressing SCT, there was a biphasic effect with early depletion of CD45 at 2 min (Fig. 2d, e), followed by enrichment at 5 min (Fig. 2e). Interestingly, an orthogonal image of the interface at 5 min revealed that, although enriched, CD45 is actually segregated from SCT (Supplementary Fig. 5). When TCR triggering was inhibited using the Src kinase inhibitor PP2, this late enrichment was not observed (Fig. 2e), suggesting that it is a consequence of TCR triggering. When T cells formed conjugates with APCs expressing SCT-CD4, there was no significant depletion of CD45 at 2 min, indicating that CD45 is less effectively excluded from such interfaces. There was also no enrichment at 5 min, consistent with poor TCR triggering by SCT-CD4. In conclusion, elongation of SCT is associated with less effective depletion of CD45 from the T cell–APC interface.
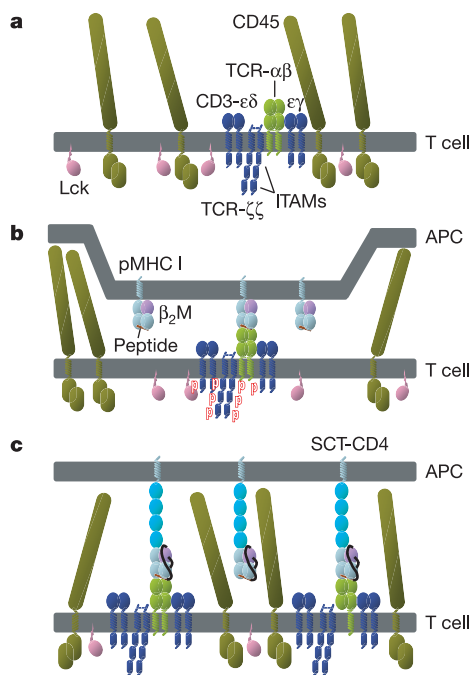
Our results suggest that elongation of pMHC abrogates TCR triggering by increasing the intermembrane separation distance, thereby allowing CD45 greater access to engaged TCR–CD3 complexes at the interface. However, it is also possible that elongation abrogates coreceptor (CD8) binding or a conformational change in the TCR–CD3 complex required for TCR triggering. To rule out these possibilities, we constructed artificial receptor–ligand systems that are similar to the TCR–pMHC in terms of dimensions and binding properties, but do not depend on coreceptor binding or mimic any putative conformational mechanism. To this end, TCR was replaced with a fusion protein comprising the ectodomain of CD2 or the related protein 2B4, fused to the transmembrane and cytoplasmic domains of TCR-ζ (Fig. 3a). The pMHC was replaced with CD48, which binds both CD2 and 2B4 (ref. 6). Structural and mutagenesis studies (reviewed in ref. 1) have shown that the CD2–CD48 complex

Figure 2 | Elongated SCT increases intermembrane distance and allows increased access of CD45 to the T-cell–APC interface. a, b, The top images are transmission electron micrographs of B3Z cells interacting with SCT- or SCT-CD4-expressing APCs (×4,000 original magnification; scale bar, 1 μm). The bottom images show segments of the corresponding interfaces at higher magnification (×36,500 original magnification; vertical scale bar, 20 nm). Intermembrane distance was measured between the outer lines (horizontal bars) in sections where both membranes had a trilaminar appearance. **c,** Intermembrane distances measured in conjugates between B3Z cells and SCT- (left) or SCT-CD4- (right) expressing APCs are grouped in 2-nm intervals. **d,** The left panels show CD45 (green) distribution on B3Z cells following incubation for 2 min at 37 °C with SCT- or SCT-CD4-expressing APCs. The B3Z–APC interfaces are indicated by arrowheads. Right panels show the corresponding differential interference contrast (DIC) images. **e,** Quantification of fold enrichment of CD45 at the B3Z–APC interface. B3Z cells were rapidly brought into contact with the indicated APCs and either fixed immediately ($t = 0$) or incubated at 37 °C for 2 or 5 min before fixation. Where indicated, B3Z cells were preincubated for 30 min with 10 μM PP2 (blue). Fold enrichment of CD45 was calculated as described in Table 1. Differences in mean fold enrichment between groups at each time point were analysed by ANOVA with correction for multiple comparisons. Each data point represents the mean ± s.e.m. for at least 30 conjugates. Selected comparisons are indicated by brackets. Asterisk, $P < 0.05$; three asterisks, $P < 0.001$.

Figure 3 | Triggering in an artificial TCR–pMHC system is also dependent on ectodomain size. a, The artificial TCR–pMHC system. 2B4–TCRζ or CD2–TCRζ chimaeras were expressed on T cells and wild-type (WT) or elongated forms of CD48 were expressed on CHO cells. CD48 was elongated with either the two IgSF domains from the human CD2 ectodomain or the three N-terminal IgSF domains from the murine CD22 ectodomain. **b,** Activation of IL-2 production by 2B4–TCRζ-expressing T cells following incubation with CHO cells expressing the indicated forms of CD48. **c,** Conjugate formation between 2B4–TCRζ-expressing T cells and CHO cells expressing the indicated forms of CD48. T cells and CHO cells loaded with different dyes were mixed together and conjugate formation was measured by flow cytometry. Conjugation levels are expressed relative to the level observed between T cells and CD48WT-expressing cells. Means ± s.d. of three experiments are shown.

and its human counterpart, the CD2–CD58 complex, span the same dimensions (~14 nm) as the TCR–pMHC complex (Fig. 3a). Given the predicted similarity between the 2B4 and CD2 ectodomains[6], the 2B4–CD48 complex is likely to span the same dimensions.

Elongated forms of CD48 were constructed and expressed as previously described[19] by inserting two or three IgSF domains from the ectodomains of human CD2 or murine CD22 into the membrane-proximal stalk region of the CD48 ectodomain (Fig. 3a and Supplementary Fig. 6a). When T cells transfected with the 2B4–TCRζ or CD2–TCRζ chimaeras were exposed to CHO cells expressing wild-type CD48, they were activated to produce IL-2, indicating that the interactions between 2B4–TCRζ or CD2–TCRζ and CD48 were able to mimic TCR triggering (Fig. 3b and Supplementary Fig. 6b). In contrast, there was a fivefold reduction in response when the same T cells were exposed to CHO cells expressing the elongated CD48–CD2 chimaera, and hardly any response to CHO cells expressing the longest CD48–CD22 chimaera. To confirm that elongation of CD48 did not reduce its ability to bind 2B4 or CD2, we assessed conjugate formation between 2B4–TCRζ- or CD2–TCRζ-expressing T cells and CD48-expressing CHO cells by flow cytometry (Fig. 3c and Supplementary Fig. 6c). Conjugate formation was significantly enhanced by elongation of CD48, indicating that elongation of CD48 enhanced, rather than diminished, CD2 and 2B4 engagement. This is

consistent with an earlier study on the effect of elongation of CD58 on CD2–CD58 interactions[20], and is presumably the result of improved access of 2B4 or CD2 to elongated forms of CD48 on the CHO cell surface. The marked inhibitory effect of CD48 elongation on triggering, despite enhanced CD2 and 2B4 ligation, demonstrates the critical importance of complex dimension, and argues against elongation affecting other mechanisms such as coreceptor binding and TCR–CD3 conformational change.

Taken together, the findings presented here support the kinetic-segregation (K-S) model of TCR triggering (Fig. 4). This model proposes that TCR triggering is the result of tethering of the TCR–CD3 complex within close-contact zones in which tyrosine phosphorylation is favoured because of size-dependent exclusion of tyrosine phosphatases such as CD45 (Fig. 4). Several other observations also support this model. First, incubation of T cells with inhibitors of tyrosine phosphatases results in increased tyrosine phosphorylation of TCR–CD3 and T-cell activation, indicating that there is a fine balance between constitutively active protein tyrosine kinases and phosphatases at the resting T cell surface[7]. Second, deleting the large ectodomains of the receptor protein tyrosine phosphatases CD45 and CD148 abrogates TCR triggering[21,22]. In the case of CD45, it was shown that a heterologous large ectodomain from CD43 was able to restore TCR triggering[21]. The CD148 ectodomain was only critical when TCR triggering was induced by superantigen or immobilized antibody, but not when soluble antibody was used. This is consistent with the ectodomain mediating size-dependent segregation from areas of close contact between T cells and APCs or solid substrates[22]. Finally, experimental[23,24] and theoretical (reviewed in ref. 25) studies support the notion that differences in ectodomain size are sufficient to drive spontaneous segregation of molecules at cell–cell interfaces. Although the K-S model provides a plausible mechanism for TCR triggering, neither the model nor the data shown here preclude a role for other mechanisms in TCR triggering, such as binding-induced conformational change[26] or multimerization of the TCR–CD3 complex[27].

Our findings that the small dimensions of the TCR–pMHC complex are important for TCR triggering raise the question as to whether size dimensions have a role in signal transduction induced by other receptor–ligand interactions. As argued elsewhere[28], it seems likely that (small) complex dimension would be important for any small receptor–ligand systems that, like the TCR, depend on tyrosine phosphorylation of their cytoplasmic portions by tyrosine kinases. These include molecules in the CD28/CTLA-4 family[29] and receptors involved in natural killer cell recognition of target cells[30].

METHODS

Please refer to the Supplementary Methods for full details.

T-cell activation assays. T cells/hybridomas were incubated with irradiated SCT- or elongated SCT-expressing CHO cells at ratios of 0.01–3.3 in optimal culture conditions. APCs were sorted for comparable surface expression of SCTs and elongated SCTs using the 25-D1.16 monoclonal antibody. SCD-expressing CHO cells were used as a negative control. IL-2 release was assayed as previously described[19].

Immunoprecipitation and western blotting. Equal numbers of B3Z cells and CHO APCs expressing SCT or SCT-CD4 were rapidly brought into contact by pulse centrifugation. After incubation at 37 °C for 1, 5 or 10 min, cells were lysed and TCR-ζ immunoprecipitated with an anti-CD3 monoclonal antibody. Samples separated by SDS–PAGE were transferred to nitrocellulose membranes and probed using anti-TCRζ or anti-phosphotyrosine (anti-pY) antibodies.

Conjugate formation assay. T-cell hybridomas and CHO APCs were loaded with Cell-Tracker Orange CMTMR and Vybrant CFDA SE dyes, respectively. Loaded cells were mixed with equal numbers of T-hybridomas and brought into contact by gentle centrifugation at 4 °C. Cells were incubated for 10 min at 37 °C, resuspended and analysed by flow cytometry. The percentage of FL1hi/FL2hi events was taken as an estimate of conjugate formation.

Immunofluorescence microscopy. Samples were prepared as for the conjugate formation assay, but with only B3Z cells loaded with dye. Cells were placed onto microscope slides (on ice) and lightly fixed. Cell surface SCT was detected using a FITC-conjugated anti-H2Kb monoclonal antibody. For analysis of CD45

Figure 4 | Kinetic-segregation (K-S) model of TCR triggering[6,7]. Cell-surface molecules involved in T-cell antigen recognition are drawn approximately to scale on the basis of structural studies (reviewed in refs 6, 7 and 11). The K-S model proposes that in a resting T cell (**a**), there is a balance between constitutively active tyrosine kinases (such as Lck) and phosphatases (such as CD45), with the latter maintaining TCR–CD3 ITAMs and other tyrosine kinase substrates in a dephosphorylated state. The model proposes that when T cells contact APCs (**b**), close-contact areas form within the interface, from which molecules with large ectodomains (such as CD45) are excluded. This increases phosphorylation of tyrosine kinase substrates such as TCR–CD3 ITAMs within the close-contact region. The increase would be transient unless TCR–CD3 is prevented from diffusing out of the close-contact region by binding to pMHC. Triggering occurs when tyrosine kinase substrates remain phosphorylated for long enough to initiate downstream signalling events. **c**, When the pMHC is elongated (SCT-CD4), TCR engagement no longer occurs in regions of close-contact. As a result, phosphatases such as CD45 are not adequately segregated from TCR–CD3 and phosphorylation of tyrosine kinase substrates such as TCR–CD3 ITAMs is abrogated.

accumulation, conjugates were incubated at 37 °C for 2 or 5 min before fixation. For the zero time point, samples were fixed immediately after conjugate formation. CD45 was detected using a pan-anti-CD45 monoclonal antibody and a FITC-conjugated secondary.

Transmission electron microscopy. Samples were prepared as for the conjugate formation assay but without tracker dye. Following light fixation, samples were treated with 1% osmium tetroxide plus 0.5% uranyl acetate. Embedded ultrathin sections were double stained with uranyl acetate and lead citrate, and examined under a Zeiss Omega 912 electron microscope. Intermembrane distance measurements were recorded from images captured at ×36,500 original magnification and 2048 × 2048 pixel resolution.

1. van der Merwe, P. A. & Davis, S. J. Molecular interactions mediating T cell antigen recognition. *Annu. Rev. Immunol.* **21**, 659–684 (2003).
2. Davis, M. M. *et al.* Dynamics of cell surface molecules during T cell recognition. *Annu. Rev. Biochem.* **72**, 717–742 (2003).
3. van der Merwe, P. The TCR triggering puzzle. *Immunity* **14**, 665–668 (2001).
4. Davis, M. M. A new trigger for T cells. *Cell* **110**, 285–287 (2002).
5. Trautmann, A. & Randriamampita, C. Initiation of TCR signalling revisited. *Trends Immunol.* **24**, 425–428 (2003).
6. Davis, S. J. & van der Merwe, P. A. The structure and ligand interactions of CD2: implications for T-cell function. *Immunol. Today* **17**, 177–187 (1996).
7. van der Merwe, P. A., Davis, S. J., Shaw, A. S. & Dustin, M. L. Cytoskeletal polarization and redistribution of cell surface molecules during T cell antigen recognition. *Semin. Immunol.* **12**, 5–21 (2000).
8. Shaw, A. S. & Dustin, M. L. Making the T cell receptor go the distance: a topological view of T cell activation. *Immunity* **6**, 361–369 (1997).
9. Weiss, A. & Samelson, L. E. in *Fundamental Immunology* (ed. Paul, W. E.) 321–364 (Lippincott Williams & Wilkins, Philadelphia, 2003).
10. Springer, T. A. Adhesion receptors of the immune system. *Nature* **346**, 425–434 (1990).
11. Barclay, A. N. *et al.* *The Leucocyte Antigen Factsbook* (Academic, London, 1997).
12. Yu, Y. Y., Netuschil, N., Lybarger, L., Connolly, J. M. & Hansen, T. H. Cutting edge: single-chain trimers of MHC class I molecules form stable structures that potently stimulate antigen-specific T cells and B cells. *J. Immunol.* **168**, 3145–3149 (2002).
13. Cresswell, P., Bangia, N., Dick, T. & Diedrich, G. The nature of the MHC class I peptide loading complex. *Immunol. Rev.* **172**, 21–28 (1999).
14. Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R. & Germain, R. N. Localization, quantitation, and *in situ* detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* **6**, 715–726 (1997).
15. Dillon, S. R., Jameson, S. C. & Fink, P. J. V beta 5+ T cell receptors skew toward OVA+H-2Kb recognition. *J. Immunol.* **152**, 1790–1801 (1994).
16. Kersh, E. N., Shaw, A. S. & Allen, P. M. Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation. *Science* **281**, 572–575 (1998).
17. Garboczi, D. N. *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134–141 (1996).
18. Garcia, K. C. *et al.* Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* **279**, 1166–1172 (1998).
19. Wild, M. K. *et al.* Dependence of T cell antigen recognition on the dimensions of an accessory receptor-ligand complex. *J. Exp. Med.* **190**, 31–41 (1999).
20. Chan, P.-Y. & Springer, T. S. Effect of lengthening lymphocyte function-associated antigen 3 on adhesion to CD2. *Mol. Biol. Cell* **3**, 157–166 (1992).
21. Irles, C. *et al.* CD45 ectodomain controls interaction with GEMs and Lck activity for optimal TCR signalling. *Nature Immunol.* **4**, 189–197 (2003).
22. Lin, J. & Weiss, A. The tyrosine phosphatase CD148 is excluded from the immunologic synapse and down-regulates prolonged T cell signalling. *J. Cell Biol.* **162**, 673–682 (2003).
23. Dustin, M. L. *et al.* A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. *Cell* **94**, 667–677 (1998).
24. Davis, D. M. *et al.* The human natural killer cell immune synapse. *Proc. Natl Acad. Sci. USA* **96**, 15062–15067 (1999).
25. Coombs, D., Dembo, M., Wofsy, C. & Goldstein, B. Equilibrium thermodynamics of cell-cell adhesion mediated by multiple ligand-receptor pairs. *Biophys. J.* **86**, 1408–1423 (2004).
26. Gil, D., Schamel, W. W., Montoya, M., Sanchez-Madrid, F. & Alarcon, B. Recruitment of Nck by CD3 epsilon reveals a ligand-induced conformational change essential for T cell receptor signalling and synapse formation. *Cell* **109**, 901–912 (2002).
27. Krogsgaard, M. *et al.* Agonist/endogenous peptide–MHC heterodimers drive T cell activation and sensitivity. *Nature* **434**, 238–243 (2005).
28. Davis, S. J. *et al.* The nature of molecular recognition by T cells. *Nature Immunol.* **4**, 217–224 (2003).
29. Greenwald, R. J., Freeman, G. J. & Sharpe, A. H. The B7 family revisited. *Annu. Rev. Immunol.* **23**, 515–548 (2005).
30. Lanier, L. L. NK cell recognition. *Annu. Rev. Immunol.* **23**, 225–274 (2005).

LETTERS

Functional coordination of intraflagellar transport motors

Guangshuo Ou[1], Oliver E. Blacque[2], Joshua J. Snow[1], Michel R. Leroux[2] & Jonathan M. Scholey[1]

Cilia have diverse roles in motility and sensory reception, and defects in cilia function contribute to ciliary diseases such as Bardet–Biedl syndrome (BBS). Intraflagellar transport (IFT) motors assemble and maintain cilia by transporting ciliary precursors, bound to protein complexes called IFT particles, from the base of the cilium to their site of incorporation at the distal tip[1–3]. In *Caenorhabditis elegans*, this is accomplished by two IFT motors, kinesin-II and osmotic avoidance defective (OSM)-3 kinesin, which cooperate to form two sequential anterograde IFT pathways that build distinct parts of cilia[4–7]. By observing the movement of fluorescent IFT motors and IFT particles along the cilia of numerous ciliary mutants, we identified three genes whose protein products mediate the functional coordination of these motors. The BBS proteins BBS-7 and BBS-8 are required to stabilize complexes of IFT particles containing both of the IFT motors, because IFT particles in *bbs-7* and *bbs-8* mutants break down into two subcomplexes, IFT-A and IFT-B, which are moved separately by kinesin-II and OSM-3 kinesin, respectively. A conserved ciliary protein, DYF-1, is specifically required for OSM-3 kinesin to dock onto and move IFT particles, because OSM-3 kinesin is inactive and intact IFT particles are moved by kinesin-II alone in *dyf-1* mutants. These findings implicate BBS ciliary disease proteins and an OSM-3 kinesin activator in the formation of two IFT pathways that build functional cilia.

Intracellular transport systems have critical roles in subcellular organization but little is known about how they are coordinated and controlled[8–11]. For example, two sequential pathways of IFT are used to build sensory cilia on *C. elegans* neurons: a 'middle segment' pathway, which builds the 4-µm-long microtubule doublets extending from the basal body, and a 'distal segment' pathway that forms the 2.5-µm-long distal microtubule singlets[7]. This process depends on the movement of IFT particles by two members of the kinesin-2 family[12]: heterotrimeric kinesin-II and homodimeric OSM-3 kinesin[4,6]. Surprisingly, each particle is moved along the middle segment by both motors working together, and then along the distal segment by OSM-3 kinesin alone[7]. To identify proteins that control the functional coordination of these motors, we reasoned that loss-of-function mutations in corresponding genes might be present among existing sensory ciliary mutants[7,13–16] and might phenocopy mutations in the motors themselves. For example, impaired IFT in *osm-3* mutants causes a marked loss of distal axoneme segments and defects in osmotic avoidance, chemotaxis and dye-filling capacity (*osm*, *che* and *dyf* phenotypes)[7,13–16]. Here, we screened both characterized and uncharacterized ciliary mutants for similar phenotypes to *osm-3*.

Among previously characterized ciliary mutants, our attention was drawn to *osm-12(n1606)* and the phenotypically similar *bbs-8(nx77)*, which encode homologues of BBS-7 and BBS-8, respectively[17]. Mutations in BBS genes are associated with the heritable human

ciliary disease Bardet–Biedl syndrome and, in *C. elegans*, lead to a partial loss of distal ciliary segments with compromised IFT; however the roles of these proteins in IFT or their relationship to kinesin-2 motors have not been explored. We address these issues below.

Among previously uncharacterized ciliary mutants, we applied a new screen to the set of dye-filling mutants (*dyf-1* to *dyf-13*) by introducing into each the green fluorescent protein (GFP) IFT particle subunit (OSM-6::GFP) and testing the progeny for loss of distal ciliary segments by fluorescence microscopy[7] (Fig. 1a). Of these 13 mutants, only *dyf-1(mn335)* and *dyf-13(mn396)* are likely to have defects in components of the OSM-3 kinesin IFT pathway, as they displayed IFT particle fluorescence extending only 4–5 µm from the base of the cilium in both the ciliated amphid and phasmid sensory neurons (Fig. 1b), as with the *osm-3* mutant. Deficiency mapping leaves open the question of whether *dyf-1(mn335)* is a severe loss-of-function allele[15], but it does perturb the function of all cilia in a uniform manner, whereas the ciliary defects of *dyf-13(mn396)* suggest low penetrance (Fig. 1b), which compromised our analysis of the role of DYF-13 in IFT. We therefore confined our attention to DYF-1.

The uncloned *dyf-1* gene was previously mapped to chromosome I at -0.53 ± 0.177 cM (ref. 15). Cosmid B0321, which also maps to this locus, rescued the *dyf-1* phenotype[18] (Fig. 1c), but it contains >20 open reading frames, one of which, F54C1.5, contains a canonical 'X-box' motif within its promoter at position -105. Because this motif is diagnostic of cilia-specific gene expression in response to DAF-19 RFX transcription factor binding[19,20], its presence suggested that F54C1.5 may encode *dyf-1* (Fig. 1c, d). Sequencing of the *dyf-1* mutant revealed a TACA insertion within the thirteenth exon that is expected to truncate the corresponding protein at amino acid 589, leading to loss of a domain required for distal segment assembly. Furthermore, *dyf-1(mn335)* animals expressing OSM-6::GFP and the F54C1.5 coding region were completely rescued with respect to dye-filling defects in four out of four transgenic lines. Moreover, OSM-6::GFP IFT particles moved along both the middle and restored distal ciliary segments (Fig. 1f; data not shown) at characteristic wild-type rates[7] (Table 1). The identification of the molecular lesion in the F54C1.5 locus combined with mutant rescue by the corresponding wild-type fragment demonstrates that F54C1.5 corresponds to *dyf-1*.

Analysis of the *dyf-1* gene suggests that it encodes two transcripts due to alternative splicing (Fig. 1d), and produces two proteins containing tetratricopeptide repeat- and prenyltransferase domains, which we speculate might allow DYF-1 to bind and prenylate target proteins for association with ciliary membranes (Fig. 1e; F54C1.5a, 656 amino acids, 75.5 kDa; F54C1.5b, 597 amino acids, 68.5 kDa). Notably, phylogenetic analysis identified 14 homologues of DYF-1 (Supplementary Fig. S1) and revealed that, similar to BBS-7, BBS-8 and the kinesin-2 motors[8,17], it is highly conserved among ciliated

[1]Center for Genetics and Development, Section of Molecular and Cellular Biology, University of California, Davis, California 95616, USA. [2]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.
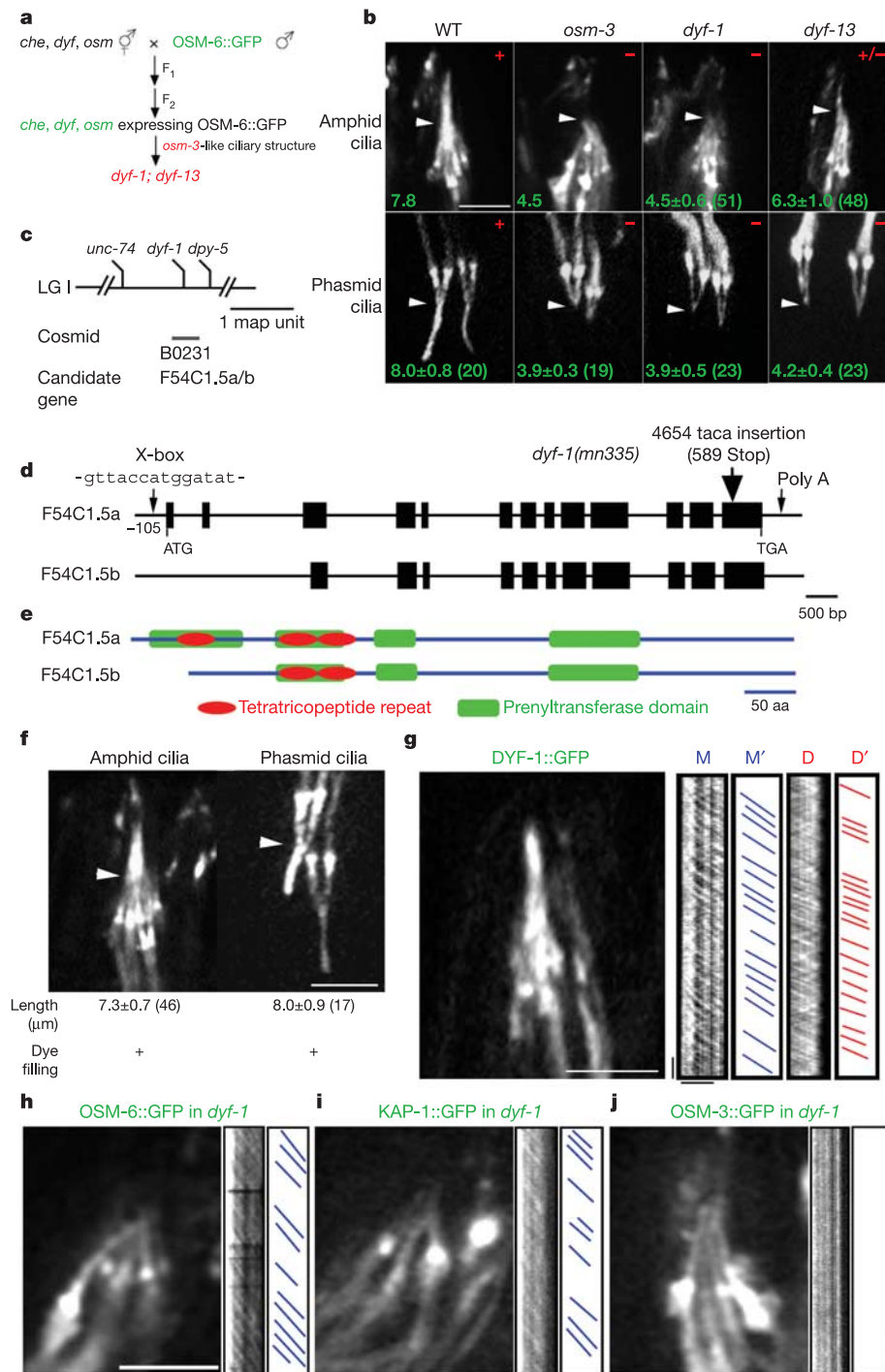
Figure 1 | Genetic screen of ciliary mutants and characterization of *dyf-1*.
a, Ciliary mutant hermaphrodites were crossed to wild-type OSM-6::GFP (IFT subcomplex B subunit) males to produce ciliary mutants expressing GFP-labelled IFT particles. **b**, Screen of progeny to identify sensory ciliary length defects. Fluorescence micrographs showing distribution of OSM-6::GFP along amphid (upper panels) and phasmid (lower panels) cilia in wild type (approximately 7.5 μm full length; positive in dye-filling assays (+)), *osm-3* (missing distal segment; no dye filling (−)), *dyf-1* (also missing distal segment; no dye filling (−)) and *dyf-13* mutants (distal segment partially missing in amphid; partial dye filling (+/−); completely missing in phasmid; no dye filling (−)). The middle–distal segment junction is indicated by an arrowhead. Ciliary lengths are indicated in green (μm) and dye-filling assay results are indicated in red. **c**, Genetic map of *dyf-1* locus on

chromosome I. **d**, The *dyf-1* promoter contains an X-box motif indicative of ciliary function and encodes two alternative transcripts. **e**, The deduced DYF-1 protein contains prenyltransferase domains and tetratricopeptide repeats (TPRs). **f**, Rescued ciliary length and dye filling in *dyf-1* mutants by the wild-type *dyf-1* gene. **g**, Motility of DYF-1::GFP within rescued cilia. Fluorescence micrograph of cilia (left panel) and kymographs (right panels) with corresponding cartoons showing selected DYF-1 particle trajectories along middle (M, M′ blue lines) and distal (D, D′ red lines) segments. **h–j**, Motility of IFT particles and motors along *dyf-1* mutant middle segments. OSM-6::GFP (**h**) and kinesin-II (KAP-1::GFP) (**i**) move together at 0.5 μm s⁻¹. OSM-3 kinesin (OSM-3::GFP) (**j**) localizes to middle segments of *dyf-1* cilia but displays no motility. Scale bars, 5 μm. Kymograph horizontal bar, 2.5 μm, vertical bar, 5 s.

Table 1 | **IFT velocity of IFT particles, motors and DYF-1 in wild type and various mutants**

Anterograde motility	Strain	Average velocities (μm s^{-1})			
		Middle segment	n	Distal segment	n
IFT particles					
OSM-6::GFP	dyf-1(mn335)	0.49 ± 0.06	264	None	
OSM-6::GFP	dyf-1 rescue	0.70 ± 0.08	95	1.22 ± 0.15	101
IFT motors					
OSM-3::GFP	dyf-1(mn335)	None		None	
KAP-1::GFP	dyf-1(mn335)	0.50 ± 0.08	221	None	
DYF-1::GFP	dyf-1(mn335)	0.70 ± 0.10	248	1.24 ± 0.15	240
	osm-3(p802)	0.50 ± 0.06	248	None	
	klp-11(tm324)	1.24 ± 0.17	247	1.27 ± 0.16	252

n, Number of GFP particles. Movies of IFT are available in Supplementary Information.

organisms but is not found in organisms lacking cilia. The functions of these DYF-1 proteins are unknown.

To test the hypothesis that DYF-1 and BBS proteins participate in the coordinated kinesin-II and OSM-3 kinesin-dependent IFT pathways, we observed corresponding GFP fusion proteins moving along sensory cilia by time-lapse spinning-disc microscopy[7,21,22]. Existing functional GFP-tagged BBS-7 and BBS-8 fusion proteins[17] were used and new DYF-1::GFP strains were constructed in which the fusion protein was functional based on its ability to fully rescue dye-filling and ciliary length in dyf-1(mn335) mutants (Fig. 1g). We found that fluorescent DYF-1, BBS-7 and BBS-8 puncta all moved anterogradely along the full length of wild-type sensory cilia: first along the middle segments at 0.7 μm s^{-1}, intermediate between kinesin-II and OSM-3 kinesin motor activity, then accelerating to 1.3 μm s^{-1} along the distal segment, characteristic of the OSM-3 kinesin motor alone[7] (Fig. 1g and Tables 1 and 2; see also Supplementary Movies S1–S3 and Fig. S2a, d). These two transport rates are identical to those of other IFT particle components[7], indicating that BBS-7, BBS-8 and DYF-1 dock onto IFT particles that are moved by the coordinate action of kinesin-II and OSM-3 kinesin.

To dissect the functions of BBS proteins and DYF-1 in IFT, time-lapse confocal transport assays were performed in various mutant backgrounds. We investigated the functions of BBS-7 and BBS-8 by first examining the motility of IFT motors in bbs mutants. In wild-type animals, kinesin-II and OSM-3 kinesin move the same IFT particles along the middle segment at 0.7 μm s^{-1} then OSM-3 kinesin alone moves IFT particles along the distal segment at a

faster rate of 1.3 μm s^{-1} (ref. 7; Table 2). In bbs mutants, the motility of GFP-tagged KAP-1, a subunit of heterotrimeric kinesin-II, is restricted to the middle segment as in wild type, but its velocity decreases to 0.5 μm s^{-1}, characteristic of kinesin II alone (Table 2; see also Supplementary Fig. S2b, e, Movie S4 and data not shown). OSM-3 kinesin, however, moves along the middle segments and a few residual distal segments at a constant rate of 1.3 μm s^{-1} and accumulates at the tips of both the middle and distal segments (Table 2; see also Supplementary Fig. S2c, f, Movie S5 and data not shown). The disparate transport velocities of KAP-1::GFP and OSM-3::GFP along the middle segments of bbs mutant sensory cilia suggests that the two kinesin-2 motors move separately rather than together and that BBS proteins normally mediate the physical association of kinesin-II and OSM-3 kinesin within the middle segments.

How do BBS proteins mediate the association of the two kinesin-2 motors? To answer this question, we re-examined the transport of IFT-A and IFT-B subcomplex components in bbs mutant backgrounds. In wild-type animals, CHE-11::GFP (subcomplex A) and CHE-2::GFP or OSM-5::GFP (subcomplex B) move at the characteristic transport velocities for both the middle and distal ciliary segments (Table 2; see also Supplementary Fig. S2 g, j, m, Movies S6 and S8 and data not shown). However, in bbs mutants CHE-11::GFP only moves along the middle segment at decreased velocities of 0.5–0.6 μm s^{-1}, and fails to enter the distal segment, a behaviour reminiscent of KAP-1::GFP (Table 2; see also Supplementary Fig. S2 h, i, Movie S7 and data not shown). In contrast, CHE-2::GFP or OSM-5::GFP move at the fast rate of 1.1–1.3 μm s^{-1} along middle segments and a few residual distal segments in bbs mutants, and accumulate at the middle and distal segment tips, as with OSM-3 kinesin (Table 2; see also Supplementary Fig. S2k, l, n, Movie S9 and data not shown).

These observations suggest that BBS proteins stabilize moving motor–cargo complexes. In bbs mutants, IFT particles dissociate into subcomplexes A and B, which are then moved by kinesin-II and OSM-3 kinesin alone, respectively. OSM-3 kinesin and subcomplex B components accumulate at the middle and distal segment tips, perhaps because subcomplex A components required for the turn-around and recycling of the anterograde IFT machinery cannot efficiently reach these tips[17].

DYF-1 has a different role in coordinating the motility of kinesin-II and OSM-3 kinesin. In a dyf-1 mutant background, IFT particles (OSM-6::GFP) move along the residual middle ciliary segment at a

Table 2 | **Velocity of IFT particles, BBS proteins and motors in wild type or bbs mutant animals**

Anterograde motility	Strain	Average velocities (μm s^{-1})			
		Middle segment	n	Distal segment	n
IFT particles					
CHE-11::GFP	WT	0.72 ± 0.11	263	1.33 ± 0.18	267
CHE-11::GFP	bbs-7(n1606)	0.58 ± 0.07	241	None	
CHE-11::GFP	bbs-8(nx77)	0.57 ± 0.09	240	None	
CHE-2::GFP	WT	0.71 ± 0.09	257	1.27 ± 0.19	244
CHE-2::GFP	bbs-7(n1606)	1.11 ± 0.18	246	1.25 ± 0.17	83 (4/11)*
CHE-2::GFP	bbs-8(nx77)	1.07 ± 0.16	243	1.31 ± 0.13	91 (4/10)*
OSM-5::GFP	WT	0.67 ± 0.08	249	1.27 ± 0.18	247
OSM-5::GFP	bbs-7(n1606)	1.14 ± 0.18	207	1.31 ± 0.20	75 (3/13)*
IFT motors					
OSM-3::GFP	WT (ref. 7)	0.70		1.30	
OSM-3::GFP	bbs-7(n1606)	1.17 ± 0.18	226	1.20 ± 0.17	78 (3/11)*
OSM-3::GFP	bbs-8(nx77)	1.17 ± 0.15	221	1.25 ± 0.17	82 (4/13)*
KAP-1::GFP	WT (ref. 7)	0.70		None	
KAP-1::GFP	bbs-7(n1606)	0.54 ± 0.07	205	None	
KAP-1::GFP	bbs-8(nx77)	0.53 ± 0.07	211	None	
BBS proteins					
BBS-7::GFP	bbs-7(n1606)	0.67 ± 0.12	247	1.29 ± 0.15	252
BBS-8::GFP	bbs-8(nx77)	0.70 ± 0.08	251	1.26 ± 0.18	267

WT, wild type; n, Number of GFP particles. Movies of IFT are available in Supplementary Information.
*a out of b animals observed have detectable IFT in the remaining partial distal segments of bbs-7 or bbs-8 mutants.

slow rate of ∼0.5 μm s^{-1}, the same as in an *osm-3* mutant background (Fig. 1h and Table 1; see also Supplementary Movie S10), suggesting that the fluorescent protein associates with IFT particles that are moved along the middle segment by kinesin-II alone rather than by kinesin-II and OSM-3 kinesin together[7]. Fluorescent kinesin-II (KAP-1::GFP) also moved at its characteristically slow rate of ∼0.5 μm s^{-1} along the middle segment of *dyf-1* mutants (Fig. 1i and Table 1; see also Supplementary Movie S11), whereas OSM-3 kinesin (OSM-3::GFP) localized to the residual middle segment but no transport events could be detected (Fig. 1j and Table 1; see also Supplementary Movie S12). This suggests that, after loss of DYF-1 function, OSM-3 kinesin can still enter the cilium and diffuse along it, but is unable to move because DYF-1 normally activates OSM-3 kinesin-driven motility. The function of DYF-1 is specific because other GFP-tagged IFT particle proteins, including OSM-1, OSM-5, CHE-2, CHE-11 and CHE-13, displayed the same behaviour as OSM-6::GFP in *dyf-1* mutants (data not shown), suggesting that abrogation of DYF-1 function does not interfere with the assembly or function of IFT components in general but selectively affects OSM-3 kinesin function.

How might DYF-1 influence the motility of OSM-3 kinesin? First, the pattern of motility of DYF-1::GFP within the cilia of wild type, OSM-3 kinesin (*osm-3(p802)*) and kinesin-II (*klp-11(tm324)*) mutants (Table 1; see also Supplementary Movie S1, S13 and S14) is consistent with it being bound to IFT particles and moved along the middle segment by kinesin-II and OSM-3 kinesin together, and then along the distal segment by OSM-3 kinesin alone[7]. Second, the observation that OSM-3 kinesin displays no movement in *dyf-1* mutants suggests that it is dissociated from its IFT cargo, otherwise it would be moved passively along the middle segment at the slow

rate characteristic of kinesin-II. We therefore propose that DYF-1 may serve to dock OSM-3 kinesin onto IFT particles. Finally, the abrogation of OSM-3 kinesin movement in *dyf-1* mutants leads us to hypothesize that DYF-1 also serves as a positive regulator of OSM-3 kinesin motility. In contrast to studies of the KAP subunit of heterotrimeric kinesin-2 and its cofactor p150/dynamitin[23,24], we have so far been unable to detect any direct protein–protein interactions between DYF-1 and OSM-3 kinesin, possibly because DYF-1 activates the cargo-binding and motility properties of OSM-3 kinesin through an unidentified intermediary (which could itself be targeted to ciliary membranes by DYF-1-catalysed prenylation). Further work on the molecular mechanism of DYF-1-mediated activation of OSM-3 kinesin is required.

Thus, time-lapse microscopy of IFT in living ciliary mutants has uncovered a complex set of interactions that generates two sequential IFT pathways within *C. elegans* sensory cilia (Fig. 2). We propose a model for the functional coordination of two IFT motors in which BBS-7 and BBS-8 serve to link the activities of kinesin-II and OSM-3 kinesin by stabilizing the association between IFT particle subcomplexes A and B, whereas DYF-1 functions as a cofactor for OSM-3 kinesin but not for kinesin-II. Although IFT-A and IFT-B can be separated biochemically[5], our results provide a demonstration that they can move separately *in vivo*, and suggest a specific role for the BBS proteins in stabilizing their interactions—a process that may be modulated physiologically under specific circumstances or in association with ciliary disease (Fig. 2a, c). In this model, in the absence of OSM-3 kinesin or its activator, DYF-1, movement of intact IFT particles along the sensory cilia of *C. elegans* is driven by kinesin-II alone, as observed for the cilia of many organisms[1,2] (Fig. 2b). The DYF-1-mediated activation of OSM-3 kinesin augments this canoni-
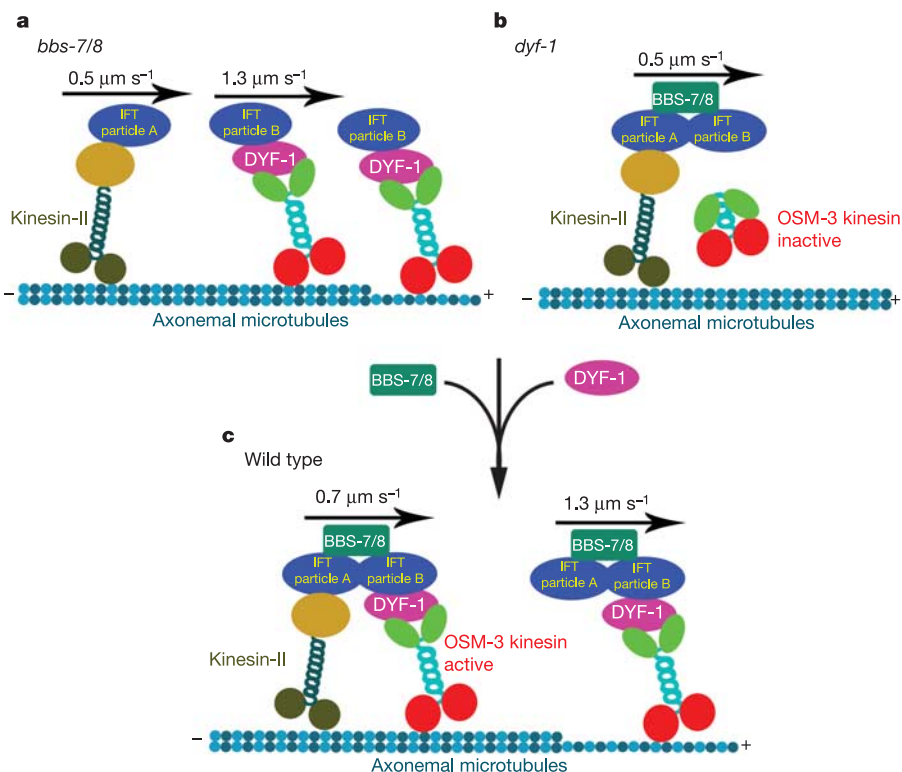
Figure 2 | Model for the functional coordination of the two IFT motors by BBS proteins and DYF-1. a, In *bbs* mutants, IFT particles separate into subcomplex A, which is moved along the middle segment at 0.5 μm s^{-1} by kinesin-II alone, and subcomplex B, which is moved along the middle and distal segments at 1.3 μm s^{-1} by OSM-3 kinesin alone. **b**, In *dyf-1* mutants, IFT particles are moved at 0.5 μm s^{-1} along the middle segment by kinesin-II alone and OSM-3 kinesin attaches to them very weakly or not at all. **c**, In

wild-type cilia, BBS proteins stabilize interactions between subcomplexes A and B, thereby linking kinesin-II and OSM-3 kinesin on intact IFT particles, whereas DYF-1 activates OSM-3 kinesin and docks it onto its cargo. IFT particles are now moved along middle segments at intermediate speeds of 0.7 μm s^{-1} by kinesin-II and OSM-3 kinesin together, then along distal segments to the ciliary tip at 1.3 μm s^{-1} by OSM-3 kinesin alone.

cal IFT pathway to create two sequential pathways (Fig. 2c). The pattern of evolutionary conservation of the proteins involved in creating the two IFT pathways suggests that they may be deployed to build middle and distal ciliary segments in a broad range of ciliated organisms[25–27]. Moreover, mutations in human homologues of some of these proteins, including DYF-1, may cause ciliary diseases, as has been established for BBS-7 and BBS-8. In a broader context, our studies provide insights into the intricate mechanisms by which motor-dependent transport pathways can be combined and integrated to build complex intracellular transport networks.

METHODS

Fluorescence microscopy. Intraflagellar transport was assayed as described previously[7,21,22]. Images were collected using an Olympus microscope equipped with a ×100, 1.35 NA objective and an Ultraview spinning disc confocal head at 0.3 s frame^{-1} for 2–3 min, with the transgenic nematodes anaesthetized with 10 mM levamisole, mounted on agar pads and maintained at 21 °C. Kymographs and movies were created using Metamorph software.

Creation and maintenance of *bbs* mutant animals. Transgenic animals expressing *kap-1::gfp* and *osm-3::gfp* were crossed with *bbs-7/osm-12(n1606)* or *bbs-8(nx77)* to create *bbs-7(n1606)III; Ex[kap-1::gfp]*, *bbs-8(nx77)V; Ex[kap-1::gfp]*, *bbs-7(n1606)III; Ex[osm-3::gfp]* or *bbs-8(nx77)V; Ex[osm-3::gfp]* progenies.

Genetic screens and characterization of the *dyf-1* mutant. To create *che*, *dyf* and *osm* mutants expressing IFT particle::GFP transgenes, males expressing *osm-6::gfp* were mated with homozygous-mutant hermaphrodites, and the F_1 heterozygotes were self-mated to obtain homozygous mutants with *mnIs17[osm-6::gfp]*. The length distribution of OSM-6::GFP along sensory cilia was determined by densitometry of confocal images of live transgenic animals.

Molecular analysis of the *dyf-1* gene. Genomic DNA of F54C1.5 was amplified from *dyf-1(mn335)* nematodes using PCR, and the PCR fragments were sequenced to identify the molecular lesion. A 5.9-kilo base, wild-type PCR-generated fragment of F54C1.5, which includes a 454-base pair (bp) promoter and 577-bp 3′ untranslated region (UTR), was co-injected with the plasmid pRF4 containing the semidominant marker mutation *rol-6(su1006)* into *dyf-1(mn335)* animals expressing *osm-6::gfp*, and heritable lines were selected. Rescue of the *dyf-1* mutant phenotype was assayed by dye filling as previously described[7,14,15]. Expressed sequence tag (EST) clones yk1508h08 and yk808g01, which encode F54C1.5a and F54C1.5b, respectively, were obtained from Y. Kohara and sequenced. Protein domain analysis was performed at http://www.ebi.ac.uk/InterProScan/.

Construction of *dyf-1::gfp* transgene and transformation. *dyf-1::gfp* constructs were generated using a PCR-fusion-based approach[28] by placing its own 454-bp promoter and coding sequence adjacent to a DNA fragment containing GFP and the UNC-54 3′ UTR amplified from the GFP expression vector pPD95.77. Transgenic lines carrying extrachromosomal arrays of the *dyf-1::gfp* expression constructs were generated by germline transformation of the *dyf-1::gfp* construct with the *rol-6* marker into homozygous *dyf-1(mn335)* nematodes. Transgenic nematodes expressing *dyf-1::gfp* were crossed with *osm-3(p802)* or *klp-11(tm324)* to create *osm-3(p802)III; Ex[dyf-1::gfp]* or *klp-11(tm324)III; Ex[dyf-1::gfp]* progenies.

1. Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nature Rev. Mol. Cell Biol.* **3,** 813–825 (2002).
2. Scholey, J. M. Intraflagellar transport. *Annu. Rev. Cell Dev. Biol.* **19,** 423–443 (2003).
3. Pan, J., Wang, Q. & Snell, W. J. Cilium-generated signalling and Cilia-related disorders. *Lab. Invest.* **85,** 452–463 (2005).
4. Cole, D. G. *et al.* Novel heterotrimeric kinesin-related protein purified from sea urchin eggs. *Nature* **366,** 268–270 (1993).
5. Cole, D. G. *et al. Chlamydomonas* kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in *Caenorhabditis elegans* sensory neurons. *J. Cell Biol.* **141,** 993–1008 (1998).
6. Signor, D., Wedaman, K. P., Rose, L. S. & Scholey, J. M. Two heteromeric kinesin complexes in chemosensory neurons and sensory cilia of *Caenorhabditis elegans. Mol. Biol. Cell* **10,** 345–360 (1999).
7. Snow, J. J. *et al.* Two anterograde intraflagellar transport motors cooperate to build sensory cilia on *C. elegans* neurons. *Nature Cell Biol.* **6,** 1109–1113 (2004).
8. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112,** 467–480 (2003).
9. Miki, H., Setou, M., Kaneshiro, K. & Hirokawa, N. All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl Acad. Sci. USA* **98,** 7004–7011 (2001).
10. Goldstein, L. S. B. & Yang, Z. Microtubule-based transport systems in neurons. The roles of kinesins and dyneins. *Annu. Rev. Cell Dev. Biol.* **23,** 39–71 (2000).
11. Mallik, R. & Gross, S. P. Molecular motors: strategies to get along. *Curr. Biol.* **14,** R971–R982 (2004).
12. Lawrence, C. J. *et al.* A standardized kinesin nomenclature. *J. Cell Biol.* **167,** 19–22 (2004).
13. Tabish, M., Siddiqui, Z. K., Nishikawa, K. & Siddiqui, S. Exclusive expression of *C. elegans* osm-3 kinesin gene in chemosensory neurons open to the external environment. *J. Mol. Biol.* **247,** 377–389 (1995).
14. Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans. Dev. Biol.* **117,** 456–487 (1986).
15. Starich, T. A. *et al.* Mutations affecting chemosensory neurons of *Caenorhabditis elegans. Genetics* **139,** 171–188 (1995).
16. Scholey, J. M., Ou, G., Snow, J. J. & Gunnarson, A. Intraflagellar transport motors in *Caenorhabditis elegans* neurons. *Biochem. Soc. Trans.* **32,** 682–684 (2004).
17. Blacque, O. E. *et al.* Loss of *C. elegans* BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. *Genes Dev.* **18,** 1630–1642 (2004).
18. De Riso, L., Ristoratore, F., Sebastiano, M. & Bazzicalupo, P. Amphid defective mutant of *Caenorhabditis elegans. Genetica* **94,** 195–202 (1994).
19. Swoboda, P., Adler, H. T. & Thomas, J. H. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans. Mol. Cell* **5,** 411–421 (2000).
20. Li, J. B. *et al.* Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* **117,** 541–552 (2004).
21. Orozco, J. T. *et al.* Movement of motor and cargo along cilia. *Nature* **398,** 674 (1999).
22. Signor, D. *et al.* Role of a class DHC1b dynein in retrograde transport of IFT motors and IFT raft particles along cilia, but not dendrites, in chemosensory neurons of living *Caenorhabditis elegans. J. Cell Biol.* **147,** 519–530 (1999).
23. Mueller, J., Perrone, C. A., Bower, R., Cole, D. G. & Porter, M. E. The FLA3 KAP subunit is required for localization of kinesin-2 to the site of flagellar assembly and processive anterograde intraflagellar transport. *Mol. Biol. Cell* **16,** 1341–1354 (2005).
24. Deacon, S. W. *et al.* Dynactin is required for bidirectional organelle transport. *J. Cell Biol.* **160,** 297–301 (2003).
25. Awan, A., Bernstein, M., Hamasaki, T. & Satir, P. Cloning and characterization of Kin5, a novel *Tetrahymena* ciliary kinesin II. *Cell. Motil. Cytoskel.* **58,** 1–9 (2004).
26. Cole, D. G. The intraflagellar transport machinery of *Chlamydomonas reinhardtii. Traffic* **4,** 435–442 (2003).
27. Mesland, D. A. M., Hoffman, J. L., Caligor, E. & Goodenough, U. W. Flagellar tip activation stimulated by membrane adhesions in *Chlamydomonas* gametes. *J. Cell Biol.* **84,** 599–617 (1980).
28. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans. Biotechniques* **32,** 728–730 (2002).

LETTERS

Optimality and evolutionary tuning of the expression level of a protein

Erez Dekel[1] & Uri Alon[1]

Different proteins have different expression levels. It is unclear to what extent these expression levels are optimized to their environment. Evolutionary theories suggest that protein expression levels maximize fitness[1–11], but the fitness as a function of protein level has seldom been directly measured. To address this, we studied the *lac* system of *Escherichia coli*, which allows the cell to use the sugar lactose for growth[12]. We experimentally measured the growth burden[13,14] due to production and maintenance of the Lac proteins (cost), as well as the growth advantage (benefit) conferred by the Lac proteins when lactose is present. The fitness function, given by the difference between the benefit and the cost, predicts that for each lactose environment there exists an optimal Lac expression level that maximizes growth rate. We then performed serial dilution evolution experiments at different lactose concentrations. In a few hundred generations, cells evolved to reach the predicted optimal expression levels. Thus, protein expression from the *lac* operon seems to be a solution of a cost–benefit optimization problem, and can be rapidly tuned by evolution to function optimally in new environments.

The expression level of a protein can change over evolutionary timescales by two main processes[15]. The first is neutral evolutionary drift[7], and the second is selection of mutations that increase fitness. The latter is often viewed as an optimization process[1–11,16–18]. Many cases of optimization have been demonstrated on the level of the organism phenotype. For example, bacteria evolve to increase their growth rates in a wide variety of laboratory conditions[1,19,20]. Recently, laboratory evolution experiments showed evolutionary adaptation of *E. coli* towards optimal metabolic fluxes[3,21].

Here we ask whether evolutionary optimization can predict the expression level of a protein in a given environment. We use the lactose operon of *E. coli*, which encodes LacZ, which cleaves the sugar lactose for use as an energy and carbon source, and LacY, which transports lactose into the cell[12,22–24]. Decades of study have provided a quantitative characterization of this system, making it an excellent starting point for a theoretical and experimental study.

The present study had three stages: first, we measured the cost and benefit of Lac protein expression in wild-type *E. coli*. Second, we found that the cost and benefit functions predict that there is an optimal expression level that maximizes growth in a given lactose environment. Third, we monitored the evolution of *E. coli* in different lactose environments and compared its Lac expression to the predicted optimum in each environment (Fig. 1).

To form a cost–benefit theory, we began with direct measurement of the cost and benefit of Lac protein expression in wild-type *E. coli*. The cost is defined as the relative reduction in growth rate due to the burden placed on the cells by production and presence of the Lac
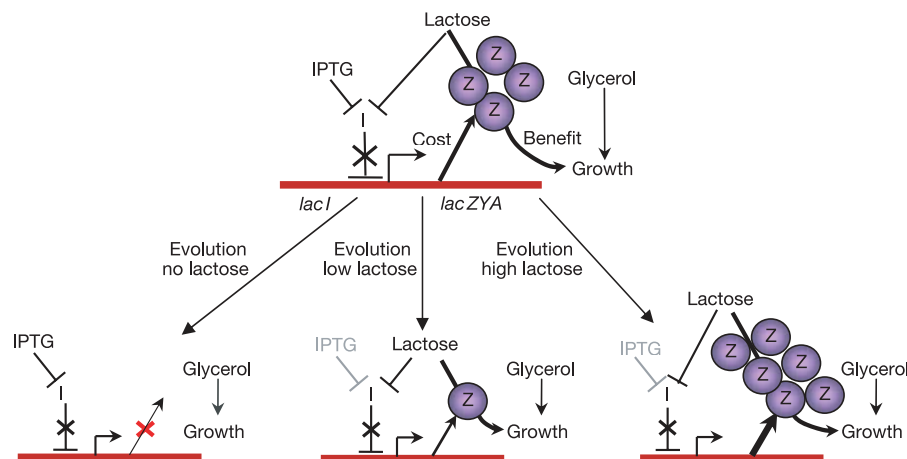
Figure 1 | The *lac* operon of *E. coli* and the experimental design in the present study. The *lac* operon encodes the enzyme LacZ, denoted Z, which uses the sugar lactose, L, to increase growth rate. In the experiments, we consider environments with constant concentrations of lactose. The repressor of the *lac* system LacI, denoted I, was deactivated by the presence of lactose. To examine evolution at zero lactose under the same conditions, we added the non-metabolized inducer IPTG, which also deactivates LacI, and glycerol as an additional carbon source, in all cases. Measurements on the wild-type strain provided estimates of the reduction in growth due to the burden of Lac protein production or maintenance (cost η) and the increase in growth generated by Lac proteins in the presence of lactose (benefit B). Evolution experiments were performed by serial dilution of cells growing in tubes for several hundred generations. Cells undergoing evolution in the absence of lactose were predicted to lose *lac* expression. Cells undergoing evolution at low (high) lactose concentrations were predicted to evolve to optimal low (high) Lac protein levels relative to wild type.

[1]Department of Molecular Cell Biology and Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot 76100, Israel.

proteins[13,14]. To measure the cost of Lac expression, we measured the growth rate at various concentrations of the inducer isopropyl-β-D-thiogalactoside (IPTG; Fig. 2a) in the absence of lactose. IPTG induces the *lac* system, but no benefit is gained because IPTG is not metabolized. We found a nonlinear, concave dependence of the cost on the level of expression (Fig. 2a). The cost per protein seems to increase with the amount of proteins produced.

The nonlinear cost function may reflect the fact that the cell has limited resources, so that production of Lac proteins reduces the capacity to produce other important proteins. For example, rapidly growing cells are known to devote most of their ribosomal capacity to synthesizing ribosomal proteins and other translation factors, and most of their transcription capacity to produce ribosomal related RNA[25]. To describe the cost as a function of protein level we used two nonlinear functions. The first is a quadratic function

$$\eta_1(Z) = \eta_0 Z + \eta_0' Z^2 \qquad (1)$$

where Z is the expression level of the Lac proteins and η_0 and η_0' are parameters. The second cost function assumes that there is an effective maximal capacity M for producing non-essential proteins. Production of Lac proteins at levels that approach M would inhibit production of essential systems, so that cell growth would be significantly slowed down. Perhaps the simplest phenomenological cost function that increases strongly as an effective limit M is approached is (see Methods for a derivation of this function):

$$\eta_2(Z) = \frac{\eta_0 Z}{1 - Z/M} \qquad (2)$$

The parameters of cost function 2 that describe the measured reduction in growth rate are $\eta_0 Z_{WT} = 0.02 \pm 0.003$, where Z_{WT} is the fully induced expression of the wild-type Lac proteins and $M = (1.8 \pm 0.3)Z_{WT}$. We note that the parameter M is expected to depend on the growth rate of the cells: at slower growth rates, more resources are known to be free for protein expression[25] and M is higher (see Supplementary Information for more details). We also note that a linear cost function does not fit the direct cost measurement of the wild-type strain (Fig. 2a) and is not able to explain the results presented below. We find that the cost of fully induced wild-type expression of the *lac* system is about 4.5%.

In addition to the cost of Lac proteins, we also directly measured their benefit. For this purpose, we provided saturating IPTG for full induction and various concentrations of the sugar lactose. The non-metabolized inducer IPTG kept the Lac protein levels constant, and hence the cost constant, allowing measurement of the benefit due to use of lactose. The growth rate of the cells increased with lactose concentration, reflecting the benefit gained by the action of the Lac proteins that transport and use this sugar (Fig. 2b).

We find that the form of the benefit function is well described by the established transport and catabolism kinetics of this system (see
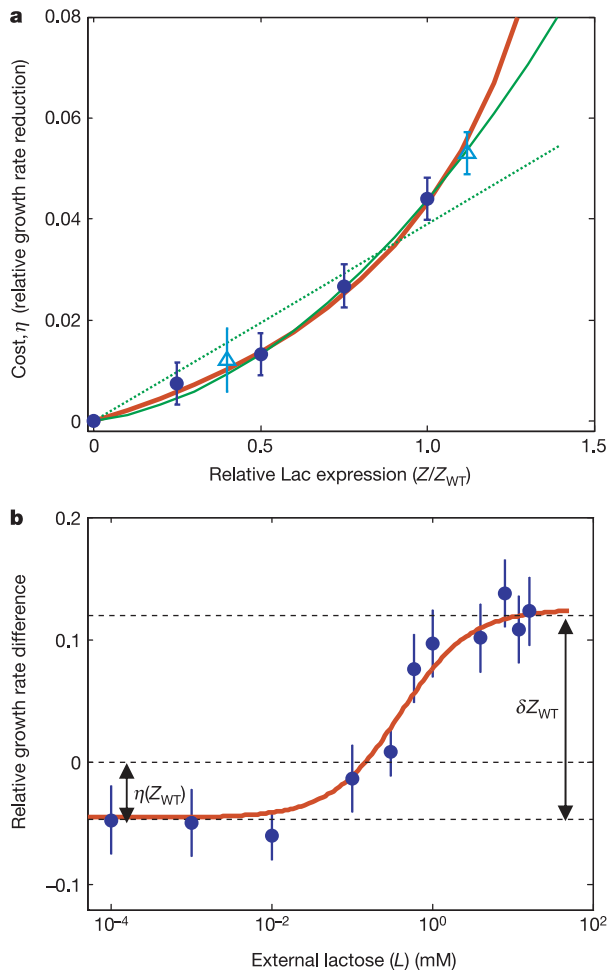
Figure 2 | Cost and benefit functions of *lac* expression in wild-type *E. coli*. a, Cost, defined as relative reduction in growth of *E. coli* wild-type cells grown in defined glycerol medium with varying amounts of IPTG relative to cells grown with no IPTG. The *x* axis is LacZ protein level relative to LacZ protein level at saturating IPTG (Z_{WT}). Also shown are the cost of strains evolved at 0.2 mM lactose for 530 generations (data point at the 0.4 point of the *x* axis; open triangle) and 5 mM lactose for 400 generations (data point at the 1.12 point of the *x* axis; open triangle). The green solid line represents cost function 1 (equation (1) with $\eta_0 Z_{WT} = 0.09 \pm 0.01$ and $\eta_0' Z_{WT}^2 = 3 \pm 0.1$). The red line represents cost function 2 (equation (2)). The green dotted line represents a linear cost function. **b,** Benefit of Lac proteins as a function of lactose. Cells were grown with saturating levels of IPTG and varying levels of lactose. The growth rate difference is shown relative to cells grown with no IPTG or lactose. $\eta(Z_{WT})$ is the cost of the *lac* system at zero lactose, and δZ_{WT} is the benefit of *lac* induction at saturating lactose levels. The red line indicates the theoretical growth rate, using $g = -\eta(Z_{WT}) + \delta Z_{WT} L/(K_Y + L)$, with $K_Y = 0.4$ mM, $\eta(Z_{WT}) = 0.044$ and $\delta Z_{WT} = 0.17$ (equation (5)). Error bars are the experimental standard errors.
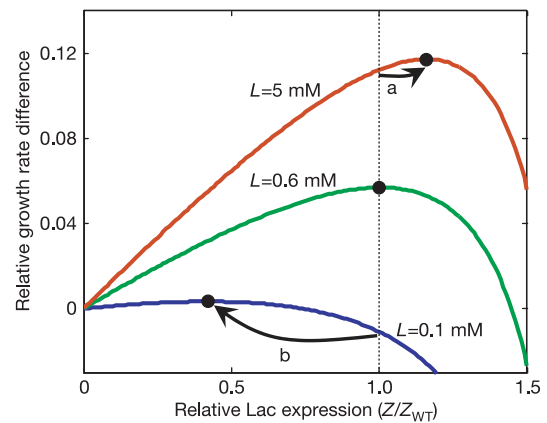
Figure 3 | 3 Predicted relative growth rate of cells (the fitness function) as a function of Lac protein expression. The fitness function, given by the difference of cost and benefit, is shown for different concentrations of lactose. The *x* axis is the ratio of protein level to the fully induced wild-type protein level, Z/Z_{WT}. Shown are relative growth rate differences with respect to the uninduced wild-type strain for environments with lactose levels $L = 0.1$ mM (blue line), $L = 0.6$ mM (green line) and $L = 5$ mM (red line), according to equation (5). The dot on each line is the predicted optimal expression level that provides maximal growth (equation (6)). Cells grown in lactose levels above 0.6 mM are predicted to evolve to increased Lac protein expression (arrow a), whereas cells grown at lactose levels lower than 0.6 mM are predicted to evolve to decreased Lac protein expression (arrow b).

Methods and Supplementary Information). The increase in growth rate is proportional to the rate of lactose use, and hence to the enzymatic rate of LacZ:

$$B(Z) = \delta[ZL_{in}] \qquad (3)$$

where L_{in} is the concentration of lactose in the cell, $[ZL_{in}]$ is the concentration of LacZ bound to lactose, and δ is the relative growth advantage per LacZ molecule at saturating lactose concentration. Using the known biochemical kinetic parameters of LacY and LacZ yields good agreement with our measured growth rates, and provides an experimental estimate of the benefit parameter δ under the present conditions: $\delta = 0.17\, Z_{WT}^{-1}$. Thus, full induction of the *lac* system confers a 17% growth advantage at saturating lactose relative to cells with no lactose.

These experiments defined the cost and benefit as a function of Lac

protein levels in the wild-type bacterium. The growth rate relative to the growth rate with no Lac expression is given by the difference between the cost and benefit functions

$$g = -\eta(Z) + B(Z) \qquad (4)$$

We find that the cost and benefit functions are such that at each level of lactose in the environment there exists a well-defined optimal protein expression level that maximizes growth. The reason for this optimum is that benefit increases linearly with expression level, whereas cost increases nonlinearly and thus high expression levels become unfavourable. The optimal protein expression level depends on the concentration of lactose in the environment, L (Fig. 3). To compute the optimum we calculated the protein level Z_{opt}, which maximizes growth (see equation (6) in Methods for an analytical solution); that is, where $dg/dZ = 0$.

We find that the predicted optimal protein level Z_{opt} increases with the lactose level in the environment (Fig. 3). In the absence of lactose, the optimal expression is $Z_{opt} = 0$, because the Lac proteins bear only a cost and no benefit. At high lactose levels, Z_{opt} is greater than the wild-type expression level Z_{WT}, because increased *lac* expression brings additional growth benefit to the cells. The cost–benefit analysis predicts that the wild-type unrepressed expression level is optimal ($Z_{opt} = Z_{WT}$) at a lactose concentration of $L_0 = 0.6 \pm 0.1$ mM.

Does *E. coli* actually evolve to these optimal expression levels if supplied with a constant lactose environment? To test this, we performed laboratory evolution experiments using serial dilution protocols[19]. Cells were grown in 10-ml cultures in a defined medium supplemented with a given lactose concentration. Each day, the cells were diluted 100-fold into a new tube with fresh medium, resulting in $\log_2 100 = 6.6$ generations of growth per day. Experiments at seven lactose concentrations were performed in parallel for over 500 generations.

At different times during the experiment we measured LacZ activity using an enzymatic assay (ONPG assay; see Supplementary Information) and LacZ protein level using quantitative gel electrophoresis (see Supplementary Information). We found that, in all cases, the activity per LacZ enzyme did not measurably change over the course of the evolutionary experiments. Hence, changes in LacZ activity were proportional to changes in LacZ protein level (Supplementary Fig. 6).

We found that LacZ activity and protein level of the cells changed heritably during the course of the experiment, and approached a new adapted state after 300–500 generations (Fig. 4). Cells grown at low lactose concentrations ($L = 0$, 0.1 mM and 0.2 mM) showed a decrease in LacZ activity and expression (Fig. 4). Cells grown in the absence of lactose lost enzyme expression and activity altogether due to mutations such as a 764-base-pair deletion that included the entire *lac* promoter (see Supplementary Information). Cells growing at $L = 0.5$ mM showed little change in LacZ activity and expression. Cells growing at lactose concentrations of 1 mM, 2 mM and 5 mM showed an increase in LacZ activity and expression, although the increase saturated at about 12% of the wild-type level.

The evolved cells reached LacZ expression levels close to the predicted optimal levels (Fig. 4). All cost functions give good predictions at low to intermediate lactose levels ($L < 1$ mM) (Fig. 4b), predicting, for example, that the wild-type protein levels are optimal in an environment with $L \approx 0.5$ mM, as observed. The cost function that best explains the data at high lactose levels is cost function 2 (equation (2)). This cost function shows a large cost per protein at high expression levels, and thus limits the optimal protein levels in high lactose environments. Hence, the cost and benefit fitness function measured in the wild-type strain can be used to evaluate the optimal protein level reached by evolutionary selection in different environments.

The rate at which the cells converged to their adapted expression level is shown in Fig. 4a. The adapted expression is reached within
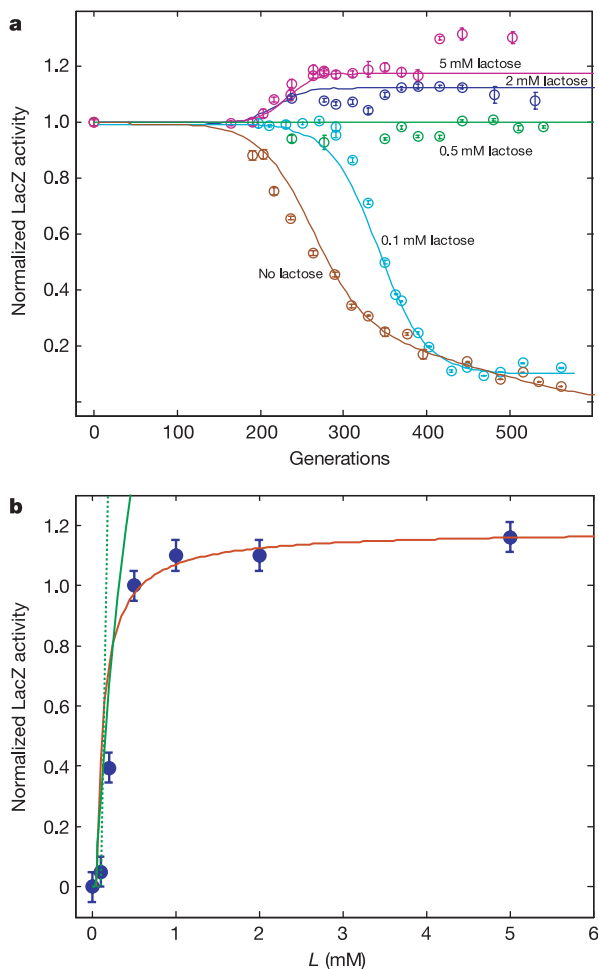
Figure 4 | Experimental evolutionary adaptation of *E. coli* cells to different concentrations of lactose. a, LacZ activity, relative to wild-type cells, of cells grown for 530 generations in serial dilution experiments with different lactose levels is shown as a function of generation number. Cells were grown in 0, 0.1 mM, 0.5 mM, 2 mM and 5 mM lactose in a glycerol minimal medium supplemented with 0.15 mM IPTG. Lines are population genetics simulations. The only fitting parameter in these simulations is the probability per generation per cell of a mutation that yields the predicted optimal LacZ level (see Methods). **b**, Adapted LacZ activity of cells in serial dilution experiments as a function of lactose concentration, L, relative to wild-type cells. Data are for more than 530 generations, except for the data point at 5 mM lactose, which is at generation 400. The red line indicates the theoretical prediction for optimal expression level using cost function 2 (equation (6)). The green solid line indicates the quadratic cost function. The green dotted line indicates the linear cost function. Error bars are experimental standard errors.

about 350 generations for high lactose levels and within 400–500 generations at low lactose levels. These rates are well described by a population genetics model in which cells grow exponentially and produce, with probability p, mutants that have the predicted protein level and the measured adapted growth rate. The only free parameter in the theoretical curves shown in Fig. 4a is the probability per cell per generation, p, of producing a mutant, which ranged between 10^{-7} and 3×10^{-7}. The simulations suggest that the mutation occurred and was lost during dilution many times during the experiment until it was fixed in the population, and that the dynamics can be explained by a single mutation that changed lac expression rather than a combination of several mutations[11]. The curves in every run of the stochastic simulation are nearly identical, showing that the selection dynamics can be described by a deterministic approximation (see Supplementary Information). Because the mutation rate in $E. coli$ is on the order of 10^{-9} per base pair per generation, the observed mutant probabilities, $p \approx 10^{-7}$, suggest that there is an effective 'mutational target' on the order of 100 different base pairs that can give rise to each of the optimal phenotypes.

At the highest level of lactose, 5 mM, cells first converged to the predicted optimal expression level, and then, around generation 400, showed a second jump[1] in LacZ expression. This jump may represent mutations in other systems that provide an advantage in the present growth conditions, or a change in the parameter M in cost function 2. Indeed, direct measurements of the cost function of this mutant suggest a threefold increase in the parameter M (Supplementary Fig. 8). In contrast, direct measurement of the cost functions of strains that evolved at lower lactose levels shows a cost function that is the same as that of the wild-type strain (Fig. 2a).

This study provides evidence that cells can rapidly evolve towards protein expression levels that are optimal solutions of a cost–benefit problem. The cost and benefit fitness function measured in the wild-type strain was found to predict optimal protein levels in different lactose environments, which were then reached in direct evolutionary experiments. It would be intriguing to test whether cost–benefit principles apply to more complex gene regulation circuits, to attempt to form and test a theory of optimal gene circuit design[6,8,26–28]. This could explain the observed convergence of different organisms to similar network motifs[16,28,29]. More generally, this study suggests that small expression differences between microorganisms can be due to selected biological functionality rather than random drift. Optimality principles might contribute to a fundamental description of the interaction between biological circuitry, evolution and the environment.

METHODS

Strains and media. $E. coli$ MG1655 ($E. coli$ genetic stock centre) was used. All experiments were in M9 defined medium consisting of M9 salts, 1 mM MgSO$_4$, 0.1 mM CaCl$_2$, 0.05% casamino acids, 0.1% glycerol, 0.15 mM IPTG and specified concentrations of D-lactose (Sigma). This concentration of IPTG had no measurable effect on growth of a strain deleted for the lac operon (data not shown). Use of glycerol in the medium allowed study of evolution also in the absence of lactose. The bacteria used lactose and glycerol simultaneously throughout growth (Supplementary Fig. 1). Previous studies have shown that cells grown in a chemostat with limiting lactose evolve to greatly increase the expression of the Lac proteins[20]. In those experiments lactose was the sole carbon source and hence the lac system affected growth rate much more strongly than in the present experiments.

Growth rate measurements. The exponential growth rate difference of two strains was measured by comparing 48 cultures of each strain grown in a checkerboard pattern on a 96-well plate, yielding an accuracy of about 0.8% (see Supplementary Information for more details). Multiple assays were averaged in cases where lower standard errors were required. Exponential growth rate in the 96-well plate is the same as the growth rate in 50-ml tubes (Supplementary Figs 3 and 4).

Serial dilution experiments. Ten-millilitre cultures were grown in 50-ml tubes shaken at 220 r.p.m. at 37 °C. After a day of growth, cells were diluted 1:100 into a fresh tube. Samples were frozen (-80 °C) every 3 days. Lactose levels in five of the tubes (5 mM, 2 mM, 1 mM, 0.5 mM, 0.2 mM) were high enough to cause full

induction of the lac system (Supplementary Fig. 7). To make induction equal and maximal in all tubes, including those with 0 mM and 0.1 mM lactose, each tube also contained a saturating level of IPTG.

β-Galactosidase activity and protein level measurements. A substrate of β-galactosidase (LacZ) that gives optical readout (ONPG) was used in a multi-well plate reader to obtain high accuracy measurements of activity (see Supplementary Information for details). LacZ protein levels were quantified using SDS gel electrophoresis, and relative expression was quantified by comparison to control lanes with appropriate mixtures of induced and uninduced wild-type cell lysates (Supplementary Fig. 6).

Derivation of cost function 2. A simple derivation of cost function 2 is based on a growth rate that is a saturated function[30] of an internal resource R, given by $g = \beta R/(K + R)$ where β is the maximal growth rate and K is the resource level for half maximal growth. Production of a unit of protein Z reduces the resource R by ε. Hence, $g(Z) = \beta(R - \varepsilon Z)/(K + R - \varepsilon Z)$. This yields a cost $\eta(Z) = (g(0) - g(Z))/g(0) = \eta_0 Z/(1 - Z/M)$, as in equation (2), where $\eta_0 = \varepsilon K/R(K + R)$ and $M = (K + R)/\varepsilon$. Note that the cost function can not diverge but rather reaches a maximum value of $\eta(Z) = 1$ when $\varepsilon Z = R$, resulting in $g = 0$.

Analytical approximation to the optimal expression level. A detailed transport model of the lac system[22–24] (see Supplementary Information) suggests that, to a good approximation, the rate of LacZ action is equal to the rate of lactose influx through the LacY permease at all but the highest lactose concentrations:

$$V_z[ZL_{in}] \approx V_Y[YL] \approx V_Y \frac{YL}{K_Y + L}$$

where V_z is the velocity of LacZ, V_Y is the velocity of LacY and $K_Y = 0.4$ mM is the Michaelis constant of LacY[24]. This yields a cost–benefit fitness function, using cost function 2 and using Y proportional to Z (so that the measured benefit parameter δ includes the ratio Y/Z as well as V_Y and V_z):

$$g = -\frac{\eta_0 Z}{1 - Z/M} + \delta \frac{ZL}{K_Y + L} \qquad (5)$$

The optimum of this, $dg/dZ = 0$, occurs at an optimal protein level of

$$Z_{opt} = M\left(1 - \sqrt{\frac{\eta_0}{\delta}\frac{L + K_Y}{L}}\right) \qquad (6)$$

Note that $Z_{opt} = 0$ for lactose levels lower than $L_c = K_Y(\delta/\eta_0 - 1)^{-1} \approx 0.057$ mM because cost exceeds benefit. The parameters are $M = 1.8 Z_{WT}$, $\eta_0/\delta = 0.02/0.17 = 0.12$ and $K_Y = 0.4$ mM.

Similar results for Z_{opt} were found by numerically solving the detailed transport model and finding the optimal expression by varying LacY and LacZ levels (Supplementary Fig. 9). Note that the analytical approximation (equations (5) and (6)) underestimates the growth advantage of the adapted strains relative to the wild-type strain in the same environment, except near $L = 0$ and $L = 0.6$ mM, where predicted growth advantage matches the experimental results. The detailed simulation gives better predictions for the growth advantage of the adapted strains (Supplementary Fig. 9) at all lactose levels.

Simulations of evolution rate. Population genetics simulations were performed as follows[7] (see Supplementary Information for details): a population of wild-type cells grew exponentially at growth rate g_0, growing each simulated day from $N_0 = 10^8$ cells to $N_f = 10^{10}$ cells. Mutants were formed with a probability p per generation per cell. The mutants grew at rate $g_0 + \Delta g$, with relative LacZ expression ΔA. The parameter Δg was set equal to the measured growth rates of the adapted strains (Supplementary Fig. 9), and ΔA was used from the optimum of the cost and benefit fitness function (equations (5) and (6)). At the end of each simulated day, 1/100 of the population was passed to the next simulated day, of which the fraction of mutants was determined by a random binomial process. The resulting dynamics show that the mutants eventually take over the population. The simulations have only one free parameter, p, which was fitted to the data of Fig. 4a, resulting in $p = 6.5 \times 10^{-6} \pm 2 \times 10^{-6}$, $3 \times 10^{-7} \pm 1 \times 10^{-7}$, $3 \times 10^{-7} \pm 1 \times 10^{-7}$ and $3 \times 10^{-7} \pm 1 \times 10^{-7}$ for $L = 0$ mM, 0.1 mM, 2 mM and 5 mM, respectively. Simulations suggested that adaptation was due to a single mutation, except at $L = 0$, in which two mutations occurred: one that affected LacZ protein level and the other that increased growth rate in glycerol[3], in agreement with measurements on this adapted strain (Supplementary Fig. 11).

1. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).

2.   Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
3.   Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186–189 (2002).
4.   Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
5.   Rosen, R. *Optimality Principles in Biology* (Butterworths, London, 1967).
6.   Savageau, M. A. *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, Massachusetts, 1976).
7.   Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer, Sunderland, Massachusetts, 1997).
8.   Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Chapman and Hall, New York, 1996).
9.   Maynard Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1997).
10. Hartl, D. L. & Dykhuizen, D. E. The population genetics of *Escherichia coli*. *Annu. Rev. Genet.* **18**, 31–68 (1984).
11. Liebermeister, W., Klipp, E., Schuster, S. & Heinrich, R. A theory of optimal differential gene expression. *Biosystems* **76**, 261–278 (2004).
12. Muller-Hill, B. *The lac Operon: a Short History of a Genetic Paradigm* (Walter de Gruyter, New York, 1996).
13. Koch, A. L. The protein burden of lac operon products. *J. Mol. Evol.* **19**, 455–462 (1983).
14. Nguyen, T. N., Phan, Q. G., Duong, L. P., Bertrand, K. P. & Lenski, R. E. Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli* K12. *Mol. Biol. Evol.* **6**, 213–225 (1989).
15. Fay, J. C., McCullough, H. L., Sniegowski, P. D. & Eisen, M. B. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R26 (2004).
16. Conant, G. C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genet.* **34**, 264–266 (2003).
17. Stephanopoulos, G. & Kelleher, J. Biochemistry. How to make a superior cell. *Science* **292**, 2024–2025 (2001).
18. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15112–15117 (2002).
19. Cooper, T. F., Rosen, D. E. & Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 1072–1077 (2003).
20. Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic flux and fitness. *Genetics* **115**, 25–31 (1987).
21. Honisch, C., Raghunathan, A., Cantor, C. R., Palsson, B. O. & van den Boom, D. High-throughput mutation detection underlying adaptive evolution of *Escherichia coli*-K12. *Genome Res.* **14**, 2495–2502 (2004).
22. Kremling, A. *et al.* The organization of metabolic reaction networks. III. Application for diauxic growth on glucose and lactose. *Metab. Eng.* **3**, 362–379 (2001).
23. Wong, P., Gladney, S. & Keasling, J. D. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* **13**, 132–143 (1997).
24. Yildirim, N., Santillan, M., Horike, D. & Mackey, M. C. Dynamics and bistability in a reduced model of the lac operon. *Chaos* **14**, 279–292 (2004).
25. Bremer, H. & Dennis, P. P. in *Escherichia coli* and *Salmonella* (ed. Neidhardt, F. C.) 1553 (American Society for Microbiology, Washington DC, 1996).
26. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
27. Endy, D., You, L., Yin, J. & Molineux, I. J. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl Acad. Sci. USA* **97**, 5375–5380 (2000).
28. Dekel, E., Mangan, S. & Alon, U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.* **2**, 81–88 (2005).
29. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
30. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).

LETTERS

Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants

Duo Wang[1], Scott Kennedy[1]*†, Darryl Conte Jr[2]*, John K. Kim[1], Harrison W. Gabel[1], Ravi S. Kamath[1], Craig C. Mello[2,3] & Gary Ruvkun[1]

Caenorhabditis elegans homologues of the retinoblastoma (Rb) tumour suppressor complex specify cell lineage during development[1,2]. Here we show that mutations in Rb pathway components enhance RNA interference (RNAi) and cause somatic cells to express genes and elaborate perinuclear structures normally limited to germline-specific P granules. Furthermore, particular gene inactivations that disrupt RNAi reverse the cell lineage transformations of Rb pathway mutants. These findings suggest that mutations in Rb pathway components cause cells to revert to patterns of gene expression normally restricted to germ cells. Rb may act by a similar mechanism to transform mammalian cells.

A mutation in the *C. elegans* gene *lin-15* causes enhanced sensitivity to RNAi (Fig. 1a). We tested four double-stranded (ds)RNAs that induce weaker RNAi phenotypes in wild-type animals than the corresponding loss-of-function mutant phenotype. All these dsRNA induce stronger loss-of-function phenotypes in *lin-15(n765)* animals (Fig. 1a and Table 1). The *lin-15* locus comprises two genes, *lin-15A* and *lin-15B*, that are inactivated by the *n765* mutation (ref. 3). RNAi is enhanced in *lin-15B(n744)* but not in the strong loss-of-function *lin-15A(n767)* or *lin-15A(n433)* alleles (Fig. 1a). *lin-15B* is a class B synthetic multivulva (synMuv B) gene[3]. This class of genes encodes homologues of the mammalian tumour suppressor Rb[1] and Rb complex components LIN-53 (mammalian RbAp48) (ref. 1) and DPL-1 (mammalian DP)[2], as well as nuclear proteins that function in the Rb pathway, such as LIN-13, LIN-9, LIN-36 and TAM-1 (ref. 4). SynMuv B genes function redundantly with synthetic multivulva A (synMuv A) genes, such as *lin-15A*, to antagonize Ras signalling in the specification of vulval cell lineages[1–3]. Animals defective in both the synMuv A and synMuv B pathways display multiple vulva structures due to cell lineage changes, whereas animals defective in only one of the synMuv pathways have normal cell lineages in these tissues[1–3].

Other members of the *C. elegans* Rb pathway also negatively regulate RNAi. Loss-of-function mutations or RNAi inactivation of six other Rb pathway genes—*lin-35* (*Rb*), *dpl-1* (*DP*), *lin-53*, *lin-9*, *lin-13* and *hpl-2* (*HP1*); where homologous mammalian genes are denoted in parentheses—markedly enhance RNAi (Fig. 1a; see also Supplementary Fig. S1). Null or strong alleles of three other synMuv B genes, *lin-36*, *tam-1* and *lin-37*, do not enhance RNAi (Fig. 1a), suggesting that some synMuv B genes can affect vulval cell lineages without enhancing RNAi of these tester genes. Strong or null alleles of the four known synMuv A genes, *lin-15A*, *lin-8*, *lin-56*, and *lin-38* do not enhance RNAi (Fig. 1a). The enhanced RNAi of synMuv B
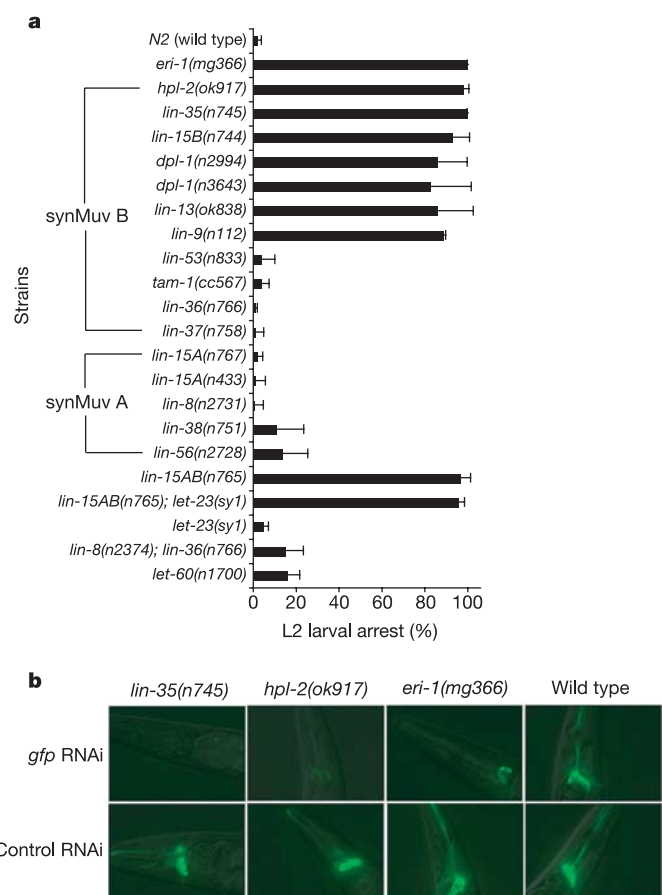
Figure 1 | Rb pathway mutants are more sensitive to RNAi than wild-type animals. a, Animals of the indicated genotype were grown on bacteria expressing *cel-1* dsRNA and the percentage of their progeny arresting at the L2 larval stage was determined. Wild-type N2 animals and *eri-1(mg366)*[5] mutants were used as negative and positive controls. Error bars represent standard deviation among the data collected from at least three independent experiments. **b**, Neuronally expressed *tub-1::gfp* is susceptible to *gfp* RNAi in the *lin-35* and *hpl-2* mutants. *lin-35(n745)*, *hpl-2(ok917)*, *eri-1(mg366)* and wild-type animals, all carrying the integrated *tub-1::gfp* transgene, were grown on bacteria expressing *gfp* dsRNA (*gfp* RNAi) or an empty vector L4440 (control RNAi). Fluorescence microscopy was performed on representative adult progeny.

[1]Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA. [2]Program in Molecular Medicine, and [3]Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. †Present address: Department of Pharmacology, University of Wisconsin, Madison, Wisconsin 53706, USA.
*These authors contributed equally to this work.

pathway mutants is not dependent on whether the animals have ectopic pseudovulvae: for example, the *lin-15AB(n765); let-23(sy1)* double mutant has enhanced RNAi but is vulvaless due to a defect in the receptor tyrosine kinase *let-23* (ref. 3) (Fig. 1a). Conversely, mutants with ectopic pseudovulvae, such as *lin-8(n2374); lin-36(n766)* or the Ras gain-of-function mutant *let-60(n1700)* have wild-type RNAi response (Fig. 1a).

Consistent with the enhanced RNAi phenotype of Rb pathway mutants, target messenger RNA abundance is markedly decreased in these mutants. The induction of movement defects by feeding nematodes bacteria that express dsRNA targeting the muscle gene *unc-22* is enhanced in Rb pathway mutants. In three independent experiments, *unc-22* RNAi in wild-type animals caused a 50% reduction in the level of *unc-22* mRNA, whereas in the *lin-35* mutant, *unc-22* mRNA was reduced by 70–80% (Supplementary Fig. S2).

RNAi is particularly enhanced in the nervous system of Rb pathway mutants. Neurons are generally refractory to RNAi[5]. In *lin-35* (*Rb*), *dpl-1* (*DP*), *hpl-2* (*HP1*), *lin-15B*, *lin-9* and *lin-13* mutants, but not in wild-type animals, dsRNA targeting *gfp* silences neuronally expressed *tub-1::gfp* (Fig. 1b). In the *lin-35* mutant, but not wild type, feeding RNAi targeting endogenous neuronal genes, such as *egl-8*, *egl-30*, *unc-17* and *unc-2*, causes resistance to the acetylcholinesterase inhibitor aldicarb, similar to mutations in these loci (Table 1; see also Supplementary Fig. S3). However, RNAi of these genes does not induce significant movement defects observed in the corresponding mutants, suggesting partial gene inactivation or inactivation only in a particular subset of neurons.

The Rb pathway regulates RNAi synergistically with other negative regulators of RNAi, such as the 3′ exonuclease ERI-1 (ref. 5) and the RNA-dependent RNA polymerase RRF-3 (ref. 6) that both act in the same pathway[5]. Rb pathway null mutants and *eri-1* (or *rrf-3*) null mutants are hypersensitive to RNAi of distinct sets of target genes (Table 1). The Rb pathway and *eri-1* have distinct responses to injected short interfering RNAs (siRNAs): whereas an *eri-1* mutant is more responsive to injection of *unc-22* siRNA (23 base pairs long) than wild type, *lin-35* and *lin-15B* mutants are not more responsive (Table 1), suggesting that the Rb pathway acts at a step upstream of siRNA production. Strains carrying mutations in both *eri-1* and *lin-15b* were synergistically sensitive to RNAi. The release from negative regulation of RNAi in strains carrying mutations in both the Rb and *eri-1*/*rrf-3* pathways increases the intensity of RNAi in the *C. elegans* nervous system. These double mutants allow RNAi to be used for the functional genomic dissection of neuronal functions (see

the accompanying paper (ref. 7)) as well as for the identification of neural components in physiological circuits, such as the regulation of metabolism, feeding and ageing.

The enhanced RNAi of Rb pathway mutants depends on many of the canonical RNAi pathway genes, with one exception that gave an indication of how the Rb pathway may enhance RNAi. Double mutants between *lin-35* and the RNAi-defective genes *rde-1*, *rde-4* and *mut-7*, and between *lin-15b* and *mut-16*, are unresponsive to *nsf-1* dsRNA; they are RNAi defective, like the single *rde-1*, *rde-4*, *mut-7*, or *mut-16* mutants (Supplementary Table S1). However, although animals with a mutation in the RNA-dependent RNA polymerase (RdRP) gene *rrf-1* are unresponsive to *unc-22* and *nsf-1* dsRNAs, *lin-15b; rrf-1* double mutant animals are responsive to both dsRNAs (Supplementary Table S2). RdRPs enhance RNAi by allowing siRNAs from the primary response to dsRNAs to prime further replication on target mRNAs, which in turn amplify the response[8]. The RRF-1 RdRP is essential for RNAi of somatic genes[8], whereas the EGO-1 RdRP is enriched in the germ line and is required for RNAi of genes expressed in the germ line[8,9]. One model for the relaxed requirement for *rrf-1* in the Rb pathway mutants is ectopic expression in somatic tissues of the germline-specific EGO-1 RdRP, making *rrf-1* and *ego-1* redundant genes.

The model of misexpression of RNAi components in Rb pathway mutants is supported by the somatic expression of normally germ-line-specific P-granule-like structures in the *Rb* pathway mutants (Fig. 2a). *pgl-1* encodes an essential RNA-binding component of P granules that mark the *C. elegans* germ plasm in wild type[10]. In the *lin-35* (*Rb*) mutant, PGL-1 is ectopically expressed in intestine, hypoderm and particular cells near the nerve ring, where it accumulates into perinuclear structures that closely resemble the normally germline-restricted P granules (Fig. 2a). Similar PGL-1-containing structures are also expressed in the somatic cells of *lin-13*, *dpl-1*, *hpl-2*, *lin-15B* and *lin-9* mutants (data not shown). In contrast, ectopic PGL-1 is not detected or is present at much lower levels in the soma of the synMuv B mutants that do not enhance RNAi— *lin-53(n833)* or *lin-37(n758)*—or the synMuv A mutants (data not shown). *pgl-1* is also misexpressed in the somatic cells of the synMuv B NuRD component mutants *let-418* (*Mi2*) and *mep-1* (ref. 11). The fact that the somatic PGL-1 in Rb pathway mutants accumulates in perinuclear structures as in germline cells suggests that additional P granule components and germline-specific factors are also ectopically expressed to allow the assembly of P granules[11].

C. elegans P granules are normally assembled in the germline

Table 1 | Rb pathway mutants display enhanced RNAi distinct from the *eri-1* and *rrf-3* mutants

Strains	Feeding RNAi							Injection of *unc-22* siRNA (% twitchers)#
	hmr-1 (% Emb)*	*unc-15* (Unc/Prz)†	*unc-17* (resistant to aldicarb)‡	*egl-8* (resistant to aldicarb)‡	*myo-2* (% L1 let.)§	*dpy-13* (Dpy) ‖	*daf-2* (% dauer)¶	
Wild type (N2)	11 ± 12	+	−	−	0	+	0	19 ± 2
lin-15AB(n765)	99 ± 1	+++	+	++	0	+	0	NT
lin-15A(n767)	15 ± 6	+	−	−	0	+	0	NT
lin-15B(n744)	71 ± 10	+++	+	++	0	+	0	21 ± 6
lin-35(n745)	82 ± 9	+++	+	++	0	+	0	0
eri-1(mg366)	98 ± 2	+++	−	+	89 ± 8	+++	0	73 ± 7
rrf-3(pk1426)	97 ± 4	+++	−	+	92 ± 5	+++	0	NT
lin-35(n745); eri-1(mg366)	100 ± 0	+++	++	+++	90 ± 4	+++	95 ± 7	NT
rrf-3(pk1426); eri-1(mg366)	100 ± 0	+++	−	+	87 ± 11	+++	0	NT

Nematodes of the indicated genotype were grown on bacteria expressing the indicated dsRNAs (Feeding RNAi) or microinjected with synthetic *unc-22* siRNA (23 bp with two-nucleotide overhangs)[5] at 20 °C. All data were collected from at least three independent experiments.
*Progeny were scored for the percentage of dead embryos (Emb).
†Progeny were scored for uncoordinated (Unc) (+) or paralysed (Prz) (+++) phenotype.
‡Progeny were scored for no (−), weak (+) or strong resistance (+++) to 1.0 mM of the acetylcholinesterase inhibitor aldicarb.
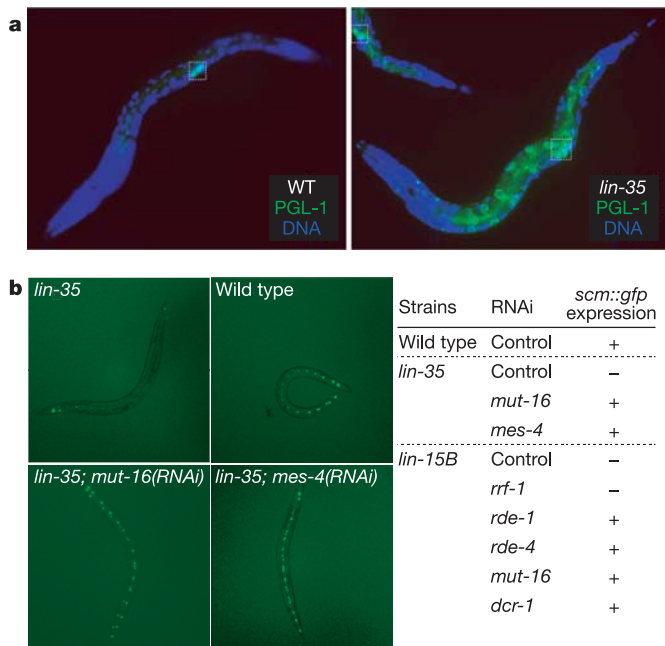§Progeny were scored for L1 larval lethality (% L1 let.).
‖Nematodes were scored as having a slightly (+) or severe (+++) dumpy body shape (Dpy).
¶Nematodes were scored for the percentage of animals arrested as dauer larvae (% dauer).
#Young adult nematodes were incubated in 1.0 mM of the nicotinic agonist levamisole, and the percentage of twitching animals was counted 3 min later (% twitchers). NT indicates that certain strains were not tested for the experiment.
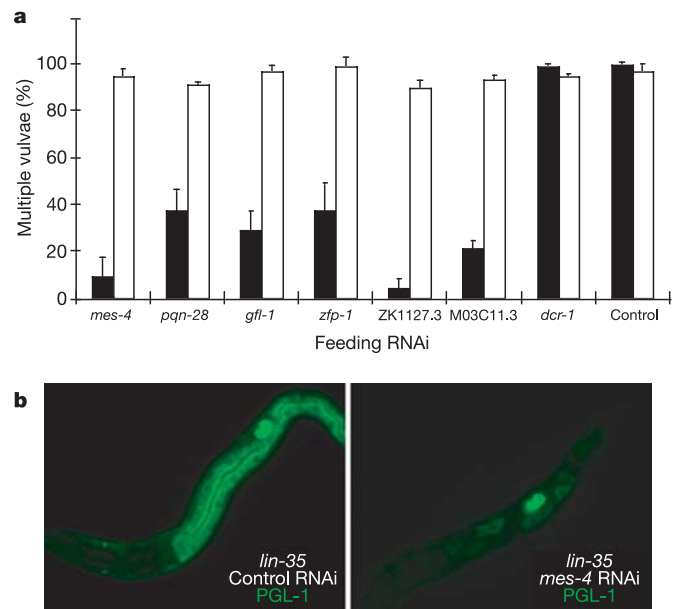
blastomeres from maternally expressed gene products, and are segregated exclusively to the germline lineage during early development. P granule components continue to be expressed only in the developing germline where they are assembled into complex structures at the nuclear pores[12]. Their perinuclear localization may reflect a nuclear transport function for the RNAs that accumulate in the P granules. The P granules are homologous to polar granules in *Drosophila*; both are composed of homologous helicases and RNA-binding proteins, as well as a number of mRNAs that specify germline fates and somatic patterning[10,12].

Germ lines are more protected than somatic cells from foreign genetic elements, such as transposons and viruses[13–15]. An increased level of RNAi in the *C. elegans* germ line is suggested by the silencing of transgenes and transposons specifically in the germ line and by the activation of transgenes and transposition by mutations that attenuate germline RNAi[14,15]. In Rb pathway mutants, germline-specific RNAi components may be de-repressed in somatic cells, which in turn enhances RNAi. Consistent with this model, Rb pathway mutations cause silencing of a repetitive *gfp* transgene in somatic cells (Fig. 2b) similar to germline-specific transgene silencing normally observed in the wild type[4]. Inactivation of *mut-16* by RNAi restores somatic GFP expression in the *lin-35* or *lin-15B* mutants (Fig. 2b), and RNAi in Rb pathway mutants requires the germline RNAi factor MUT-7 (Supplementary Table S1). PGL-1 is required for germline co-suppression[15], suggesting that the somatic PGL-1 expression in Rb pathway mutants could also be necessary for the enhanced RNAi.

Because mutations in the Rb pathway affect the cell divisions of vulval cells as well as enhance RNAi, we asked whether genes implicated as possible components in the RNAi pathway mediate the cell lineage defects of Rb pathway mutants. *C. elegans* RNAi gene inactivations that in turn attenuate RNAi have been identified[13]. We tested whether any of 36 viable RNAi-suppressing gene inactivations could also suppress the multivulvae phenotype of Rb pathway and other multivulvae mutants. Six gene inactivations suppress the multivulvae phenotype of three Rb pathway synMuv A and synMuv B double mutants: *lin-15AB(n765)*, *lin-35(n745); lin-8(n2731)*, and *lin-36(n766); lin-8(n2374)*. These six gene inactivations are *mes-4* (SET domain, trithorax class gene), *pqn-28* (SIN3 component), *gfl-1* (chromatin protein, orthologue of mammalian glioma-amplified sequence 41 (GAS41)), *zfp-1* (PHD domain chromatin protein, orthologue of mammalian AF10), M03C11.3 (weakly homologous to a chromatin-associated protein) and ZK1127.3 (novel protein) (Fig. 3a and data not shown). None of these 36 gene inactivations suppressed the multivulvae phenotype of the Ras gain-of-function mutant *let-60(n1700)* or the ETS transcription factor loss-of-function mutant *lin-1(e1777)* (Fig. 3a and data not shown), showing that these clones are not general multivulvae suppressors or Ras pathway genes. In a control RNAi screen of a hand-picked library of 434 clones encoding a variety of known RNAi, RNA-binding and RNA-processing factors, only *pqn-28* (SIN3)—the only one out of the six multivulvae-suppressing clones in this library—suppressed the multivulvae phenotype of *lin-35(n745); lin-8(n2731)* animals. Thus, a high hit rate was specific to the RNAi-attenuating clones. Neither RNAi of *dcr-1* nor null mutations in other genes necessary for RNAi, such as *rde-1* and *rde-4*, restore the wild-type vulval cell lineage in the Rb pathway mutants (Fig. 3a; see also Supplementary Table S1), suggesting that excessive RNAi is

Strains	RNAi	scm::gfp expression
Wild type	Control	+
lin-35	Control	−
	mut-16	+
	mes-4	+
lin-15B	Control	−
	rrf-1	−
	rde-1	+
	rde-4	+
	mut-16	+
	dcr-1	+

Figure 2 | Mutations in the Rb pathway cause soma to germline transformation and transgene silencing. a, The *lin-35* null mutant *lin-35(n745)* exhibits somatic expression of the germline-specific gene *pgl-1*. Wild-type and *lin-35(n745)* L1 larvae were fixed and immunostained with antibodies against the germline-specific P granule protein PGL-1 (green). The germ cells Z2/Z3 are marked in dashed boxes. DNA was counterstained with DAPI (blue). **b,** Transgene silencing in the soma of Rb pathway mutants. A seam cell-specific *gfp* transgene (*scm::gfp*) is silenced in *lin-35(n745)* and *lin-15B(n744)*, but not in wild type; inactivation of *mut-16* or *mes-4* by feeding RNAi restores expression of *scm::gfp* in *lin-35(n745)* or *lin-15B(n744)* animals. Fluorescence microscopy was performed on representative animals at the L4 larval stage. + and − represent the expression and silencing of the *scm::gfp* transgene, respectively. Bacteria expressing an empty vector L4440 are used as the control feeding RNAi.

Figure 3 | Inactivation of six RNAi factors suppresses the multiple vulvae phenotype of the Rb pathway mutants. a, *lin-35(n745);lin-8(n2731)* (filled columns) and the Ras gain-of-function mutant *let-60(n1700)* (open columns) were grown on bacteria expressing the indicated dsRNA (feeding RNAi) or an empty vector L4440 (control) at 22 °C. The percentage of progeny with multiple vulvae was scored. The error bars represent standard deviation among the data collected from at least three independent experiments. **b,** RNAi of *mes-4* suppresses the somatic expression of *pgl-1* in *lin-35(n745)* animals. The *lin-35(n745)* animals were fed bacteria expressing *mes-4* dsRNA (*mes-4* RNAi) or an empty vector L4440 (control RNAi) at 22 °C. Their progeny at larval stage L1 were fixed and immunostained with antibodies against PGL-1 (green).

not the cause of the cell lineage transformations of Rb pathway mutants.

Five of the six multivulvae-suppressing clones are annotated as chromatin factors that directly interact with each other and with synMuv B pathway proteins in other species. *Drosophila* homologues of MES-4 interact with SWI/SNF and NuRD complexes and antagonize Polycomb complexes[16]. MES-4 is homologous to the human MLL protein that is the fusion partner to AF10, the orthologue of ZFP-1, in leukaemic translocations[17]. The *Drosophila* homologue of ZFP-1, dAF10, binds to HP1, an orthologue of *C. elegans* HPL-2 (ref. 18). Mammalian AF10 binds the GFL-1 orthologue GAS41. GAS41 interacts with the SWI/SNF complex[17] that contains mSIN3 histone deacetylase, the orthologue of PQN-28, and is a direct target of the oncogene *Myb*[19]. Mammalian SIN3 interacts directly with Rb chromatin complexes[20]. Myb co-purifies with Rb and E2F in the *Drosophila* dREAM complex, and seven out of eight dREAM subunits are orthologues of *C. elegans* synMuv B genes[21]. The dREAM complex binds to particular chromosomal domains on the *Drosophila* polytene chromosomes, suggesting that, similar to the Polycomb complex, an antagonism between Rb-repressing and MES-4-activating complexes may take place at particular chromatin loci[21]. MES-4 coats the autosomes in the *C. elegans* germ line; the MES-2, MES-3 and MES-6 proteins antagonize the binding of MES-4 to the X chromosome to mediate the silencing of the X chromosome in the *C. elegans* germ line[22].

We suggest that in the absence of the antagonistic Rb pathway complex, a MES-4 chromatin remodelling complex that includes the five other proteins identified above is inappropriately active in the soma to activate the expression of P granule genes like *pgl-1*. In the hypodermis, where ectopic P granules are manifest in the Rb mutant, inappropriate activity of the MES-4 complex may promote the expression of vulva-specific genes in the adjacent vulval precursor cells[23]. The loss of Rb pathway function may activate latent programmes of gene expression characteristic of more primitive, less differentiated cells. Consistent with this model, *mes-4* RNAi suppresses the transgene silencing (Fig. 2b), vulval cell lineage defect (Fig. 3a) and somatic misexpression of PGL-1 (Fig. 3b) in the *lin-35* (*Rb*) mutant. A mutation in *mes-4* also suppresses the lethality and ectopic P granule phenotypes of *mep-1* and *let-418* (ref. 11), and *mes-4* is required for germline co-suppression[13,15]. Two of the other five multivulvae-suppressing clones—*gfl-1* (*GAS41*) and ZK1127.3—also reverse the transgene silencing of Rb pathway mutants, but none of the other five suppresses somatic P granule formation as markedly as *mes-4* RNAi (data not shown). Consistent with the activity of these genes normally in the germ line, inactivation of these genes causes decreased brood size or sterility[13]. Because these proteins may form a biochemical complex, they probably regulate the choice between germ line and soma via the same molecular mechanism. However, *mes-4* has an additional function in X chromosome inactivation and could be more pleiotropic[22].

A related model that incorporates precedents for Rb control of cell cycle and endoreduplication[21] is that the Rb pathway chromatin factors inhibit the endoreduplication of particular target loci such as the ones that generate the abundant P granule components in the germ line. *mes-4*, *zfp-1*, *gfl-1*, *pqn-28*, ZK1123.3 and M03C11.3 may be these amplified targets in the normal germ line. In the absence of Rb, these genome regions may be amplified in soma to activate biogenesis of P granules; RNAi of the *mes-4* complex genes may interfere with these amplifications or with the assembly of P granules and other germline components from these amplified gene products. This model is favoured by the observation that mammalian homologues of *mes-4*, *zfp-1*, *gfl-1* and *pqn-28* are amplified in some tumours that may also carry Rb mutations[24,25]. In support of an endoreduplication model, five RNAi regulatory genes[13] including ZK1127.3 are clustered in one genomic region (ZK1127.3, ZK1127.7, ZK1127.4, ZK1127.9/6, ZK1127.5), and *gfl-1* (M04B2.3) clusters

with its neighbour, the synMuv B gene *mep-1* (M04B2.1); these clusters may correspond to domains of amplification regulated by the Rb pathway chromatin remodelling factors.

Rb pathway mutations may also enhance RNAi by inactivating a chromatin silencing pathway that competes with the RNAi machinery for shared silencing components. The methylated histone-binding protein HP1 associates with RNAi components on yeast[26] and fly[27] heterochromatin. In *C. elegans*, mutants lacking the HP1 orthologue *hpl-2* display an enhanced RNAi phenotype as well as vulval cell lineage defects when combined with a synMuv A mutation, showing that *hpl-2* is a synMuv B gene with enhanced RNAi (ref. 28) (Fig. 1a, b). Consistent with a role in transcriptional gene silencing, the *hpl-2* mutation causes transgene de-silencing[28]. In the absence of Rb and its associated chromatin remodelling complex, including HP1, RNAi factors such as DCR-1 and Argonaute proteins may become available to mediate RNAi in the cytoplasm, leading to enhanced RNAi. However, PGL-1 is also misexpressed in the soma of the *hpl-2* mutant, suggesting that this mutant does not just release RNAi factors from chromatin to enhance RNAi. The expression of germline components in somatic cells and a release of RNAi factors from their role specifying heterochromatin regions may conspire to activate RNAi in Rb pathway mutants.

It may be significant to the action of human retinoblastoma that *C. elegans* Rb pathway mutations cause a somatic to germline cell transformation, and that some gene inactivations that disrupt RNAi can reverse the cell lineage transformations of Rb pathway mutants. The mammalian cell cycle transformations induced by Rb inactivation may also be triggered by expression of germ line or other stem-cell-like programmes in somatic cells. Gene microarrays have revealed increased expression in *Drosophila* Rb mutant cells of genes annotated to function in oocyte development, including the polar granule RNA helicase *vasa*[29]. Similarly, the RNA helicase DDX1 is amplified in Rb tumours and neuroblastomas, and is a marker for stem cell fate in the mouse[30,31]. If Rb pathway mutations cause amplification of particular chromatin regions, gene dosage studies of tumours could reveal those target genes, such as the *MLL, GAS41* and *AF10* genes. Inactivation of the mammalian homologues of the four multivulvae-suppressing genes that have good mammalian homologues—*mes-4*, *pqn-28*, *gfl-1* and *zfp-1*—for example by specific siRNAs or pharmaceuticals, might reverse the cell cycle defects and genome instability of Rb mutant tumours and cell lines. Conversely, activation of these Rb-regulated genes may confer more totipotent cell fates, which could be useful in inducing cells to recapitulate developmental pathways.

METHODS

Immunofluorescence staining. L1 larvae were permeablized using a freeze–crack method by freezing between a polylysine-coated slide and a coverslip followed by rapid removal of the coverslip. Slides were immediately immersed in ice-cold 100% methanol for 10 min followed by ice-cold 100% acetone for 5 min. Larvae were stained using monoclonal (1:20 dilution) or affinity purified polyclonal (1:2,000 dilution) anti-PGL-1 antibodies[10] (gifts from S. Strome) followed by TRITC-conjugated donkey anti-mouse IgM (1:200 dilution; Jackson Immunoresearch) or Alexa Fluor-conjugated goat anti-rabbit IgG (1:1,000 dilution; Molecular Probes).

Behavioural test. For the aldicarb resistance test, young adult animals were transplanted to agar plates containing 1.0 mM of the acetylcholinesterase inhibitor aldicarb, and the percentage of animals paralysed over a time course of 140 min was measured. For the levamisole resistance test, young adult

nematodes were incubated in M9 solution with 1.0 mM of the nicotinic agonist levamisole, and the percentage of twitching animals was counted 3 min later.

1. Lu, X. & Horvitz, H. R. lin-35 and lin-53, two genes that antagonize a *C. elegans* Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell* **95**, 981–991 (1998).
2. Ceol, C. J. & Horvitz, H. R. dpl-1 DP and efl-1 E2F act with lin-35 Rb to antagonize Ras signalling in *C. elegans* vulval development. *Mol. Cell* **7**, 461–473 (2001).
3. Huang, L. S., Tzou, P. & Sternberg, P. W. The lin-15 locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol. Biol. Cell* **5**, 395–411 (1994).
4. Hsieh, J. *et al.* The RING finger/B-box factor TAM-1 and a retinoblastoma-like protein LIN-35 modulate context-dependent gene silencing in *Caenorhabditis elegans*. *Genes Dev.* **13**, 2958–2970 (1999).
5. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
6. Simmer, F. *et al.* Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12**, 1317–1319 (2002).
7. Sieburth, D. *et al.* Systematic analysis of genes required for synapse structure and function. *Nature* doi:10.1038/nature03809 (this issue).
8. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
9. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
10. Kawasaki, I. *et al.* PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in *C. elegans*. *Cell* **94**, 635–645 (1998).
11. Unhavaithaya, Y. *et al.* MEP-1 and a homolog of the NURD complex component Mi-2 act together to maintain germline-soma distinctions in *C. elegans*. *Cell* **111**, 991–1002 (2002).
12. Pitt, J. N., Schisa, J. A. & Priess, J. R. P granules in the germ cells of *Caenorhabditis elegans* adults are associated with clusters of nuclear pores and contain RNA. *Dev. Biol.* **219**, 315–333 (2000).
13. Kim, J. K. *et al.* Functional genomic analysis of RNA interference in *C. elegans*. *Science* **308**, 1164–1167 (2005).
14. Sijen, T. & Plasterk, R. H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
15. Robert, V. J., Sijen, T., van Wolfswinkel, J. & Plasterk, R. H. Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. *Genes Dev.* **19**, 782–787 (2005).
16. Klymenko, T. & Muller, J. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).
17. Debernardi, S. *et al.* The MLL fusion partner AF10 binds GAS41, a protein that interacts with the human SWI/SNF complex. *Blood* **99**, 275–281 (2002).
18. Linder, B., Gerlach, N. & Jackle, H. The *Drosophila* homolog of the human AF10 is an HP1-interacting suppressor of position effect variegation. *EMBO Rep.* **2**, 211–216 (2001).
19. Braas, D., Gundelach, H. & Klempnauer, K. H. The glioma-amplified sequence 41 gene (GAS41) is a direct Myb target gene. *Nucleic Acids Res.* **32**, 4750–4757 (2004).
20. Lai, A. *et al.* RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumour suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. *Mol. Cell. Biol.* **21**, 2918–2932 (2001).
21. Lewis, P. W. *et al.* Identification of a *Drosophila* Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev.* **18**, 2929–2940 (2004).
22. Fong, Y., Bender, L., Wang, W. & Strome, S. Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. *Science* **296**, 2235–2238 (2002).
23. Myers, T. R. & Greenwald, I. lin-35 Rb acts in the major hypodermis to oppose ras-mediated vulval induction in *C. elegans*. *Dev. Cell* **8**, 117–123 (2005).
24. Fischer, U., Meltzer, P. & Meese, E. Twelve amplified and expressed genes localized in a single domain in glioma. *Hum. Genet.* **98**, 625–628 (1996).
25. Streubel, B. *et al.* Amplification of the MLL gene on double minutes, a homogeneously staining region, and ring chromosomes in five patients with acute myeloid leukemia or myelodysplastic syndrome. *Genes Chromosom. Cancer* **27**, 380–386 (2000).
26. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
27. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
28. Couteau, F., Guerry, F., Muller, F. & Palladino, F. A heterochromatin protein 1 homologue in *Caenorhabditis elegans* acts in germline and vulval development. *EMBO Rep.* **3**, 235–241 (2002).
29. Black, E. P. *et al.* Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res.* **63**, 3716–3723 (2003).
30. Godbout, R., Packer, M. & Bie, W. Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines. *J. Biol. Chem.* **273**, 21161–21168 (1998).
31. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. Stemness: transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597–600 (2002).

Author Contributions D.W. conducted experiments and analysed data for all figures and tables, except for Figs 2a and 3b; S.K. made the initial observation of enhanced RNAi in *lin-15* and synergistic enhancement of RNAi in *eri-1*; *lin-15*; D.C. did immunofluorescence staining of PGL-1 in Figs 1a and 3b; J.K.K., H.W.G. and R.S.K. identified 36 RNAi-defective genes that were tested by D.W. for their interaction with the Rb pathway; C.C.M. and G.R. were principal investigators.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R. (ruvkun@molbio.mgh.harvard.edu).

● FOCUS
● SPOTLIGHT
● RECRUITMENT
● ANNOUNCEMENTS
● EVENTS
●

naturejobs

India's changing face

Jitu Major returned to his native India in 1995 after a PhD and postdoc in New York. But he was concerned that in India his career would be threatened by "obscurantist science, an ossified research set-up and cash-strapped institutes", he says. Yet after ten years at the National Centre for Biological Sciences in Bangalore, Major is happy to admit that he was wrong.

India's scientific reputation is not really one of innovation. Its inexpensive labour force compared with the West meant that the subcontinent became an outsourcing centre first in information technology and, more recently, in medicinal chemistry. These contracts with Western drug and IT companies have provided jobs for Indian scientists, albeit at the less glamorous end of the spectrum and governed from abroad. But there are now signs that India is ready to host its own scientific revolution (see page 477).

The many changes that are taking place within Indian science are doing a lot to lay Major's concerns to rest. Obscurantist science? Researchers at the Indian Institute of Science in Bangalore have consulted with 161 companies around the world, with three professors spinning out companies in the past five years. And in the same time the country has produced more than 200 papers in high-impact journals. Ossified equipment? The Centre for Genomic Application now has a supercomputer as fast as any at the Wellcome Trust Sanger Institute in Britain or at the Institute for Molecular Science in Japan.

These examples are all improvements that can help to lure expats back home. What is still missing is funding for basic research, higher salaries for university professors and a real postdoc scheme. Money from global pharmaceutical companies and foundations such as the Wellcome Trust are helping to bridge the gap. But greater investment from within the country will be necessary to see a flood of Indian expat scientists join Major in coming home.

Paul Smaglik, *Naturejobs* editor

MOVERS

Floyd Bloom, chief scientific officer and chairman of the board, Neurome, La Jolla, California

2000-05: Chief executive, Neurome, La Jolla, California
1989-2000, 2002-05: Chairman, neuropharmacology department, Scripps Research Institute, La Jolla, California
1995-2000: Editor-in-chief, *Science*, Washington DC
1975-83: Director of behavioural neurobiology, Salk Institute for Biological Studies, La Jolla, California

Floyd Bloom isn't afraid to take risks. As a medical student, his interest in pharmacology led him to the National Institute of Mental Health (NIMH) Clinical Neuropharmacology Research Center at St Elizabeth's Hospital in Washington DC. He ended up in a brain-research lab studying the membrane biophysics of nerve conduction.

At a time when the term 'neuroscience' wasn't yet in use, he went next to Yale University to learn histochemistry and improve his understanding of the nervous system. Ever eager to use new tools, he was one of the first neurobiologists to use modern molecular-biology techniques to characterize brain-specific genes. He also pioneered the use of computers to create neuro-anatomic databases.

Bloom's early career looks as though it was heavier on breadth than depth. But he believes that his wide-ranging pursuit of knowledge helped him to identify the role of specific neurotransmitter systems and how they work together, as well as making breakthroughs in detailing how nerve cells communicate. "I never played it safe," he says, admitting that he took the jump from physiology to electron microscopy and anatomy without knowing whether he would discover anything useful.

By all measures, his risk-taking has paid off. After stints as director of behavioural neurobiology at the Salk Institute for Biological Studies in La Jolla, California, and chief of the neuropharmacology laboratory at the NIMH, Bloom took the chair of the neuropharmacology department at the Scripps Research Institute in La Jolla, in 1989. Six years later he also became editor-in-chief of *Science*, where he stayed for five years, leading the journal into the electronic age. "We went online in ways that are still quite novel and useful," he says. From 2002 to 2003 he was president of the American Association for the Advancement of Science.

After his stint in publishing, Bloom decided to tackle the corporate world, founding a biotech company called Neurome to seek therapies for neurodegenerative diseases such as Alzheimer's and Parkinson's. Neurome is set up around Bloom's hypothesis that finding and mapping where genes are expressed in the brain will improve understanding of diseases and foster drug development.

Bloom recently stepped down from his chair position at Scripps, preferring to focus his energies on his roles as Neurome's chairman of the board and chief scientific officer. He says this stage is the most fun of his entire career because he doesn't "need to ask anyone's permission" to do what he wants.

RECRUITERS & ACADEMIA

Software solutions

Many scientists spend a good proportion of their time using software to organize and analyse data, which means they also spend time writing, debugging and maintaining software. But only a handful of researchers have ever been taught how to do this efficiently. After an introductory course in C, Java or Fortran, many scientists have to rediscover or reinvent the rest of programming on their own. As a result, they spend too much time wrestling with software when they would be better off pursuing their research — studying gene expression in mustard plants or the molecular dynamics of nanotubes, for example.

What is potentially worse is that most scientists don't know how reliable their software is, and are often unable to trace the results they publish back to the programs that produced them.

To address these issues, I am working with the Python Software Foundation, a non-profit organization devoted to advancing open-source technology, to develop a course that will teach scientists and engineers the 10% of software engineering they need to solve 90% of their problems. Our goal is not to turn bench researchers into computer scientists, but to introduce them to some open-source tools and working practices that can reduce the amount of time they spend programming by up to 25%.

Some of the topics the course addresses are the equivalent of good laboratory practice, such as using version-control systems to track changes to documents and data files, and automated unit testing to ensure that once a bug is fixed, it stays fixed. The course also looks at how to read and manipulate common data formats, from plain text to XML, how to integrate new data with legacy code (such as the tangled Fortran program you just inherited from your PhD supervisor), and how web-based tools such as wikis and weblogs can be used to streamline and coordinate geographically distributed research teams.

A draft version of the course notes will soon be online at www.third-bit.com/swc. All of the material will be freely available for personal, academic and commercial use under an open-source licence, and questions and comments are welcome.

Greg Wilson, a computer scientist, divides his time between consulting, writing and teaching at the University of Toronto, Canada. For more information and feedback e-mail: gvwilson@third-bit.com

Crisis of faith

When I began my PhD, my undergraduate research supervisor said to me: "Welcome on board. You are no longer a student, you are now one of my colleagues." That made a big impact on me, because my supervisor is a well-known professor who has years of scientific experience. I feel privileged to have her as a co-worker on some of my projects. When she welcomed me, I felt as though I had finished one chapter of my life and advanced to the next.

But within less than a year, I was disappointed to learn that not all senior scientists see me in the same light. When I attended a recent event at an institution that is collaborating with some of our labs, I felt as though I was being judged more on the basis of my age and position in the laboratory than on what I was contributing to the projects.

As a result I have began to question whether I have, in fact, turned a new page and am advancing in my career. Was this unfortunate experience just an insignificant incident? Or was I really being downgraded to the status I held as an undergraduate? If the latter, what then does it mean to be a PhD researcher? Is it just another stage of student education or a genuine scientific promotion for somebody who already possesses some knowledge?

For me this matter is still unclear. Maybe my encounter was an exception — I certainly hope it is not the general perception among more senior academics. Because I can't imagine my undergraduate supervisor 'promoting' me and sending me into a collaboration if she didn't think I had anything to offer.
Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.

The forever kitten

Beyond nine lives.

Peter F. Hamilton

The mansion's garden was screened by lush trees. I never thought I'd be so entranced by anything as simple as horse chestnuts, but that's what 18 months in jail on remand will do for your appreciation of the simple things.

Joe Gordon was waiting for me; the venture capitalist and his wife Fiona were sitting on ornate metal chairs in a sunken patio area. Their five-year-old daughter, Heloise, was sprawled on a pile of cushions, playing with a ginger kitten.

"Thanks for paying my bail," I said.

"Sorry it took so long, Doctor," he said. "The preparations weren't easy, but we have a private plane waiting to take you to the Caribbean — an island the EU has no extradition treaty with."

"I see. Do you think it's necessary?"

"For the moment, yes. The Brussels Bioethics Commission is looking to make an example of you. They didn't appreciate how many regulations you violated."

"They wouldn't have minded if the treatment had worked properly."

"Of course not, but that day isn't here yet, is it? We can set you up with another lab out there."

"Ah well, there are worse places to be exiled. I appreciate it."

"Least we could do. My colleagues and I made a lot of money from the Viagra gland you developed."

I looked at Heloise again. She was a beautiful child, and the smile on her face as she played with the kitten was angelic. The ball of ginger fluff was full of rascally high spirits, just like every two-month-old kitten. I kept staring, shocked by the familiar pattern of marbling in its fluffy light fur.

"Yes," Joe said with quiet pride. "I managed to save one before the court had the litter destroyed. A simple substitution; the police never knew."

"It's three years old now," I whispered.

"Indeed. Heloise is very fond of it."

"Do you understand what this means? The initial stasis-regeneration procedure is valid. If the kitten is still alive and maintaining itself at the same biological age after this long, then in theory it can live for ever, just as it is. The procedure stabilized its cellular structure."

"I understand perfectly, thank you, Doctor. Which is why we intend to keep on funding your research. We believe that human rejuvenation is possible."

I recognized the greed in his eyes: it wasn't pleasant. "It's still a long way off. This procedure was just the first of a great many. It has no real practical application, we can't use it on an adult. Once a mammal reaches sexual maturity its cells can't accept such a radical modification."

"We have every confidence that in the end you'll produce the result we all want."

I turned back to the child with her pet, feeling more optimistic than I had in three years. "I can do it," I said through clenched teeth. "I can." Revenge, it is said, is best served cold. I could see myself looking down on the gravestones of those fools in the Bioethics Commission in, say…oh, about 500 years' time. They'd be very cold indeed by then.

Joe's affable smile suddenly hardened. I turned, fearing the police had arrived. I'm still very twitchy about raids.

It wasn't the police. The teenage girl coming out from the house was dressed in a black leather micro-skirt and very tight scarlet T-shirt. She would have been attractive if it wasn't for the permanent expression of belligerence on her face; the tattoos weren't nice either. The short sleeves on the T-shirt revealed track marks on her arms. "Is that…"

"Saskia," Joe said with extreme distaste.

I really wouldn't have recognized his older daughter. Saskia used to be a lovely girl. This creature was the kind of horror story that belonged on the front page of a tabloid.

"Whatcha starin' at?" she demanded.

"Nothing," I promised quickly.

"I need money," she told her father.

"Get a job."

Her face screwed up in rage. I really believed she was going to hit him. I could see Heloise behind her on the verge of tears, arms curling protectively around the kitten.

"You know what I'll do to get it if you don't," Saskia said.

"Fine," Joe snapped. "We no longer care."

She made an obscene gesture and hurried back through the mansion. For a moment I thought Joe was going to run after her. I'd never seen him so angry. Instead he turned to his wife, who was frozen in her chair, shaking slightly. "Are you all right?" he asked tenderly.

She nodded bravely, her eyes slowly refocusing.

"What happened?" I asked.

"I don't know," Joe said bitterly. "We didn't spoil her, we were very careful about that. Then about a year ago she started hanging out with the wrong sort: we've been living in a nightmare ever since. She's quit school; she's got a drug habit, she steals from us constantly, I can't remember how many times she's been arrested for joyriding and shoplifting."

"I'm sorry. Kids, huh!"

"Teenagers," he said wretchedly. "Fiona needed two Prozac gland implants to cope."

I smiled over at Heloise, who had started playing with the kitten again. "At least you've got her."

"Yes." Joe seemed to make some kind of decision. "Before you leave, I'd like you to perform the cellular stasis-regeneration procedure for me."

"I don't understand. I explained before, it's simply the first stage of verifying the overwrite sequence we developed."

His attitude changed. "Nevertheless, you will do it again. Without my help you will be going back to prison for a long time."

"It's of no use to adults," I said helplessly. "You won't become young, or even maintain your current age."

"It's not for me," he said.

"Then who…" I followed his gaze to Heloise. "Oh."

"She's perfect just the way she is," he said quietly. "And that, Doctor, is the way she's going to stay." ▪

JACEY